Philippine Journal of Science
150 (S1): 17-28, Special Issue on Biodiversity
ISSN 0031 - 7683
Date Received:27 Jul 2020

Genetic Diversity and Structure of *Oryza rufipogon* Griff. Populations in the Philippines

Sandy Jan E. Labarosa[1]*, **Neah Rosandra Sevilla**[2], **Dindo Agustin A. Tabanao**[2], **Nenita B. Baldo**[1], **Helen L.V. Ebuña**[1], and **Joy M. Jamago**[1]*

[1]Department of Agronomy and Plant Breeding, College of Agriculture
Central Mindanao University, Musuan, Maramag, Bukidnon 8710 Philippines
[2]Plant Breeding and Biotechnology Division, Philippine Rice Research Institute
Science City of Muñoz, Nueva Ecija 3119 Philippines

***Oryza rufipogon* Griff., or "Rufi" in the Philippines, was previously known to be found only in Lake Apo, Bukidnon. However, a new population was identified in Lake Napalit in the same province. A better understanding of the genetic diversity of both Rufi populations using molecular methods may be beneficial to further ascertain its usefulness in rice breeding and in the development of effective conservation strategies. Population genetic analysis was conducted to estimate the degree of genetic diversity and population structure of the two Philippine Rufi populations using 98 genome-wide simple sequence repeat (SSR) markers. Four *Oryza sativa indica* and three *O. sativa japonica* cultivars were added for comparison. Results indicate that both Rufi populations exhibit low genetic diversity but with significant population structure and differentiation. Low genetic diversity suggests that both populations might be in a genetic bottleneck, perhaps due to observed unsustainable farming practices near their habitat and lack of awareness among locals of their importance. Also, geographical isolation that prevented gene flow between the two populations, as well as the unique climatic conditions between the two lakes might have contributed to the significant population structuring and differentiation. Thus, *in situ* and *ex situ* conservation should be observed for both Rufi populations in the Philippines.**

Keywords: conservation genetics, crop wild relatives, diversity analysis, *Oryza sativa*, simple sequence repeats

INTRODUCTION

Oryza rufipogon Griff. (nicknamed "Rufi", AA genome, 2n = 2x = 24) is the ancestor of the cultivated Asian rice (*Oryza sativa* L.) (Song *et al.* 2005). For breeding, Rufi and 20 other wild rice species offer a reservoir of genetic diversity as essential sources of resistance and tolerance to some biotic and abiotic stresses, among others (Jacquemin *et al.* 2013). Rufi has resistance mechanisms to various pests and diseases, such as bacterial leaf streak (He *et al.* 2012), rice tungro disease (Shibata *et al.* 2007), and other biotic stresses (Shibata *et al.* 2007). It also has tolerance mechanisms to several environmental stresses like acid sulfate soils (Bui and Nguyen 2017) and chilling stress (Cen *et al.* 2018).

Bon and Borromeo (2003) reported that the only natural populations of Rufi in the Philippines were in Lake Apo, Barangay Guinoyoran, Valencia City, Bukidnon. However, in a class excursion by Jamago and her students in 2012, they recorded the existence of new natural populations along Lake Napalit, Barangay Pigtauranan, Pangantucan,

*Corresponding Authors: sandyjanlabarosa@gmail.com
f.joy.jamago@cmu.edu.ph

Bukidnon. This new population might harbor novel alleles useful for current and future rice breeding programs. Balos and Jamago (2013) compared some morpho-ecological parameters of Rufi populations in both lakes *in situ*. Their findings showed that Lake Apo populations (Apo) had longer leaves, culms, and awns than the Lake Napalit populations (Napalit). Also, they reported that Lake Napalit had higher average rainfall (73.50 mm) and was relatively colder (23.5 °C) than Lake Apo (average rainfall: 26.33 mm; average temperature: 25.3 °C). These variations in morphological and ecological parameters between the two populations offer an opportunity for researchers and breeders to explore for their possible use in rice breeding programs.

This study estimated the magnitude of genetic diversity and determined the population structure of *in situ* populations of *O. rufipogon* Griff. in the Philippines using SSRs. Further, allelic patterns and degree of population differentiation across the populations were also calculated. Information on Rufi's genetic diversity and population structure would provide more clarity on its potential for utilization and even insights for its conservation.

MATERIALS AND METHODS

Leaf Sampling and Genotyping

Twenty-seven and 14 Rufi strains from Lake Napalit (~ 1,041 masl) and Lake Apo (~ 640 masl) populations, respectively, were sampled *in situ* in 2014 spaced at least 30 m to minimize the collection of leaves from possible clonal plants. Leaf sampling was done by cutting 4–5 in from tips of flag leaves. Leaf cuttings per strain were secured in individual 200 ml centrifuge tubes that were placed immediately in an icebox before storing at –20 °C until DNA extraction. Total genomic DNA was extracted from lyophilized leaf tissues following the modified cetyltrimethylammonium bromide method of Perez and colleagues (2012), then dissolved in tris-ethylenediaminetetraacetic acid buffer. DNA samples were quality-checked through electrophoresis on 1% agarose gel. DNA stocks of four *O. sativa indica* (*i.e.* NSIC Rc192, PSB Rc14, PSB Rc18, and PSB Rc82) and three *O. sativa japonica* cultivars (*i.e.* Azucena, Li-Jiang-Xin-Tuan-Hei-Gu, and Nipponbare) from the Philippine Rice Research Institute were added for comparison.

One-hundred twenty-eight polymorphic genome-wide SSR markers for *O. sativa* were initially selected based on a 3-mb bin system to allow uniform marker distribution in the chromosomes. These were used to evaluate the genetic variation of 41 Rufi samples, including seven *O. sativa* cultivars, for comparison. Polymerase chain reaction (PCR) was conducted per 6-µl reaction solution composed of 1.55 µl sdH_20, 1.5 µl 10x PCR buffer, 0.6 µl 25 mM magnesium chloride, 0.35 µl 5mM deoxynucleoside triphosphates, 0.5 µl forward primer, 0.5 µl reverse primer, 1.0 µl 1:100 (concentrated suspension: storage buffer) *Taq* DNA polymerase, and 1.5 µl diluted DNA.

Each locus was amplified using a PTC-100® Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA, USA) for 2 h and 30 min, following the PCR profile: 5 min initial denaturation at 94 °C, followed by 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 58 °C, and 2 min extension at 72 °C; then a final extension at 72 °C for 5 min. PCR products were subjected to 6% polyacrylamide gel electrophoresis in 1x TBE buffer at 90 V for 1–2 h depending on the size of PCR products. Gels were stained using an ethidium bromide solution for 10–15 m. DNA bands were visualized using the AlphaImager® HP system (ProteinSimple, San Jose, CA, USA). Detected allelic bands were scored from the fastest to the slowest migrating alleles by letters *a*, *b*, *c*, and so on.

Data Analysis

A total of 128 SSR markers were initially selected to assess the genetic diversity and population structure of 41 Rufi samples along with seven rice cultivars for comparison, which represented the two *O. sativa* subspecies, *i.e. japonica* and *indica*. SSR markers with > 20% missing band across all samples were excluded and only 98 markers were retained for further analysis (Appendix I). Rufi samples per lake were examined for genetic diversity per population. Genetic diversity indices per population were assessed by calculating the mean number of alleles, observed heterozygosity (H_o), expected heterozygosity (H_e), Shannon information index (I), and fixation index (F). Calculations were performed using GenAlEx ver. 6.5 (Peakall and Smouse 2012). In addition, allelic richness (AR) and private allelic richness (PAR) were calculated using the rarefaction method implemented in HP-Rare v1.1 (Kalinowski 2005). This accounts for the differences in sample size between the pre-defined population groups of Rufi and *O. sativa*.

Moreover, the population structure of Rufi and rice populations were evaluated by phylogenetic analysis. The pairwise genetic distance matrix for all 48 genotypes was constructed using Nei's genetic distance (Nei 1972). Neighbor-joining (NJ) tree was used for cluster analysis. A boot function from "poppr" package (Kamvar *et al*. 2015) was used to build and visualize the unrooted dendrogram following 1000 bootstraps. The analysis was done using R (R Core Team 2020). Two models were implemented to determine the patterns of the genetic structure of populations. Model 1 only analyzed the Apo and Napalit Rufi populations, whereas Model 2 included the *japonica*

and *indica* rice cultivars. A systemic Bayesian clustering approach applying Markov Chain Monte Carlo (MCMC) estimation was performed using STRUCTURE ver. 2.3.2 (Pritchard *et al*. 2000). The analysis was set using the admixture ancestry model, with the number of assumed populations (K) set from 1–10, and each was run 10x. Each run started with 500,000 burn-in periods followed by 500,000 MCMC iterations. MCMC was ran using a correlated allele frequency model based on the default frequency model information under the advanced option of STRUCTURE. Results from STRUCTURE were collated and imported to the web-based application Structure Harvester (Earl and von Holdt 2012) to calculate for Evanno's delta K (Evanno *et al*. 2005), which determined the optimal number of genetic cluster groups.

Locus-by-locus analysis of molecular variance (AMOVA) with 1000 permutations for Models 1 and 2 was calculated to determine the primary source of genetic variation at each locus. Further, pairwise F_{st} analysis (Weir and Cockerham 1984) for Model 2 was used to determine the degree of genetic divergence among populations. Both AMOVA and pairwise F_{st} were calculated using Arlequin ver. 3.5.2.2 (Excoffier and Lischer 2010). Pairwise F_{st} results were visualized using the R script developed by Wong (2017).

RESULTS

Genetic Diversity Analysis

AR and PAR corrected for sample size, H_o, H_e, I, and F were calculated to describe the genetic diversity of Rufi populations in comparison to *O. sativa* accessions. Results suggested that Napalit populations of Rufi were more genetically diverse than Apo populations (Table 1). However, the genetic diversity of Rufi in both populations was lower compared to the cultivated rice accessions in both the *japonica* and *indica* group (Table 1). This higher diversity observed among *O. sativa* compared to Rufi might be due to its diverse countries of origin, and which were also bred for different ecotypes and came from different genetic background. The former is different from the natural populations of Rufi – which were geographically isolated, with restricted gene flow, and might have been predominantly inbred – resulting in heterozygote deficiency.

Population Structure and Differentiation

Population genetic structure was analyzed using the NJ tree based on Nei's genetic distance. Unrooted NJ tree showed a clear separation between the two *Oryza* species, where both Rufi populations were genetically closer to each other than to the rice cultivars (Figure 1a). However, Nap13 from Napalit appeared to be an intermediate between Rufi and *O. sativa* accessions, suggesting that this strain is either a hybrid between the two species or a weedy variant.

Furthermore, Bayesian clustering results showed a major peak of Evanno's ΔK at $K = 2$ for Model 1 (without *O. sativa* accessions; Appendix II) and a major peak at $K = 8$ for Model 2 (with *O. sativa* accessions; Appendix II). Two genetic clusters were inferred for Model 1, as represented by the colors green and blue. Most of the Apo samples belonged to the green cluster, with five admixed individuals (*i.e.* Apo1, Apo5, Apo6, Apo7, and Apo8) detected (Figure 1b). On the other hand, Napalit samples had cluster membership of more than 70% strains in the blue cluster – with Nap3, Nap5, and Nap19 as admixed genotypes (Figure 1b).

In contrast, eight genetic clusters were inferred for Model 2, each represented by a color (Figure 1c). These results show that *indica* cultivars have at least 98% cluster membership per strain to the red cluster, whereas

Table 1. Genetic diversity parameters showing that Rufi populations have lower diversity compared to the *O. sativa* cultivars.

Population		AR	PAR	Ho	He	I	F
Apo	Mean	1.17	0.21	0.079	0.167	0.277	0.487
	SE			0.020	0.021	0.034	0.060
Napalit	Mean	1.19	0.25	0.079	0.191	0.321	0.549
	SE			0.019	0.021	0.033	0.053
Japonica	Mean	1.37	0.31	0.024	0.211	0.308	0.863
	SE			0.012	0.024	0.035	0.042
Indica	Mean	1.57	0.59	0.026	0.446	0.699	0.938
	SE			0.012	0.020	0.035	0.025

Note: SE – standard error; AR – allelic richness corrected for sample size; PAR – private allelic richness corrected for sample size; H_o – observed heterozygosity; H_e – expected heterozygosity; I – Shannon Information Index; F – fixation/inbreeding index

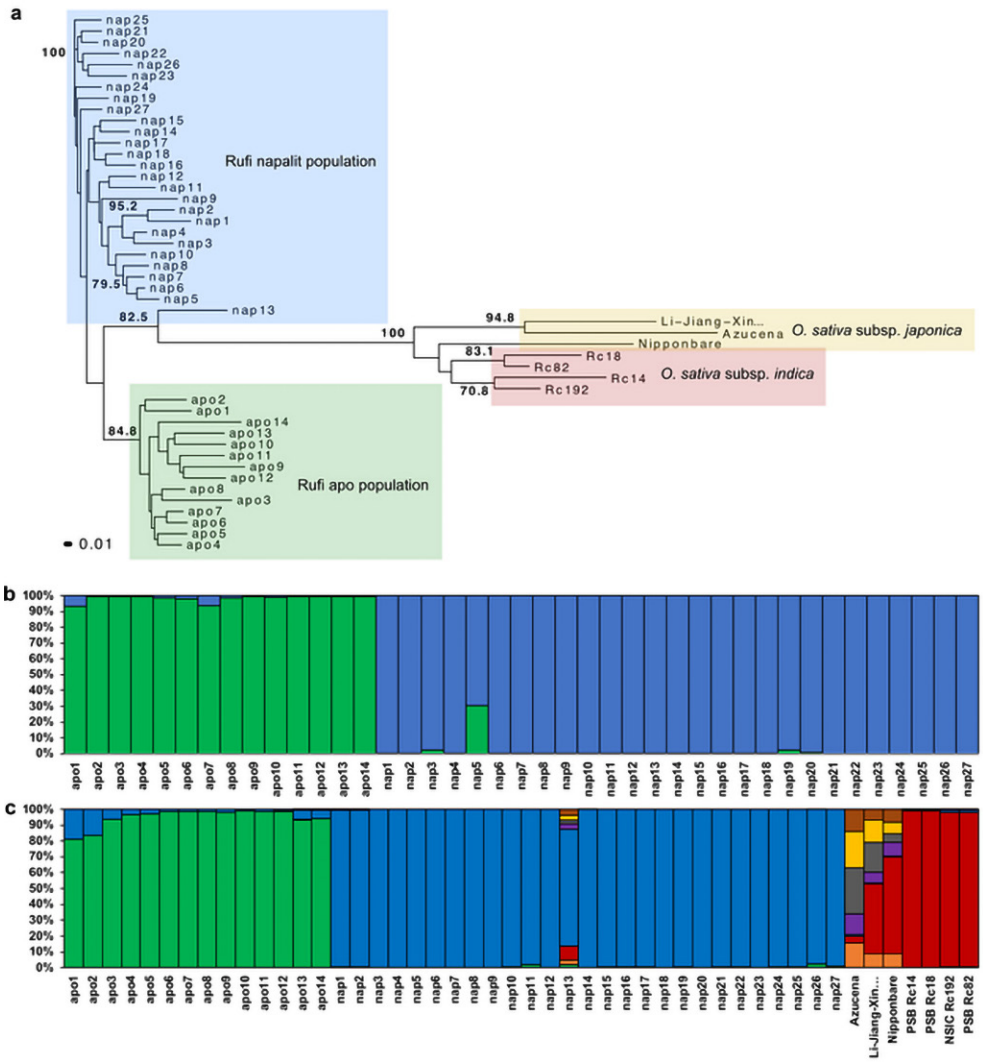

Figure 1. (a) Unrooted NJ tree based on Nei's genetic distance from 98 SSR loci. Numbers above branches indicate bootstrap support values above 70%. (b–c) Bayesian clustering showed the estimated proportion of membership (%) to inferred genetic cluster for Model 1 (b) and Model 2 (c). Each color represents one genetic cluster.

japonica cultivars are admixed genotypes of six clusters: brown, yellow, gray, purple, red, and orange. Such might actually show that the tested modern *indica* cultivars were developed from a common parent, whereas the traditional *japonica* cultivars showed diverse origins from an evolutionary perspective. In addition, most of the Apo strains had a membership of more than 80% genotypes to the green cluster and all were admixed with the blue cluster. On the other hand, most of Napalit strains had a cluster membership of more than 80% to the blue cluster. Moreover, Nap11 and Nap26 were admixed individuals to the green cluster, whereas Nap13 was admixed with the other clusters. This is also consistent with the result of the NJ tree. This further suggests that Nap13 could be either a hybrid between Rufi and *O. sativa*, specifically the *japonica* group, or a weedy variant.

However, the geographical distribution of genetic clusters in Lake Napalit (Figure 2) showed that the nearest rice field from Nap13 is ~ 200–300 m. Also, Nap4, Nap5, and Nap6 – which had direct connection to the rice field *via* a small creek – did not show any admixture gradient with *O. sativa*. This indicates that direct gene flow from this rice field to the Rufi population is probably non-existent, and the admixture might be due to other unknown causes for now. When combined with the NJ tree results, Bayesian clustering indicated a robust genetic structuring of all Rufi individuals, with both Rufi populations distinct from each other and from *O. sativa* groups.

AMOVA for Model 1 determined that variation within

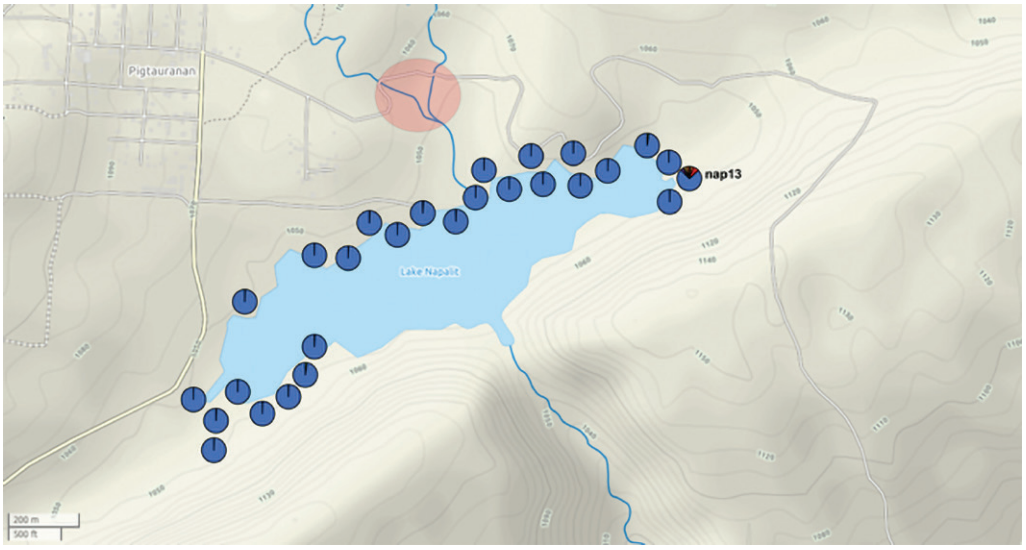

Figure 2. Geographic distribution of the STRUCTURE clusters of Rufi samples for Model 2 (K = 8) in Lake Napalit (n = 27). Red area indicates the approximate location of the nearest rice field in Lake Napalit. Maps were generated using http://umap.openstreetmap.fr/en/.

populations contributed greater than the variation among populations in differentiating the two Rufi populations (Table 2). This might indicate that specific adaptations to the local conditions per lake or specific mutations among individuals within the population have caused this distinction. For Model 2 AMOVA, variation among populations contributed higher to the genetic separation of different Rufi and *O. sativa* groups compared to variation within populations (Table 2). This suggests that the domestication process, plant breeding, and perhaps genetic drift contributed to this differentiation.

Pairwise F_{st} of all defined populations, including *O. sativa* groups, showed that Napalit and Apo Rufi populations are very genetically distinct ($P < 0.05$) from the *japonica* (*vs.* Apo: $F_{st} = 0.68956$; *vs.* Napalit: $F_{st} = 0.64159$) and *indica* (*vs.* Apo: $F_{st} = 0.6454$; *vs.* Napalit: $F_{st} = 0.61761$) groups (Figure 3). Moreover, of the two Rufi populations, Napalit is genetically closer to the *O. sativa* groups, perhaps because of Nap13, specifically to the *japonica* group.

DISCUSSION

Genetic Diversity and Differentiation

This is the first study to report on the molecular genetic diversity of the two in situ Rufi populations in the Philippines. Both Rufi populations possess low SSR variation compared to Rufi populations from China (Na = 3.7, Ho = 0.57, He = 0.46; Wang et al. 2020) and Malaysia

Table 2. Locus-by-locus AMOVA summary table for both Model 1 and Model 2.

Source of variation	Sum of squares	Variance component	Variation (%)	*P*-value
Model 1 (average for 78 loci)				
Among populations	247.380	6.89851	42.97198	0.00000
Within populations	693.712	9.15500	57.02802	
Total	941.091	16.05350		
Model 2 (average for 95 loci*)				
Among populations	726.356	13.13295	56.01069	0.00000
Within populations	898.641	10.31427	43.98931	
Total	1624.997	23.44723		

*RM12460 and RM473 have no amplicon for all individuals within the japonica and indica groups while RM219 has no amplicon for individuals within the japonica group; thus, these three loci were excluded from the analysis.

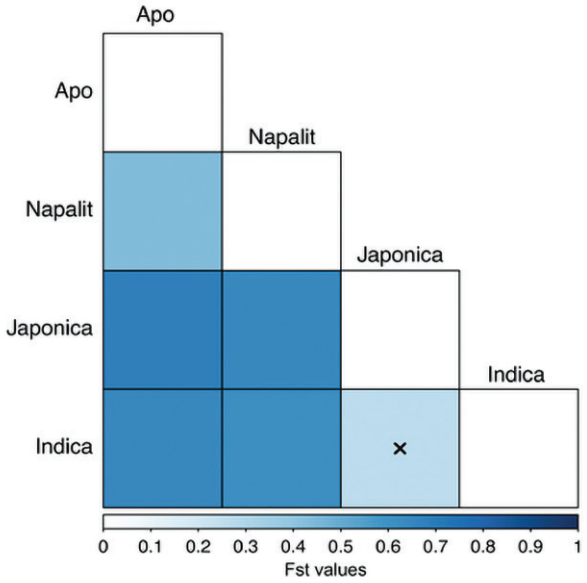

Figure 3. Pairwise F_{st} information between the pre-defined populations for Rufi and *O. sativa* showed that Napalit and Apo Rufi populations are very genetically distinct from *O. sativa* groups. Note: the darker the color, the higher the F_{st} values, the more genetically distinct the population is to its pair. Box marked with "x" is not significantly different at ($P < 0.05$).

(Na = 3.98, Ho = 0.583, He = 0.673; Ngu et al. 2010). In addition, Rufi populations in the Philippines might be experiencing or had experienced genetic bottlenecks which may explain its low genetic diversity. Habitat fragmentation of both Rufi populations may have had negatively impacted its genetic diversity by limiting gene flow between the two populations (Chen et al. 2017), increased incidence of inbreeding (Gao and Gao 2016), and rapid genetic erosion due to observed human activities such as fishing, grazing of buffalo, and tourism along the lakes.

The present study also detected a high level of population differentiation despite having low genetic diversity. This high differentiation can be attributed to the increased incidence of inbreeding – probably selfing, resulting in genetic drift within the population (Fan *et al*. 2016) – as evident in low heterozygosity and high F values for the two populations (Bower and Hipkins 2017). However, the result of AMOVA for Model 1 indicates that variation within the population has the greatest contribution to the genetic differentiation between the two populations. This means that natural selection and adaptation due to differences in elevation, temperature, and other parameters in each lake ecosystem may be the primary force in the genetic differentiation of the two populations (Gao and Gao 2016).

In addition, the study also reported high genetic differentiation between populations of Rufi and *O. sativa* groups based on the AMOVA results for Model 2 and pairwise F_{st} values. These results corroborated with the results of past studies conducted in Rufi populations in China that also showed high genetic differentiation (Wang *et al*. 2020). F_{st} results indicate that Rufi populations, especially Napalit, is genetically closer to the *japonica* group, which is also similar to the results of Xie and colleagues (2010).

Implication for Conservation and Management of Rufi

Genetic diversity of crops and its crop wild relatives is an important evolutionary aspect of a species to cope with a changing environment (He *et al*. 2019), as well as its utilization in plant breeding programs (Lopes *et al*. 2015). This study had revealed that both Philippine Rufi populations have low genetic diversity and might be in a genetic bottleneck. If undesirable human activities are left unchecked and unregulated, these *in situ* populations may eventually be lost. *Ex situ* conservation for Rufi from Lake Apo has been around for years now in the genebank of the International Rice Research Institute (accession numbers: IRGC 1010785, IRGC 105568, and IRGC 80774) and from Lake Napalit by the Genetic Resources Division of the Philippine Rice Research Institute. However, *ex situ* conservation for threatened populations alone is not enough, as Xie and colleagues (2010) reported that *ex situ* collection of Rufi in China exhibited lower genetic diversity compared to *in situ* counterparts – despite suffering from population decline over the years. Thus, a combination of *in situ* and *ex situ* conservation strategy must be in place for both populations. Moreover, Xie and colleagues (2010) suggested that reintroduction of germplasm from an existing *ex situ* collection may promote the long-term maintenance of genetic diversity and improve the gene flow of *in situ* populations; hence, this may be done for the Apo population due to its smaller population size and genetic diversity than in Lake Napalit. Nonetheless, spatial isolation of both Rufi populations from cultivated rice must also be carefully considered to effectively maintain their genetic integrity.

CONCLUSION

Our present study on the genetic diversity and structure of Rufi populations in the Philippines revealed that: 1) Rufi populations in the Philippines exhibit low genetic diversity, perhaps due to geographical isolation, increased incidence of inbreeding, and genetic drift; 2) the two populations of Rufi are very distinct from each other, probably because of natural selection and specific adaptation to the lake's unique climatic conditions; 3) the Rufi population from Napalit is closely related to *O. sativa japonica*; and 4) both Rufi populations are in a serious

genetic bottleneck and implementation of both *ex situ* and *in situ* conservation approaches may prevent local extinction of this species in the Philippines.

ACKNOWLEDGMENTS

We would like to thank the local government units of Brgy. Guinoyoran, Valencia City, and Brgy. Pigtauranan, Pangantucan in Bukidnon for the consent to establish and conduct the study in Lake Apo and Lake Napalit, respectively. We would also like to express our gratitude to the Philippine Rice Research Institute – Central Experiment Station for supporting and funding this research project.

STATEMENT ON CONFLICT OF INTEREST

We declare no conflict of interest.

NOTES ON APPENDICES

The complete appendices section of the study is accessible at https://1drv.ms/u/s!AsD1AfPqmMeCgugfaF9VrecxJ5BdHg?e=dv5jyI

REFERENCES

BALOS JL, JAMAGO JM. 2013. Comparative characterization of in situ *Oryza rufipogon* Griff. populations in Lakes Apo and Napalit, Bukidnon. In: Dayrit FM ed. Transactions of the National Academy of Science and Technology at the 35th NAST Annual Scientific Meeting; 2013 10–11 July; Manila, Philippines, 14p.

BON SG, BORROMEO TH. 2003. Discovery & Re-Discovery Of Wild Rice Populations In The Philippines. Philippine Journal of Crop Science 27(1): 53–58. https://www.cabi.org/gara/FullTextPDF/2009/20093019283.pdf

BOWER AD, HIPKINS V. 2017. Genetic Diversity and Population Structure in the Rare, Endemic Baker Cypress (*Hesperocyparis bakeri*). Madroño 64(2): 71–82. https://doi.org/10.3120/0024-9637-64.2.71

BUI C, NGUYEN T. 2017. QTL analysis on rice genotypes adapted to acid sulfate soils in the Mekong river delta, Vietnam. Vietnam Journal of Science, Technology, and Engineering 59(4): 26-31.

CEN W, LIU J, LU S, JIA P, YU K, HAN Y, LI R, LUO J. 2018. Comparative proteomic analysis of QTL CTS-12 derived from wild rice (*Oryza rufipogon* Griff.), in the regulation of cold acclimation and de-acclimation of rice (*Oryza sativa* L.) in response to severe chilling stress. BMC Plant Biology 18(1): 1–17. https://doi.org/10.1186/s12870-018-1381-7

CHEN Y, LIU Y, FAN X, LI W, LIU Y. 2017. Landscape-Scale Genetic Structure of Wild Rice *Zizania latifolia*: The Roles of Rivers, Mountains and Fragmentation. Frontiers in Ecology and Evolution 5: 17. https://doi.org/10.3389/fevo.2017.00017

EARL DA, VON HOLDT BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2): 359–361. DOI: 10.1007/s12686-011-9548-7

EVANNO G, REGNAUT S, GOUDET J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14: 2611–2620. DOI: 10.1111/j.1365-29X.2005.02553.x

EXCOFFIER L, LISCHER HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564–567

FAN XR, REN XR, LIU YL, CHEN YY. 2016. Genetic structure of wild rice *Zizania latifolia* and the implications for its management in the Sanjiang Plain, Northeast China. Biochemical Systematics and Ecology 64: 81–88. https://doi.org/10.1016/j.bse.2015.11.017

GAO L-Z, GAO C-W. 2016. Lowered Diversity and Increased Inbreeding Depression within Peripheral Populations of Wild Rice *Oryza rufipogon*. PLOS ONE 11(3). https://doi.org/10.1371/journal.pone.0150468

HE F, PASAM R, SHI F, KANT S, KEEBLE-GAGNERE G, KAY P, FORREST K, FRITZ A, HUCL P, WIEBE K, KNOX R, CUTHBERT R, POZNIAK C, AKHUNOVA A, MORRELL PL, DAVIES JP, WEBB SR, SPANGENBERG G, HAYES B, …AKHUNOV E. 2019. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nature Genetics 51(5): 896–904. https://doi.org/10.1038/s41588-019-0382-2

HE WA, HUANG DH, LI RB, QIU Y. FU, SONG, JD, YANG HN, ZHENG JX, HUANG YY, LI XQ, LIU C, ZHANG YX, MA ZF, YANG Y. 2012. Identification of a Resistance Gene bls1 to Bacterial Leaf Streak in Wild Rice *Oryza rufipogon* Griff. Journal of Integrative Agriculture 11(6): 962–969. https://doi.org/10.1016/S2095-3119(12)60087-2

JACQUEMIN J, BHATIA D, SINGH K, WING RA. 2013. The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol 16: 147–156. 10.1016/j.pbi.2013.02.014

KALINOWSKI ST. 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5: 187–189.

KAMVAR ZN, BROOKS JC, GRÜNWALD NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics 6: 208. https://doi.org/10.3389/fgene.2015.00208

LOPES MS, EL-BASYONI I, BAENZIGER PS, SINGH S, ROYO C, OZBEK K, AKTAS H, OZER E, OZDEMIR F, MANICKAVELU A, BAN T, VIKRAM P. 2015. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. Journal of Experimental Botany 66(12): 3477–3486. https://doi.org/10.1093/jxb/erv122

NEI M. 1972. Genetic distance between populations. American Naturalist 106: 283–392.

NGU MS, SABU KK, LIM LS, ABDULLAH MZ, WICKNESWARI R. 2010. Genetic Structure of *Oryza rufipogon* Griff. Natural Populations in Malaysia: Implications for Conservation and Genetic Introgression of Cultivated Rice. Tropical Plant Biology 3(4): 227–239. https://doi.org/10.1007/s12042-010-9060-3

PEAKALL R, SMOUSE PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537–2539.

PEREZ LM, PASTOR HM, DOMINGO JM, MANANGHAYA TE, TABANAO DA, MANIGBAS NL. 2012. Detecting Hybrid Admixtures Through DNA Analysis. Philippine Journal of Crop Science 37(1): 42–46.

PRITCHARD T, JAMJOD S, DONNELLY P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

R CORE TEAM. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

SHIBATA Y, CABUNAGAN RC, CABAUATAN PQ, IL-RYONG C. 2007. Characterization of *Oryza rufipogon*-derived resistance to tungro disease in rice. Plant Disease - PLANT DIS 91: 1386–1391. 10.1094/PDIS-91-11-1386

SONG ZP, LI B, CHEN JK, LU BR. 2005. Genetic diversity and conservation of common wild rice (*Oryza rufipogon*) in China. Plant Species Biol 20: 83–92.

WANG J, SHI J, LIU S, SUN X, HUANG J, QIAO W, CHENG Y, ZHANG L, ZHENG X, YANG Q. 2020. Conservation recommendations for *Oryza rufipogon* Griff. in China based on genetic diversity analysis. Scientific Reports 10(1): 14375. https://doi.org/10.1038/s41598-020-70989-w

WEIR BS, COCKERHAM CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

WONG W. 2017. Extract FST Mat data from Arlequin XML output. Retrieved on 04 Oct 2020 from https://nicolawongwaiyee.wordpress.com/.

XIE J, AGRAMA HA, KONG D, ZHUANG J, HU B, WAN Y, YAN W. 2010. Genetic diversity associated with conservation of endangered Dongxiang wild rice (*Oryza rufipogon*). Genetic Resources and Crop Evolution 57(4): 597–609. https://doi.org/10.1007/s10722-009-9498-z

Appendix I. Marker information of the selected 98 polymorphic simple sequence repeats (SSR) loci. Note: CN = chromosome number, EPP = expected PCR product size (bp), AT = annealing temperature (ºC), and RM = repeat motif. All information here were downloaded from Gramene Markers Database (https://archive.gramene.org/markers/).

SSR Loci	Chromosome number	Expected PCR product size	Annealing temperature (ºC)	Primer sequence		Repeat motif
				Forward	Reverse	
RM101	12	324	55	gtgaatggtcaagtgacttaggtggc	acacaacatgttccctcccatgc	(CT)37
RM105	9	134	55	gtcgtcgacccatcggagccac	tggtcgaggtggggatcgggtc	(CCT)6
RM107	9	189	67	agatcgaagcatcgcgcccgag	actgcgtcctctgggttcccgg	(GA)7
RM12300	2	156		AGTCAGGCAGACAACACCATCG	TGCTGTTCATCTGTGTCAGTCAGC	(CT)14
RM12460	2	168		TGGCACTACAGTGACAACAAACC	AGGGACTTTATCCAAAGGACACG	(TC)16
RM134	7	93	55	acaaggccgcgagaggattccg	gctctccggtggctccgattgg	(CCA)7
RM13500	2	183		CCTTAACCACTTCATCACTTCTCTCG	CCAACACCAACCATTAAGCAAAGG	(CT)16
RM146	5	345	55	ctattattccctaacccccataccctcc	agagccactgcctgcaaggccc	(CT)11-(CT)7
RM15254	3	148		GCACTCAGTCTCAGCTAGCTTTCAGC	AGGGCGGTCGGTTACTTGACG	(CAGC)5
RM16426	4	197		agaacaccatgtccacccacacc	agctcgccaagatgctgatcg	(CCG)7
RM16434	4	152		aggccatccaacaaactagtcacc	gtcagctatttaggctccacttgc	(TC)11
RM16446	4	176		gacagcgagcggtacaggttgg	gtcgtcgtcgtcgtctaggaagg	(CCG)8
RM171	10	328	55	AACGCGAGGACACGTACTTAC	ACGAGATACGTACGCCTTTG	(GATG)5
RM172	7	159	55	tgcagctgcgccacagccatag	caaccacgacaccgccgtgttg	(AGG)6
RM17305	4	145		gagatcgaagacgatcgagatgc	gggatttgtccagtactcctcttcc	(GA)12
RM178	5	117	67	tcgcgtgaaagataagcggcgc	gatcaccgttccctccgcctgc	(GA)5(AG)8
RM181	11	240	67			(CT)13(AT)19
RM184	10	219	55	atcccattcgccaaaaccggcc	tgacacttggagagcggtgtgg	(CA)7
RM19	12	226	55	caaaaacagagcagatgac	ctcaagatggacgccaaga	(ATC)10
RM192	7	267	61	gcggcggatcatgaattgcgag	cttgttccccggcgtcgagtcc	(TGG)5
RM193	6	189	61	cgcctcttcttcctcgcctccg	cgggtccatcccccctctcctc	(GCT)5
RM197	6	106	61	gatccgtttttgctgtgccc	cctcctctccgccgatcctg	(ACC)7
RM205	9	122	55	ctggttctgtatgggagcag	ctggccttcacgtttcagtg	(CT)25
RM206	11	147	55	cccatgcgtttaactattct	cgttccatcgatccgtatgg	(CT)21
RM209	11	1134	55	atatgagttgctgtcgtgcg	caacttgcatcctcccctcc	(CT)18
RM212	1	136	55	ccactttcagctactaccag	cacccatttgtctctcattatg	(CT)24
RM213	2	139	55	atctgtttgcaggggacaag	aggtctagacgatgtcgtga	(CT)17

SSR Loci	Chromosome number	Expected PCR product size	Annealing temperature (°C)	Primer sequence		Repeat motif
RM215	9	148	55	caaaatggagcagcaagagc	tgagcacctccttctctgtag	(CT)16
RM219	9	202	55	cgtcggatgatgtaaagcct	catatcggcattcgcctg	(CT)17
RM223	8	165	55	gagtgagcttgggctgaaac	gaaggcaagtcttggcactg	(CT)25
RM229	11	116	55	cactcacacgaacgactgac	cgcaggttcttgtgaaatgt	(TC)11(CT)5C33(CT)5
RM23654	9	186		CTCCGATGCCTTCTTCCTCTTGC	AAAGGGAGTAGCAAGCCGAGTGG	(CTT)9
RM241	4	138	55	gagccaaataagatcgctga	tgcaagcagcagatttagtg	(CT)31
RM25121	10	133		GGTATGACATGTGGCTCCTACCG	CGATCTAAGAGCGTTTGCAGTGG	(TCTA)5
RM256	8	127	55	gacagggagtgattgaaggc	gttgatttcgccaagggc	(CT)21
RM257	9	147	55	cagttccgagcaagagtactc	ggatcggacgtggcatatg	(CT)24
RM258	10	148	55	tgctgtatgtagctcgcacc	tggcctttaaagctgtcgc	(GA)21(GGA)3
RM26	5	112	55	gagtcgacgagcggcaga	ctgcgagcgacggtaaca	(GA)15
RM26547	11	459		TTGATTCGTTTCTTGCGCTACC	GCTCCAGGAGAGTTAATAGAGCTTCG	(AG)10
RM26801	11	165		GCCTTCATCCGTAAATCCATAAGC	GAGTACCACATGGCATTATGAGAGC	(TCTA)7
RM27191	11	372		CGGGAAGTCCTTGTAGGTGATGG	TCTGGATCATCATGTCACCATGC	(ACG)8
RM277	12	124	55	cggtcaaatcatcacctgac	caaggcttgcaagggaag	(GA)11
RM28048	12	93		TTCAGCCGATCCATTCAATTCC	GCTATTGGCCGGAAAGTAGTTAGC	(CGC)8
RM281	8	138	55	accaagcatccagtgaccag	gttcttcatacagtccacatg	(GA)21
RM28130	12	175		CAGCAGACGTTCCGGTTCTACTCG	AGGACGGTGGTGGTGATCTGG	(GAG)7
RM282	3	136	55	ctgtgtcgaaaggctgcac	cagtcctgtgttgcagcaag	(GA)15
RM284	8	141	55	atctctgatactccatccatcc	cctgtacgttgatccgaagc	(GA)8
RM287	11	118	55	ttccctgttaagagagaaatc	gtgtatttggtgaaagcaac	(GA)21
RM288	9	125	55	ccggtcagttcaagctctg	acgtacggacgtgacgac	(GA)7G6(GA)7
RM289	5	108	55	ttccatggcacacaagcc	ctgtgcacgaacttccaaag	G11(GA)16
RM300	2	121	55	gcttaaggacttctgcgaacc	caacagcgatccacatcatc	(GTT)14
RM301	2	153	55	ttactctttgtgtgtgtgag	ctacgacacgtcatagatgacc	(GT)5G2(GT)8T2(GT)3
RM310	8	105	55	ccaaaacatttaaaatatcatg	gcttgttggtcattaccattc	(GT)19
RM3143	1	98	55	AGCCTGGATAAGATGGTTCG	CGAGAAGACCCAGTTTCTGC	(CA)17
RM337	8	192	55	gtaggaaaggaagggcagag	cgatagatagctagatgtggcc	(CTT)4-19-(CTT)8
RM349	4	136	55	ttgccattcgcgtggaggcg	gtccatcatccctatggtcg	(GA)16
RM3549	2	168	55	GATCCTCCACACCCAACAAC	AACGAACGACCAACTCCAAG	(GA)12

SSR Loci	Chromosome number	Expected PCR product size	Annealing temperature (°C)	Primer sequence		Repeat motif
RM3658	4	169	55	AGGCGAGGGTACGGATCTC	TTTCTCTCTCCTCCTTCCCC	(GA)14
RM3735	4	138	55	GCGACCGATCAGCTAGCTAG	ATAACTCCTCCCTTGCTGCC	(GA)16
RM3843	4	172	55	ACCCTACTCCCAACAGTCCC	GGGGTCGTACGCTCATGTC	(GA)23
RM402	6	133	55	gagccatggaaagatgcatg	tcagctggcctatgacaatg	(ATA)7
RM407	8	172	55	gattgaggagacgagccatc	ctttttcagatctgcgctcc	(AG)13
RM416	3	114	55	gggagttagggttttggagc	tccagtttcacactgcttcg	(GA)9
RM4194						
RM422	3	385	55	ttcaacctgcatccgctc	ccatccaaatcagcaacagc	(AG)30
RM432	7	187	55	ttctgtctcacgctggattg	agctgcgtacgtgatgaatg	(CATC)9
RM434	9	152	55	gcctcatccctctaaccctc	caagaaagatcagtgcgtgg	(TC)12
RM441	11	189	55	acaccagagagagagagagagag	tctgcaacggctgatagatg	(AG)13
RM444	9	162	55	gctccacctgcttaagcatc	tgaagaccatgttctgcagg	(AT)12
RM447	8	111	55	cccttgtgctgtctcctctc	acgggcttcttctccttctc	(CTT)8
RM461	6	195	55	gagaccggagagacaactgc	tgatgcggtttgactgctac	(AAAC)6
RM470	4	83	55	tcctcatcggcttcttcttc	agaacccgttctacgtcacg	(CTT)14
RM473	7	97	55	tatcctcgtctccatcgctc	aaggatgtggcggtagaatg	(TCTA)14
RM486	1	104	55	ccccctctctctctctctc	tagccacatcaacagcttgc	(CT)14
RM510	6	122	55	aaccggattagtttctcgcc	tgaggacgacgagcagattc	(GA)15
RM511	12	130	55	cttcgatccggtgacgac	aacgaaagcgaagctgtctc	(GAC)7
RM524	9	198	55	tgaagagcaggaaccgtagg	tctgatatcggttccttcgg	(AT)11
RM529	1	273	55	ccctcccttctgtaagctcc	gaagaacaatggggttctgg	(CT)12
RM530	2	161	55	gcactgaccacgactgtttg	accgtaacccggatctatcc	(GA)23
RM536	11	243	55	tctctcctcttgtttggctc	acacaccaacacgaccacac	(CT)16
RM544	8	248	55	tgtgagcctgagcaataacg	gaagcgtgtgatatcgcatg	(TC)9
RM546	3	268	55	gagatgtagacgtagacggcg	gatcatcgtccttcctctgc	(CCT)7
RM551	4	192	55	agcccagactagcatgattg	gaaggcgagaaggatcacag	(AG)18
RM5515	9	123	50	CGCACAGGAGAAAATCCAAC	TAGCATTATGGGGTGGAAGG	(TG)12
RM552	11	195	55	cgcagttgtggatttcagtg	tgctcaacgtttgactgtcc	(TAT)13
RM556	8	93	55	actccaaacctcactgcacc	tagcacactgaacagctggc	(CCAG)6
RM5639	3	123	55	GGAAGAACAGAGTTGCTCGG	GTGCCATTTATTTCCGTCCC	(AAG)13

SSR Loci	Chromosome number	Expected PCR product size	Annealing temperature (°C)	Primer sequence		Repeat motif
RM5777	9	198	55	AAAGATCGAGATGGCGAGAG	TCTCACCGGTCAAAATCCTC	(AGA)11
RM590	10	137	55	catctcggctctccatgc	ggagttgggggtcttgttcg	(TCT)10
RM6	2	163	55	gtcccctccacccaattc	tcgtctactgttggctgcac	(AG)16
RM6063	7	157	55	GTCGCTAGTTCACTGCCAAG	TGGGGGAGGATAGGATTAGG	(CCT)8
RM6142	10	100	55	TCTTCCTCACCTGCTTCTCC	TACAGAGGCTACTACCACGACG	(CGC)9
RM6365	4	191	50	GCTTGCATGGAGGAGAGAAC	TGGAAGAACATCTCAAGCCC	(GAA)13
RM6374	2	127	55	TGAGGACGCTGATTGTCAAC	GCTGCCCCTATTATTTCACC	(GAA)16
RM6407	1	145	50	TGAAATGGTGGAGTCCAAGG	ACGGAGCCACTGACAGGTC	(GAG)8
RM6776	7	169	50	AGCCCGGACATGCAAAAC	GAAGCAGGCGAAATCTCCTC	(TCC)9
RM7193	6	139	50	ATGTGGGAATTTCTAGCCCC	CCCTAGTTTTCCAAATGGCC	(ATAG)7
RM7417	10	142	55	CGCAGCAGCTGTTCTACATG	TGCTTGCAAGGTCATCTGAG	(GCTA)6

Appendix II. Estimated number of optimal clusters (K) calculated using the Evanno method which showed a major peak at *K*=2 for Model 1 (a) and at *K*=8 for Model 2 (b).