

Oryza Longistaminata-like Cryptic Species Found in the Mekong Delta Probably Involved in the Origin of O. Rufipogon Indica Ancestor

Htet Aung Htut

Chiba University: Chiba Daigaku

So Makabe

Bex Co. Ltd.

Hiroko Takahashi

Bex Co. Ltd.

Bui Chi Buu

Institute of Agricultural Science for Southern Vietnam: Vien Khoa Hoc Ky Thuat Nong Nghiep Mien Nam

Nguyen Thi Lang

High Agricultural Technology Research Institute

Poku Aduse Samuel

Chiba University: Chiba Daigaku

Ryuji Ishikawa

Hirosaki University: Hirosaki Daigaku

Ikuo Nakamura (✉ inakamur@faculty.chiba-u.jp)

Chiba University: Chiba Daigaku <https://orcid.org/0000-0002-3640-2971>

Original article

Keywords: cryptic species, deletion, indica ancestor, Mekong Delta, plastid, O. longistaminata, O. rufipogon, PolA1, single-copy gene, hybrid speciation

Posted Date: July 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-717529/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Oryza sativa L. contains two cultivated subspecies, *indica* and *japonica*, which are considered to be derived from wild rice relative, *O. rufipogon*. Although domestication scenario of *O. sativa* has been extensively studied, it is still inconclusive and contradictory. Moreover, it remains to be revealed why *O. rufipogon* had acquired wide range of genetic diversity to fit various habitat and life cycle of plant from annual to perennial.

Results

Although *O. rufipogon* distributes in southeast Asia from Indonesia to India, perennial wild rice strains were distributed in deep-water area of the Mekong Delta in Southern Vietnam. Some of wild rice strains collected in Dong Thap were heterologous in the length of intron 20 sequence of single-copy nuclear *PolA1* gene, encoding for the largest subunit of RNA polymerase I. Thus, these wild rice strains were probably considered as interspecific hybrids, because the S-type (0.14 kb) intron 20 was found in *O. rufipogon*, *O. glumaepatula*, and *O. barthii* while L-type (1.5 kb) was shared by *O. longistaminata* and *O. meridionalis*. Therefore, we identified two wild rice strains that contained homologous L-type intron 20 of *PolA1* gene in the CLRRRI collection. These two strains, named as MD1 and MD2, had fertile spikelet with long anthers which were typical feature of African *O. longistaminata*. In addition, MD1 strain and *O. longistaminata* shared the same protein tag (Ptag) sequence, which was different to that of *O. rufipogon* and *O. sativa*. Interestingly, MD1 strain contained plastid DNA with a 69-bp deletion in ORF100 which was shared with *indica* and annual *O. rufipogon*, but not with *japonica* and perennial *O. rufipogon*.

Conclusion

We have identified that MD strains found in the Mekong Delta were probably *O. longistaminata*-like cryptic species and contributed to enlarge genetic diversity of *O. rufipogon*, and is probably involved in the origin of *O. rufipogon indica* ancestor of as the maternal lineage.

Background

Asian cultivated rice, *O. sativa* L., have been classified into two distinct subspecies, *indica* and *japonica* (Kato 1930). The domestication scenario of two subspecies from a common wild relative *O. rufipogon* have long been controversial issue in many evolutionary studies (Kovach et al. 2007, Vaughan et al. 2008, Izawa 2008, Kovach and McCouch 2008). Single-origin hypothesis indicates that cultivated rice was domesticated once and differentiated into *indica* and *japonica* (Vaughan et al. 2008) because domestication-controlling genes, such as seed shattering and pericarp color, were shared between two subspecies (Sang and Ge 2007, Sweeney and McCouch 2007, Zhang et al. 2009). Moreover, single domestication event was supported by demographic modeling of SNPs (Molina et al. 2011).

In contrast, two-origin hypothesis states that *indica* and *japonica* cultivars were independently originated from distinct populations of *O. rufipogon* (Oka 1988, Yamanaka et al. 2003). Many molecular data showed *indica* and *japonica* formed two independent clusters together with different groups of *O. rufipogon* (Second 1982, Cheng et al. 2003, Londo et al. 2006). In addition, new intermediate hypothesis between single-origin and two- or multiple-origin hypothesis was proposed that *japonica* cultivar was first evolved from *O. rufipogon* as a single domestication and was subsequently hybridized with another group of *O. rufipogon* to generate *indica* cultivar (Huang et al. 2012, Huang and Hun 2016, Choi et al. 2017).

Oryza rufipogon is native to East to South Asia, and its habitat and life cycle is differentiated from deep swamp to temporary marshes on higher lands, and from perennial to annual (Oka 1988). Although annual group of *O. rufipogon* is classified as *O. nivara*, these two species were regarded as a single species (Morishima et al. 1992, Vaughan et al. 2003, Zhu et al. 2007), because artificial hybridization of these two species have not developed an apparent reproductive barrier. Wide range of genetic diversity controlling many kinds of characteristics of *O. rufipogon* had contributed to produce various ecotypes, aus, aman, boro, bulu, and tjereh (Ueno et al. 1990). However, it remains to be revealed why *O. rufipogon* had acquired such large genetic diversity. Even genome wide sequence analyses could not answer this issue.

Single-copy nuclear gene, *PolA1*, encodes the largest subunit (194 kDa) of RNA Polymerase I holoenzyme which plays an essential role for the synthesis of 45S rRNA precursors (Russell et al 2016). The *PolA1* gene (ca. 15 kb) contains 21 exons on chromosome 6 of rice genome (Kawahara et al. 2013). Phylogenetic studies for the intron 19 and exon 20 sequences of *PolA1* gene were useful to analyze relationships in *Petunia* (Zhang et al. 2008), *Aegilops* (Nakamura et al. 2009), *Oryza* (Takahashi et al. 2009), *Triticum-Aegilops* (Takahashi et al. 2010), *Triticum-Hordeum* (Rai et al. 2012), and *Brassica* (Fareed et al. 2015, 2016).

Recently, we found that intron 20 sequence of *PolA1* gene were differentiated into short (S)-type (0.14 kb) and long (L)-type (1.5 kb) in *Oryza* AA-genome species. Except some exceptions, *O. longistaminata* and *O. meridionalis* had L-type while *O. rufipogon*, *O. barthii*, *O. glumaepatule* contained S-type. Because all *Oryza* species containing other than AA genome had L-type, this result provided a direct evidence for the speciation from L-type species to S-type species within AA genome species (Htut et al. submitted).

Nakamura (2010) found that a particular protein tag (Ptag) sequence (ca. 400 aa) showed a species-specific variation, encoded by exons 19 – 21 of *PolA1* gene in land plants. The Ptag sequence is useful to classify species in *Triticum-Hordeum* (Rai et al, 2012), *Brassica* (Fareed et al. 2016), and *Trichophyton* fungi (Yamanishi et al. 2017). Therefore, Ptag sequence was determined in order to discriminate species-level differences of wild rice strains found in the Mekong Delta.

A 69-bp deletion (D) in open reading frame 100 (ORF100) on rice chloroplast DNA (ca. 134 kb) was found in *indica* whereas *japonica* contained non-deletion (ND)-type ORF100 (Kanno et al. 1993). Also, D-type

and ND-type ORF100 were predominant in annual and perennial strains of *O. rufipogon*, respectively (Chen et al. 1993, Garriss et al. 2005, Takahashi et al. 2008). Because other *Oryza* AA-genome wild species, *O. longistaminata*, *O. barthii*, *O. glumaepatula*, and *O. meridionalis*, contained ND-type ORF100 (Takahashi et al. 2008), distribution of D-type ORF100 in *O. sativa* and *O. rufipogon* became good DNA marker to uncover the domestication pathway of maternal lineage from *O. rufipogon* to *O. sativa* (Garriss et al. 2005).

Morishima (1994) first reported that the observations of wild rice species in deep-water area along the Mekong delta, and wild rice populations in South Vietnam were all perennial type of *O. rufipogon*. During wild rice survey in the Mekong Delta at 2010, we found wild rice strains with sterile spikelets and big stature (Ac7 strain in this paper) and were specially interested in their origins. Although genetic diversity of nuclear and plastid markers was extensively studied among wild rice strains in the Mekong Delta (Ishii et al. 2011, Lam et al. 2019), their evolutionary relationships were difficult to be resolved by comparing combinations of SNPs and DNA markers. Therefore, in this study, we were interested in the deletions of large DNA fragments found in the intron 20 of single-copy nuclear *PolA1* gene and in ORF100 of plastid DNA to study evolutionary relationships among wild rice strains in the Mekong Delta.

Results

Wild rice strains collected in the Mekong delta of Vietnam

Wild rice survey was performed in the Mekong Delta of Vietnam conducted by Dr. R. Ishikawa (Hiroshima University, Japan) during 2010 – 2015 (Fig. 1). At first, we found a wild rice strain (Ac7) in a canal along a road in front of the Cuu Long Delta Rice Research Institute (CLRRI), Can Tho. This clonal propagated wild rice strain with sterile spikelet showed big stature like hybrid vigor with unusually bigger leaf blades (width: 3.3 cm, length: more than 1 m) than ordinary *O. rufipogon* (Fig. 2). Thus, six wild rice strains (Ac1 – Ac6) were collected in the Tram Chim National Park, upstream of the Mekong river close to Cambodian border (Fig. 1).

DNA analysis of wild rice strains in the Mekong Delta

Although *japonica* 'Nipponbare' (NB) contained S-type intron 20 of *PolA1* gene, four strains (Ac1, 3, 4, 5) out of six Dong Thap strains were found to be heterologous of S-type and L-type (1.5 kb) intron 20 (Fig. 3C). Dong Thap Ac2 and Ac6, and Can Tho Ac7 strains had homologous S-type intron 20. Therefore, two wild rice strains, named as MD1 and MD2, having homologous L-type intron 20 were identified among 100 strains of the germplasm collection in CLRRI. The MD1 and MD2 strains from Dong Cat village contained a 69-base deletion (D)-type ORF100 of plastid DNA like *indica* cultivars (Fig. 3D). Five Dong Thap strains had D-type ORF100 while Ac5 (Dong Thap) and Ac7 (Can Tho) strains contained ND-type ORF100 like *japonica* cultivar.

Morphological characteristics of the MD1 strain

The MD1 plant was grown with three strains (W0137, W0130, W1244) of *O. rufipogon* or two strains (IRGC101198, IRGC101205) of *O. longistaminata* (Fig. 4A). Although their growth condition was insufficient in Chiba, Japan, MD1 strain was probably taller than *O. rufipogon* and more branchy than *O. longistaminata*. The MD1 plant possibly produced rhizomatous tissues like *O. longistaminata* (data not shown). Wild rice strain (not MD strain) having homologous L-type intron 20 was found at My Tho in Vietnam but we could not observe its flowers (Fig. 4B).

Floret, anthers, and pollen viability of MD1 strain were compared with those of *O. rufipogon* (W0137) and *O. longistaminata* (IRGC101198) (Fig. 5A). Protruding anther and large purple stigma was clearly seen in MD1 strain. Anther length of MD1 strain was comparable to those of *O. longistaminata* and two times longer than those of *O. rufipogon* (Fig. 5B). African *O. longistaminata* was known to have semi-sterile spikelet and pollen viability of IRGC101198 strain was very low (Fig. 5C). Whereas MD1 strain showed higher pollen viability like W0137 strain of *O. rufipogon*.

Phylogenetic analysis of the MD1 strain

The intron 20 and Ptag-coding sequences of *PolA1* gene were determined in MD1 strain. Phylogenetic Neighbor-Joining (N-J) tree of the L-type intron 20 sequences (*ca.* 1.5 kb) of *PolA1* gene showed that MD1 strain was closely related to two strains of *O. longistaminata* and distantly related to two strains of *O. meridionalis* (Fig. 6A). And as shown in Fig. 6B, Ptag-coding sequence (1230 bp) of MD1 strain was identical to that of *O. longistaminata* and different to that of *O. rufipogon*, *indica*, and *japonica*, and distantly related to that of *O. glumaepatula* and *O. meridionalis*.

PCR product containing ORF100 sequence of MD1 strain was smaller than that of two *O. longistaminata* strains (Fig. 6C). Sequence analysis revealed that the same 69-bp deletion (D-type) in ORF100 was found in MD1 strain as well as *indica* cultivar. Whereas African *O. longistaminata* had ND-type ORF100 like *japonica* cultivar. We found that *O. rufipogon* possessed either ND-type or D-type ORF100 (Chen et al. 1993).

According to the 3000 rice genome project of IRRI, Ptag sequences (410 aa) of 2858 rice cultivars were classified into two major types, *japonica* (1618 cultivars, 57%) and *indica* (1083 cultivars, 38%). Ptag sequences of *japonica*-type and *indica*-type were represented by the combination of three amino acids, as $^{204}\text{D-}^{257}\text{L-}^{375}\text{G}$ and $^{204}\text{G-}^{257}\text{L-}^{375}\text{E}$, respectively (Fig. 6D). Recombination of Ptag-sequence between *japonica*-type and *indica*-type was probably suppressed because frequency of the recombinant-type ($^{204}\text{G-}^{257}\text{L-}^{375}\text{G}$) was very low (157 cultivars, 5%). MD1 strain and *O. longistaminata* contained the same Ptag-sequence ($^{204}\text{D-}^{257}\text{S-}^{375}\text{E}$), which was different to that of *japonica*-type and *indica*-type (Fig. 6D).

Discussion

During wild rice survey in the Mekong Delta (Fig. 1), we found that a wild rice strain (Ac7) with sterile spikelet in Can Tho propagated clonally by shooting from each node and showed unusually big stature (Fig. 2). As Ac7 plant was considered as showing heterosis, six wild rice strains (Ac1 – Ac6) were

collected in the Tram Chim National Park, Dong Thap, upstream the Mekong Delta. Four Dong Thap (Ac1, Ac3, Ac4, Ac5) strains were considered as interspecific hybrids because they contained both S- and L-type intron 20 of *PolA1* gene (Fig. 3C). Although two Dong Thap (Ac1 and Ac6) and one Can Tho (Ac7) strains possessed homologous S-type, their maternal lineages were differentiated into D-type and ND-type, respectively (Fig. 3D). Two strains (MD1, MD2) from the Dong Cat village had homologous L-type intron 20 and D-type ORF100 (Fig. 3D).

In this study, we focused large deletions of DNA fragment found in the intron 20 of single-copy nuclear *PolA1* gene and in ORF100 region on plastid DNA. Because a large insertion/deletion of DNA fragment is generally considered as irreversible event, the direction of event could be estimated either insertion or deletion if related species could be analyzed. The S-type intron 20 (0.14 kb) of *PolA1* gene was found in *O. sativa* and *O. rufipogon* while *O. longistaminata* and *O. meridionalis* contained L-type intron 20 (1.5 kb) (Htut et al. submitted). Because *Oryza* species containing other than AA-genome contained long intron 20 (1.1 – 2.6 kb) of *PolA1* gene, the S-type intron 20 was probably derived from L-type intron 20, where a 1.4 kb DNA fragment was deleted between two tandem TTTTGC repeats by the intramolecular homologous recombination. (Htut et al. submitted). And because AA-genome species except annual *O. rufipogon* and *indica* contained ND-type ORF100 of plastid DNA (Takahashi et al. 2009), D-type ORF100 found in annual *O. rufipogon* and *indica* was probably originated from ND-type ORF100, of which a 69-bp sequence was deleted between two tandem TTACTTTTTTCA repeats (Kanno et al. 1993).

We first identified MD strain as *O. longistaminata*-like cryptic species because anther length of MD1 strain was comparable to that of African *O. longistaminata* and twice longer than that of *O. rufipogon* (Fig. 5A, B). The L-type intron 20 sequence of MD1 strain was closely related to that of *O. longistaminata* (Fig. 6A). And MD1 strain and *O. longistaminata* shared the same Ptag-coding sequence which was different to that of *O. rufipogon* and other AA-genome species (Fig. 6B). Although phenotypic characters of MD1 strain should be compared to those of *O. rufipogon* and *O. longistaminata* under tropical condition (Fig. 4), these results obtained in this study clearly indicated that MD strain was believed to be Asian-type *O. longistaminata*.

Interestingly, the MD1 strain contained D-type ORF100 on plastid DNA while African *O. longistaminata* showed ND-type ORF100 (Fig. 6C). This result indicated that MD strain was probably maternal parent of five Dong Thap strains (Fig. 3D) and maternal ancestor for annual *O. rufipogon* and *indica* cultivars (Chen et al. 1993). Perennial *O. rufipogon* with ND-type ORF100 probably provided its plastid DNA to Ac5 (Dong Thap) and Ac7 (Can Tho) strains (Fig. 3D).

Analysis of public data from the 3000 rice genome project in IRRI revealed that Ptag sequences of 2858 rice cultivars were classified into two major type: *japonica*-type^{204D-257L-375G} (57%) and *indica*-type^{204G-257L-375E} (38%) (Fig. 6D). Although classification based on Ptag sequence did not exactly match to that based on phenotypic characteristics, a relatively strong reproductive barrier was present between cultivars containing *japonica*-type and *indica*-type Ptag sequence because frequency of recombinant-type

²⁰⁴G-²⁵⁷L-³⁷⁵G was very low (5%). In *O. rufipogon*, most perennial strains contained *japonica*-type Ptag sequence whereas most annual strains had *indica*-type Ptag sequence (data not shown).

The finding of MD strain as *O. longistaminata*-like cryptic species provided us a new idea for the origin of *O. rufipogon indica* ancestor (Fig. 7). MD strain hybridized with perennial *O. rufipogon* to produce F₁ hybrid like Ac1 strain (Table 1). The F₁ hybrid was again backcrossed with perennial *O. rufipogon* to produce perennial BC₁ hybrid like Ac2 strain. Although most of BC₁ hybrids might contain homologous S-type intron 20, D-type ORF100, and homologous Ptag sequences (²⁰⁴D-²⁵⁷L-³⁷⁵G), a few BC₁ hybrids might have homologous S-type intron 20, D-type ORF100, and heterologous Ptag sequences that ²⁰⁴D-²⁵⁷S-³⁷⁵E and ²⁰⁴D-²⁵⁷L-³⁷⁵G were respectively derived from MD strain and perennial *O. rufipogon*. Then, offspring of the BC₁ hybrid might evolve to *O. rufipogon indica* ancestor under the homoploid hybrid speciation, where life cycle might be changed from perennial to annual. And the heterologous Ptag sequences might be changed to *indica*-type ²⁰⁴G-²⁵⁷L-³⁷⁵E by means of gene conversion because ²⁵⁷L and ³⁷⁵E were probably derived from perennial *O. rufipogon* and MD strain, respectively.

As annual *O. rufipogon* and *indica* cultivars shared D-type ORF100, *O. rufipogon indica* ancestor was probably annual because change of plant life cycle from perennial to annual was also happened during speciation from *O. longistaminata* to *O. barthii* (Zhu and Ge 2005). And the domestication from perennial *O. rufipogon* to *japonica* cultivar was occurred as supported by many researches (Oka 1988, Yamanaka et al. 2003, Second 1982, Cheng et al. 2003, Londo et al. 2006). But the direct domestication from *O. rufipogon indica* ancestor to *indica* cultivar was obscure. Probably *indica* cultivar was originated from the hybridization between *O. rufipogon indica* ancestor and *japonica* cultivar. Subsequent hybridizations between *indica* and *japonica* cultivars produced different five ecotypes of rice cultivars (Huang et al. 2012, Choi et al. 2017).

Genome wide association studies (GWAS) using many strains of *O. sativa* and *O. rufipogon* revealed that two types of *O. rufipogon* were involved in the origin of *indica* and *japonica* cultivars, respectively (Huang et al. 2012, Choi et al. 2017). But any reports using GWAS could not predict the presence and involvement of *O. longistaminata*-like cryptic species. Therefore, it should be afraid of the delusion that the true speciation of rice and other plants could be revealed by using only GWAS.

The wild relatives of cultivated rice had a great potential for crop improvement due to their wide genetic diversity (Jena and Khush 1990). It was necessary to understand wild relatives on phylogenetic relationships to assess inter- and intraspecific genetic diversity for effective use of desirable gene pool (Aggarwal et al. 1999). Because MD strain (*O. longistaminata*-like cryptic species) found in this study could hybridize with *indica* and *japonica* cultivars (Dr. K. Ichitani personal communication), it will serve as genetic resources to improve various agronomic traits, such as deep-water resistance and acid sulfate tolerance (Khush and Virk 2005, Can and Lang 2007, Lang and Buu 2011). Also, the MD strain will be useful to reveal the mechanism underlying the origin of *indica* cultivar by reproducing *O. rufipogon indica* ancestor as a model case of the hybrid speciation.

Conclusion

In this study, we identified MD strain as *O. longistaminata*-like cryptic species found in the Mekong Delta of Vietnam. The MD strain had long anthers of spikelet as well as African *O. longistaminata*. Molecular phylogenetic analysis also showed that MD strain was closely related to *O. longistaminata*. Because MD strain, unlike *O. longistaminata*, contained D-type ORF100 of plastid DNA specific to annual *O. rufipogon* strains and *indica* cultivars, we proposed a hypothesis that MD strain was involved, as maternal parent, in the origin of *O. rufipogon indica* ancestor and subsequently *indica*. Therefore, MD strain will contribute to resolve the mechanism underlying hybrid speciation of *O. rufipogonindica* ancestor and the domestication process of *indica* cultivar, as well as to improve agronomical traits of rice cultivars as genetic resources.

Materials And Methods

Plant materials

As listed in Table 1, six wild rice strains (Ac1 – Ac6) were collected at 2012 in the Tram Chim Sanctuary, National Park, in Dong Thap (Fig. 1). Ac7 strain shown in Fig. 2 was collected in a canal near CLRRRI at Can Tho, Vietnam. Two wild rice strains (MD1, MD2) from Dong Cat village were selected among 100 strains in the germplasm collection of CLRRRI. Three strains of *O. rufipogon* and two strains of *O. longistaminata* were obtained from the Gene Bank of National Institute Genetics, Shizuoka, Japan, and of the International Rice Research Institute (IRRI), Laguna, Philippines, respectively. Pollen viability of wild rice strain was tested by I₂KI staining.

DNA extraction

Total genomic DNA was extracted from 100 mg leaf materials in 2-ml plastic tubes that were frozen with liquid nitrogen and crushed into fine powder using Multi-Bead Shocker (Yasui Kikai Co., Japan). The CTAB method (Doyle and Doyle 1987) was used for DNA extraction.

PCR amplification and direct sequencing

As shown in Fig. 3A, DNA fragments containing Ptag-coding sequence and intron 20 (S-type and L-type) were amplified by PCR using various pairs of nine primers (Table 2). PCR products containing ORF100 region of plastid DNA were amplified using a pair of primers 9 and 10 (Fig. 3B). The primers were designed based on the sequence of rice *PolA1* mRNA (LOC9270399) and of whole plastid genome (X15901, Hiratsuka et al. 1989) on *japonica* rice 'Nipponbare'. Sequencing primer 6 was designed based on sequence result of the intron 20 in this study.

The PCR amplification was carried out *ExTaq* DNA polymerase (Takara, Shiga, Japan) according to manufacturer's instruction. The PCR condition were 40 cycles of 94°C for 1 min, 58 °C for 1 min for annealing, and 72 °C for 2 min for elongation in a PTC200 thermocycler (MJ Research, Waltham, MA,

USA). The amplified PCR products were subjected to 1.5% agarose gel electrophoresis and purified using a PCR purification kit (QIAquick; Qiagen, CA, USA). The purified PCR products of Ptag-coding sequence (1230 bp) and ORF100 (657 bp) were determined by direct sequencing with the same primer as used for PCR amplification in an automated DNA sequencer ABI310 (Applied Biosystems, CA, USA) with a Big Dye Terminator Cycle Sequencing kit (Applied Biosystems, USA).

The intron 20 and Ptag-coding sequences of *PoIA1* gene used for phylogenetic analysis were determined in *Oryza* AA-genome species (Htut et al. submitted).

Phylogenetic analysis

The intron 20 and Ptag-coding sequences of *PoIA1* gene were analyzed using the NCBI web-based Blast server (Altschul et al. 1990) and aligned by using CLUSTAW (Thompson et al. 1994). The alignment was then manually adjusted using Genetyx Software ver. 6.0, Software Development Co., Tokyo, Japan. The phylogenetic tree of the intron 20 and Ptag-coding sequences were constructed using Neighbor-Joining method with bootstrap estimate from 1,000 replicates in the MEGA7 software (Kumar et al. 2017). *O. officinalis* was used as an out-group. Ptag sequences of 2858 rice cultivars were obtained from web site (<http://iric.irri.org/resources/3000-genomes-project>) of the 3000 rice genome project of IRRI.

Abbreviations

CTAB: cetyltrimethylammonium bromide, ²⁰⁴*G*: Glycine at the position 204, *MEGA7*: Molecular Evolutionary Genetics Analysis version 7, *PCR*: polymerase chain reaction

Declarations

Acknowledgments

Wild rice strains in the genus *Oryza* were provided from the National Institute of Genetics, Mishima, Japan

Authors' Contributions

DTL and BCB managed collection of perennial wild rice. BCB, NTL, RI, and IN surveyed natural populations. HAH, PAS, HT and SM contributed to phenotypic and molecular analyses. HAH and IN wrote manuscript

Funding

This study was partially supported by a Grant-in-aid B (No. 16H05777) to R.I.

Availability of Data and Materials

The datasets supporting the conclusions of this article are included within the article. The nucleotide sequences of the intron 20 and Ptag-coding sequences and ORF100 sequences have been deposited to DNA Data Bank of Japan under the accession numbers of LC638405 to LC638446.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors have no conflicts of interest to declare.

Author details

1. Graduate School of Horticulture, Chiba University, Matsudo, Chiba 271-8510, Japan. 2. BEX Co Ltd., Itabashi-ku, Tokyo 173-0004, Japan. 3. Institute of Agricultural Science for Southern Vietnam, Ho Chi Minh City, Vietnam. 4. High Agricultural Technology Research Institute, Cantho, Vietnam. 5. Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki, Japan.

References

- Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP makers. *Theor Appl Genet* 98: 1320–1328. <https://doi.org/10.1007/s001220051198>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Can NT, Lang NT (2007) Identification of some promising varieties for salinity soil and phosphorous deficient areas in the Mekong Delta. *Omonrice* 15: 179–184.
- Chen WB, Nakamura I, Sato Y-I, Nakai H (1993) Distribution of deletion type in cpDNA of cultivated and wild rice. *Jap J Genet* 68: 597–603.
- Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: Based on the interspersed pattern of *SINEs*. *Mol Biol Evol* 20: 67–75. <https://doi.org/10.1093/molbev/msg004>
- Choi JY, Platts AE, Fuller DQ, Hsing T-I, Wing RA, Purugganan, MD (2017) The rice paradox: multiple origins but single domestication in Asian rice. *Mol Biol Evol* 34: 969–979. <https://doi.org/10.1093/molbev/msx049>

- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11–15.
- Fareed A, Shindo H, Takahashi H, Nakamura I (2015) Analysis of *PolA1* intron 19 and Ntag sequences reveals two ancestral lineages in the origin of the *Brassica rapa* complex and Chinese cabbage (*B. rapa* var. *pekinensis*). *J Hort Sci Biotech* 90: 273–278. <https://doi.org/10.1080/14620316.2015.11513182>
- Fareed A, Shindo H, Takahashi H, Nakamura I (2016) Phylogeny of *PolA1* Gene Consistent with the Relationships of U's Triangle in *Brassica*. *Hort J* 85: 55–62. <https://doi.org/10.2503/hortj.MI-052>
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* 163: 1631–1638. <https://doi.org/10.1534/genetics.104.035642>
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto, Mori, M, Kondo C, Honji Y, Sun CR, Meng BY, et al (1989) The complete sequence of rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol Gen Genet* 217: 185–194.
- Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501. <https://doi.org/10.1038/nature11532>
- Huang X, Han B (2016) Rice domestication occurred through single origin and multiple introgressions. *Nature Plants* 2: 15207. <http://doi.org/10.1038/nplant.2015.207>
- Ishii T, Hiraoka T, Kanzaki T, Akimoto M, Shishido R, Ishikawa R (2011). Evaluation of genetic variation among wild populations and local varieties of rice. *Rice* 4: 170–177. <https://doi.org/10.1007/s12284-011-9067-x>
- Izawa T (2008) The process of rice domestication: A new model based on recent data. *Rice* 1: 127–134. <https://doi.org/10.1007/s12284-008-9014-7>
- Jena KK, Khush GS (1990) Introgression of genes from *Oryza officinalis* Well ex Watt to cultivated rice, *O. sativa* L. *Theor Appl Genet* 80: 737–745.
- Kanno A, Watanabe N, Nakamura I, Hirai A (1993) Variations in chloroplast DNA from rice (*Oryza sativa*): Differences between deletions mediated by short direct-repeat sequences within a single species. *Theor Appl Genet* 86: 579–584.
- Kato, S. (1930) On the affinity of the cultivated varieties of rice plant, *Oryza sativa* L. *J Dept Agric Kyushu Univ* 2: 242–276.
- Kawahara Y, Bastide MDL, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next

generation sequence and optical map data. *Rice* 6: 3–10. <https://doi.org/10.1186/1939-8433-6-4>

Khush GS, Virk PS (eds.) (2005). *IR varieties and their Impact* (p. 163). Los Banos, Philippines: International Rice Research Institute.

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trend Genet* 23: 578–587. <https://doi.org/10.1016/j.tig.2007.08.012>

Kovach MJ, McCouch SR (2008) Leveraging natural diversity: back through the bottleneck. *Curr Opin Plant Biol* 11: 193–200. <https://doi.org/10.1016/j.pbi.2007.12.006>

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Evol Genet Anal* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>

Lam DT, Buu BC, Lang NT, Toriyama K, Nakamura I, Ishikawa R (2019) Genetic diversity among perennial wild rice *Oryza rufipogon* Griff., in the Mekong Delta. *Ecol Evol* 9: 2964–2977. <https://doi.org/10.1002/ece3.4978>

Lang NT, and Buu BC (2011) Rice genetic resource conservation and utilization in the Mekong Delta. *Omonrice* 18: 22-23.

Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci U S A* 103: 9578–9583. <https://doi.org/10.1073/pnas.0603152103>

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan M D. (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci U S A* 108: 8351–8356. <https://doi.org/10.1073/pnas.1104686108>

Morishima H, Sano Y, Oka HI (1992) Evolutionary studies in cultivated rice and its wild relatives. *Oxf Surv Evol Biol* 8: 135–184.

Morishima H (1994) Rice in the Mekong Delta. Y-I sato (ed.), *Ecological-genetic studies on wild and cultivated rice in tropical asia* (4th survey), pp 211-215.

Nakamura I, Rai B, Takahashi H, Kato K, Sato Y-I, Komatsuda T (2009) *Aegilops* section *Sitopsis* species contains the introgressive *PolA1* gene with a closer relationship to that of *Hordeum* than *Triticum*–*Aegilops* species. *Breed Sci* 59: 602–610. <https://doi.org/10.1270/jsbbs.59.602>

Nakamura I (2010) Method of identifying eukaryotic species. JP 2010088398-A/274, 22 Apr 2010.

Oka H-I (1988) *Origin of Cultivated Rice*. Tokyo: Japan Scientific Societies Press.

- Rai B, Takahashi H, Kato K, Sato Y-I, Nakamura I (2012) Single-copy nuclear *PolA1* gene sheds light on the origin of *S* genome with relationships to *B* and *G* genomes of polyploid wheat species. *Genet Resour Crop Evol* 59: 1713–1726. <https://doi.org/10.1007/s10722-012-9793-y>
- Russell J, Zomerdijs, Joost CBM (2006) *The RNA polymerase I transcription machinery. Biochem Soc Symp* 73: 203–16. <https://doi.org/10.1042/bss0730203>
- Sang T, Ge S (2007) **The puzzle of rice domestication.** *J Integr Plant Biol* 49: 760–768. <https://doi.org/10.1111/j.1744-7909.2007.00510.x>
- Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jap J Genet* 57: 25–57.
- Sweeney G, McCouch SR (2007) The complex history of the domestication of rice. *Ann. Bot* 100: 951–957. <https://doi.org/10.1093/aob/mcm128>
- Takahashi H, Sato Y-I, Nakamura I (2008) Evolutionary analysis of two plastid DNA sequences in cultivated and wild species of *Oryza*. *Breed Sci* 58: 225–233. <https://doi.org/10.1270/jsbbs.58.225>
- Takahashi H, Sato T, Sato Y-I, Nakamura I (2009) Genome-type-specific variation of the 19th intron sequence within the RNA polymerase I largest subunit gene in the genus *Oryza*. *Plant Syst Evol* 282: 21–29. <https://doi.org/10.1007/s00606-009-0172-x>
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acid Res* 22: 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- Ueno K, Sato T, Takahashi N (1990) The *indica-japonica* classification of Asian rice ecotypes and Japanese lowland and upland rice (*Oryza sativa* L.). *Euphytica* 46: 161–164.
- Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. *Curr Opin Plant Biol* 6: 139–146. [https://doi.org/10.1016/S1369-5266\(03\)00009-8](https://doi.org/10.1016/S1369-5266(03)00009-8)
- Vaughan DA, Lu B-R, Tomooka N (2008) The evolving story of rice evolution. *Plant Sci* 174: 394–408. <https://doi.org/10.1016/j.plantsci.2008.01.016>
- Yamanaka S, Nakamura I, Nakai H, Sato Y-I (2003) Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. *Genet Resour Crop Evol* 50: 529–538. <https://doi.org/10.1023/A:1023926802198>
- Yamanishi C, Alshahni MM, Sano A, Nakamura I, Makimura K (2017) A new marker sequence for systematics of medically important fungi based on amino acid sequence of the largest subunit of RNA polymerase I. *Med Mycol* 55: 555–562. <https://doi.org/10.1093/mmy/myw098>

Zhang X, Takahashi H, Nakamura I Mii M (2008) Molecular discrimination among taxa of *Petunia axillaris* complex and *P. integrifolia* complex based on *PolA1* sequence analysis. Breed Sci 58: 71–75. <https://doi.org/10.1270/jsbbs.58.71>

Zhu Q, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol 167: 249– 265. <https://doi.org/10.1111/j.1469-8137.2005.01406.x>

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24: 875–888. <https://doi.org/10.1093/molbev/msm005>

Tables

Table 1 Wild rice strains used in this study

Strain	Origin	PI20 ¹⁾	ORF100 ²⁾	A/P ³⁾
MD1	Dong Cat (CLRRI) ⁴⁾	LL	D	P
MD2	Dong Cat (CLRRI)	LL	D	P
Ac1	Dong Thap (Tram Chim Sanctuary)	SL	D	P
Ac2	Dong Thap (Tram Chim Sanctuary)	SS	D	P
Ac3	Dong Thap (Tram Chim Sanctuary)	SL	D	P
Ac4	Dong Thap (Tram Chim Sanctuary)	SL	D	P
Ac5	Dong Thap (Tram Chim Sanctuary)	SL	ND	P
Ac6	Dong Thap (Tram Chim Sanctuary)	SS	D	P
Ac7	Can Tho (Canal near CLRRI)	SS	ND	P
<i>O. sativa</i>	'Nipponbare'	SS	ND	P
<i>O. rufipogon</i>	W0137 (NIG) ⁵⁾	SS	D	A
<i>O. rufipogon</i>	W0130 (NIG)	SS	ND	P
<i>O. rufipogon</i>	W1244 (NIG)	SS	D	A
<i>O. longistaminata</i>	IRGC101198 (IRRI) ⁶⁾	LL	ND	P
<i>O. longistaminata</i>	IRGC101205 (IRRI)	LL	ND	P

1) L-type (ca. 1.5 kb) or S-type (0.14 kb) intron 20 (PI20) of *PolA1* gene.

- 2) 69-bp deletion (D) or non-deletion (ND) in ORF100 of plastid DNA.
- 3) Annual (A) or perennial (P).
- 4) Germplasm Collection of Cuu Long Delta Rice Research Institute (CLRRI).
- 5) National BioResources Project, National Institute of Genetics (NIG), Japan.
- 6) International Rice Gnebank, International Rice Research Institute (IRRI).

Table 2 Primers used in this study

Name	Sequence
1	5'-CTCGCTGGACGGGGTGAGATGAATG-3'
2	5'-CCTTGAGAACTGTTTTTATTGATG-3'
3	5'-AATGAATGAAGCTGAGGATGAAAA-3'
4	5'-AAAACAGTTCTCAAGGTTTCATGG-3'
5	5'-CATAATCCATCTCATCTGTTTCTTG-3'*
6	5'-ACAACCTTCTCCACCAACATTCTCT-3'
7	5'-CTTACAGGCCTTGACAAAAACAGA-3'
8	5'-TGAAATCCGCAATCAAGTTCAGATG-3'
9	5'-GCCGCTTTAGTCCACTCAGCCATC-3'
10	5'-TCAATGCCTTTTTTCAATGGTCTC-3'

* this primer was designed based on the sequence of *Arabidopsis PolA1* gene for the initial study. Positions of primers were shown in Fig. 3A, B.

Figures

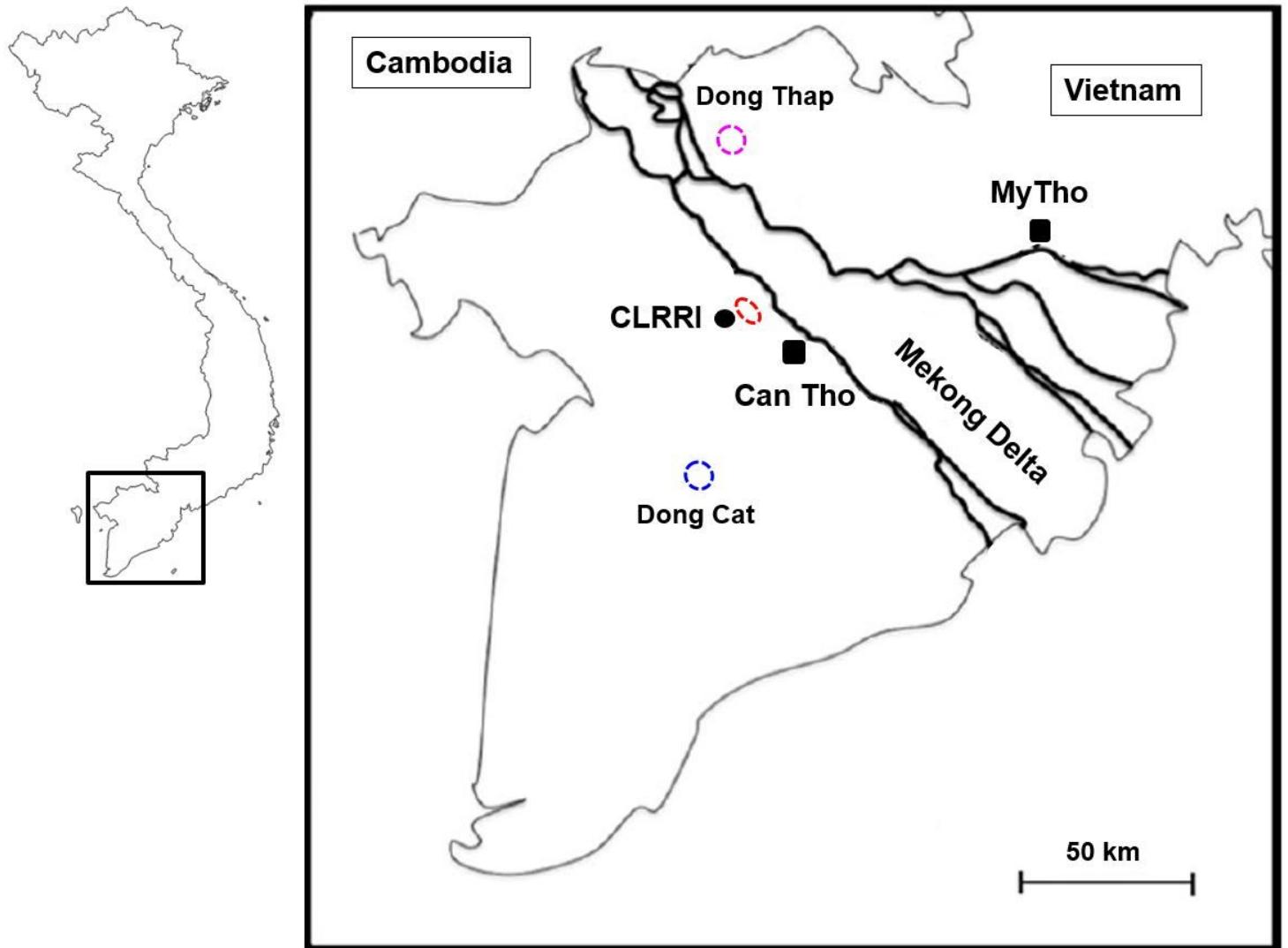


Figure 1

Wild rice survey during 2010 – 2015 in the Mekong Delta. Wild rice strains were collected in the Tram Chim Sanctuary of Dong Thap (Ac1–Ac6) and in a canal near CLRRRI of Can Tho (Ac7). Two wild rice strains (MD1, MD2) from Dong Cat village were selected from the wild rice collection in CLRRRI.

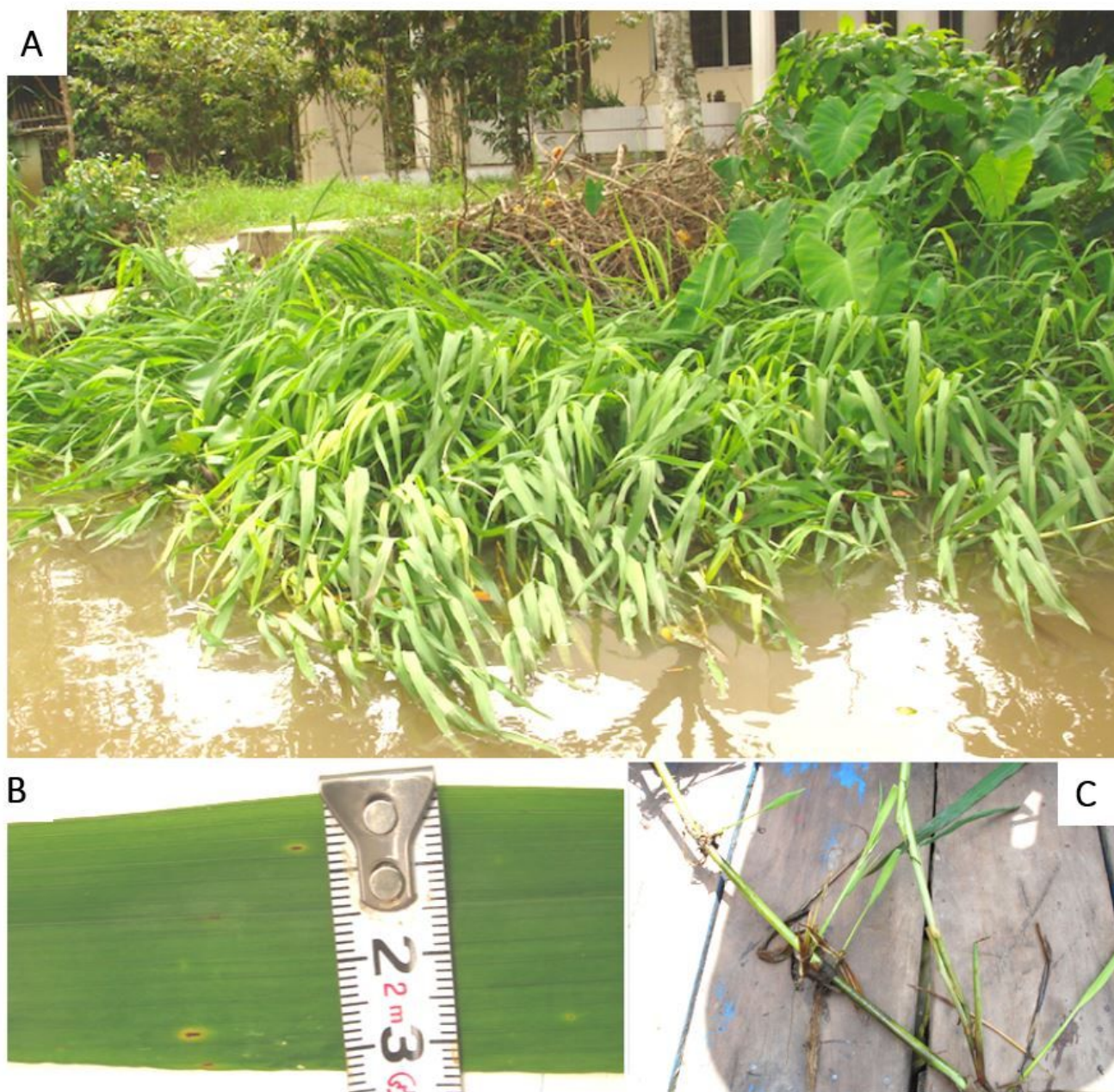


Figure 2

Wild rice strain (Ac7) found in a canal near CLRRRI. A: overview of Ac7 strain. B: Leaf blade of Ac7 strain. C: Ac7 strain was clonally propagated by shooting from nodes.

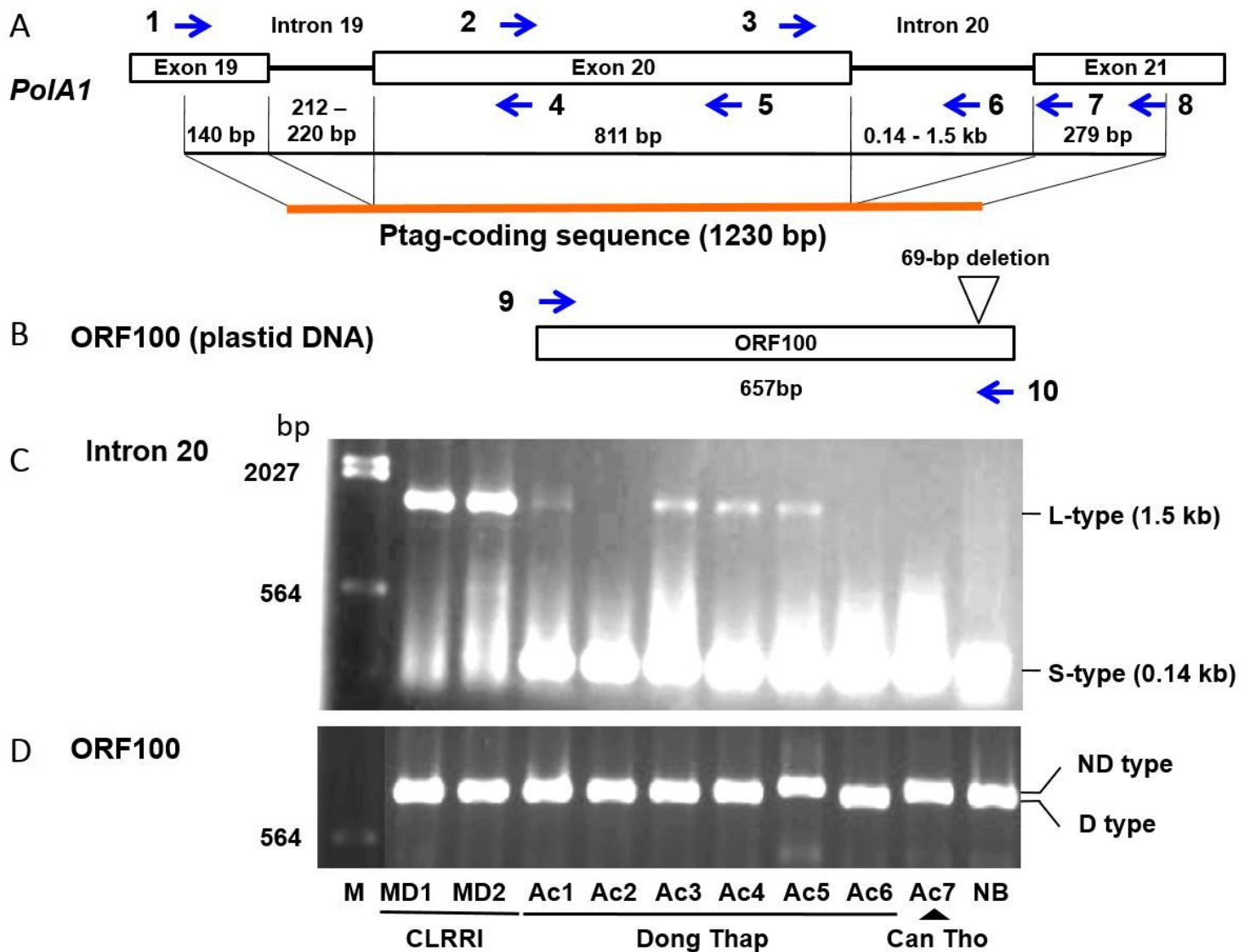


Figure 3

Analysis of intron 20 of *PolA1* gene and ORF100 on plastid DNA in wild rice strains. A: DNA fragment containing the intron 20 was amplified by PCR using a pair of primers 3 and 7, listed in Table 1. The intron 20 (ca. 0.14–1.5 kb) and Ptag-coding sequence (1230 bp) were analyzed by direct sequencing using various pairs of 8 primers. B: DNA fragment containing ORF100 (657 bp) was amplified and sequenced using a pair of primers 9 and 10. C: PCR amplification of the intron 20 in two strains (MD1, MD2) from CLRRI, six strains (Ac1–Ac6) from Dong Thap, and one strain (Ac7) from Can Tho. NB: *O. sativa* 'Nipponbare'. D: PCR amplification of ORF100 using the same samples above.

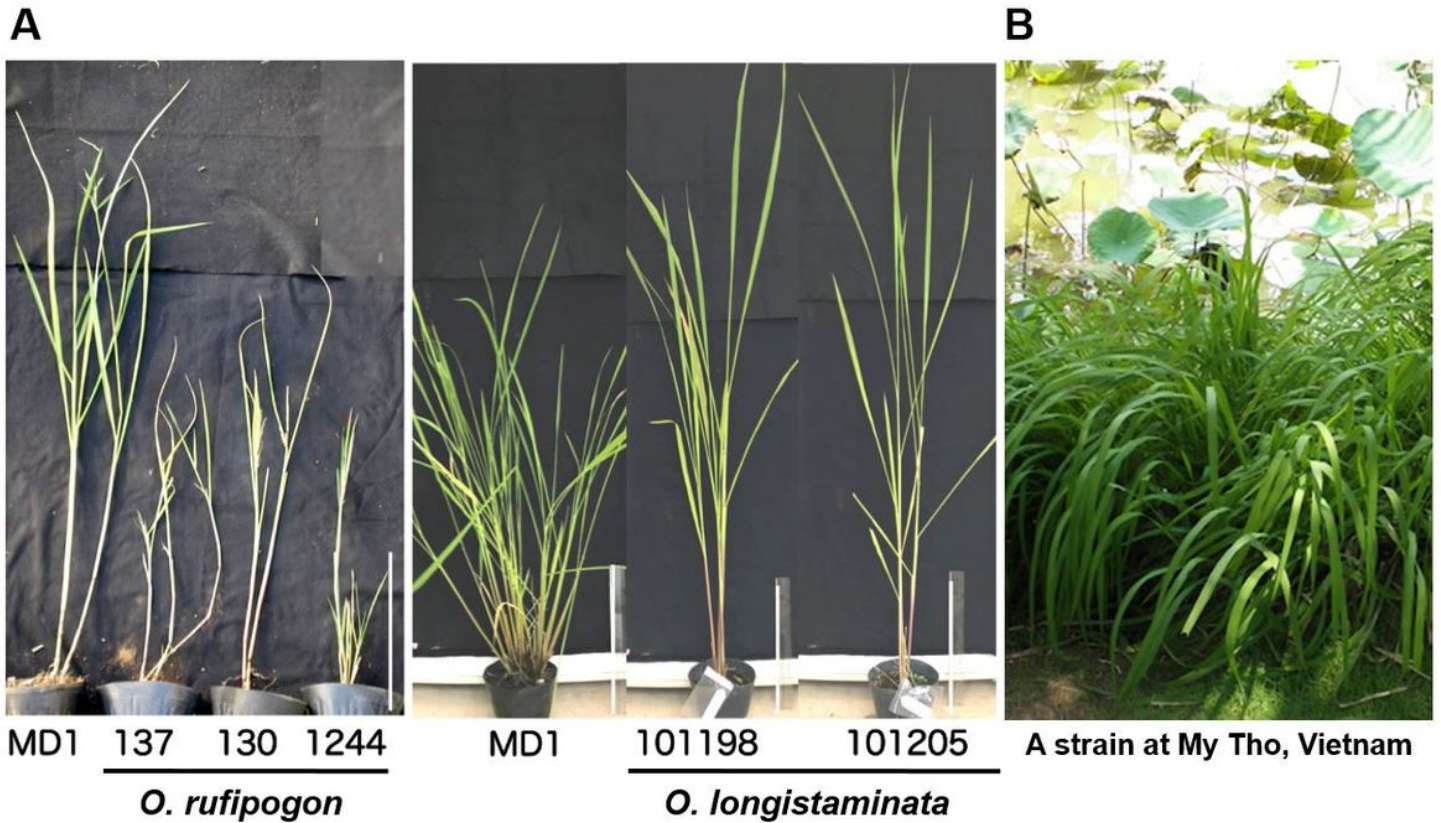


Figure 4

Identified MD1 strain containing homologous L-type intron 20 of PolA1 gene. A: MD1 strain was grown together with three strains (W0137, W0130, W1244) of *O. rufipogon* or two strains (IRGC101198, IRGC101205) of *O. longistaminata*. B: A wild rice strain, not MD strain, having the homologous L-type intron 20 found in My Tho, Vietnam (Fig. 1).

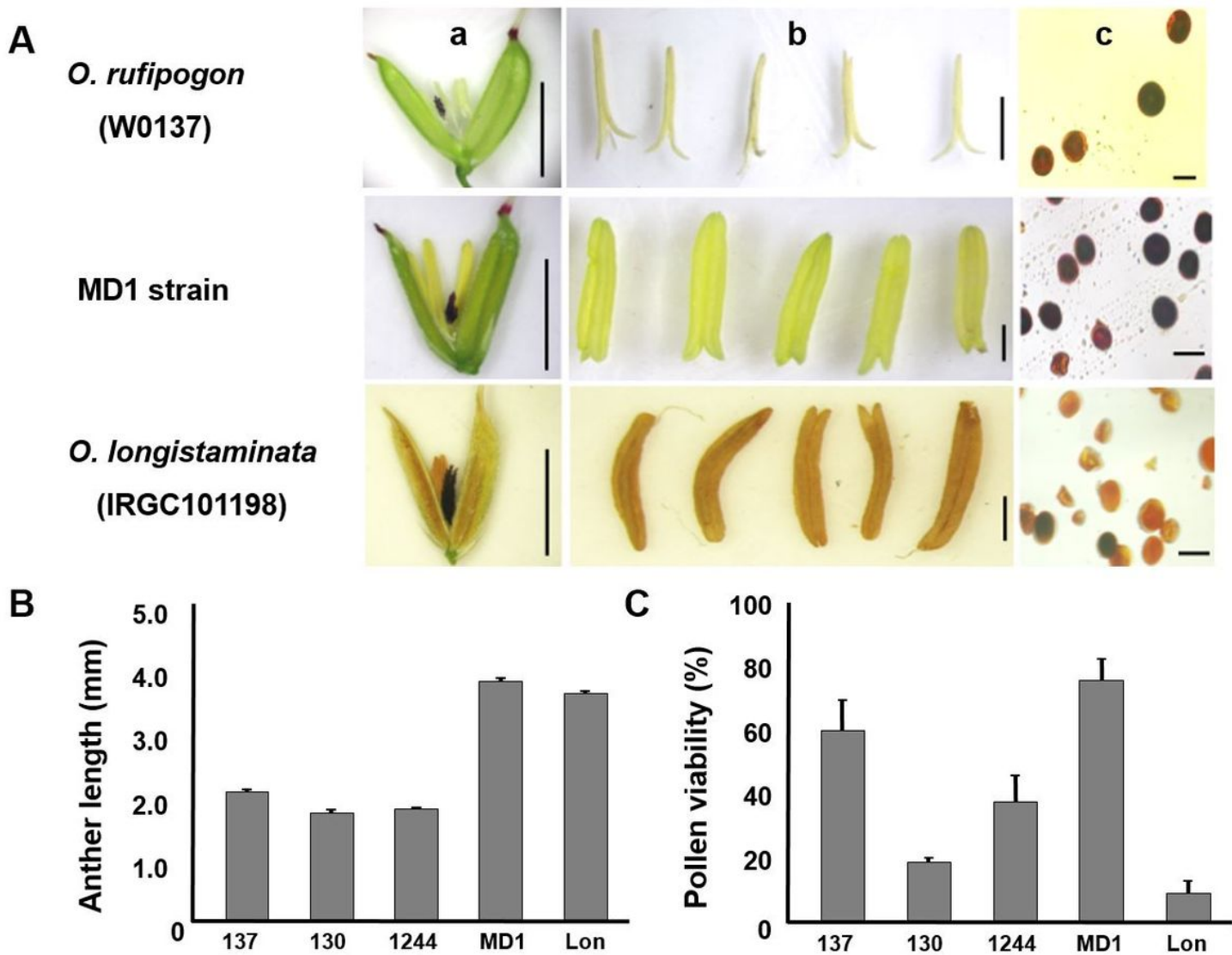


Figure 5

Flower characteristics of MD1 strain. A: floret (a, scal bar: 5 mm), anther (b, scal bar: 1 mm), and pollen viability (c, scal bar: 50 μ m) were compared among MD1 strain, *O. rufipogon* (W0137), and *O. longistaminata* (IRGC101198). B: Anther length of MD1 strain, *O. rufipogon* three strains (W0137, W0130, W1244), and *O. longistaminata* (Lon: IRGC101198). C: Pollen viability of the same samples above using I2KI staining.

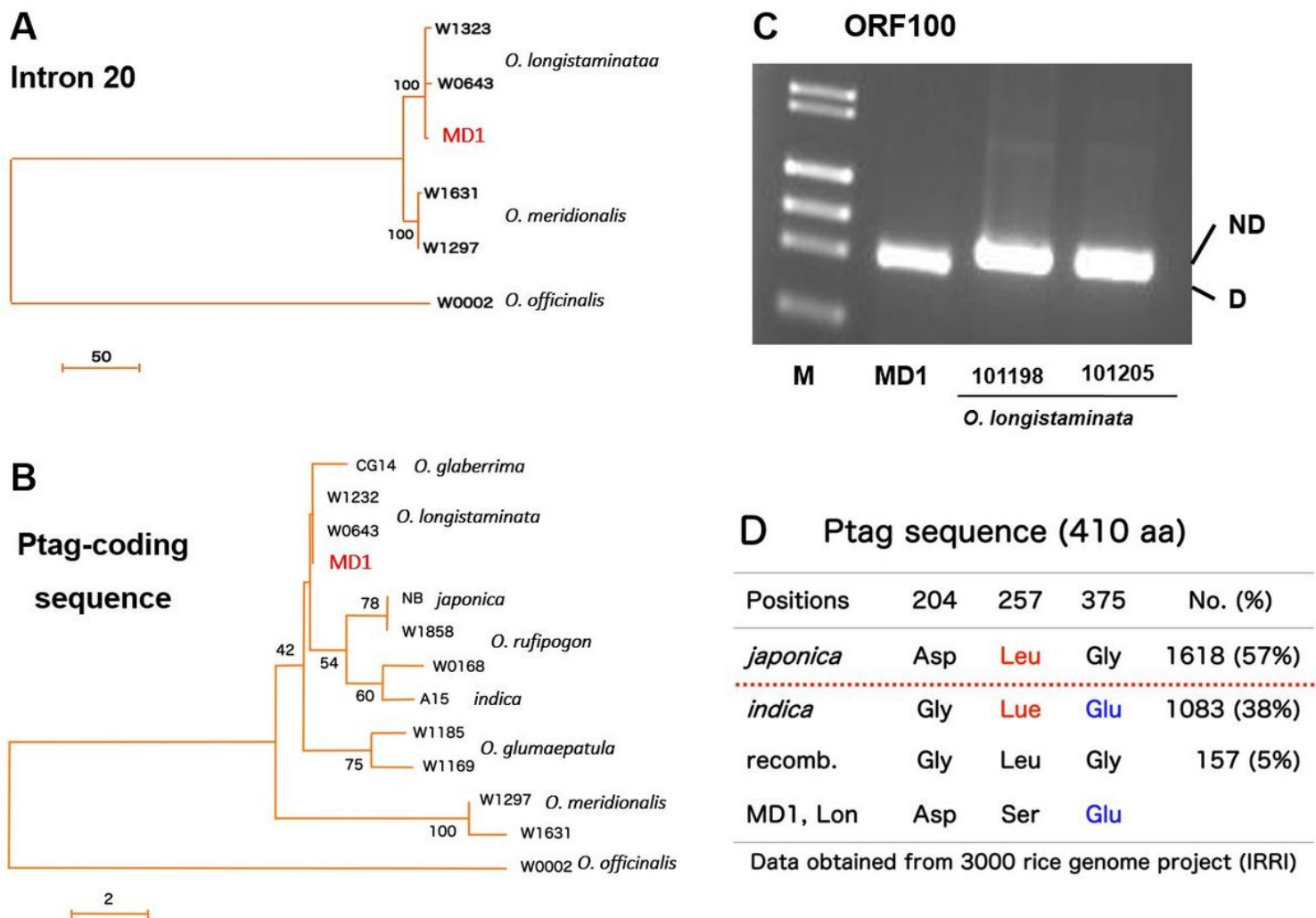
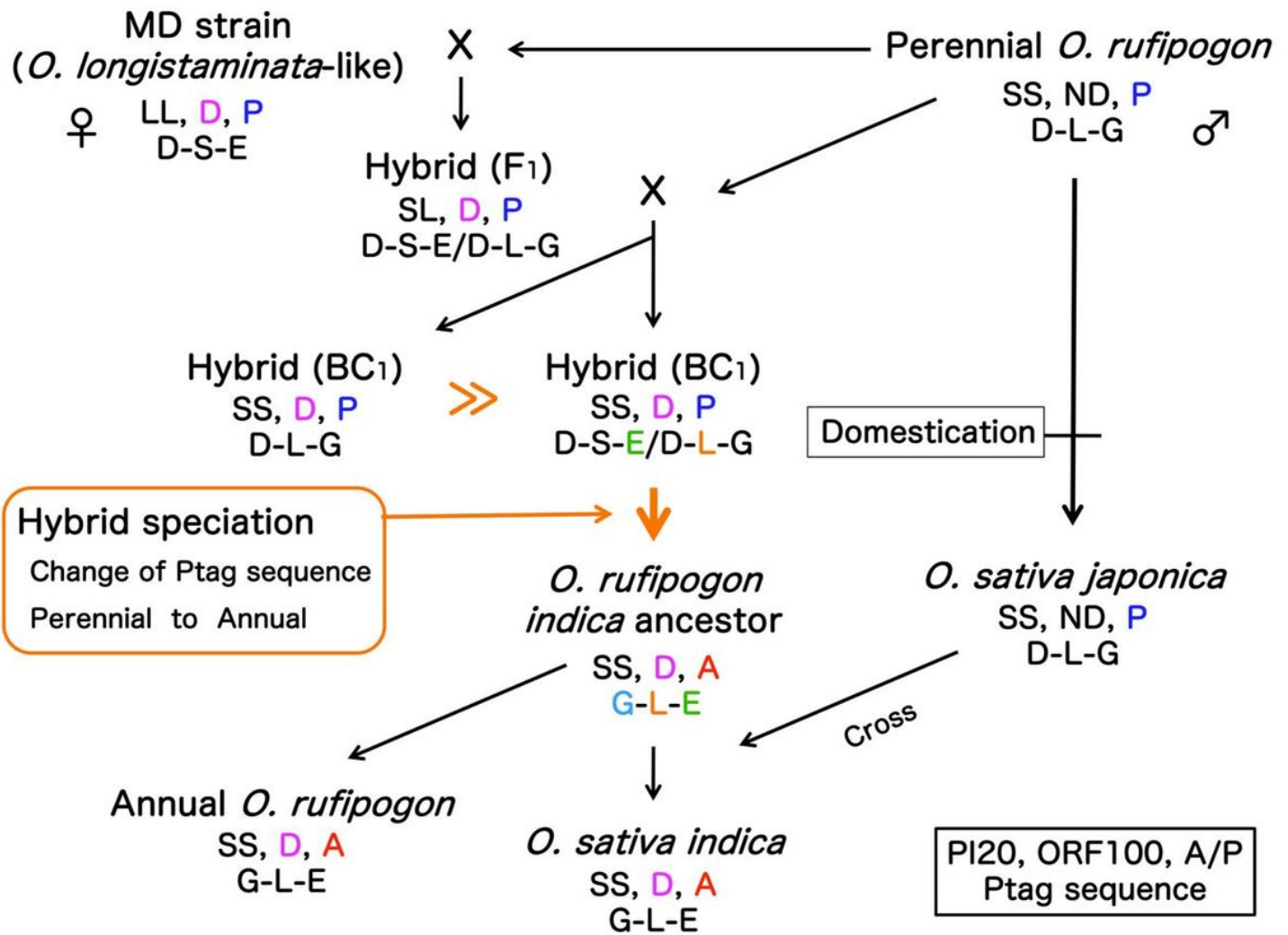


Figure 6

Phylogenetic analyses of MD1 strain. A: Neighbor-Joining (N-J) phylogenetic tree based on the L-type intron 20 sequences of PolA1 gene. *O. officinalis* was used as an out-group. B: N-J phylogenetic tree of Ptag-coding sequences. MD1 strain and *O. longistaminata* contained the same Ptag-coding sequence. C: PCR products containing ORF100 in MD1 and two strains (IRGC101198, IRGC101205) of *O. longistaminata*. D: Ptag sequences (410 aa) of 2858 rice cultivars obtained from the 3000 rice genome project (IRRI) were classified into two major types, *japonica* type (1618, 57%) and *indica* type (1083, 38%) based on the combination of three amino acids (positions 204, 257, 375). Frequency of recombinant type between *japonica* and *indica* was very low (157, 5%). The same Ptag sequence was shared with MD1 strain and *O. longistaminata* (Lon).



←/→

Figure 7

Hypothesis on the origin of *O. rufipogon indica* ancestor through the hybrid speciation. MD1 strain and perennial *O. rufipogon* might be involved in the origin of *O. rufipogon indica* ancestor as maternal and paternal lineage, respectively. PI: S-type or L-type intron 20, ORF100: ND-type or D-type of ORF100, A/P: annual (A) or perennial (P), and combinations of three amino acids of Ptag sequences were shown. See the text for details.