1 **Short Communication**

2 **Genome-wide development of lncRNA-derived-SSR markers for Dongxiang wild**
3 **rice (*Oryza rufipogon* Griff.)**

4 Wanling Yang[1#], Yuanwei Fan[2, 3#], Yong Chen[1], Gumu Ding[1], Hu Liu[4], Jiankun Xie[1*]
5 and Fantao Zhang[1*]

6 [1]*Jiangxi Provincial Key Lab of Protection and Utilization of Subtropical Plant*
7 *Resources, College of Life Sciences, Jiangxi Normal University, Nanchang 330022,*
8 *China*

9 [2]*College of Biological Sciences and Biotechnology, Beijing Forestry University,*
10 *Beijing 100083, China*

11 [3]*Current address: Department of Biology and Center for Engineering*
12 *Mechanobiology, Washington University in St.Louis, St.Louis, MO 63130, USA*

13 [4]*Huaian Zhuxiang Ecological Agriculture CO. LTD, Huaian 223299, Jiangsu*
14 *Province, China*

15 [#]Contributed equally to this work

16 [*]Corresponding authors. E-mail: zhang84004@163.com and xiejiankun11@163.com
17

18

19

Summary

Dongxiang wild rice (*Oryza rufipogon* Griff., DXWR) is the northernmost distributed common wild rice found in the world. It contains a large number of agronomically valuable genes, which makes it a natural gene pool for rice breeding. Molecular markers, especially simple repeat sequence (SSR) markers, play important roles in crop breeding. Although a large number of SSR markers have been developed, most of them are derived from the genome coding sequences, rarely from non-coding sequences. Meanwhile, long non-coding RNAs (lncRNAs), which are derived from the transcription of non-coding sequences, play vital roles in plant growth, development and stress responses. In this study, 1878 SSR loci were detected from the lncRNA sequences of DXWR, and 1258 lncRNA-derived-SSR markers were developed on the genome-wide scale. To verify the validity and applicability of these markers, 72 pairs of primers were randomly selected to test 44 rice materials. The results showed that 42 (58.33%) pairs of primers have abundant polymorphism among these rice materials; the polymorphism information content (PIC) values ranged from 0.04 to 0.87 with an average of 0.50; the genetic diversity index of SSR loci varied from 0.04 to 0.88 with an average of 0.56; and the number of alleles per marker ranged from 2 to 11 with an average of 4.36. Thus, we concluded that these lncRNA-derived-SSR markers are a very useful source for future basic and applied research, including genetic diversity analysis, QTL mapping, and molecular breeding programs, to make good use of the elite lncRNA genes from DXWR.

Keywords: common wild rice, genetic marker, germplasm resource, long non-coding RNA, simple repeat sequence

Introduction

Dongxiang wild rice (*Oryza rufipogon* Griff., DXWR) is a common wild rice species found in Dongxiang County, Jiangxi Province, China. It is by far the northernmost (28°14' N) distributed common wild rice found in the world (Mao *et al.*, 2015). Previous studies have showed that DXWR possesses many stress tolerance and high yield-related genes, which have been lost in modern cultivated rice (Zhang *et al.*, 2016). Therefore, DXWR is an elite genetic resource to breed new rice varieties.

In eukaryotes, over 90% of the genome is transcribed to large amounts of non-coding RNAs (ncRNAs) that are not translated into proteins (Wierzbicki, 2012). Long non-coding RNAs (lncRNAs) are an important type of ncRNAs. In plants, lncRNAs act as key regulators of plant growth and development, as well as in response to various biotic and abiotic stresses (Xu *et al*., 2017). Molecular marker-assisted selection (MAS) can make great use of the valuable genetic resources in wild rice by precisely selecting genes for various traits, therefore producing superior germplasm (Collard and Mackill, 2008; Akhtamov *et al*., 2020). Simple repeat sequences (SSRs), also known as microsatellites, are DNA segments with 1-6 nucleotides short base-pair motif repeated several times in tandem (Kozlowski *et al*., 2010). They have been extensively used in genetic diversity analysis and MAS breeding because of their high reproducibility, abundant polymorphisms, co-dominant

62 inheritance, high genome coverage and simple analysis methods (Xie *et al*., 2017;
63 Singh *et al*., 2018). Although recently we developed SSR markers from drought-stress
64 responsive miRNA and identified seed storability-related loci in DXWR using
65 whole-genome sequencing (Zhao *et al*., 2021; Chen *et al*., 2021), there are no reports
66 on the development of lncRNA-derived-molecular markers in common wild rice so
67 far, which greatly limits the discovery and utilization of the elite lncRNA genes from
68 the important and endangered germplasm resource. Therefore, the main objectives of
69 this study were to: (i) use the lncRNAs sequencing data to develop a set of
70 lncRNA-derived-SSR markers and (ii) test their stabilities and polymorphisms for
71 DXWR.

72 **Experimental**

73 The information of rice materials used in this study was presented in Supplementary
74 Table S1. To develop more lncRNA-derived-SSR markers, SSR Hunter software was
75 used to identify all possible di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs with a
76 minimum set of three repeats, respectively (Li and Wan, 2005). Subsequently, primers
77 were designed based on the flanking sequences of the SSRs using Primer3.0 software
78 (Untergasser *et al*., 2012). The main parameters of the primer design were as follows:
79 the primer length was 18-25 bp, GC content was 40-60 %, melting temperature was
80 55-65 ^{o}C, the expected length of the amplification product was 100-250 bp.

81 Genomic DNA was extracted using the Plant Genomic DNA Rapid Extraction
82 Kit (Sangon Biotech Co., Ltd). The PCR reaction system was 10 μL, including 1 μL

83 (200 ng/μL) genomic DNA, 5 μL 2 × FastTaq Premix (Tolo Biotech Co., Ltd), 1 μL
84 (0.01 nmol/μL) primers and 3 μL ddH_2O. The PCR amplification reaction program
85 was as follows: pre-denaturation at 95°C for 5 min; denaturation at 95°C for 30 s,
86 annealing at 55°C for 45 s, extension at 72°C for 45 s for 30 cycles; and a final
87 extension at 72°C for 10 min. The PCR products were run on 9 % denaturing
88 polyacrylamide gel with 0.5 × TBE buffer. After electrophoresis, the gels were
89 visualized using the silver staining method (Cook *et al*., 2004). The genetic
90 parameters were analyzed using Powermarker software (Liu and Muse, 2005).

91 **Discussion**

92 In our previous study, a total of 1655 lncRNA transcripts were obtained from DXWR
93 using strand-specific RNA sequencing (Qi *et al.*, 2020). In this study, to develop a set
94 of lncRNA-derived-SSR markers for DXWR, SSR Hunter software was used to
95 search for SSRs present in these lncRNA transcripts. Totally, 1878 SSRs were
96 detected. Among them, the dinucleotide (1291, 68.74%) and trinucleotide (498,
97 26.52%) repeat motifs were the most abundant types (Table 1). Moreover, the types of
98 AG/CT (505, 39.12%) and CCG/CGG (77, 15.46%) were the most predominant in the
99 dinucleotide and trinucleotide repeat types, respectively (Fig. 1). These results
100 following the same pattern as the lncRNA-derived-SSR study in wheat (Bhandawat *et*
101 *al*., 2020).

102 Based on the detected lncRNA-derived-SSR loci, we successfully developed a
103 total of 1258 molecular markers, including 885 (70.35%) for dinucleotide repeat type,

104 316 (25.12%) for trinucleotide repeat type, 35 (2.78%) for tetranucleotide repeat type,
105 16 (1.27%) for pentanucleotide repeat type and 6 (0.48%) for hexanucleotide repeat
106 type (Supplementary Table S2). These lncRNA-derived-SSR markers were present
107 throughout all of the 12 chromosomes. Of these markers, 781 (62.08%) were found to
108 be present in the first six chromosomes (Chr. 1-6); while 477 (37.92%) were present
109 in the remaining six chromosomes (Chr. 7-12). This is largely due to the identified
110 SSR loci exhibit some preference for the lncRNA transcripts located in chromosomes
111 1-6 of DXWR. It was also observed that Chr. 2 has the highest number of
112 lncRNA-derived-SSR markers (169, 13.43%) while Chr.9 has the lowest number of
113 lncRNA-derived-SSR markers (66, 5.25%).

114 To test the stabilities and polymorphisms of the developed lncRNA-derived-SSR
115 markers, we randomly selected 72 pairs of primers for further analysis. As a result, 42
116 (58.33%) pairs of primers showed abundant polymorphisms among the 44 rice
117 materials (Fig. 2). The 42 polymorphic SSRs produced a total of 183 alleles among
118 the 44 rice materials, ranging from 2 to 11, with an average 4.36 alleles per locus. The
119 polymorphic information content (PIC) of these polymorphic SSR markers ranged
120 from 0.04 to 0.87 with an average of 0.50. The genetic diversity index was observed
121 from 0.04 to 0.88, with a mean value of 0.56 (online Supplementary Table S3). The
122 mean number of alleles and PIC value of the verified lncRNA-derived-SSR markers
123 in our study were more than capsicum which were 2.50 and 0.39 respectively (Jaiswal
124 *et al*., 2020), indicating that the developed lncRNA-derived-SSR markers were highly
125 polymorphic and can be wildly applied in DXWR and modern cultivated rice.

126 In summary, this is the first report of the development and characterization of
127 lncRNA-derived-markers in common wild rice, which lays the foundation for
128 discovery and utilization of the elite lncRNA genes to further make good use of the
129 valuable and endangered wild rice germplasm resource.

130 **Acknowledgements**

131 This research was partially supported by the National Natural Science Foundation of
132 China (32070374, 31960370, 31960085), the Natural Science Foundation of Jiangxi
133 Province, China (20202ACB205002), the Foundation of Jiangxi Provincial Key Lab
134 of Protection and Utilization of Subtropical Plant Resources (YRD201903).

135 **References**

136 Akhtamov M, Adeva C, Shim KC, Lee HS, Kim SH, Jeon YA, Luong NH, Kang JW,
137 Lee JY and Ahn SN (2020) Characterization of quantitative trait loci for
138 germination and coleoptile length under low-temperature condition using
139 introgression lines derived from an interspecific cross in rice. *Genes* 11: 1200.

140 Bhandawat A, Sharma H, Pundir N, Madhawan A and Roy J (2020) Genome-wide
141 identification and characterization of novel non-coding RNA-derived SSRs in
142 wheat. *Molecular Biology Reports* 47: 6111–6125.

143 Chen Y, Fan YW, Yang WL, Ding, GM, Zhao MM, Chen YH, Xie JK and Zhang FT
144 (2021) Development and verification of SSR markers from drought

145 stress-responsive miRNAs in common wild rice. *BioRxiv*.
146 doi: https://doi.org/10.1101/2021.10.08.463621

147 Collard BC and Mackill DJ (2008) Marker-assisted selection: an approach for
148 precision plant breeding in the twenty-first century. *Philosophical Transactions*
149 *of the Royal Society B: Biological Sciences* 363: 557–572.

150 Cook GA, Prakash O, Zhang K, Shank LP, Takeguchi WA, Robbins A, Gong YX,
151 Iwamoto T, Schultz BD and Tomich JM (2004) Activity and structural
152 comparisons of solution associating and monomeric channel-forming peptides
153 derived from the glycine receptor m2 segment. *Biophysical Journal* 86:
154 1424–1435.

155 Jaiswal V, Rawoof A, Dubey M, Chhapekar SS, Sharma V and Ramchiary N (2020)
156 Development and characterization of non-coding RNA based simple sequence
157 repeat markers in *Capsicum* species. *Genomics* 112: 1554–1564.

158 Kozlowski P, de Mezer M and Krzyzosiak WJ (2010) Trinucleotide repeats in human
159 genome and exome. *Nucleic Acids Research* 38: 4027–4039.

160 Li Q and Wan JM (2005) SSR hunter: development of local searching software for
161 SSR sites. *Hereditas* 27: 808–810.

162 Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for
163 genetic marker analysis. *Bioinformatics* 21: 2128–2129.

164 Mao DH, Yu L, Chen DZ, Li L, Zhu YX, Xiao YQ, Zhang DC and Chen CY (2015)
165 Multiple cold resistance loci confer the high cold tolerance adaptation of
166 Dongxiang wild rice (*Oryza rufipogon*) to its high-latitude habitat. *Theoretical*
167 *and Applied Genetics* 128: 1359–1371.

168 Qi WD, Chen HP, Yang ZZ, Hu BL, Luo XD, Ai B, Luo Y, Huang Y, Xie JK and
169 Zhang FT (2020) Systematic Characterization of Long Non-Coding RNAs and
170 Their Responses to Drought Stress in Dongxiang Wild Rice. *Rice Science* 27:
171 21–31.

172 Singh AK, Chaurasia S, Kumar S, Singh R, Kumari J, Yadav MC, Singh N, Gaba S
173 and Jacob SR (2018) Identification, analysis and development of salt responsive
174 candidate gene based SSR markers in wheat. *BMC Plant Biology* 18: 249.

175 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen
176 SG (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Research* 40:
177 e115.

178 Wierzbicki AT (2012) The role of long non-coding RNA in transcriptional gene
179 silencing. *Current Opinion in Plant Biology* 15: 517–522.

180 Xie JK, Zhang M, Sun J and Zhang FT (2017) Genome-wide genic SSR marker
181 development for the endangered Dongxiang wild rice (*Oryza rufipogon*). *Plant*
182 *Genetic Resources* 15: 566–569.

183 Xu Q, Song ZH, Zhu CY, Tao CC, Kang LF, Liu W, He F, Yan J and Sang T (2017)
184 Systematic comparison of lncRNAs with protein coding mRNAs in population
185 expression and their response to environmental change. *BMC Plant Biology* 17:
186 42.

187 Zhang FT, Xu T, Mao LY, Yan SY, Chen XW, Wu ZF, Chen R, Luo XD, Xie JK and
188 Gao S (2016) Genome-wide analysis of Dongxiang wild rice (*Oryza rufipogon*
189 Griff.) to investigate lost/acquired genes during rice domestication. *BMC Plant*
190 *Biology* 16: 103.

191 Zhao MM, Hu BL, Fan YW, Ding GM, Yang WL, Chen Y, Chen YH, Xie JK and
192 Zhang FT (2021) Identification, analysis, and confirmation of seed
193 storability-related loci in Dongxiang wild rice (Oryza rufipogon Griff.) *Genes*
194 12:1831.

195
196
197
198
199
200
201
202
203
204
205
206

Table 1. Percentage of the identified SSR motifs in different repeat types of DXWR

Fig. 1. Frequency of different SSR repeat motif types in DXWR.

Fig. 2. Representative results of polymorphic lncRNA-derived-SSR markers amplified by the genome of DXWR and other worldwide rice cultivars. (a) Lnc-SSR-566686; (b): Lnc-SSR-353993; M: DL2000 DNA Marker; 1-3: three different populations of DXWR; 4-44: other worldwide cultivars. The information of the 44 rice materials was presented in online Supplementary Table S1.

Table S1. Dongxiang wild rice and worldwide rice cultivars used for verification of the developed lncRNA-derived-SSR markers in this study

231 **Table S2.** The information of developed lncRNA-derived-SSR markers for
232 Dongxiang wild rice

233 **Table S3.** The verified lncRNA-derived-SSR markers and their amplification results
234 with 44 rice materials in this study

235

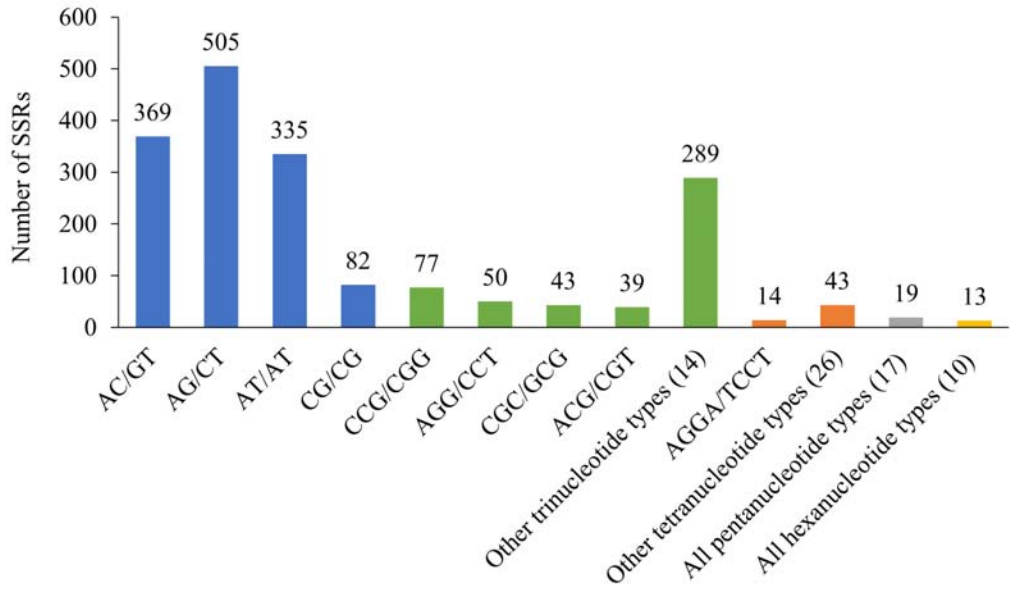

Fig. 1. Frequency of different SSR repeat motif types in DXWR.

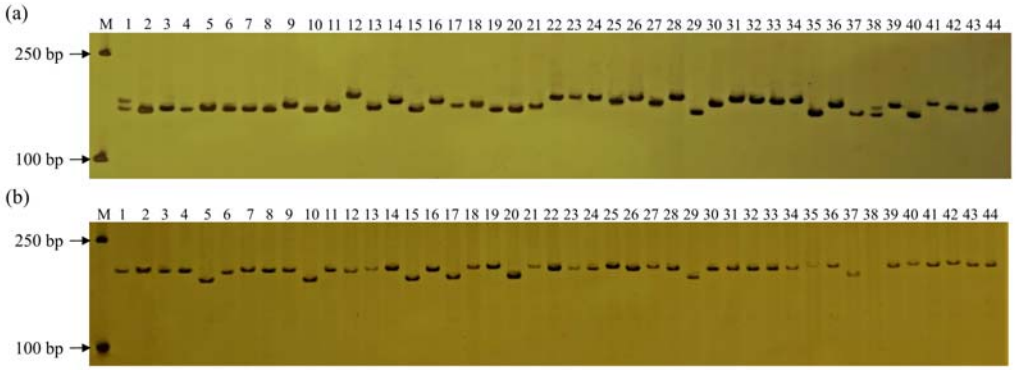

Fig. 2. Representative results of polymorphic lncRNA-derived-SSR markers amplified by the genome of DXWR and other worldwide rice cultivars. (a) Lnc-SSR-566686; (b): Lnc-SSR-353993; M: DL2000 DNA Marker; 1-3: three different populations of DXWR; 4-44: other worldwide cultivars. The information of the 44 rice materials was presented in online Supplementary Table S1.

Table 1. Percentage of the identified SSR motifs in different repeat types of DXWR

SSR motif	Repeat number					Percentage
	3	4	5	≥ 6	Total	(%)
Dinucleotide	1123	117	20	31	1291	68.74
Trinucleotide	424	47	23	4	498	26.52
Tetranucleotide	51	5	0	1	57	3.04
Pentanucleotide	18	0	1	0	19	1.01
Hexanucleotide	11	2	0	0	13	0.69
Total	1627	171	44	36	1878	100
Percentage (%)	86.63	9.11	2.34	1.92	100	