



# Reticulate Evolution in AA-Genome Wild Rice in Australia

Sharmin Hasan<sup>1,2</sup>, Agnelo Furtado<sup>1</sup> and Robert Henry<sup>1\*</sup>

<sup>1</sup> Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia,

<sup>2</sup> Department of Botany, Jagannath University, Dhaka, Bangladesh

## OPEN ACCESS

### Edited by:

Yohei Koide,  
Hokkaido University, Japan

### Reviewed by:

Mulatu Geleta,  
Swedish University of Agricultural  
Sciences, Sweden  
Peng Qi,  
University of Georgia, United States

### \*Correspondence:

Robert Henry  
robert.henry@uq.edu.au

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 31 August 2021

**Accepted:** 04 February 2022

**Published:** 11 March 2022

### Citation:

Hasan S, Furtado A and Henry R  
(2022) Reticulate Evolution  
in AA-Genome Wild Rice in Australia.  
*Front. Plant Sci.* 13:767635.  
doi: 10.3389/fpls.2022.767635

The wild rice gene pool, i.e., AA-genome, in Australia is geographically and genetically distinct from that in Asia. Two distinct taxa are found growing together in northern Australia, *Oryza meridionalis* (including annual and perennial forms) and an *Oryza rufipogon* like taxa that have been shown to have a chloroplast genome sequence that is closer to that of *O. meridionalis* than to *O. rufipogon* from Asia. Rare plants of intermediate morphology have been observed in the wild despite a reported reproductive barrier between these two species. We now report the resequencing of plants from 26 populations including both taxa and putative hybrids. A comparison of chloroplast and nuclear genome sequences indicated re-combinations that demonstrated hybridisation in both directions. Individuals with intermediate morphology had high nuclear genome heterozygosity consistent with a hybrid origin. An examination of specific genes (e.g., starch biosynthesis genes) revealed the presence of heterozygotes with alleles from both parents suggesting that some wild plants were early generation hybrids. These plants may have low cross-fertility preserving the continuation of the two distinct species. Repeated backcrossing of these rare hybrids to one parent would explain the plants exhibiting chloroplast capture. These observations suggest that reticulate evolution is continuing in wild *Oryza* populations and may have been a key process in rice evolution and domestication.

**Keywords:** hybridisation, introgression, chloroplast capture, reproductive barrier, reticulate evolution

## INTRODUCTION

Hybridisation and introgression undoubtedly play a crucial role in evolution allowing gene exchange between species resulting in genomic diversification and novel genetic combinations (Stebbins, 1959). Cytonuclear non-concordance, i.e., incongruence between organellar (plastid and mitochondrial) and nuclear genome phylogenies appears to reflect both hybridisation and successive introgression (Rieseberg, 1991; Nge et al., 2021). Recurrent hybridisation results in contrasting gene trees in two ways: (1) if recurrent unidirectional cytoplasmic gene flow from one species can invade another species in the absence of nuclear gene flow (Rieseberg and Wendel, 1993), and (2) if unidirectional nuclear gene flow but not cytoplasmic gene flow is allowed to invade a species (Petit et al., 2001). However, cytoplasmic genome invasion occurs in plants more frequently

than nuclear genome introgression (Rieseberg and Wendel, 1993). Cytoplasmic gene flow results in chloroplast capture, which may occur due to partial cytoplasmic male sterility, nuclear genome incompatibilities, or partial selfing in species (Tsitrone et al., 2003). Contrasting gene trees can be attributed to sampling error, evolutionary convergence, evolutionary rate heterogeneity, lineage sorting, and reticulation (Rieseberg and Soltis, 1991). Chloroplast capture has been used to infer reticulate evolution in many plant species such as *Pisum* (Palmer et al., 1985), Australian cotton (Wendel et al., 1991), sunflower (Rieseberg, 1991), peonies (Sang et al., 1995), *Opuntia* (Griffith, 2003), soybeans (Doyle et al., 2003); kiwifruit (Chat et al., 2004), the wheat tribe (*Elymus*) (Mason-Gamer, 2004); baobabs (Karimi et al., 2019); and *Adenanthos* (Nge et al., 2021). Much effort has been put into re-constructing reticulate evolution by means of investigating putative hybrid origins through multiple molecular techniques (Vriesendorp and Bakker, 2005). Morphological features along with geographical, ecological, biosystematics, and cytological data have been used to infer hybrid origin. However, genetic evidence provides the greatest power of inference in detecting putative interspecific hybrid origin (Griffith, 2003). Genetic evidence can be derived through implying an additivity of molecular markers (Wendel et al., 1991; Griffith, 2003), polymorphic nucleotide additivity at a single position, i.e., ITS additivity (Sang et al., 1995), and incongruence between gene trees (Mason-Gamer, 2004). Next-generation sequencing (NGS) approaches appear to be increasingly effective in addressing the causes of cytonuclear incongruence (Lemmon et al., 2012; Weitemier et al., 2014).

Rice (the genus *Oryza*) belongs to the tribe Oryzaceae—a member of the grass subfamily Ehrhartoideae. The genus *Oryza* comprises 11 genome groups of which the AA-genome group (with eight diploid species) is the most diverged group, distributed throughout the world, except for, Antarctica (Vaughan, 1989). The other genome groups are BB (one *Oryza* species), BBCC (three *Oryza* species), CC (three *Oryza* species), CCDD (three *Oryza* species), EE (one *Oryza* species), FF (one *Oryza* species), GG (three *Oryza* species), HHJJ (two *Oryza* species), HHKK (one *Oryza* species), and KKLL (one *Oryza* species) (Brar and Khush, 2018). Vicariance events and long-distance dispersal explain the evolution and divergence of AA genomes globally (Tang et al., 2010). The estimated mean divergence age of the AA genome from other *Oryza* genomes is 2.41 million years ago (Stein et al., 2018). Four *Oryza* species have been recognised in Australia (Henry et al., 2010). Of these, two wild *Oryza* species belong to the AA-genome clade: *Oryza meridionalis* Ng. and *Oryza rufipogon* Griff. The former was originally characterised by annual life history, self-pollination, and short anther length (Ng et al., 1981). More recently, the *O. meridionalis* gene pool has been shown to include both annual and perennial forms (Sotowa et al., 2013; Henry, 2019). *O. rufipogon* is considered a native perennial in Northern Queensland, Australia (Henry et al., 2010). Australian *O. rufipogon* populations have been shown to have a chloroplast genome that is closer to *O. meridionalis* than to *O. rufipogon* from Asia (Brozynska et al., 2014). This indicates that the Australian

population might need to be considered as separate taxon (Brozynska et al., 2017).

Reproductive barriers appear to be common in the AA genome *Oryza* species, especially in crosses between *O. sativa* and its wild relatives (including *O. meridionalis*), resulting in hybrid sterility (Li et al., 2007). Hybridisation between *O. meridionalis* and Asian wild rice produced low fertility hybrids (Naredo et al., 1997). Isolation due to pollen sterility was identified in *O. sativa* L. × *O. meridionalis* hybrids, resulting in partial and full abortion of male gametes (Li et al., 2018). Post-zygotic reproductive isolation happens to abort seeds in first filial generation (F1) hybrid offspring derived from hybridisation between *O. meridionalis*, and *O. sativa* (Toyomoto et al., 2019). Furthermore, geographic distributions and differences in flowering times influence intra- and interspecific crossability of *O. meridionalis* in Australia (Juliano et al., 2005).

Cytonuclear discordant in phylogenetic trees (i.e., chloroplast capture) has been observed in these wild rice populations, which were assumed to be due to hybridisation between the *O. meridionalis* and *O. rufipogon* type taxa (Moner et al., 2018). These wild hybrid plants were identified in the field due to their intermediate morphological characteristics with combinations of spike morphology and anther size, suggesting the inheritance of traits from both parental taxa. However, the exact nature of their hybrid origin was not characterised by nuclear genome analysis.

We have now reexamined the origins of these hybrids by re-sequencing and analysing the nuclear genomes of these wild rice populations including wild plants of hybrid appearance.

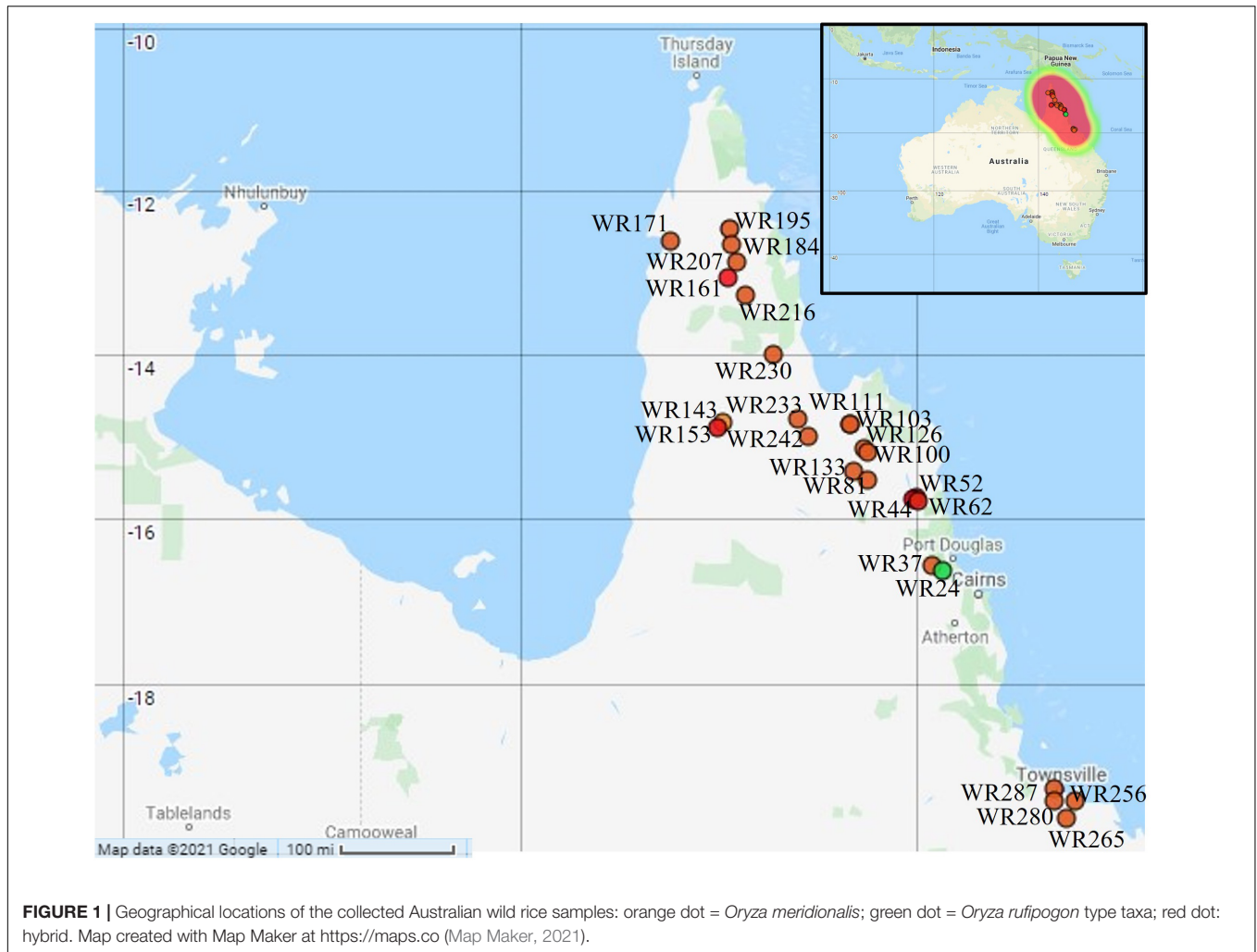
## MATERIALS AND METHODS

### Sample Collection

Australian wild rice is widely distributed from the south of Townsville to the northern parts of Cape York Peninsula, Queensland, Australia. Fresh leaves of wild rice from 26 different geographical sites were collected in the two consecutive years 2015 and 2016 by Moner et al. (2018) (Figure 1). Only one sample was collected from each site. Among these, 22 samples were collected in 2015 from 22 different sites in the northern parts of Cape York Peninsula while only four samples were collected in 2016 from the south of Townsville. The samples were kept in the dry ice all the way from the sampling sites to the laboratory until stored at  $-80^{\circ}\text{C}$  freezer.

### DNA Extractions, Sequencing, and Quality Filtering

Fresh leaves were pulverised with tissue lyser (Qiagen, United States). Samples were kept in liquid nitrogen to prevent thawing throughout the pulverisation process. DNA was extracted from pulverised leaf tissue of wild rice samples following the CTAB method (Furtado, 2014). The quality and quantity of the extracted DNA were evaluated with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Delaware, United States) as well as by electrophoresis on a 0.7% agarose gel stained with SYBR safe. The 260/280 nm ratio for 26 samples ranged from 1.11 to 2.06. Nextera DNA Flex



Libraries were prepared. Nextera DNA Library Preparation Kits were designed in an automation-compatible workflow that introduced tagmentation chemistry, which joined DNA fragmentation and adapter ligation steps into a single 15-min reaction. Then these libraries were sequenced on a NovaSeq 6000 SP and S4 Flow cell along with other samples to produce  $2 \times 150$  paired-end reads with a data yield of 20x whole genome coverage on average. A quality check was applied to the raw data using CLC Genomic Workbench (CLC-GWB) version 20 software (QIAGEN CLC Genomics Workbench 20, Denmark)<sup>1</sup> to identify any sequencing error before performing further analysis. Sequences were trimmed at both 0.05 and 0.01 quality scores to truncate low-quality reads. Based on the percentage of loss of reads and bases, 0.01 trimmed reads were considered for further analysis. The total Illumina raw reads for the 26 samples ranged from 56,349,078 to 104,325,654 bp. The total raw nucleotides ranged between 8,416,673,405 and 15,628,960,467 bp. After trimming, the total 0.01 trimmed reads ranged from 54,631,067 to 98,924,768 and nucleotides from 7,925,007,230 to 14,260,593,039 bp (**Supplementary Table 1**).

<sup>1</sup><http://www.clcbio.com>

## Illumina Reads Mapping and Genome-Wide Variant Detection

Prior to final mapping, the best mapping parameters were evaluated with the three different alignment quality thresholds: 1.0 length fraction (LF) and 0.95 similarity fraction (SF), 1.0 LF and 0.90 SF and 1.0 LF and 0.85 SF. Illumina paired-end reads were then mapped to the reference genome Os-Nipponbare-Reference-IRGSP-1, which was downloaded at Rice Annotation Project Database (Kawahara et al., 2013) with these three settings. The average mapped read depth (times of reference length), mapped reads (%), mapped bases (%), and total consensus length as a % of reference length were calculated from the output of mapped reads to make a comparison among the three different settings. After comparison, the 1 LF and 0.9 SF settings were found to be stringent enough to consider for final mapping. A total of 26 wild rice Illumina sequences quality trimmed at 0.01 (Phred equivalent  $> 20$ , 94–98% of the reads had a phred score greater than 35, average across all bases) was mapped to the Os-Nipponbare-Reference-IRGSP-1, separately using the tool “Map Reads to Reference” in CLC-GWB, with the following

parameters: masking mode = no masking; match score = 1; mismatch cost = 2; cost of insertions and deletions = linear gap cost; insertion cost = 3; deletion cost = 3; length fraction = 1; similarity fraction = 0.9; global alignment = no; auto-detected paired distances = yes; non-specific match handling = map randomly. Mapped reads was subjected to variant analysis using the tool “Basic Variant Analysis” in CLC-GWB, with the following parameters: ploidy = 2; ignore positions with coverage above = 100,000; restrict calling to target regions = not set; ignore broken pairs = no; ignore non-specific matches = no; minimum coverage = 10; minimum count = 3; minimum frequency (%) = 25; neighborhood radius = 5; minimum central quality = 20; minimum neighborhood quality = 15; read direction filter = no; read position filter = no; remove pyro-error variants = no. Total variants that included SNV (single nucleotide variant), insertions, deletions, MNV (multiple nucleotide variants), and replacement were determined. Percentage of SNV and InDels (insertions and deletions) were calculated as: (total number of SNV/InDels  $\times$  100)/total number of variants detected (1). Heterozygous SNV positions were determined by SNV loci with two alleles. Heterozygosity across the whole genome was calculated as: (total heterozygous SNP position  $\times$  100)/the whole genome size (2). Unique heterozygous SNPs were those that were distinctively located in the reads of a sample. The whole-genome size of *O. rufipogon* type taxa (= Taxon A) and *O. meridionalis* (= Taxon B) was 384.8 and 354.9 Mb, respectively (Brozynska et al., 2017). Other data resources for *O. rufipogon* and *O. meridionalis* are available at ([https://oryza-ensembl.gramene.org/Oryza\\_meridionalis/Info/Index](https://oryza-ensembl.gramene.org/Oryza_meridionalis/Info/Index), [https://oryza-ensembl.gramene.org/Oryza\\_rufipogon/Info/Index](https://oryza-ensembl.gramene.org/Oryza_rufipogon/Info/Index), <http://viewer.shigen.info/oryzagenome21detail/downloads/index.xhtml?jsessionid=16d6389e9b7080a8c4fb7378b753>). Sequence data for the Australian *O. rufipogon* like taxon is reported in Brozynska et al. (2017).

## Nuclear Geneflow and Genetic Structure Analysis

In further experiment, five natural hybrids (WR44, WR52, WR62, WR153, and WR161) were investigated to test for nuclear introgression from the putative parental populations. Among these, WR44, WR52, and WR153 were reported by Moner et al. (2018). The earlier study also identified two hybrids (WR65 and WR162), which were collected from Lakeland Cook town road and Merluna site 1, respectively. Two hybrids (WR62 and WR161), explored in this study, were also collected from the same sites but not reported as hybrids by Moner et al. (2018). These were collected from the same sites as WR65 and WR162, respectively. To ensure the chloroplast genome of WR62 and WR161 were of hybrid origin showing similarity with the previously published samples WR65 and WR162 (Moner et al., 2018), respectively, 0.01 trimmed data of WR62 and WR161 were mapped to the reference chloroplast genome of *O. meridionalis*, and the reference chloroplast genome of *O. rufipogon* type taxa obtained from Brozynska et al. (2014). Mapping was done following the reference-based mapping workflow explained by Moner et al. (2018) using CLC-GWB. Then the consensus sequences of the chloroplast genome were extracted from these

two samples (sequences S1 and S2 provided as **Supplementary Material**). Chloroplast genomes of seven samples; WR62, WR161 from this study, WR65 and WR162 (Moner et al., 2018), two reference chloroplast genomes of *O. meridionalis* and *O. rufipogon* type taxa (Brozynska et al., 2014), and *Oryza officinalis* Wall. ex Watt (Moner et al., 2018) were aligned with Multiple Alignment using Fast Fourier Transform (MAFFT) version 7.407 (Katoh and Standley, 2013) with default parameters in Geneious version 11.1.5 software.<sup>2</sup> The substitution model for the phylogenetic analysis was selected by the jModelTest2 (Darrriba et al., 2012) on the Extreme Science and Engineering Discovery Environment (XSEDE). The parameters for best selected model following the Akaike information criterion (AIC) was used in maximum likelihood tree construction: Model = TVM+I+G, partition = 012314,  $-\ln L = 1888839.3939$ ,  $K = 21$ ,  $\text{freqA} = 0.3068$ ,  $\text{freqC} = 0.1941$ ;  $\text{freqG} = 0.1957$ ;  $\text{freqT} = 0.3034$ ;  $\text{ti/tv} = -$ ;  $\text{p-inv} = 0.897$ ;  $\text{R(a)} = 1.0604$ ;  $\text{R(b)} = 2.5331$ ,  $\text{R(c)} = 0.5698$ ;  $\text{R(d)} = 0.9914$ ;  $\text{R(e)} = 2.5331$ ;  $\text{R(f)} = 1$ ;  $\text{gamma} = 0.587$ . Phylogenetic reconstruction was performed using PAUP\* version 4 software (Swofford, 2003) with maximum likelihood (ML) with heuristic searching, retain group with frequency > 50%, bisection, reconnection branch swapping method and 1,000 bootstrap replicates. *O. officinalis* was selected as an outgroup species in the phylogenetic tree.

Single nucleotide polymorphism (SNP) genotypes of six genes located on chromosomes 4, 6, and 7 from hybrids and putative parental taxa were aligned, with respect to Os-Nipponbare-Reference-IRGSP-1 with CLC-GWB to screen for contributing donor parents in genotypes of hybrid populations. These genes include *seed shattering 4 (sh4)* on chromosome 4, four starch synthesis related genes on chromosome 6 [*granule bound starch synthase I (GBSSI)*, *soluble starch synthase-I (SSI)*, *starch branching enzyme I (SBEI)*, and *alkali degeneration (SSIIa or ALK)*] and *granule bound starch synthase II (GBSSII)* on chromosome 7. Plausible hybridisation generations were determined based on the level of heterozygosity across the whole genome of these hybrids.

Four taxon (ABBA-BABA) test as a measure of geneflow was performed with Hybrid Check version 1 software (Ward and van Oosterhout, 2016). Four genes (*GBSSI*, *SSI*, *SBEI*, and *ALK*) from two putative parental populations (WR24 and WR81), one hybrid and outgroup species were first aligned with MAFFT version 7.407 (Katoh and Standley, 2013) with default parameters in Geneious version 11.1.5 software to generate an aligned sequence. Then, an aligned sequence was used to perform the four-taxon test with a topology (((P1, P2), P3) A), where a triplet comprises of a set of three aligned sequences from a hybrid (P3) and two putative parental populations (P1 and P2). *O. sativa* spp. *japonica* cv. Nipponbare was considered an outgroup species (A). We simulated a total of 5 four-taxon tests to determine biparental nuclear geneflow between hybrid and two putative parental taxa. We estimated Patterson's *D*-statistic based on the random distribution of ABBA-BABA polymorphisms in a four-taxon test (Martin et al., 2015). The *D*-statistic estimates the deviation from the distribution of ABBA-BABA polymorphisms

<sup>2</sup><https://www.geneious.com>

in a four-taxon tree. If  $D > 0$ , introgression occurs between P2 and P3, and if  $D < 0$ , introgression occurs between P1 and P3. We performed a block jackknife method using a block length of 1,000 bp to calculate  $Z$  score and  $p$ -value to test significant deviation from the null expectation ( $D = 0$ , lineage sorting).

Genome-wide SNP data from 26 wild rice samples were used to perform principal component analysis (PCA) to assess the genetic variation in wild rice populations using plink version 1.9 (Purcell et al., 2007). To ensure the SNP data was not biased with any linkages, linkage disequilibrium (LD) based pruning was done to generate a final neutral SNP data for PCA analysis. LD pruning was performed using the “-indep-pairwise” option by specifying a window of 50 kb, with a step window size of 10 bp and a threshold of linkage,  $r^2 = 0.2$  (which recursively eliminates SNPs within the sliding window if  $r^2$  is greater than 0.2). To investigate the relationship of population structure with geographic distance, a subset of genome-wide SNP data (2,987,364 SNPs of 12,247,310 SNPs) from 26 wild rice genotypes (two alleles per loci) were used to perform isolation by distance (IBD) analysis using Mantel test (Mantel, 1967) with the R package adegenet version 2.0.1 (Jombart and Ahmed, 2011). Nei (1972) pairwise genetic distances vs. geographic distances were correlated with the function mantel.randtest using a Monte Carlo simulation with 999 permutations in IBD analysis.

## RESULTS

### Mapping Outputs and Sequence Variation

Sequence variation from *O. sativa* was substantial among the 26 samples (Supplementary Table 1). The average sequence read depth (times reference length) ranged from 20.7 in WR153 to 46.9 in WR103 (Supplementary Table 1). While an average mapped sequence read depth (times reference length) varied from 8.88 in WR24 to 46.7 in WR103. The percentage of mapped reads was highest in WR44 (72.0%) and the lowest in WR24 (24.0%). Similarly, the highest and lowest percentage of mapped bases were 71.9 and 23.8% in WR44 and WR24, respectively. The total consensus sequences length ranged from 257,716,289 bp in WR37 to 322,074,565 bp in WR24. The total consensus as a percentage of the reference ranged from 67.4% in WR37 to 83% in WR44. The data for eight samples were also mapped to the lower quality draft genomes (Brozynska et al., 2017) for the two Australian taxa (Supplementary Table 2) resulting in somewhat higher genome coverage. WR 24 displayed an *O. rufipogon* type nuclear genome and showed much greater coverage of this reference confirming it was closer to *O. sativa* than the *O. meridionalis* types. Further analysis was based upon the use of the common high-quality *O. sativa* genome to align the reads to explore hybridisation and heterozygosity in the wild populations.

### Chloroplast and Nuclear Genome Characterisations

The analysis of phylogenetic relationships based on the whole chloroplast and nuclear genome had suggested that chloroplast

capture was present in Australian wild rice populations (Moner et al., 2018). The whole chloroplast and nuclear genome of both *O. rufipogon* type taxa and *O. meridionalis* were compared with the chloroplast and nuclear genome of WR62 and WR161 to determine the origin of these samples that had not been characterised in the earlier study of Moner et al. (2018). The results confirmed that these were of hybrid origin as the chloroplast genome of WR62 was grouped with the *O. meridionalis* type (Brozynska et al., 2014) while the nuclear genome was of the *O. rufipogon* type taxa (Supplementary Figure 1). While WR161 contained an *O. meridionalis* type nuclear genome and an *O. rufipogon* type chloroplast genome (Brozynska et al., 2014). This reinforced the presence of bidirectional chloroplast capture in hybridising populations in the Australian wild rice gene pool. Hybrids acquired their chloroplast genome from either the *O. rufipogon* type or *O. meridionalis* during hybridisation. The other three hybrids (WR44, WR52, WR153) contained an *O. rufipogon* type chloroplast genome and an *O. meridionalis* type nuclear genome.

### Analysis of Gene Loci

SNP were analysed for six genes on chromosomes 4, 6, and 7 were analysed to assess the genetic contribution of the putative parents to the natural hybrid populations in relation to *O. sativa* spp. *japonica* cv. Nipponbare reference genome. Of the six genes, one gene (*sh4*) on chromosome 4 is associated with seed shattering while the other five genes are starch synthesis-related genes located on chromosome 6 (*GBSSI*, *SSI*, *SBEI*, and *ALK*) and chromosome 7 (*GBSSII*). From SNP data, it was clearly shown that the six genes of WR62 received alleles from both *O. rufipogon* type taxa and *O. meridionalis* parents. *SBEI* and *GBSSII* had heterozygous SNP genotypes with both *O. rufipogon* type taxa and *O. meridionalis* alleles in all of five hybrids (Table 1 and Figure 2). However, the *sh4* alleles were heterozygous for WR44, WR62, and WR161 and homozygous for WR52 and WR153. Two hybrids (WR153 and WR161) also had heterozygous loci for *ALK* and *SSI*, respectively. Sequence alignment images of SNP genotypes of *sh4* gene (ID: Os04g0670900-1, position on chromosome 4 from 34,231,186 to 34,233,221 bp), *ALK* gene (ID: Os06g0229800-1, position on chromosome 6 from 6,748,398 to 6,753,302 bp), and *GBSSII* (ID: Os07g0412100-1, position on chromosome 7 from 12,916,883 to 12,924,202 bp) (RAPDB, 2021) showed heterozygous locus as in Figure 2. WR44 and WR62 showed heterozygous loci at the 34,231,465 and 34,231,468 base positions (Figure 2A) while WR62 and WR44 showed heterozygous loci at 6,751,298 and 12,919,934 base position, respectively (Figures 2B,C). In all cases, a heterozygote was formed by receiving an allele from *O. rufipogon* type taxa and the alternative allele from *O. meridionalis*. Hybrids with homozygous loci represented alleles mostly from *O. meridionalis*. The number of total variants across the whole genome varied among hybrids, ranging between 5,484,261 and 8,024,717. W44 and W62 are early generation hybrids as shown by the presence of alleles from both parents while others are backcrossed to recover a nuclear genome similar to the recurrent parent. The highest number of SNPs (6,910,977) was found in WR44, followed by 5,917,274 in WR62, 5,291,708 in WR52, 5,295,508 in WR161 and 4,764,266 in

**TABLE 1** | Characterisation of natural hybrids in wild rice populations.

Hybrid	SNP genotypes of six genes on chromosome 4, 6, and 7						Total SNPs	Total heterozygous SNPs	Heterozygosity across the whole genome (%)	Plausible hybrid generations
	<i>sh4</i> alleles	<i>GBSSI</i> alleles	<i>SSI</i> alleles	<i>SBEI</i> alleles	<i>ALK</i> alleles	<i>GBSSII</i> alleles				
WR44	AB	BB	BB	AB	BB	AB	6,910,977	2,464,823	0.69	Early (F1/F2)
WR52	BB	BB	BB	AB	BB	AB	5,291,708	1,137,527	0.32	Later
WR62	AB	AB	AB	AB	AB	AB	5,917,274	2,318,878	0.60	Early (F1/F2)
WR153	BB	BB	BB	AB	AB	AB	4,764,266	1,111,890	0.31	Later
WR161	AB	BB	AB	AB	BB	AB	5,295,508	1,189,627	0.34	Later

Plausible hybrid generations are shown based on the level of heterozygosity across the whole genome. AB and BB denote heterozygous and homozygous loci, respectively.

WR153 (Table 1). Similarly, the highest number of heterozygous SNPs were found in WR44 (2,464,823) and the lowest number in WR153 (1,111,890). To determine the heterozygosity across the whole genome, the genome size of *O. rufipogon* type taxa was applied to the above-mentioned formula (2) for WR62, which exhibited *O. rufipogon* type nuclear genome while the genome size of *O. meridionalis* was used for the other hybrids. A high level of genetic variation was observed in both WR44 (0.69%) and WR62 (0.6%) across the whole genome followed by WR161 (0.34%), WR52 (0.32%), and WR153 (0.31%) in relation to *O. sativa* spp. *japonica* cv. Nipponbare (Table 1). The level of heterozygosity across the whole genome of these hybrids confirmed hybridisation events and led to the conclusion that WR44 and WR62 were early generation hybrids while the other hybrids were much later generations due to recurrent backcrossing with one parent (*O. meridionalis*).

### Analysis of Genetic Introgression

A four-taxon test (as *D*-statistic) was performed to test for biallelic mutation patterns (ABBA-BABA) to determine geneflow between the hybrids and two putative perennial parental populations in a topology  $\{[(O. rufipogon \text{ type taxa}, O. meridionalis), \text{hybrid}], O. sativa \text{ spp. japonica cv. Nipponbare}\}$ . The length of a MAFFT aligned sequence was 4,473 bp with a total of four blocks: (1–1,118); (1,119–2,236); (2,237–3,354); and (3,355–4,473). The *D*-statistic value based on the distribution of AABB and BABA on each block showed significant deviation for all four-taxon trees (Table 2). The negative *D*-statistic value suggested potential genetic introgression between *O. rufipogon* type taxa (WR24) and hybrid while the positive value suggests geneflow between *O. meridionalis* (WR81) and hybrid. In all cases, *O. sativa* spp. *japonica* cv. Nipponbare was an outgroup. The *Z*-score for all four-taxon tests was significant ( $p < 0.001$ ).

### Intra- and Interspecific Variations in the Wild Rice Gene Pool

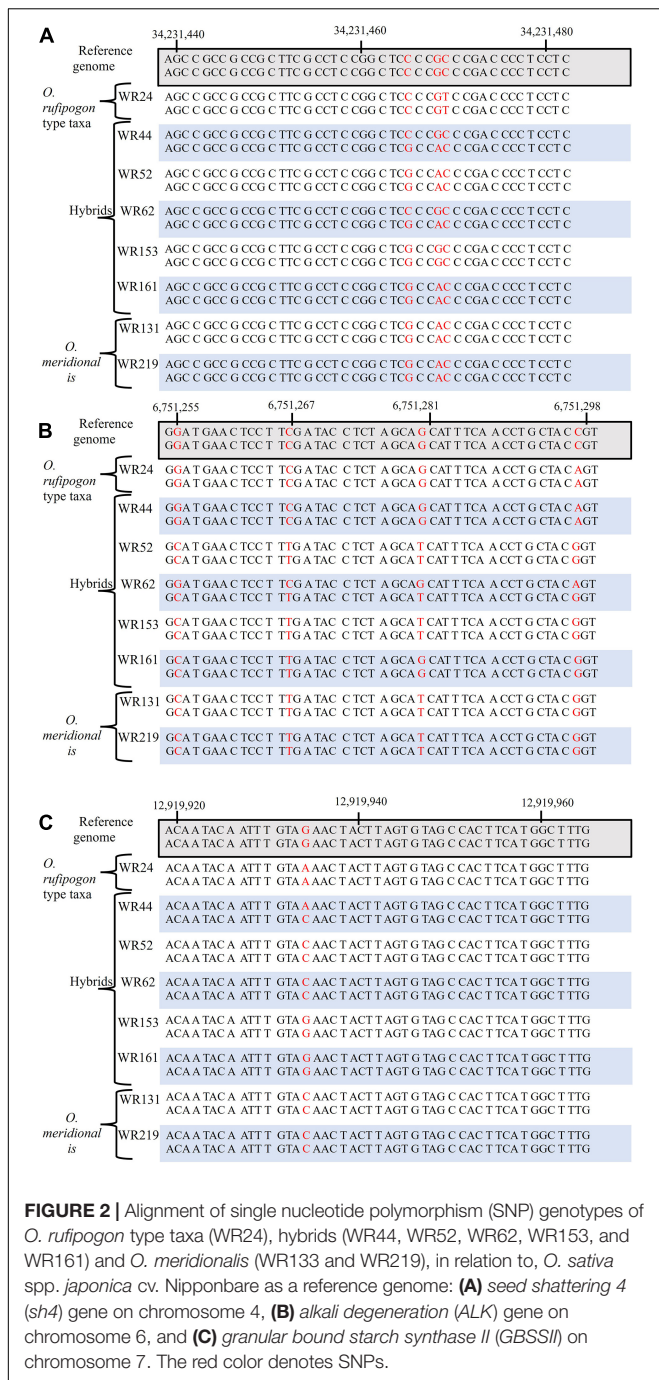
To estimate intra- and interspecific variations in the wild rice gene pool, variants comprised of five different types: SNV, insertions, deletions, MNV, and replacement were assessed across the whole genome of twenty-one putative parental taxa (including one *O. rufipogon* type taxa genotypes and 20 *O. meridionalis* genotypes). The number of allelic positions where the variants were located was determined. The total number

of variants ranged from 1,918,262 in WR24 to 7,310,928 in WR81 (Table 3). Overall, 87–90% of the total variants in an individual were SNPs for all samples. While the proportion of InDels (insertions and deletions) was low ranging between 6 and 8%. The number of heterozygous SNP positions varied among wild rice populations. The highest number of heterozygous SNP was found in WR81 (1,365,802) while the lowest number of heterozygous SNP was found in WR24 (632,272). Consequently, the level of heterozygosity across the whole genome was higher in *O. meridionalis* (0.38%), and the least heterozygosity was found in *O. rufipogon* type taxa (0.16%). The SNPs that were uniquely present in individuals were determined to define the genetic relationships between individuals. The rare or unique heterozygous SNP position (%) ranged from 4 to 48%. The highest number of unique heterozygous SNP positions was in *O. rufipogon* type taxa and the lowest number in *O. meridionalis*, indicating *O. rufipogon* type taxa might be subject to more genetic drift compared to *O. meridionalis* populations due to their smaller population size.

### Genetic Differentiation of Wild Rice Populations

Genome-wide SNP data (12,247,414 SNPs) of 26 wild rice genotypes were first processed to remove linkage for PCA analysis. A total of 490,655 SNP sites were retained that fell below the linkage threshold. PCA was performed on linkage-pruned sites to determine the genetic groupings of wild rice populations. PCA analysis showed two clear clusters: (1) one group formed with hybrids (WR44, WR62, and WR161) and one putative parental taxon (WR24) and (2) all *O. meridionalis* populations along with two plausible later generations of hybrids (WR52 and WR153) formed another group (Figure 3). The first axis (PC1) explained 64.6% of the variation, which separated out the former group. There was a very low level of genetic differentiation observed in the latter group represented by the second axis (PC2) with 22.4% variation.

Isolation by distance analysis based on a Monte-Carlo test with 999 permutations clearly showed that there was no significant relationship ( $R^2 = -0.008$ ,  $p$ -value = 0.42) between Nei's pairwise genetic distance and geographic distances (Figure 4). A total of 300 pairwise Nei's genetic distance data were plotted against pairwise geographic distance. The observed variation ( $R^2$ ) was 0.008 with a  $p$ -value of 0.42. This indicated genetic



taxa. A total of five hybrids were studied of which only one hybrid (WR62) exhibited an *O. meridionalis* type chloroplast genome while the other four hybrids displayed *O. rufipogon* type chloroplast genomes. This indicates that hybridisation in both directions is possible. The previous study of these populations using the whole chloroplast and nuclear genome revealed chloroplast capture results in discordant cytonuclear phylogenies, inferring hybrid origins (Moner et al., 2018). Chloroplast capture can be the consequence of lineage sorting, convergent evolution, and recurrent hybridisation (Tsitrone et al., 2003). Unidirectional gene flow from recurrent hybridisation paves the way to capture either cytoplasm (Takahata and Slatkin, 1984) or nuclear genes (Petit et al., 2001) from one species to another species. However, introgression promotes chloroplast capture rather than nuclear genome capture in sympatric populations as has been detected in many plant species, resulting in discordant cytonuclear phylogenies (Rieseberg and Soltis, 1991). Unlike introgression, chloroplast capture can occur in species through one single hybridisation event and species with the freshly obtained chloroplast genome can be inherited by descendant lineages over time (Nge et al., 2021). In this study, hybridisation or introgressive hybridisation may have allowed chloroplast genome capture among sympatric wild rice populations. It is also known that nuclear genome incompatibilities may stimulate chloroplast capture in a random mating population (Tsitrone et al., 2003). Brozynska et al. (2017) reported that nuclear genomes of two parental taxa were distinguished where *O. rufipogon* type taxa exhibited more sequence similarity with *O. sativa* relative to the other Australian AA genome wild rice species. Whole nuclear genome analysis revealed that the Australian *O. rufipogon* had a nuclear genome that was part of a clade with Asian cultivated rice and its wild progenitors in Asia while *O. meridionalis* was sister to this clade in a nuclear genome phylogenetic tree (Moner et al., 2018). Despite the distinct nuclear genome, the chloroplast genome of the Australian *O. rufipogon* type taxa are more closely related to *O. meridionalis* differing by 53 variants and 36 variants from different accessions of *O. meridionalis*, respectively, while these forms of *O. meridionalis* from different regions (Queensland and Northern Territory) differed from each other by 34 variants (Brozynska et al., 2014). The *O. rufipogon* in Australia was much more divergent from *O. rufipogon* in Asia (125 variants) and *O. sativa* spp. Nipponbare cv. *japonica* (125 variants). In the Australian *O. rufipogon*, ancient chloroplast capture has occurred from *O. meridionalis* (Brozynska et al., 2014), however, hybrids showed the two types of chloroplasts (*O. meridionalis* and Australian *O. rufipogon*) due to recent bidirectional chloroplast capture from these parents. The recent introgressive hybridisation demonstrated here may be an explanation of both the apparent chloroplast capture in the Australia *O. rufipogon* populations and the wild hybrids.

Analysis of nuclear genes demonstrated that the hybrids had alleles from both putative parental populations forming a heterozygote in early generations, indicating biparental nuclear gene flow from interspecific hybridisation. However, these loci become homozygous in what appear to be many later generations of hybrids that are fixed for the alleles of the recurrent parent, usually *O. meridionalis*. These observations clearly demonstrate

differentiation was not significantly increased with geographic distance, indicating genetic connectivity in wild rice populations across a wide range.

## DISCUSSION

### Reticulate Origin of AA Genome Hybrids

This study confirms two types of the chloroplast genome in hybrid populations introgressed from one of their parental

**TABLE 2 |** Estimation of geneflow in a triplet species following four-taxon tests with HybridCheck version 1.0 software.

Species in triplet			Outgroup species	Block start	Block end	numABBA	numBABA	D-statistic	Z-score	Direction of geneflow
P1	P2	P3	A							
WR24	WR81	WR44	Nipponbare	1	1,118	168	10	0.47	0.35*	P3→P1
WR24	WR81	WR52	Nipponbare	1	1,118	133	6	0.38	-0.38*	P3→P1
WR24	WR81	WR62	Nipponbare	1	1,118	66	18	0.06	-1.81*	P3→P1
WR24	WR81	WR153	Nipponbare	1	1,118	73	18	0.03	-2.14*	P3→P1
WR24	WR81	WR161	Nipponbare	1	1,118	55	0	0.33	-0.78*	P3→P1
WR24	WR81	WR44	Nipponbare	1,119	2,236	0	8	-0.28	0.35*	P2→P3
WR24	WR81	WR52	Nipponbare	1,119	2,236	0	13	-0.49	-0.38*	P2→P3
WR24	WR81	WR62	Nipponbare	1,119	2,236	0	27	-0.69	-1.81*	P2→P3
WR24	WR81	WR153	Nipponbare	1,119	2,236	0	15	-0.49	-2.14*	P2→P3
WR24	WR81	WR161	Nipponbare	1,119	2,236	0	14	-0.43	-0.78*	P2→P3
WR24	WR81	WR44	Nipponbare	2,237	3,354	9	19	-0.57	0.35*	P2→P3
WR24	WR81	WR52	Nipponbare	2,237	3,354	9	16	-0.65	-0.38*	P2→P3
WR24	WR81	WR62	Nipponbare	2,237	3,354	9	19	-0.66	-1.81*	P2→P3
WR24	WR81	WR153	Nipponbare	2,237	3,354	9	19	-0.68	-2.14*	P2→P3
WR24	WR81	WR44	Nipponbare	3,355	4,472	19	21	-0.06	0.35*	P2→P3
WR24	WR81	WR52	Nipponbare	3,355	4,472	19	21	-0.12	-0.38*	P2→P3
WR24	WR81	WR62	Nipponbare	3,355	4,472	19	21	-0.16	-1.81*	P2→P3
WR24	WR81	WR153	Nipponbare	3,355	4,472	19	21	-0.31	-2.14*	P2→P3
WR24	WR81	WR161	Nipponbare	3,355	4,472	19	21	-0.14	-0.78*	P2→P3

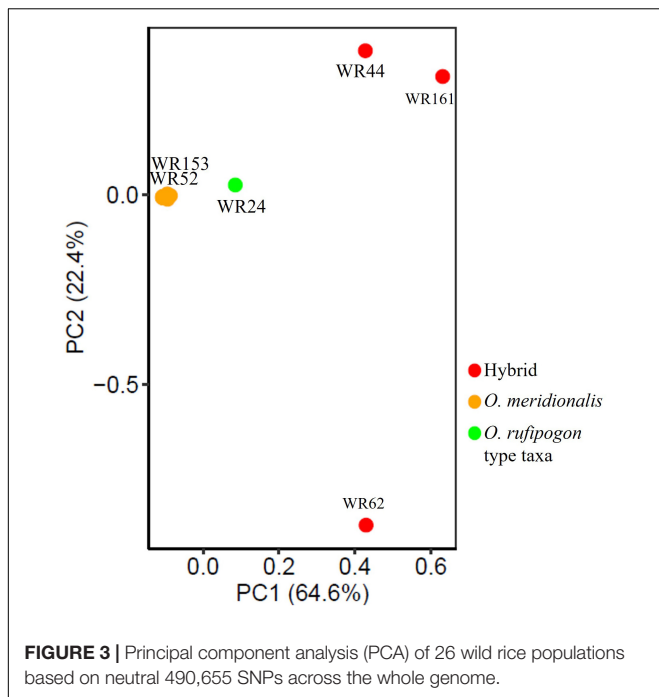
Triplets are characterised: P1 = *O. rufipogon* type taxa; P2 = *O. meridionalis*; and P3 = Hybrid. *O. sativa* spp. japonica cv. Nipponbare (A) is an outgroup species. The length of aligned nucleotide is 4,473 bp containing four blocks: (1–1,118); (1,119–2,236); (2,237–3,354); and (3,355–4,473). The biallelic patterns based on the number of ABBA-BABA in a triplet used in the Patterson’s D-statistic are shown. P3→P1 denotes geneflow between *O. rufipogon* type taxa and hybrid while P2→P3 denotes geneflow between *O. meridionalis* and hybrid. \*Z-score represents statistically significant at  $p < 0.001$ .

**TABLE 3 |** Genetic variation across the whole genome of 21 wild rice samples, excluding hybrids.

Species	Samples	Total number of variants	SNP (%)	InDel (%)	Total number of heterozygous SNV position	Heterozygous SNP variation across the whole genome (%)	*Unique heterozygous SNP (%)
<i>O. rufipogon</i> type taxa	WR24	1,918,262	90	6	632,272	0.16	48
<i>O. meridionalis</i>	WR37	3,745,572	87	8	827,689	0.23	6
	WR81	7,310,928	87	8	1,365,802	0.38	8
	WR100	6,007,510	87	8	1,175,173	0.33	4
	WR103	6,256,551	87	8	1,198,555	0.34	5
	WR111	3,899,236	88	7	894,765	0.25	5
	WR126	5,194,396	87	8	1,125,277	0.32	10
	WR133	6,545,660	87	8	1,227,038	0.35	4
	WR143	5,611,030	87	8	1,109,062	0.31	5
	WR171	3,852,279	88	7	966,368	0.27	6
	WR184	5,421,763	87	8	1,074,882	0.30	5
	WR195	5,985,028	87	8	1,168,776	0.33	6
	WR207	6,520,874	87	8	1,210,794	0.34	12
	WR219	6,951,983	87	8	1,312,291	0.37	6
	WR230	6,694,713	87	8	1,261,074	0.36	5
	WR233	3,623,110	88	7	876,065	0.25	4
	WR242	6,849,045	87	8	1,299,754	0.37	6
	WR256	5,991,356	87	8	1,119,055	0.32	6
	WR265	6,625,636	87	8	1,255,312	0.35	5
WR280	6,621,409	87	8	1,249,681	0.35	5	
WR287	5,722,665	87	8	1,105,062	0.31	5	

\*Unique heterozygous SNP: SNPs that are uniquely present in one sample but absent in other samples.

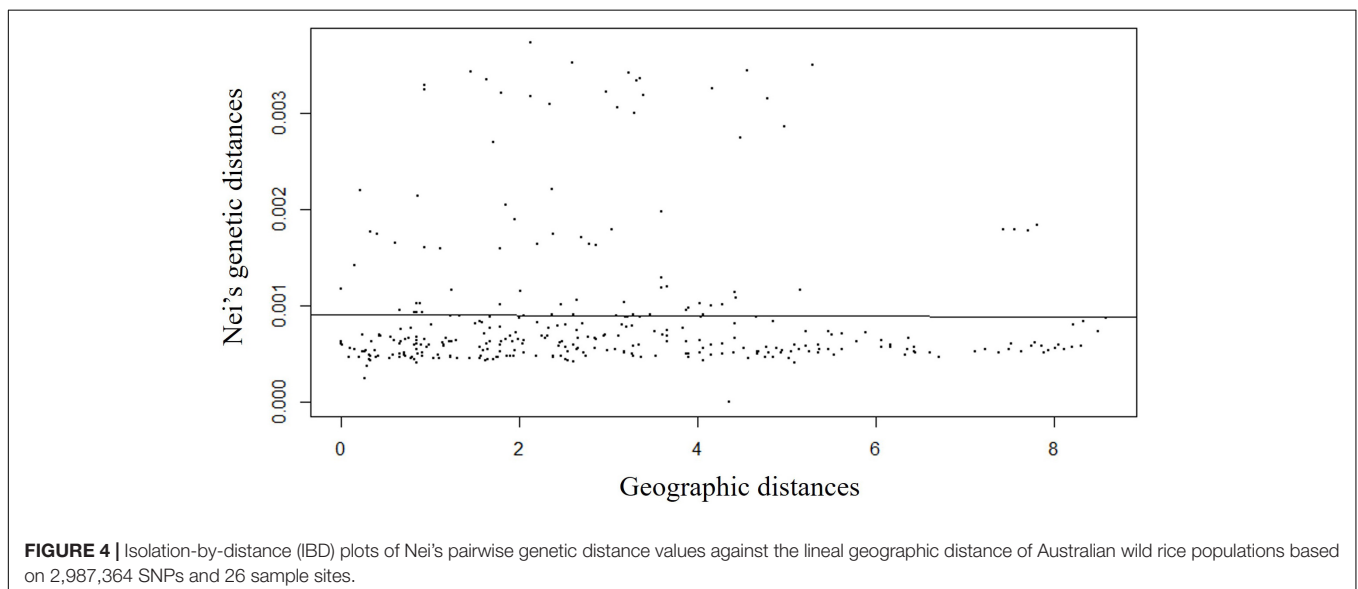




an ongoing hybridisation and introgression persisting in the wild rice populations evidenced by the presence of viable hybrids between *O. rufipogon* type taxa and *O. meridionalis* being found at very low frequency. Earlier findings postulated that interspecific crossability between an *O. meridionalis* and a perennial Asian *O. rufipogon* was modest but asymmetric (Banaticla-Hilario et al., 2013). As a result, they reported a very low percentage of fertile first filial generation (F1) (i.e., with mean F1 fertility = 17.2%) with perennial habit. However, they posited that a perennial habit enabled the F1 progeny to persist for a longer period than its annual counterpart. Like annual strains,

crosses between accessions from the Jpn2 location (known to be *O. meridionalis*- like) and *O. rufipogon* produce low fertility pollen (<10%) as found with putative annual strains (<20–40%) (Sotowa et al., 2013). Intra-specific crossability between *O. meridionalis*, within the same geographical origin, produces fertile hybrids with pollen stainability (96.7%) and panicle fertility (66.9%), however, geographical distance seems to be the reason for the development of reproductive barriers (Juliano et al., 2005). They observed fertile hybrids, which were derived from crosses between Queensland accessions, however, crosses between accessions from different geographical regions produced sterile to partially fertile hybrids. Similarly, very low pollen stainability and seed fertility were reported in intraspecific crosses between *O. meridionalis* populations, from different geographical locations (Naredo et al., 1997).

The positive and negative significant *D*-statistic values from the four-taxon test indicate a nuclear geneflow from both putative parents into hybrids. Early generation hybrids exhibited higher heterozygosity and polymorphic variations than that of the later generations of hybrids. Repeated backcrossing between F1/F2 offspring and *O. meridionalis* may lead to comparatively low heterozygous SNP variation in the later-generation of hybrids. PCA analysis also reveal that early generation hybrids were clustered with *O. rufipogon* type taxa and later generation hybrids (except for WR161) form a group with *O. meridionalis*. Reproductive barriers underpin a termination of hybridisation and thereby, introduce introgression (Rieseberg and Wendel, 1993). An interspecific reproductive barrier gene (*SEED DEVELOPMENT 1*) on chromosome 6 is detected in *O. meridionalis* and *O. sativa* (Toyomoto et al., 2019). A recessive allele of a *SEED DEVELOPMENT 1* gene (*sdv1-m*) in wild rice strain from the Jpn2 site (= *O. meridionalis*) appears to evoke segregation distortion (1:2:0) in a BC<sub>4</sub>F<sub>2</sub> progeny. Therefore, offspring carrying homozygous recessive alleles (*sdv1-m*) cause seed abortion at the seed development stage. Toyomoto et al. (2019) made an F1 offspring from a cross between *O. sativa*



and *O. meridionalis*, then backcrossed with *O. sativa* to obtain backcrossing progeny. Despite homozygous recessive alleles, offspring bear seed if they are self-fertilised. *O. rufipogon* from Asia or Australia does not have this abortive gene. Backcrossing with *O. meridionalis* along with selfing may reduce seed abortion in later-generation hybrids.

Phenotypic characteristics of hybrids observed by Moner et al. (2018) seem consistent with the nuclear gene analysis. *O. rufipogon* type taxa are distinguished with open panicles and short awns from *O. meridionalis*, which exhibits long awns with closed panicles. However, awn lengths sometimes overlapped (Brozynska et al., 2017; Moner et al., 2018). Early generation hybrids (WR44 and WR62) showed open to partially open panicles orientation like *O. rufipogon* type taxa. While all later generations hybrids exhibit panicles orientation like *O. meridionalis* type closed panicles. Interestingly, both generations of hybrids exhibited an intermediate anther length between two putative parental taxa and awn size close to *O. meridionalis*. Therefore, hybrids exhibit both intermediate and putative parental phenotypic characteristics, depending on which dominant loci of a trait are received from putative parents. Intermediate phenotypic characteristics have been observed in F1 hybrids derived from a cross between *O. rufipogon* and *O. meridionalis* (Naredo et al., 1997). Furthermore, F1 hybrids follow the lifecycle of their parental taxa, i.e., perennial F1 hybrids form when the two parental taxa are perennial in habit and annual and perennial parents produce semi-perennial to perennial offspring (Naredo et al., 1997). In contrast, Banaticla-Hilario et al. (2013) reported F1 offspring derived from a cross between Australasian perennial *O. rufipogon* and an annual *O. meridionalis* appears to be phenotypically closer to annual *O. meridionalis*. While F1 offspring with non-Australian *O. rufipogon* produce phenotypic characteristics more likely to be perennial *O. rufipogon* species. They also reported that interspecific F1 hybrids demonstrate either parental or intermediate phenotypic characteristics, i.e., some characteristics are analogous to their annual parental species and few features are intermediate between the two parents.

Reticulate evolution is a common phenomenon in vascular plants, wherein genetic changes occur through hybrid speciation, introgression, introgressive hybridisation, and horizontal gene transfer (Vriesendorp and Bakker, 2005). Hybrid origins were first proposed in the AA genome wild rice gene pool as an explanation of discordant cytonuclear phylogenies due to chloroplast capture (Moner et al., 2018). However, the disparity in cytonuclear phylogenies alone can never be an indicator of hybrid status (Vriesendorp and Bakker, 2005). Other events include incomplete lineage sorting, i.e., maintenance of ancestral polymorphisms through multiple speciation events (Andreasen and Baldwin, 2003; Goldman et al., 2004), homoplasy, and taxonomic sampling error (Wendel and Doyle, 1998) can cause such discordant patterns. The nuclear gene analysis reported here confirms hybrid events in the wild AA genome rice gene pool and confirms that the discordant phylogenies are the result of hybridisation events. The nuclear gene analysis along with chloroplast genome data clearly reveal a reticulate evolution pattern in the wild AA genome rice gene pool.

*Oryza rufipogon* from other countries shows no evidence of hybridisation with *O. meridionalis* as this species is not found in other countries. No highly divergent AA genome species that *O. rufipogon* could hybridise with is present outside Australia. This analysis is only possible in Australia where large populations of the two divergent species (*O. meridionalis* and the distinct Australian *O. rufipogon* like taxa) were searched to discover extremely rare apparent hybrids based upon their intermediate morphology. This study has confirmed the hybrid identity of these extremely rare plants. The Asian *O. rufipogon* samples that have been sequenced all show the presence of an *O. rufipogon* chloroplast that is distinct from that of *O. meridionalis* and the Australian *O. rufipogon* like taxa (Brozynska et al., 2014; Moner et al., 2018). The two AA genome species in Australia (*O. meridionalis* and the Australian *O. rufipogon* like taxa) remain as distinct populations despite the extremely rare hybridisation events between these two taxa confirmed in this study. The reproductive barrier between them resulting in low pollen fertility requires the backcrossing of the hybrids in subsequent generations.

## Australian Wild Rice Populations as a Reservoir for Novel Genetic Resources

This study reveals high levels of heterozygous SNP variation in AA genome wild rice populations, in relation to *O. sativa* ssp. *japonica* cv. Nipponbare. This finding is, in agreement with the result reported by Krishnan et al. (2014). They found that the Australian wild rice gene pool contains a high level of genetic variations, in terms of, SNPs across the whole genome, i.e., 2,418,084 SNPs in *O. meridionalis* and 2,564,013 SNPs in *O. rufipogon*, compared to Asian *O. rufipogon* (917,738 SNPs) and cultivated *O. sativa* ssp. *indica* (978,630 SNPs), with respect to *O. sativa* ssp. *japonica* cv. Nipponbare. A molecular study using restriction fragment length polymorphism (RFLP) markers revealed that cultivated rice retained 40% of the wild alleles during rice domestication (Sun et al., 2001). This study reveals the level of heterozygosity is lower in *O. rufipogon* than in *O. meridionalis*. This is also consistent with the previous results reported by Brozynska et al. (2017). They postulated *O. meridionalis* (Taxon B) had a more heterozygous and repetitive genome than *O. rufipogon* (Taxon A), resulting in a higher number of scaffolds in the egnoem of this taxon. Mating systems play an important role in acquiring genetic variations. Perennials exhibit outcrossing mating system, which results in a higher degree of genetic variation than annuals while selfing is a feature of annuals (Oka, 1988). In Australia, *O. meridionalis* is common and widespread while the Australian *O. rufipogon* populations are much smaller. Comparison of the relative diversity of the two species in Australia will require more extensive sampling of the much rarer *O. rufipogon* populations. PCA and Isolation by-distance analysis also reveals a high degree of genetic connectivity among *O. meridionalis* populations across a wide geographic range in Northern Queensland, Australia. The role of animals (especially birds) in dispersing *Oryza* species over large distances is quite evident (Tang et al., 2010; Wambugu et al., 2015).

Natural variations in genes for agronomically important traits are limited in cultivated rice germplasm. In-plant breeding, desirable germplasm is sourced from wild species, landraces, and distant relatives, or for new traits through induced mutation or genetic manipulation (Ganeshan et al., 2010). The Australian wild rice populations include the earliest branching AA genome lineages and, therefore the most genetically and geographically distinct from the other AA genome wild relatives of cultivated rice. Asian domesticated rice was introduced to Australia around 200 years ago and mainly grown far from wild rice populations (Henry et al., 2010) avoiding genetic introgression of domestication genes into the wild populations. Recent studies already reveal some interesting features of the native Australian populations, in terms of disease resistance, grain appearance, and nutritional properties (Kasem et al., 2012, 2014; Tikapunya et al., 2017) suggesting that they could be exploited to extract numerous genes for agronomically important traits for rice improvement.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, BioProject number PRJNA758754.

## AUTHOR CONTRIBUTIONS

RH conceived the project. AF and RH collected field samples. SH, AF, and RH analysed the data. SH wrote the first draft of the

manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the Cooperative Research Centre for Developing Northern Australia (CRCNA).

## ACKNOWLEDGMENTS

We would like to acknowledge the contribution of Ali M. Moner to collecting samples from the field. We also thank the Research Computer Centre at the University of Queensland for their support and advice and the CRCNA for funding.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.767635/full#supplementary-material>

**Supplementary Figure 1** | The phylogenetic tree of the chloroplast genome of four samples and two reference chloroplast genomes (*O. meridionalis* and *O. rufipogon* type taxa), with *O. officinalis* as the outgroup. Among four samples, WR62 and WR161 (red in colour) were taken from this study while the other two samples WR65 and WR162 (green in colour) were taken from Moner et al. (2018). The tree was generated using the maximum likelihood (ML) method in PAUP\* version 4 software with 1,000 bootstrap replicates. The ML bootstrap value (/40) is marked on each node.

## REFERENCES

- Andreasen, K., and Baldwin, B. G. (2003). Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (*Sidalcea*, Malvaceae). *Mol. Phylogenet. Evol.* 29, 563–581. doi: 10.1016/S1055-7903(03)00136-2
- Banaticla-Hilario, M. C. N., McNally, K. L., van den Berg, R. G., and Hamilton, S. N. R. (2013). Crossability patterns within and among *Oryza* series sativae species from Asia and Australia. *Genet. Resour. Crop Evol.* 60, 1899–1914. doi: 10.1007/s10722-013-9965-4
- Brar, D. S., and Khush, G. S. (2018). “Wild relatives of rice: a valuable genetic resource for genomics and breeding research,” in *The Wild Oryza Genomes*, eds T. K. Mondal and R. J. Henry (Cham: Springer International Publishing), 1–25.
- Brozynska, M., Copetti, D., Furtado, A., Wing, R. A., Crayn, D., Fox, G., et al. (2017). Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnol. J.* 15, 765–774. doi: 10.1111/pbi.12674
- Brozynska, M., Omar, E. S., Furtado, A., Crayn, D., Simon, B., Ishikawa, R., et al. (2014). Chloroplast genome of novel rice germplasm identified in Northern Australia. *Trop. Plant Biol.* 7, 111–120. doi: 10.1007/s12042-014-9142-8
- Chat, J., Jáuregui, B., Petit, R. J., and Nadot, S. (2004). Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *Am. J. Bot.* 91, 736–747. doi: 10.3732/ajb.91.5.736
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- Doyle, J. J., Doyle, J. L., Rauscher, J. T., and Brown, A. H. D. (2003). Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (Glycine subgenus Glycine). *New Phytol.* 161, 121–132. doi: 10.1046/j.1469-8137.2003.00949.x
- Furtado, A. (2014). “DNA extraction from vegetative tissue for next-generation sequencing,” in *Cereal Genomics: Methods and Protocols*, Vol. 1099, eds R. J. Henry and A. Furtado (Totowa, NJ: Humana Press Inc), 1–5. doi: 10.1007/978-1-62703-715-0\_1
- Ganeshan, S., Sharma, P., and Chibbar, R. N. (2010). “Functional genomics for crop improvement,” in *Molecular Techniques in Crop Improvement*, 2nd Edn, eds S. M. Jain and D. S. Brar (Dordrecht: Springer Netherlands), 63–95. doi: 10.1007/978-90-481-2967-6\_3
- Goldman, D. H., Jansen, R. K., van den Berg, C., Leitch, I. J., Fay, M. F., and Chase, M. W. (2004). Molecular and cytological examination of *Calopogon* (Orchidaceae, Epidendroideae): circumscription, phylogeny, polyploidy, and possible hybrid speciation. *Am. J. Bot.* 91, 707–723. doi: 10.3732/ajb.91.5.707
- Griffith, M. P. (2003). Using molecular evidence to elucidate reticulate evolution in *Opuntia*. *Madroño* 50, 162–169.
- Henry, R. J. (2019). Australian wild rice populations: a key resource for global food security. *Front. Plant Sci.* 10:1354. doi: 10.3389/fpls.2019.01354
- Henry, R. J., Rice, N., Waters, D. L. E., Kasem, S., Ishikawa, R., Hao, Y., et al. (2010). Australian *Oryza*: utility and conservation. *Rice* 3, 235–241. doi: 10.1007/s12284-009-9034-y
- Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521
- Juliano, A. B., Naredo, M. E. B., Lu, B.-R., and Jackson, M. T. (2005). Genetic differentiation in *Oryza meridionalis* Ng based on molecular and crossability

- analyses. *Genet. Resour. Crop Evol.* 52, 435–445. doi: 10.1007/s10722-005-2259-8
- Karimi, N., Grover, C. E., Gallagher, J. P., Wendel, J. F., Ané, C., and Baum, D. A. (2019). Reticulate evolution helps explain apparent homoplasmy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae). *Syst. Biol.* 69, 462–478. doi: 10.1093/sysbio/syz073
- Kasem, S., Waters, D. L. E., and Henry, R. J. (2012). Analysis of starch gene diversity in the wild relatives of *Oryza sativa*. *Trop. Plant Biol.* 5, 286–308. doi: 10.1007/s12042-012-9111-z
- Kasem, S., Waters, D. L. E., Ward, R. M., Rice, N. F., and Henry, R. J. (2014). Wild *Oryza* grain physico-chemical properties. *Trop. Plant Biol.* 7, 13–18. doi: 10.1007/s12042-013-9133-1
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Krishnan, S. G., Waters, D. L. E., and Henry, R. J. (2014). Australian wild rice reveals pre-domestication origin of polymorphism deserts in rice genome. *PLoS One* 9:e98843. doi: 10.1371/journal.pone.0098843
- Lemmon, A. R., Emme, S. A., and Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61, 727–744. doi: 10.1093/sysbio/sys049
- Li, J. Q., Lu, Y. G., Feng, J. H., Zhao, X. J., and Liu, X. D. (2007). The crossability and F1 hybrid fertility between *Oryza sativa* and other AA genome species. *J. Plant Genet. Resour.* 8, 1–6.
- Li, J., Zhou, J., Xu, P., Deng, X., Deng, W., Zhang, Y., et al. (2018). Mapping five novel interspecific hybrid sterility loci between *Oryza sativa* and *Oryza meridionalis*. *Breed. Sci.* 68, 516–523. doi: 10.1270/jsbbs.18001
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27(2 Pt 1), 209–220.
- Map Maker (2021). *Map Maker Make & Share Custom Maps*. Available online at: <https://maps.co/> (accessed August 12, 2021).
- Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi: 10.1093/molbev/msu269
- Mason-Gamer, R. J. (2004). Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Syst. Biol.* 53, 25–37. doi: 10.1080/10635150490424402
- Moner, A. M., Furtado, A., Chivers, I., Fox, G., Crayn, D., and Henry, R. J. (2018). Diversity and evolution of rice progenitors in Australia. *Ecol. Evol.* 8, 4360–4366. doi: 10.1002/ece3.3989
- Naredo, M. E. B., Juliano, A. B., Lu, B.-R., and Jackson, M. T. (1997). Hybridization of AA genome rice species from Asia and Australia I. Crosses and development of hybrids. *Genet. Resour. Crop Evol.* 44, 17–23. doi: 10.1023/A:1008670923628
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292.
- Ng, N. Q., Hawkes, J. G., Williams, J. T., and Chang, T. T. (1981). The recognition of a new species of rice (*Oryza*) from Australia. *Bot. J. Linn. Soc.* 82, 327–330. doi: 10.1111/j.1095-8339.1981.tb00964.x
- Nge, F. J., Biffin, E., Thiele, K. R., and Waycott, M. (2021). Reticulate evolution, ancient chloroplast haplotypes, and rapid radiation of the Australian plant genus *Adenanthos* (proteaceae). *Front. Ecol. Evol.* 8:616741. doi: 10.3389/fevo.2020.616741
- Oka, H. I. (1988). *Origin of Cultivated Rice*. Tokyo: Japan Scientific Societies Press.
- Palmer, J. D., Jorgensen, R. A., and Thompson, W. F. (1985). Chloroplast DNA variation and evolution in pismus: patterns of change and phylogenetic analysis. *Genetics* 109, 195–213. doi: 10.1093/genetics/109.1.195
- Petit, R. J., Bialozyt, R., Brewer, S., Cheddadi, R., and Combs, B. (2001). “From spatial patterns of genetic diversity to post glacial migration processes in forest trees,” in *Integrating Ecology and Evolution in a Spatial Context*, eds J. Silvertown and J. Antonovics (Oxford: Blackwell Science), 295–318.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- RAPDB (2021). *Rice Annotation Project Database*. Available online at: [https://rapdb.dna.affrc.go.jp/viewer/gbrowse\\_details/irgsp1?name=Os06g0229800;feature\\_id=50299](https://rapdb.dna.affrc.go.jp/viewer/gbrowse_details/irgsp1?name=Os06g0229800;feature_id=50299) (accessed June 19, 2021).
- Rieseberg, L. H. (1991). Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am. J. Bot.* 78, 1218–1237. doi: 10.2307/2444926
- Rieseberg, L. H., and Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* 5, 65–84.
- Rieseberg, L. H., and Wendel, J. F. (1993). “Introgression and its consequences in plants,” in *Hybrid Zones and the Evolutionary Process*, ed. R. G. Harrison (Oxford: Oxford University Press), 70.
- Sang, T., Crawford, D. J., and Stuessy, T. F. (1995). Documentation of reticulate evolution in Peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA — implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6813–6817. doi: 10.1073/pnas.92.15.6813
- Sotowa, M., Ootsuka, K., Kobayashi, Y., Hao, Y., Tanaka, K., Ichitani, K., et al. (2013). Molecular relationships between Australian annual wild rice, *Oryza meridionalis*, and two related perennial forms. *Rice* 6:26. doi: 10.1186/1939-8433-6-26
- Stebbins, G. L. (1959). The role of hybridization in evolution. *Proc. Am. Philos. Soc.* 103, 231–249.
- Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* 50, 285–296. doi: 10.1038/s41588-018-0040-0
- Sun, C. Q., Wang, X. K., Li, Z. C., Yoshimura, A., and Iwata, N. (2001). Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theor. Appl. Genet.* 102, 157–162. doi: 10.1007/s001220051631
- Swofford, D. L. (2003). *PAUP4. Phylogenetic Analysis Using Parsimony (4and Other Methods)*. Version 4. 2003. Sunderland, MA: Sinauer Associates.
- Takahata, N., and Slatkin, M. (1984). Mitochondrial gene flow. *Proc. Natl. Acad. Sci. U.S.A.* 81, 1764–1767. doi: 10.1073/pnas.81.6.1764
- Tang, L., Zou, X. H., Achoundong, G., Potgieter, C., Second, G., Zhang, D. Y., et al. (2010). Phylogeny and biogeography of the rice tribe (Oryzaceae): evidence from combined analysis of 20 chloroplast fragments. *Mol. Phylogenet. Evol.* 54, 266–277. doi: 10.1016/j.ympev.2009.08.007
- Tikapunya, T., Zou, W., Yu, W., Powell, P. O., Fox, G. P., Furtado, A., et al. (2017). Molecular structures and properties of starches of Australian wild rice. *Carbohydr. Polym.* 172, 213–222. doi: 10.1016/j.carbpol.2017.05.046
- Toyomoto, D., Uemura, M., Taura, S., Sato, T., Henry, R., Ishikawa, R., et al. (2019). Segregation distortion observed in the progeny of crosses between *Oryza sativa* and *O. meridionalis* caused by abortion during seed development. *Plants (Basel)* 8:398. doi: 10.3390/plants8100398
- Tsitrona, A., Kirkpatrick, M., and Levin, D. A. (2003). A model for chloroplast capture. *Evolution* 57, 1776–1782. doi: 10.1111/j.0014-3820.2003.tb00585.x
- Vaughan, D. A. (1989). *The Genus Oryza L.: Current Status of Taxonomy*, Vol. 138. Manila: IRRI.
- Vriesendorp, B., and Bakker, F. T. (2005). Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *New Trends Plant Syst.* 54, 593–604. doi: 10.2307/25065417
- Wambugu, P. W., Brozynska, M., Furtado, A., Waters, D. L., and Henry, R. J. (2015). Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* 5:13957. doi: 10.1038/srep13957
- Ward, B. J., and van Oosterhout, C. (2016). HYBRIDCHECK: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Mol. Ecol. Resour.* 16, 534–539. doi: 10.1111/1755-0998.12469
- Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., et al. (2014). Hyb-Seq: combining target enrichment and

genome skimming for plant phylogenomics. *Appl. Plant Sci.* 2:1400042. doi: 10.3732/apps.1400042

Wendel, J. F., and Doyle, J. J. (1998). "Phylogenetic incongruence: window into genome history and molecular evolution," in *Molecular Systematics of Plants II: DNA Sequencing*, eds D. E. Soltis, P. S. Soltis, and J. J. Doyle (Boston, MA: Springer US), 265–296. doi: 10.1007/978-1-4615-5419-6\_10

Wendel, J. F., Stewart, J. M., and Rettig, J. H. (1991). Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* 45, 694–711. doi: 10.1111/j.1558-5646.1991.tb04339.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hasan, Furtado and Henry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.