Short Communication

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

Wanling Yang^{1#}, Yuanwei Fan^{2,3#}, Yong Chen¹, Gumu Ding¹, Hu Liu⁴, Jiankun Xie^{1*} and Fantao Zhang^{1*}

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

²*College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China*

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

⁴*Huaian Zhuxiang Ecological Agriculture CO. LTD, Huaian 223299, Jiangsu Province, China*

[#]Contributed equally to this work

^{*}Corresponding authors. E-mail: zhang84004@163.com and xiejiankun11@163.com

20 **Abstract** Dongxiang wild rice (*Oryza rufipogon* Griff., DXWR) is the northernmost
 21 distributed common wild rice found in the world. It contains a large number of
 22 agronomically valuable genes, which makes it a natural gene pool for rice breeding.
 23 Molecular markers, especially simple repeat sequence (SSR) markers, play important
 24 roles in crop breeding. Although a large number of SSR markers have been developed,
 25 most of them are derived from the genome coding sequences, rarely from non-coding
 26 sequences. Meanwhile, long non-coding RNAs (lncRNAs), which are derived from
 27 the transcription of non-coding sequences, play vital roles in plant growth,
 28 development and stress responses. In this study, 1878 SSR loci were detected from the
 29 lncRNA sequences of DXWR, and 1258 lncRNA-derived-SSR markers were
 30 developed on the genome-wide scale. To verify the validity and applicability of these
 31 markers, 72 pairs of primers were randomly selected to test 44 rice materials. The
 32 results showed that 42 (58.33%) pairs of primers have abundant polymorphism among
 33 these rice materials; the polymorphism information content (PIC) values ranged from
 34 0.04 to 0.87 with an average of 0.50; the genetic diversity index of SSR loci varied
 35 from 0.04 to 0.88 with an average of 0.56; and the number of alleles per marker
 36 ranged from 2 to 11 with an average of 4.36. Thus, we concluded that these
 37 lncRNA-derived-SSR markers are a very useful source for future basic and applied
 38 research, including genetic diversity analysis, QTL mapping, and molecular breeding
 39 programs, to make good use of the elite lncRNA genes from DXWR.

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

42 **Introduction**

43 Dongxiang wild rice (*Oryza rufipogon* Griff., DXWR) is a common wild rice species
 44 found in Dongxiang County, Jiangxi Province, China. It is by far the northernmost
 45 (28°14' N) distributed common wild rice found in the world (Mao *et al.*, 2015).
 46 Previous studies have showed that DXWR possesses many stress tolerance and high
 47 yield-related genes, which have been lost in modern cultivated rice (Zhang *et al.*,
 48 2016). Therefore, DXWR is an elite genetic resource to breed new rice varieties.

49 In eukaryotes, over 90% of the genome is transcribed to large amounts of
 50 non-coding RNAs (ncRNAs) that are not translated into proteins (Wierzbicki, 2012).
 51 Long non-coding RNAs (lncRNAs) are an important type of ncRNAs. In plants,
 52 lncRNAs act as key regulators of plant growth and development, as well as in
 53 response to various biotic and abiotic stresses (Xu *et al.*, 2017). Molecular
 54 marker-assisted selection (MAS) can make great use of the valuable genetic resources
 55 in wild rice by precisely selecting genes for various traits, therefore producing
 56 superior germplasm (Collard and Mackill, 2008; Akhtamov *et al.*, 2020). Simple
 57 repeat sequences (SSRs), also known as microsatellites, are DNA segments with 1-6
 58 nucleotides short base-pair motif repeated several times in tandem (Kozłowski *et al.*,
 59 2010). They have been extensively used in genetic diversity analysis and MAS
 60 breeding because of their high reproducibility, abundant polymorphisms, co-dominant
 61 inheritance, high genome coverage and simple analysis methods (Xie *et al.*, 2017;
 62 Singh *et al.*, 2018). However, there are no reports on the development of

63 lncRNA-derived-molecular markers in common wild rice so far, which greatly limits
64 the discovery and utilization of the elite lncRNA genes from the important and
65 endangered germplasm resource. Therefore, the main objectives of this study were to:
66 (i) use the lncRNAs sequencing data to develop a set of lncRNA-derived-SSR
67 markers and (ii) test their stabilities and polymorphisms for DXWR.

68 **Experimental**

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

77 Genomic DNA was extracted using the Plant Genomic DNA Rapid Extraction
78 Kit (Sangon Biotech Co., Ltd). The PCR reaction system was 10 µL, including 1 µL
79 (200 ng/µL) genomic DNA, 5 µL 2 × FastTaq Premix (Tolo Biotech Co., Ltd), 1 µL
80 (0.01 nmol/µL) primers and 3 µL ddH₂O. The PCR amplification reaction program
81 was as follows: pre-denaturation at 95°C for 5 min; denaturation at 95°C for 30 s,
82 annealing at 55°C for 45 s, extension at 72°C for 45 s for 30 cycles; and a final
83 extension at 72°C for 10 min. The PCR products were run on 9 % denaturing

84 polyacrylamide gel with $0.5 \times$ TBE buffer. After electrophoresis, the gels were
85 visualized using the silver staining method (Cook *et al.*, 2004). The genetic
86 parameters were analyzed using Powermarker software (Liu and Muse, 2005).

87 **Discussion**

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

98 Based on the detected lncRNA-derived-SSR loci, we successfully developed a
99 total of 1258 molecular markers, including 885 (70.35%) for dinucleotide repeat type,
100 316 (25.12%) for trinucleotide repeat type, 35 (2.78%) for tetranucleotide repeat type,
101 16 (1.27%) for pentanucleotide repeat type and 6 (0.48%) for hexanucleotide repeat
102 type (Supplementary Table S2). These lncRNA-derived-SSR markers were present
103 throughout all of the 12 chromosomes. Of these markers, 781 (62.08%) were found to
104 be present in the first six chromosomes (Chr. 1-6); while 477 (37.92%) were present

105 in the remaining six chromosomes (Chr. 7-12). This is largely due to the identified
106 SSR loci exhibit some preference for the lncRNA transcripts located in chromosomes
107 1-6 of DXWR. It was also observed that Chr. 2 has the highest number of
108 lncRNA-derived-SSR markers (169, 13.43%) while Chr.9 has the lowest number of
109 lncRNA-derived-SSR markers (66, 5.25%).

110 To test the stabilities and polymorphisms of the developed lncRNA-derived-SSR
111 markers, we randomly selected 72 pairs of primers for further analysis. As a result, 42
112 (58.33%) pairs of primers showed abundant polymorphisms among the 44 rice
113 materials (Fig. 2). The 42 polymorphic SSRs produced a total of 183 alleles among
114 the 44 rice materials, ranging from 2 to 11, with an average 4.36 alleles per locus. The
115 polymorphic information content (PIC) of these polymorphic SSR markers ranged
116 from 0.04 to 0.87 with an average of 0.50. The genetic diversity index was observed
117 from 0.04 to 0.88, with a mean value of 0.56 (online Supplementary Table S3). The
118 mean number of alleles and PIC value of the verified lncRNA-derived-SSR markers
119 in our study were more than capsicum which were 2.50 and 0.39 respectively (Jaiswal
120 *et al.*, 2020), indicating that the developed lncRNA-derived-SSR markers were highly
121 polymorphic and can be widely applied in DXWR and modern cultivated rice.

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

126 **Acknowledgements**

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

131 **References**

- 132 Akhtamov M, Adeva C, Shim KC, Lee HS, Kim SH, Jeon YA, Luong NH, Kang JW,
133 Lee JY and Ahn SN (2020) Characterization of quantitative trait loci for
134 germination and coleoptile length under low-temperature condition using
135 introgression lines derived from an interspecific cross in rice. *Genes* 11: 1200.
- 136 Bhandawat A, Sharma H, Pundir N, Madhawan A and Roy J (2020) Genome-wide
137 identification and characterization of novel non-coding RNA-derived SSRs in
138 wheat. *Molecular Biology Reports* 47: 6111–6125.
- 139 Collard BC and Mackill DJ (2008) Marker-assisted selection: an approach for
140 precision plant breeding in the twenty-first century. *Philosophical Transactions*
141 *of the Royal Society B: Biological Sciences* 363: 557–572.
- 142 Cook GA, Prakash O, Zhang K, Shank LP, Takeguchi WA, Robbins A, Gong YX,
143 Iwamoto T, Schultz BD and Tomich JM (2004) Activity and structural
144 comparisons of solution associating and monomeric channel-forming peptides

145 derived from the glycine receptor m2 segment. *Biophysical Journal* 86:
146 1424–1435.

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

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

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

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

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

160 Qi WD, Chen HP, Yang ZZ, Hu BL, Luo XD, Ai B, Luo Y, Huang Y, Xie JK and
161 Zhang FT (2020) Systematic Characterization of Long Non-Coding RNAs and
162 Their Responses to Drought Stress in Dongxiang Wild Rice. *Rice Science* 27:
163 21–31.

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

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

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

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

175 Xu Q, Song ZH, Zhu CY, Tao CC, Kang LF, Liu W, He F, Yan J and Sang T (2017)
176 Systematic comparison of lncRNAs with protein coding mRNAs in population
177 expression and their response to environmental change. *BMC Plant Biology* 17:
178 42.

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

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

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

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

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

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

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

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

223

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

SSR motif	Repeat number				Total	Percentage (%)
	3	4	5	≥ 6		
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

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.