

A NEO-DIVERSITY GROUP OF *ORYZA SATIVA* L. SSP. *JAPONICA* IN NORTHEASTERN CHINA EVALUATED BY LARGE-EFFECT SNPs

HONG GAO^{1,2,3}, QUAN XU², MING ZHONG², ZHENGJIN XU^{2*} AND GUOMIN SUI^{1*}

¹Liaoning Academy of Agricultural Sciences, Shenyang 110000, China

²Rice Research Institute / Key Laboratory of Rice Biology and Breeding, Ministry of Agriculture / Key Laboratory of Northern Japonica Super Rice Breeding, Ministry of Education, Shenyang Agricultural University, Shenyang 110866, China

³Liaoning Rice Research Institute, Shenyang 110101, China

*Corresponding authors: Zhengjin Xu at xuzhengjin@126.com, and Guomin Sui at guomin666@126.com

Abstract

Rice breeding on the subspecies *Oryza sativa* L. ssp. *japonica* in northeastern China has been successful over a relatively short time. However, little is known about how the genetic structure of *japonica* has diverged as a result of artificial selection. In the present study, 189 single nucleotide polymorphisms (SNP) markers from gene coding regions were selected to characterize the diversity, phylogeny and population structure of four breeding groups in northeastern China, viz. modern *japonica*, historical *japonica*, exotic *japonica*, and *indica* varieties. Results suggested that the neo-diversity of *japonica* has formed gradually through introgression from other gene resources during the breeding selection. The genetic diversity (π) of modern *japonica* was lowest compared with historical *japonica*, exotic *japonica* and *indica* varieties, which implies a genetic bottleneck caused by the breeding selection. Population structure and phylogeny analysis proved that the modern *japonica* was independent of the historical *japonica* and exotic *japonica* groups.

Key words: Breeding selection, Genetic diversity, Rice, SNP marker.

Abbreviations: SNP-single nucleotide polymorphisms; SINE-short interspersed nuclear element; SSR-simple sequence repeat; Fst-F-statistics; RFLP-restriction fragment length polymorphisms; MCMC- Markov chain Monte Carlo; FDR-false discovery rate.

Introduction

The domestication of Asian cultivated rice (*Oryza sativa* L.), which occurred more than 10,000 years ago from divergent populations of the wild rice species *Oryza rufipogon*, is one of the most important events in human history (Oka, 1988; Cheng *et al.*, 2003; Huang *et al.*, 2012). To investigate population structure and subpopulation relationships in *O. sativa*, genome-wide molecular markers, such as restriction fragment length polymorphisms (RFLPs), simple sequence repeat (SSR) and SNPs have been widely used (Garris *et al.*, 2005; Lu *et al.*, 2009; Zhao *et al.*, 2009). Molecular phylogenetic analyses based on short interspersed nuclear element (SINE) insertion patterns indicated that *Oryza sativa* L. ssp. *japonica* (hereafter *japonica*) is closely related to *O. rufipogon*, whereas *O. sativa* ssp. *indica* (hereafter *indica*) is more closely related to *O. nivara* (Chen *et al.*, 2003). Phylogeographic research suggested that cultivated rice was domesticated at least twice, in two different geographic regions (Londo *et al.*, 2006). An analysis of low diversity regions (LDRs) shared by the *indica* and *japonica* genomes, revealed that the genealogical history of these overlapping LDRs was distinct from the genomic background, suggesting independent origins of two subspecies (He *et al.*, 2011). This hypothesis is also supported by intensive genome resequencing and a molecular evolutionary analysis of functional genes in cultivated and wild rice (Wei *et al.*, 2012; Xu *et al.*, 2012).

Demographic modeling of resequencing data for 630 gene fragments from wild and domesticated rice accessions resulted in the identification of putative selective sweeps and sequence variations in sequence-tagged sites (STSs) at the genome-wide level, thus

providing molecular evidence for a single domestication origin for cultivated rice (Wei *et al.*, 2011). Some domestication genes, such as *sh4*, *rc*, and *bh4*, appear to be fixed in both subspecies with the same alleles, also supporting a single domestication origin (Li *et al.*, 2006; Sang *et al.*, 2007; Sweeney *et al.*, 2007; Zhang *et al.*, 2009; Zhu *et al.*, 2011). An alternative explanation for these allele types appearing in both subspecies is that they originated only once as a result of introgression after selection (Caicedo *et al.*, 2007; He *et al.*, 2011; Mao *et al.*, 2018).

Sequence analyses have determined that *japonica* was first domesticated from *O. rufipogon*, with *indica* subsequently developing from crosses between *japonica* and local wild rice (Li *et al.*, 2006). A wide range of genetic and genomic studies have been performed to examine the process of rice domestication, leading to divergent hypotheses. Most studies have focused on landraces, domestic regional varieties that are unlikely to be influenced by gene flow from artificial hybridization and thus maintain their genetic identities (Huang *et al.*, 2012; Ali *et al.*, 2017).

To characterize the diversity and genetic structure of modern Japanese rice varieties, comparisons of genomic sequences among several varieties with large genetic differentiation can be used to develop fine-resolution single nucleotide polymorphism (SNP) microarrays, thus enabling the detection of genomic polymorphisms and haplotype updated with breeding selection (Ebana *et al.*, 2010; Nagasaki *et al.*, 2010; Yamamoto *et al.*, 2010; Arai-Kichise *et al.*, 2011; Yonemaru *et al.*, 2012). In northeastern China, introducing *indica* germplasm in *japonica* rice breeding has been largely responsible for the increased rice yield in this region over the last five decades (Sun *et al.*, 2012).

Japonica rice breeding in northeastern China has been over a relatively short time (Sun *et al.*, 2013). The coverage rate of the super *japonica* rice variety now accounts for more than 60% of the total area in rice cultivation, with a yield increase of 750-1113 kg per hectare in northeastern China. In order to reveal the genetic structure in gene function angle for the *japonica* rice in northeastern China, we tried to characterize the diversity, phylogeny and population structure based on SNP in coding regions in the present study.

Materials and Methods

Plant materials: Historical and exotic *japonica* varieties are collectively referred to as the ancestral populations of modern varieties. To assess the impact of breeding practices in northeastern China on the *japonica* genome, we carried out population-level genetic analyses based on SNP variations in 109 accessions, categorized into four groups: 33 modern varieties, 26 historical varieties, 25 exotic varieties from Japan and the Korean Peninsula, and 25 Chinese *indica* varieties. Seeds of the Liaoning varieties were obtained from the Liaoning Rice Research Institute and the Liaoning Provincial Saline-Alkali Land Utilization and Research Institute. Seeds of the Jilin varieties were obtained from the Jilin Academy of Agricultural Sciences. Seeds of the Heilongjiang varieties

were obtained from the Rice Research Institute, Heilongjiang Academy of Agricultural Sciences and the Rice Research Institute, Heilongjiang Academy of Land Reclamation Sciences. Seeds of exotic *japonica* and *indica* varieties were selected from germplasm pool of Shenyang Agricultural University. Plants were grown at the Shenyang Agricultural University, Liaoning Province, China. The detailed information on plant material is provided in Fig. 1 and Supplementary Table 1.

DNA isolation: The genomic DNA was extracted from the fresh leaves using a Rapid DNA Extraction Kit (Tiangen Biotech Co., Beijing, China).

Selection and genotyping of SNP loci: Previously, the genome resequencing of 517 Chinese original core landraces (including *indica*, *aus*, tropical *japonica*, temperate *japonica* and intermediate types) identified 3,625 SNPs in gene coding regions that were predicted to cause amino acid changes by termination and initiation codons and splicing sites, and insertions of new termination codons (Huang *et al.*, 2010). Among these SNPs, 190 evenly distributed on rice chromosomes were selected to estimate population divergence from the perspective of functional variation. Information on these SNPs is shown in Supplementary Table 2.

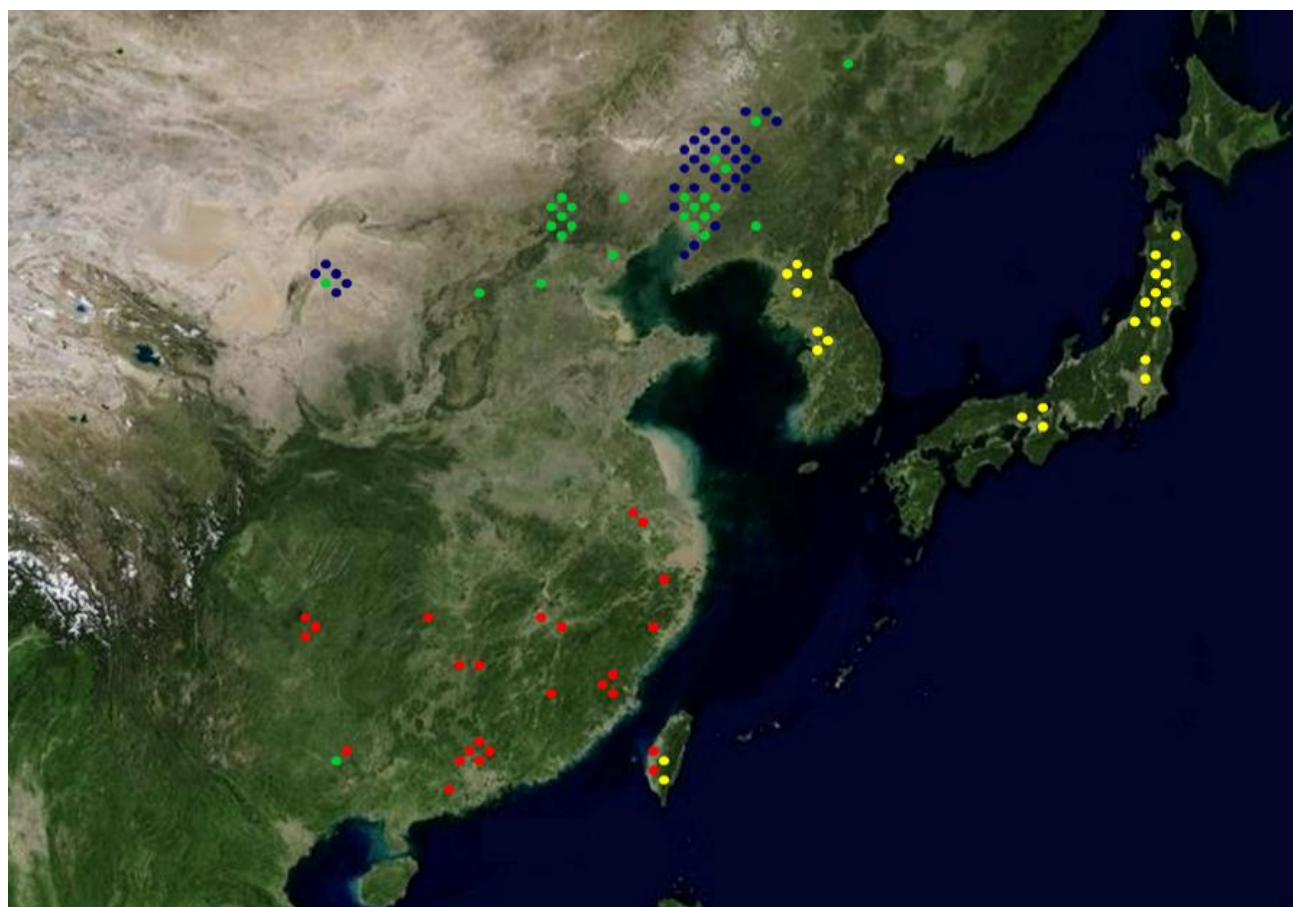


Fig. 1. Geographic locations of samples used in this study
Blue: modern *Japonica*, Green: historical *japonica*, Yellow: exotic *japonica*, and Red: *indica*. Base map from <http://map.baidu.com>.

Table S1. 109 different breeding phase population samples.

Accession name	Registration time	Origin	Subspecies	Latitude	Longitude
Nipponbare	1957	Japan	<i>O.japonica</i>	N35°0'	E137°13'
Toyonishiki	1974	Japan	<i>O.japonica</i>	N39°22'	E140°31'
Akitakomachi	2000	Japan	<i>O.japonica</i>	N39°43'	E140°5'
Hitomebore	1991	Japan	<i>O.japonica</i>	N38°25'	E140°54'
Chunyang	NA	Japan	<i>O.japonica</i>	N35°1'	E137°12'
Aoyu392	NA	Japan	<i>O.japonica</i>	N39°34'	E140°46'
Juanqi	NA	Japan	<i>O.japonica</i>	N34°55'	E137°10'
Guxiang	NA	Japan	<i>O.japonica</i>	N34°38'	E135°33'
Fuxiang	NA	Japan	<i>O.japonica</i>	N36°6'	E136°13'
Zhongguang	NA	Japan	<i>O.japonica</i>	N34°55'	E135°28'
Qiandaijin	1990	Japan	<i>O.japonica</i>	N35°41'	E139°45'
Chuxing	NA	Japan	<i>O.japonica</i>	N35°46'	E140°0'
Akihikari	1970	Japan	<i>O.japonica</i>	N40°46'	E140°49'
Sasanishiki	1983	Japan	<i>O.japonica</i>	N39°11'	E140°9'
LimingB	NA	Japan	<i>O.japonica</i>	N39°41'	E140°7'
Huadongdao	NA	Korean peninsula	<i>O.japonica</i>	N37°21'	E127°2'
Renyuedao	NA	Korean peninsula	<i>O.japonica</i>	N37°27'	E126°42'
Dazhendao	NA	Korean peninsula	<i>O.japonica</i>	N36°20'	E127°23'
Dengxi52	NA	Korean peninsula	<i>O.japonica</i>	N39°26'	E125°43'
Xiongji3	NA	Korean peninsula	<i>O.japonica</i>	N42°23'	E130°31'
Xiannan23	NA	Korean peninsula	<i>O.japonica</i>	N41°54'	E129°26'
Pyongyang1	NA	Korean peninsula	<i>O.japonica</i>	N39°1'	E125°45'
Xiannan1	NA	Korean peninsula	<i>O.japonica</i>	N41°54'	E129°26'
Taidongludao328	NA	Taiwan	<i>O.japonica</i>	N23°2'	E120°58'
Taizhong65	NA	Taiwan	<i>O.japonica</i>	N24°15'	E120°56'
Longhuamaohulu	NA	Hebei	<i>O.japonica</i>	N41°24'	E117°43'
Gaoyangdiandahongmang	NA	Hebei	<i>O.japonica</i>	N38°42'	E115°46'
Zhonglou1	NA	Shanxi	<i>O.japonica</i>	N38°8'	E112°33'
Xingguo	NA	Jilin	<i>O.japonica</i>	N43°36'	E124°49'
Baimaodao	NA	Heilongjiang	<i>O.japonica</i>	N46°42'	E130°3'
Weiguo	1960	Liaoning	<i>O.japonica</i>	N40°10'	E122°7'
Dandongludao	NA	Liaoning	<i>O.japonica</i>	N40°1'	E124°20'
Ying9443	NA	Liaoning	<i>O.japonica</i>	N40°37'	E122°30'
Panjin782	NA	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Longpan 5	2006	Liaoning	<i>O.japonica</i>	N41°2'	E122°4'
Liaoyan158	2009	Liaoning	<i>O.japonica</i>	N41°6'	E122°4'
Liaohe12	NA	Liaoning	<i>O.japonica</i>	N41°2'	E122°4'
Liaojing287	1988	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'
Liaojing5	1981	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'
2428	NA	guangxi	<i>O.japonica</i>	N22°51'	E108°17'
Zhonghua15	NA	Beijing	<i>O.japonica</i>	N40°0'	E116°20'
Zhongzuo9059	2004	Beijing	<i>O.japonica</i>	N40°12'	E116°13'
Handao227	NA	Beijing	<i>O.japonica</i>	N40°4'	E116°10'
Han526	NA	Beijing	<i>O.japonica</i>	N40°4'	E116°10'
Han529	NA	Beijing	<i>O.japonica</i>	N40°4'	E116°10'
Han535	NA	Beijing	<i>O.japonica</i>	N40°4'	E116°10'
Yufeng7	NA	Liaoning	<i>O.japonica</i>	N41°42'	E123°19'
Yutian301	NA	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Ningdao216	1998	Ningxia	<i>O.japonica</i>	N38°16'	E106°14'
Jinzhul	NA	Tianjin	<i>O.japonica</i>	N39°7'	E117°11'
Zhongzuo9128	1999	Beijing	<i>O.japonica</i>	N40°4'	E116°10'
Liaoyan16	1994	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Liaojing294	1998	Liaoning	<i>O.japonica</i>	N41°38'	E123°19'
Liaojing326	1992	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'

Table S1. (Cont'd.).

Accession name	Registration time	Origin	Subspecies	Latitude	Longitude
Liaojing454	1996	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'
Chaochan1	1999	Jilin	<i>O.japonica</i>	N43°55'	E125°13'
Jijing88	2005	Jilin	<i>O.japonica</i>	N43°55'	E125°13'
Changbai17	2006	Jilin	<i>O.japonica</i>	N43°55'	E125°13'
Yanjing218	2009	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Yanjing68	2003	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Yanfeng47	2006	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Tianfeng201	2006	Liaoning	<i>O.japonica</i>	N41°8'	E122°4'
Shennong9903	2009	Liaoning	<i>O.japonica</i>	N41°49'	E123°33'
Shennong606	2003	Liaoning	<i>O.japonica</i>	N41°48'	E123°31'
Shennong265	2001	Liaoning	<i>O.japonica</i>	N41°48'	E123°31'
Shenjing4311	2006	Liaoning	<i>O.japonica</i>	N41°51'	E123°26'
Shendao29	2009	Liaoning	<i>O.japonica</i>	N41°49'	E123°33'
Liaoxing17	2007	Liaoning	<i>O.japonica</i>	N41°38'	E123°15'
Liaoxing16	2006	Liaoning	<i>O.japonica</i>	N41°39'	E123°20'
Liaonong49	2006	Liaoning	<i>O.japonica</i>	N41°38'	E123°19'
Liaojing534	2002	Liaoning	<i>O.japonica</i>	N41°38'	E123°20'
Liaojing371	2002	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'
Huadan995	2009	Liaoning	<i>O.japonica</i>	N42°11'	E123°1'
Chenhe1	2009	Liaoning	<i>O.japonica</i>	N40°26'	E122°19'
Liaoxing1	2003	Liaoning	<i>O.japonica</i>	N41°49'	E123°30'
Liaojing9	2005	Liaoning	<i>O.japonica</i>	N41°40'	E123°20'
Shendao3	2005	Liaoning	<i>O.japonica</i>	N41°49'	E123°29'
Shendao6	2005	Liaoning	<i>O.japonica</i>	N41°49'	E123°33'
Shendao9	2005	Liaoning	<i>O.japonica</i>	N41°49'	E123°33'
Ningjing36	2006	Ningxia	<i>O.japonica</i>	N38°16'	E106°14'
Ningjing37	2006	Ningxia	<i>O.japonica</i>	N38°28'	E106°10'
Ningjing38	2006	Ningxia	<i>O.japonica</i>	N38°28'	E106°10'
Ningjing39	2006	Ningxia	<i>O.japonica</i>	N38°28'	E106°10'
Hua96	2007	Ningxia	<i>O.japonica</i>	N38°16'	E106°14'
9311	2002	Jiangsu	<i>O.indica</i>	N35°0'	E119°9'
Zhongyouzao8	1996	Zhejiang	<i>O.indica</i>	N30°4'	E119°55'
Qishanzhan	1991	Guangdong	<i>O.indica</i>	N23°8'	E113°20'
Aisimiao	NA	Guangdong	<i>O.indica</i>	N23°15'	E113°47'
Zaoxiang1	2001	Guangxi	<i>O.indica</i>	N22°49'	E108°21'
Chenghui448	2007	Sichuan	<i>O.indica</i>	N30°37'	E104°6'
Honglika	NA	Guangdong	<i>O.indica</i>	N23°21'	E116°37'
Hetianxiangdao	NA	Fujian	<i>O.indica</i>	N25°39'	E116°37'
Zhenshan97B	NA	Zhejiang	<i>O.indica</i>	N27°59'	E116°37'
Dijiaowujian	NA	Taiwan	<i>O.indica</i>	N23°12'	E120°16'
Jiangxisimiao	NA	Jiangxi	<i>O.indica</i>	N28°13'	E116°48'
Aijiaonante	NA	Guangdong	<i>O.indica</i>	N23°16'	E116°35'
Nanjing6	NA	Jiangsu	<i>O.indica</i>	N32°6'	E118°47'
TN1	NA	Taiwan	<i>O.indica</i>	N23°9'	E120°14'
Minghui63	1980	Fujian	<i>O.indica</i>	N26°19'	E117°39'
II -32B	NA	hunan	<i>O.indica</i>	N28°12'	E113°4'
Xiangyaxiangzhan	2006	Guangdong	<i>O.indica</i>	N22°23'	E112°40'
KqingB	NA	sichuan	<i>O.indica</i>	N31°6'	E104°22'
Jiafuzhan	NA	Fujian	<i>O.indica</i>	N26°3'	E119°15'
Gang96B	2002	Sichuan	<i>O.indica</i>	N29°58'	E102°59'
R99	NA	hubei	<i>O.indica</i>	N30°19'	E109°27'
Nantehao	1934	Jiangxi	<i>O.indica</i>	N28°33'	E115°56'
Lucaihao	1946	Fujian	<i>O.indica</i>	N26°17'	E117°37'
Guangluai4	1983	Guangdong	<i>O.indica</i>	N23°8'	E113°20'
Shenglixian	NA	Hunan	<i>O.indica</i>	N28°29'	E112°58'

Table S2. The list of genome-wide large-effect SNPs.

Marker ID	Chr	Pos	Nipponbare allele	Tother allele	Gene ID in RAP-DB
S1	1	152340	G	A	Os01g0102800
S2	1	2000988	T	A	Os01g0137300
S3	1	4061128	G	C	Os01g0178100
S4	1	8032973	G	A	Os01g0245700
S5	1	1E+07	T	C	Os01g0281100
S6	1	1.2E+07	C	T	Os01g0316600
S7	1	1.4E+07	C	T	Os01g0350500
S8	1	1.6E+07	G	A	Os01g0380700
S9	1	1.9E+07	G	A	Os01g0496900
S10	1	2.2E+07	A	G	Os01g0549250
S11	1	2.4E+07	C	T	Os01g0578500
S12	1	2.6E+07	A	C	Os01g0613900
S13	1	2.8E+07	T	G	Os01g0650200
S14	1	3E+07	G	A	Os01g0686100
S15	1	3.2E+07	A	G	Os01g0729600
S16	1	3.4E+07	C	T	Os01g0768400
S17	1	3.6E+07	A	T	Os01g0810533
S18	1	3.8E+07	T	A	Os01g0845900
S19	1	4E+07	C	T	Os01g0881400
S20	1	4.2E+07	T	C	Os01g0924600
S21	1	4.4E+07	G	T	Os01g0957600
S22	2	98322	G	A	Os02g0101800
S23	2	2087512	G	A	Os02g0139200
S24	2	4121585	C	T	Os02g0175400
S25	2	6463017	T	C	Os02g0215000
S26	2	8207347	T	A	Os02g0244700
S27	2	1E+07	T	A	Os02g0274900
S28	2	1.5E+07	T	C	Os02g0437200
S29	2	1.6E+07	C	T	Os02g0457500
S30	2	1.8E+07	T	G	Os02g0493300
S31	2	2E+07	T	G	Os02g0526201
S32	2	2.2E+07	C	T	Os02g0563301
S33	2	2.4E+07	A	T	Os02g0599100
S34	2	2.6E+07	A	T	Os02g0629400
S35	2	2.8E+07	C	A	Os02g0670700
S36	2	3E+07	G	A	Os02g0708100
S37	2	3.2E+07	A	C	Os02g0742800
S38	2	3.4E+07	T	C	Os02g0782800
S39	3	91995	A	T	Os03g0101200
S40	3	2045189	A	T	Os03g0137400
S41	3	4027409	T	G	Os03g0176100
S42	3	6015791	T	C	Os03g0214200
S43	3	8016343	A	C	Os03g0251100
S44	3	1E+07	C	A	Os03g0292900
S45	3	1.2E+07	G	A	Os03g0328200
S46	3	1.4E+07	T	A	Os03g0356582
S47	3	1.6E+07	A	T	Os03g0387900
S48	3	1.8E+07	A	T	Os03g0421000

Table S2. (Cont'd.).

Marker ID	Chr	Pos	Nipponbare allele	Tother allele	Gene ID in RAP-DB
S49	3	2.1E+07	A	G	Os03g0556900
S50	3	2.2E+07	C	T	Os03g0582000
S51	3	2.4E+07	G	A	Os03g0615300
S52	3	2.6E+07	T	A	Os03g0653900
S53	3	3E+07	G	T	Os03g0721001
S54	3	3.2E+07	A	G	Os03g0755600
S55	3	3.4E+07	G	A	Os03g0797550
S56	3	3.6E+07	G	A	Os03g0840400
S57	4	235187	A	G	Os04g0103601
S58	4	2877914	C	T	Os04g0142400
S59	4	4162260	A	T	Os04g0160500
S60	4	8035823	A	G	Os04g0220500
S61	4	9998438	T	C	Os04g0252400
S62	4	1.2E+07	C	T	Os04g0286500
S63	4	1.4E+07	G	T	Os04g0311400
S64	4	1.6E+07	C	T	Os04g0339000
S65	4	1.8E+07	C	A	Os04g0376600
S66	4	2E+07	C	T	Os04g0405100
S67	4	2.2E+07	A	C	Os04g0442950
S68	4	2.4E+07	T	A	Os04g0476000
S69	4	2.6E+07	T	A	Os04g0517766
S70	4	2.8E+07	C	T	Os04g0553500
S71	4	3E+07	C	T	Os04g0587900
S72	4	3.2E+07	G	A	Os04g0621900
S73	4	3.4E+07	A	T	Os04g0663100
S74	4	3.6E+07	A	G	Os04g0692200
S75	5	123891	C	A	Os05g0102300
S76	5	2007838	G	C	Os05g0134700
S77	5	4055727	C	T	Os05g0167600
S78	5	6119765	G	T	Os05g0198700
S79	5	8092731	A	T	Os05g0233000
S80	5	1.1E+07	G	A	Os05g0268400
S81	5	1.2E+07	C	T	Os05g0286200
S82	5	1.4E+07	G	A	Os05g0309300
S83	5	1.6E+07	C	G	Os05g0342600
S84	5	1.8E+07	C	T	Os05g0376000
S85	5	2E+07	G	A	Os05g0412300
S86	5	2.2E+07	G	A	Os05g0452900
S87	5	2.4E+07	T	C	Os05g0487300
S88	5	2.6E+07	A	G	Os05g0521500
S89	5	2.8E+07	G	T	Os05g0561800
S90	5	3E+07	C	T	Os05g0595950
S91	6	369569	C	T	Os06g0105400
S92	6	2113037	C	T	Os06g0140300
S93	6	4011083	T	A	Os06g0181200
S94	6	6037710	C	T	Os06g0218150
S95	6	8011144	A	T	Os06g0254700
S96	6	1E+07	G	A	Os06g0286310

Table S2. (Cont'd.).

Marker ID	Chr	Pos	Nipponbare allele	Tother allele	Gene ID in RAP-DB
S97	6	1.4E+07	T	C	Os06g0349700
S98	6	1.7E+07	G	A	Os06g0472000
S99	6	1.8E+07	T	A	Os06g0493600
S100	6	2E+07	A	C	Os06g0521000
S101	6	2.2E+07	A	T	Os06g0555400
S102	6	2.4E+07	G	A	Os06g0591450
S103	6	2.6E+07	C	T	Os06g0624900
S104	6	2.8E+07	C	A	Os06g0661000
S105	6	3E+07	C	A	Os06g0698900
S106	6	3.2E+07	T	A	Os06g0731200
S107	7	299710	G	T	Os07g0105200
S108	7	2011864	A	G	Os07g0137500
S109	7	4025330	A	G	Os07g0175900
S110	7	6076251	G	A	Os07g0211200
S111	7	8002056	C	T	Os07g0243000
S112	7	1E+07	T	A	Os07g0270900
S113	7	1.2E+07	C	A	Os07g0296800
S114	7	1.4E+07	G	T	Os07g0418700
S115	7	1.6E+07	A	G	Os07g0451101
S116	7	1.8E+07	T	C	Os07g0481300
S117	7	2E+07	C	T	Os07g0510300
S118	7	2.2E+07	A	T	Os07g0540600
S119	7	2.4E+07	C	T	Os07g0578333
S120	7	2.6E+07	C	T	Os07g0615900
S121	7	2.8E+07	G	T	Os07g0652800
S122	7	3E+07	T	A	Os07g0689900
S123	8	30699	G	C	Os08g0100400
S124	8	1920216	G	A	Os08g0133700
S125	8	4054686	T	C	Os08g0169300
S126	8	6002126	C	T	Os08g0203300
S127	8	8146859	G	A	Os08g0233900
S128	8	1E+07	T	G	Os08g0265300
S129	8	1.2E+07	G	A	Os08g0296600
S130	8	1.4E+07	T	C	Os08g0320800
S131	8	1.6E+07	G	A	Os08g0351250
S132	8	2E+07	C	T	Os08g0417100
S133	8	2.2E+07	C	T	Os08g0448900
S134	8	2.4E+07	G	A	Os08g0487300
S135	8	2.8E+07	C	A	Os08g0559000
S136	9	197973	A	C	Os09g0101200
S137	9	2611118	G	T	Os09g0135100
S138	9	4056967	T	C	Os09g0246500
S139	9	6029162	C	T	Os09g0272800
S140	9	8078602	A	T	Os09g0301700
S141	9	1E+07	C	T	Os09g0325100
S142	9	1.2E+07	G	A	Os09g0356000
S143	9	1.4E+07	T	C	Os09g0393200
S144	9	1.6E+07	G	T	Os09g0425700

Table S2. (Cont'd.).

Marker ID	Chr	Pos	Nipponbare allele	Tother allele	Gene ID in RAP-DB
S145	9	1.8E+07	C	A	Os09g0460300
S146	9	2E+07	T	A	Os09g0500600
S147	9	2.2E+07	A	C	Os09g0540600
S148	9	2.4E+07	T	C	Os09g0572900
S149	10	24497	G	T	Os10g0100200
S150	10	2035993	C	T	Os10g0132700
S151	10	4315958	C	T	Os10g0167200
S152	10	6165575	T	C	Os10g0193100
S153	10	8984749	A	T	Os10g0321700
S154	10	1E+07	C	T	Os10g0343000
S155	10	1.2E+07	T	C	Os10g0371600
S156	10	1.4E+07	C	T	Os10g0405500
S157	10	1.6E+07	A	T	Os10g0439700
S158	10	1.8E+07	C	T	Os10g0474933
S159	10	2E+07	T	A	Os10g0512100
S160	10	2.2E+07	G	A	Os10g0549850
S161	10	2.4E+07	G	A	Os10g0580900
S162	11	1396	G	A	Os11g0100150
S163	11	2063383	T	C	Os11g0145600
S164	11	4092896	T	A	Os11g0181900
S165	11	6056685	A	G	Os11g0216300
S166	11	8103943	G	A	Os11g0250000
S167	11	1E+07	G	T	Os11g0283500
S168	11	1.2E+07	G	A	Os11g0308601
S169	11	1.5E+07	C	T	Os11g0417266
S170	11	1.6E+07	T	C	Os11g0438700
S171	11	1.8E+07	A	G	Os11g0471200
S172	11	2E+07	A	T	Os11g0513900
S173	11	2.2E+07	T	A	Os11g0550500
S174	11	2.4E+07	G	A	Os11g0588400
S175	11	2.6E+07	G	T	Os11g0620800
S176	11	2.8E+07	G	T	Os11g0654800
S177	11	3E+07	C	T	Os11g0691500
S178	12	35199	C	T	Os12g0101000
S179	12	2130198	T	G	Os12g0143400
S180	12	6564588	G	A	Os12g0221400
S181	12	8164034	G	A	Os12g0246700
S182	12	1E+07	T	A	Os12g0273960
S183	12	1.2E+07	C	T	Os12g0407200
S184	12	1.4E+07	G	T	Os12g0435000
S185	12	1.6E+07	C	T	Os12g0458100
S186	12	2E+07	T	G	Os12g0516900
S187	12	2.2E+07	G	A	Os12g0543800
S188	12	2.4E+07	C	T	Os12g0577600
S189	12	2.6E+07	T	A	Os12g0611200
S190	12	2.8E+07	T	G	Os12g0640900

The genotyping for the SNPs was based on the commercially available Sequenom MassARRAY platform. The process consisted of an initial locus-specific PCR reaction, followed by single base extension using mass-modified dideoxynucleotide terminators of an oligonucleotide probe primer that annealed immediately upstream of the polymorphic site of interest (Gabriel *et al.*, 2009). The distinct mass of the extended primer was identified using MALDI-TOF mass spectrometry; genotyping results were obtained using TYPER 4.0 software. Information on the locus-specific PCR and probe primers are presented in Supplementary Table 3.

Diversity and genetic structure analysis: After obtaining SNP genotypes from those populations, the DNAsp 5.0 software was used to estimate sequence diversity (π). The population structure was then assessed with Structure 2.3.1 (Pritchard *et al.*, 2000). Ten replicates were performed for each value of K , with the number of clusters considered using a burn-in length of 10,000 steps followed by a run length of 100,000 Markov chain Monte Carlo (MCMC) replicates. The number of subgroups from $k = 1$ to 20 were tested under a simulation model that assumed admixture and correlated allele frequencies. The Evanno's ΔK method was used to determine the most probable number of clusters (Evanno *et al.*, 2005). Phylogeny reconstruction was built by the neighbor-joining method and distance-p model on MEGA 6.0 (Tamura *et al.*, 2013). Bootstrap values correspond to 1,000 replications.

Identification of outlier SNPs: SNPs are assumed to cause large-effect changes on gene functions; however it is unknown whether the 190 SNPs chosen for the analysis were subject to selection. Consequently, the outlier SNPs were analyzed using the Bayesian method proposed by Foll and Gaggiotti (Foll & Gaggiotti, 2008), and implemented in the program Bayesian 2.01. The estimation of model parameters was set as 10 pilot runs of 5,000 iterations each, followed by 100,000 iterations. We used the R (v. 5.0) statistical software to plot outliers (<http://cran.r-project.org/>).

Results

SNP genotyping and neutrality test: Based on allele peaks observed in the mass spectra using the MassARRAY platform (Fig. 2), we performed genotyping for 109 samples from four rice groups (modern *japonica*, historical *japonica*, exotic *japonica* introduced from Japan and Korea, and *indica* varieties) at 190 genome-wide functional SNP loci.

Distinguishing neutral SNP markers helped to increase the accuracy of the population genetics analysis. Thus, we filtered the outlier loci before the analysis with the Bayesian 2.01 software. According to Jeffreys' scale of evidence, a $\log_{10}BF$ of 1.5-2.0 is interpreted as 'strong evidence' of selection based on Bayesian factor (Jeffreys, 1961). After constructing the expected distribution of *Fst*, only one SNP (SNP22) was consistently identified as an

outlier from 10 independent iterations by controlling false discovery rate (FDR) < 0.05. The remaining 189 loci were considered as neutral, and used for the subsequent analysis (Fig. 3).

Diversities of four groups of varieties: We estimated the genetic diversities for the four rice groups based on genotypes of 189 SNPs using the DnaSP 5.0 software (Librado & Rozas, 2009). Sequence diversity (π) was estimated at 0.022 for modern *japonica* varieties, and 0.033, 0.063 and 0.207 for historical *japonica*, exotic *japonica* and *indica* varieties, respectively.

Genetic structure and phylogenetic analyses of four groups based on SNP variations: The population structure of the sampled groups was assessed twice, with and without *indica* groups, respectively, using the Structure 2.3.1 software (Figs. 4A and 5) (Evanno *et al.*, 2005).

For the full set of populations, a peak ΔK value was observed when $k = 2$ according to the Evanno's ΔK method (Fig. 6A). This result indicated that the best fitting model contained two clusters, with the three *japonica* populations in one cluster and the *indica* in the other. For $k = 3$, the modern *japonica* was separated into its own cluster, independent of the other two *japonica* groups. For $k = 4$, the population differentiation was evident in historical *japonica*, whereas modern *japonica* was in a single cluster. For $k = 5$, further population differentiation occurred only in the historical *japonica*. These demonstrate that (1) historical varieties and exotic varieties share a common *japonica* gene pool in their genomic constitutions, and (2) compared with the ancestral groups (historical and exotic *japonica*), the genetic structure of modern *japonica* varieties in northeastern China is a well-defined genetic group of *O. sativa*.

For the population structure and Evanno's ΔK analysis without *indica*, a peak ΔK value was also observed at $k = 2$ (Fig. 6B) and the modern *japonica* was independent of the other two *japonica* groups.

To further reveal the impact of the introduction of *indica* and selection on the genetic structure of modern *japonica* varieties in northeastern China, we constructed a bootstrap neighbor-joining tree inferred from 1,000 replicates based on 189 functional SNP variations. Most of *indica* samples were clustered together in one portion of the tree except three intermediate samples, whereas all the exotic and historical varieties were intermixed in the other parts of the tree (Fig. 4B). Modern *japonica* varieties were more genetically close to the main cluster of *indica* compared with exotic and historical *japonica*. The neighbor-joining tree, combined with the genetic structure analysis, revealed that modern *japonica* varieties in northeastern China have gradually separated from ancestral *japonica*, indicating that partial introgression with *indica* genes may contribute to its genetic divergence. Thus, a neo-diversity of modern *japonica* varieties could have gradually arisen through breeding selection by the introduction of limited *indica* lineages in northeastern China.

Table S3. Information of the locus-specific PCR and probe primer.

SNP_ID	2nd-PCR	1st-PCR	UEP_SEQ
S1	ACGTTGGATGTGGCACACCTTTACACGCAAC	ACGTTGGATGCAGCCCTTTGACAAATGCAAC	CCAGAAAGTGGGAGTGCAGT
S2	ACGTTGGATGCATAACCAACCTTCTCCAAGC	ACGTTGGATGGGATCCCTAGCTCCAAAAG	CCACCTTCTCCAAGCTTATTACTCAG
S3	ACGTTGGATGACATCTGACGACCGACTTC	ACGTTGGATGGCTGTTTCCGTGACTTGTG	CGGTTTCCTAGCCTTATTAT
S4	ACGTTGGATGTCGAGCACGATCTCTGGTGAC	ACGTTGGATGAAGGAGCTCGGCGTGGTGT	cccGTTCCGCGAGGTGATCT
S5	ACGTTGGATGCAACCAAGAAGAACTGAAGAAC	ACGTTGGATGCTGTTCTTCGCAAGAAAGGG	GAACACAAAACCTGCATG
S6	ACGTTGGATGAGAACGCCTAGAAAACCTC	ACGTTGGATGCAGCTCATCATGCACTAAG	CCTAGAAAACACCTCCACGTC
S7	ACGTTGGATGCTGCAAAAACCTCGGTACTG	ACGTTGGATGACCCATTCTCTGTGACCTTG	GTACTTTCAGCTTCTGAATCTTACT
S8	ACGTTGGATGTGAGAGTGGTGCATTCGAAG	ACGTTGGATGTTTCTCAGGGTGTCTGGGTG	GTTGTCAGAAAGTTTGGAG
S9	ACGTTGGATGTGTTCTCATAGCGCTATCG	ACGTTGGATGGTTCAGGCACAGTGGCGAAC	AGGGGCACCTCTACATGGGTTTGGCGT
S10	ACGTTGGATGTTGGTGTGCCCTCCCTC	ACGTTGGATGATCAGGCACAGTGGCCGAAAC	CTCTCCGAGGACCATGA
S11	ACGTTGGATGCGGGTGACATGTTAAAAGCA	ACGTTGGATGGTCAAGTGAATGTTTTGC	CTCTCCGACGACAGGGGCTAC
S12	ACGTTGGATGGAAGTCTGTGTGGTATTGC	ACGTTGGATGGATATGAGCCAGAGACAGG	TGCAAAATTCACCTTAAAAGAATC
S13	ACGTTGGATGTTCCCTTCCGCGGAGT	ACGTTGGATGGTCTGGGAAAAGGCATAC	GTCTGTTGTGGTATTGCGAGAACT
S14	ACGTTGGATGTCGACCCACAGTTGCTTTAC	ACGTTGGATGTTTCGGCCAAAACCAAGGG	CCCTTCCGCGGAGTCCCTGCCT
S15	ACGTTGGATGTCACCAAGATTATTGTGG	ACGTTGGATGGATGGTACTACTAAATATTG	GTTGCTTACTTCCCTATG
S16	ACGTTGGATGGATGGCAACAGAAATCCCG	ACGTTGGATGTGCTGAACGTGCTGCATTTG	AAGATTATTGGTACCACAC
S17	ACGTTGGATGTCATGTTGACAGAGCGAC	ACGTTGGATGAACCGTGCCTTCTTTGCAAG	CCAGGATCATCCTAGTG
S18	ACGTTGGATGGCACTCACC AATAGGTGAAG	ACGTTGGATGGAAGATGATTGTCAGCGTG	TGAAGAGAGCTCTTGGG
S19	ACGTTGGATGAGTCTTGAGTCCAACTTCC	ACGTTGGATGTGACTCCCTCCTGATCCAAAC	CGGATTGAGAGTTTTCTAAAT
S20	ACGTTGGATGGAGGAGTATGCATCATAAT	ACGTTGGATGTCTGAGCCAGCAGCGAA	GATGTAGCAACTATTATGTTGTGC
S21	ACGTTGGATGCTCCACGACCAAAAAGAGAAG	ACGTTGGATGTTCCGACTGTTGACAGACCCAC	AAGAAAAGTAAATTTGAATTTATCA
S22	ACGTTGGATGCAATACCTTGTGACACCGC	ACGTTGGATGAAAAGCCCATCTACGTACC	AGAAGCTACTACTGGGCTC
S23	ACGTTGGATGGAACCTGCTGTATGCTG	ACGTTGGATGGAATTTCTGCTTTGAGGG	CCTTTGTGCACCGCAGCATC
S24	ACGTTGGATGATAATGCATGCAAGAAAGCCC	ACGTTGGATGGAAATTTCTCAGCCAAAG	GAGCTGTTCCACACCTT
S25	ACGTTGGATGGGATCCCTTCCAGAAATATC	ACGTTGGATGGATCTGAGAAAGTTAGTGCCC	TTC AATTTATTTTCCCATTGCTTGC
S26	ACGTTGGATGTGAAAAGAGCGGATGATGGAG	ACGTTGGATGCACACAGAAATGGAAAACAC	AATATCCTTGTGACCCGGCAACC
S27	ACGTTGGATGGCGCTAAGACATATATGGTG	ACGTTGGATGCATCATCCTGCTTTGCTTC	TGGTCGCAACTTTAATGG
S28	ACGTTGGATGGGCAAGAAGATTTCCAGCCC	ACGTTGGATGAATCCTCAAAATGCTGCTCG	ATTATCTTGTGTGTGCTC
S29	ACGTTGGATGCCAACAGAACTAATAAGGG	ACGTTGGATGTTTAGAGGAGGCAACCCAAAC	TTGCGTGGGTTTCCAGCAA
S30	ACGTTGGATGGCAATGCCCTGCAATTAATC	ACGTTGGATGCATCTGGTGCATCGTAAAC	ATTTACATCAGCACAGCC
S31	ACGTTGGATGCCCTCCTCAAGAAAACCC	ACGTTGGATGGTGCATGACGCTGTGCGAT	TGCATTACTCCCATCCA
S32	ACGTTGGATGCCCCCTGCATTAAGCTTATC	ACGTTGGATGCCGGATTAATCAATTTGGAC	ACATCATCTACGCTTTG
S33	ACGTTGGATGCAAGTGCAGGTACGTA AAC	ACGTTGGATGCATGAAAAGCCGTGATGGTCC	GAATGCTTTTTGCTGT
S34	ACGTTGGATGGAAGGTTGCTTACCCTATG	ACGTTGGATGCAGGTTTCTCCCATGTTTC	TATCTGAAGCCTTGTGTGCCAACCTT
S35	ACGTTGGATGCTCCCAACACCAAGTAATACG	ACGTTGGATGGTCTTTTCTGTTACGTATTG	GCTTACCCATGGACATGGAGAT
S36	ACGTTGGATGTGGTCTGAAAGATGGTCAAC	ACGTTGGATGGTCAATTTGGATATCACCCAC	CCAACACCAAGTAATACGGAAAGC
S37	ACGTTGGATGAGCTGTGTTGGATCTTTCCC	ACGTTGGATGTGGTGCATGATACATAAG	ACATTTGCTATGCGAAAGG
S38			GGATCTTTCCCATATAATATGCTAGG

Table S3. (Cont'd).

SNP_ID	2nd-PCR	1st-PCR	UEP_SEQ
S39	ACGTTGGATGGCGCTTAATCAAACTGACCTC	ACGTTGGATGCAGTTTGTGACACGTCGTTG	AGGTGAAAGCTCAAAGTAT
S40	ACGTTGGATGGGACCAACTGGACCATAATC	ACGTTGGATGTGATCTGTTGGCTCCAATCC	CAACTGGACCATAATCATCTTTC
S41	ACGTTGGATGTCAAAGACCTCTTGTGGCC	ACGTTGGATGCTTAAACCTGCAGACAAGAC	CTTCATTTATTTGATCTGCAGTT
S42	ACGTTGGATGACAGCAAGTCCCTCGAGAAG	ACGTTGGATGAGAGCTTGTGTGATTTGCTG	GTCCCTCGAGAAGCCCTCCAT
S43	ACGTTGGATGCATACITTTAAGGATGTCGGG	ACGTTGGATGACCGCTGCTGTCTGCTCTTG	ATTGACGAGAGCATTAACCTTG
S44	ACGTTGGATGATGGCCAGCAGCTGTATAAG	ACGTTGGATGTGTCAAAGCTGTTTCGCCCTC	TTAAAAGCATTCAGCAGTAATTT
S45	ACGTTGGATGCAACAGGGAGCAATTCATGG	ACGTTGGATGGCTGGAGCGAAAGCTTATTAC	TGGGAGACACAGGCAGCTTAGGAAG
S46	ACGTTGGATGCATCTTTGCGTTGCAAAAAC	ACGTTGGATGTAGCACAAGCTAAGGTGTATGG	ACAATGGTTCCAAGAGTTCCAAAAG
S47	ACGTTGGATGACAGTACACAGTTGTTGAAGC	ACGTTGGATGTACATTCACCAAGAGGGCAG	TAACTTCAGTTGTTCTGCAAGATT
S48	ACGTTGGATGCCTCCAAATGCGAAACAGTC	ACGTTGGATGTGCCCATTTGGAAAATTTCCG	AGTAGGAGTGACAAATCGTGCTTA
S49	ACGTTGGATGATCTCTGCCTCTTCCAGAG	ACGTTGGATGGAACACAGTGGAGAAGGGCAG	GTCAACCACCTTCCACCC
S50	ACGTTGGATGGACTGTGGAGATTAGGTTTC	ACGTTGGATGAGAGGGGAGAGCTGAAGAAGG	TACCTTGTCTGTGTGCTTTT
S51	ACGTTGGATGAAGCCGCTTTTAGGACCGATAAC	ACGTTGGATGTCTCCGCGCTACTCAAAG	GTTCCTCCCTTTTGGCT
S52	ACGTTGGATGTGGCTTTTCCCACTTGACCATC	ACGTTGGATGGTGAACCCGGGATCAAAC	CTTGATAGGGTTCACAGAT
S53	ACGTTGGATGTTCTCCCACTTGACCATC	ACGTTGGATGTTACCTCCCTGCTGTGTG	TGACCATCTCCCAATCTA
S54	ACGTTGGATGACCGAGCCATCTTGTATTGC	ACGTTGGATGGCTGTGAATTTGCAATGCAGG	CTCTTTTTTATTATTAATGTTCTC
S55	ACGTTGGATGTTCTCCGACCCCTACACCGA	ACGTTGGATGATCGACATGACCCCGTCCAT	CTACACCGACGGCCCGTCAAC
S56	ACGTTGGATGAATACCCCACTGAACCTGAG	ACGTTGGATGGGTTGTTGGATCAAACGG	GTCTAACACCTTAAGCGATTACCCCT
S57	ACGTTGGATGATTCAGTCGAGAGGATAGCG	ACGTTGGATGTCCTCTCCGGTGTGTG	CGAGGCCACCGAATTAACCCCA
S58	ACGTTGGATGTAATGCCTCAATCTTCTG	ACGTTGGATGCGATCCATTTTCGCCATTAG	TTGCTAACTTCATCAITTAATTCA
S59	ACGTTGGATGGTGTGTTGGTGTCTCAAATC	ACGTTGGATGGAGCTGTCCAATGAGAAGAG	ACTCCAGTGTGGAGGACACAT
S60	ACGTTGGATGTGCAATCTTGGCCAGCCAAA	ACGTTGGATGTCCATTTCAATCCACGCATAG	CATCTTGCCAGCCAAACCAAGG
S61	ACGTTGGATGTGCAATACACTGTTCCCTC	ACGTTGGATGGTAAATATACCTGGCTATCG	CTTAATTCATTTGTGCTTTATGCTTGC
S62	ACGTTGGATGACTCCCTCCTACTCAATC	ACGTTGGATGGTGAGGCTTGGTAAAATG	CTCAACTTCTCCCAATCA
S63	ACGTTGGATGTCAAAGACAAAGAGGAGAC	ACGTTGGATGTGATCCATGGTCTCCTCAG	AAAGAGGAGACTAATCCTATCCTACT
S64	ACGTTGGATGCATTTGTGATTGGATCGGC	ACGTTGGATGAGCAAGTCAATGGCATGCTC	GGGATCGGGAGCCGCTTCCCTTCT
S65	ACGTTGGATGCAATCCAGGTAATGGCTGAG	ACGTTGGATGGCAGCATCACGATACTAAG	GAGAACTTGTGTTGGACTCT
S66	ACGTTGGATGGTCTTGCAAAAAGTCTGTG	ACGTTGGATGGTGTGGAACAGCTGTACAA	ATCAATGAAACTAAATGCTACATA
S67	ACGTTGGATGACTTTCCCACTATGAGCC	ACGTTGGATGGATGCCCTTGTGTTTATGG	TGGCATCAAGGGCCCAT
S68	ACGTTGGATGCTTGGTATACGTGTGATGG	ACGTTGGATGGATGCGGATAATGGAGACC	ATTTGCTCATCTTCTGGT
S69	ACGTTGGATGTTCTATCTGCCATGAAAGG	ACGTTGGATGCCGATTAGTCCATGCTCTC	TGTTGGTGAAGCCACATA
S70	ACGTTGGATGTTGCATGCCGGCATTTGGTG	ACGTTGGATGTAGCAACAACAGCTGACGC	CCGGAGCCGTGTTCTGC
S71	ACGTTGGATGCGAGGGCTACTGTGATGAAG	ACGTTGGATGGACAATAGAGTAGTACAACAC	AAGATTCACTCAGAATATGTTGCTC
S72	ACGTTGGATGAATTTCTGTAGGACCTCCGC	ACGTTGGATGCCCTTCTCGAATGCTTTG	TGCAGCATCAGAGATAG
S73	ACGTTGGATGGGACAGGAAAAGCCTGATT	ACGTTGGATGAAAAGGTAGTTCGGCTCAAGG	gAAATCTGAGGATCTTATGTG
S74	ACGTTGGATGTAATGACGGGTAAGGAATGC	ACGTTGGATGTGCTCTTCGTCTATGGCAC	ACGGTAAGGAATGCCCTCTGTGAAC
S75	ACGTTGGATGTCTTTAAGAGCCCCGTAAAC	ACGTTGGATGACGAATGGAATCTTGGAAAG	CATCAGCCATCACCATT
S76	ACGTTGGATGGAGATCAGGAAGACTTGCAG	ACGTTGGATGTTCCGCTGACTCGATCGTTG	CAGATGCAACCAACCACTACTACTA

Table S3. (Cont'd.).

SNP_ID	2nd-PCR	1st-PCR	UEP_SEQ
S77	ACGTTGGATGTCAGCTTAGTATGGCCCAAG	ACGTTGGATGCGATGCTGTGTTGATTTGAG	CGGAAGAAAAAGAAAAAGACAG
S78	ACGTTGGATGTGGATTGTTGGTGCATGAC	ACGTTGGATGGGACCTGTGTATTCATGTG	GATTTGGTGGCATGACATTTCTA
S79	ACGTTGGATGCCGATCAATTTGCATCTGTC	ACGTTGGATGCCACTGTGACAGAGAAAAAGC	GGAAAGTCAACTCCTCT
S80	ACGTTGGATGGTCCACAGCTTCTGAGAAATC	ACGTTGGATGGCTCTAAAACCCAGCTTCTC	CTGAGAAATCTGTAGTTACATATT
S81	ACGTTGGATGTCCCATCATCCAAGCTTTC	ACGTTGGATGCCTTCTCGATCTCTCTCAAC	TCCAAGCTTTCATCCATCC
S82	ACGTTGGATGTGAACTTCAACAGGCTATCC	ACGTTGGATGTGGTCAGGATAAAGTACGGG	CCACTTCGACCACTTGAATATTCT
S83	ACGTTGGATGCCTAGAAATCTGCTACCTC	ACGTTGGATGTATGCTTTTCCGTTGAACC	TCGTACCTCTCCATTCGGAA
S84	ACGTTGGATGTACGGTATTGCCATGCCATC	ACGTTGGATGCCAGTGGAAATAGTAGCAGG	GCCTACAATAAGAACCTCAGTA
S85	ACGTTGGATGATCCGGCCTAGAGCTCGATG	ACGTTGGATGGAGCTCGGATTTCTCCGCC	GCGGGTGGCGCCCGCGGACAT
S86	ACGTTGGATGGTATATTCATGGCTGGCTG	ACGTTGGATGGGAAAGGCCCATAAATCCAAG	GCATCGACCTGGTCCACTTCTG
S87	ACGTTGGATGATGACCCCTCCCAAGGTCC	ACGTTGGATGACTACTACAAGCTGTTCCAG	CCCAAAGTCTCTGGGAGAGCA
S88	ACGTTGGATGCCTAATTCCTGATATCAGAG	ACGTTGGATGACGAGGCTAGAGATGTGTTG	ATTCCTGATATCAGAGATACCGAC
S89	ACGTTGGATGCTAGACTTCTGCAGCTTG	ACGTTGGATGTTGCTGCTGCAGTAGGAAC	TGGAACCTCACGGAGTA
S90	ACGTTGGATGCTTTCATGCTGCCATCTC	ACGTTGGATGTCAGGATGCTCAACTACCTC	TGCTGCCGTTGAGCTTGTTC
S91	ACGTTGGATGAACGCCCTTCATCCATCTGC	ACGTTGGATGGATTCCCATAGCTTGTCTG	CATCTGCACCAATGGCAC
S92	ACGTTGGATGTTCTTCTGAGACATCCAAGGC	ACGTTGGATGGAGCACAGATCAAGAGCAAG	TAATTTGATGAGAAGTTCCAGAG
S93	ACGTTGGATGGCTAATGGTCTCTTCCGAC	ACGTTGGATGATGGCACTGAACAGCTAC	GTTCTCTCGACAAAGCAT
S94	ACGTTGGATGGAAACACACTCTGGTCTTAG	ACGTTGGATGATACCCAAAGGATTCACACC	ACCTTACCCCTTAAAACATAAC
S95	ACGTTGGATGTCACAAGGATCTGTCCAGGG	ACGTTGGATGACTGCTTCTTGTCTCCATC	GGATACTGCAGGGCAGTCTA
S96	ACGTTGGATGAGCTCATGCACATCTTGG	ACGTTGGATGAGAATAATCATGGCAAAGC	TCAAAGGATATCCTGAAGGCG
S97	ACGTTGGATGTACACTGCTCCCTCGCGGT	ACGTTGGATGTCCCCACAGCGCTCGTC	GCATCTGGATCAACAAC
S98	ACGTTGGATGAAACGACCGGACTCCTCTC	ACGTTGGATGTGCTCTGAACTGGAGGAGG	TCTCGAGCCGAAAGAAAATGG
S99	ACGTTGGATGCCAGTAGTAAGTAGGAAAG	ACGTTGGATGCAAGACTATGGCTACTGCAC	GGTAGGAAGGATACTGCATAAGCC
S100	ACGTTGGATGGAGTGGACATTTGCTCCATC	ACGTTGGATGCTTGGAGGAACTTGAGAAGC	CTCCATCACTTCCACTT
S101	ACGTTGGATGGTAGACAAAGGCTGGAGATG	ACGTTGGATGGAACACATCAAGACTGCAAG	GCTGGAAAGAAAGATG
S102	ACGTTGGATGATCTTCAAAGTCTCGGAGC	ACGTTGGATGAACTGGTGGACCCGGCTATA	AGGTGTTCTGGCTAGAAAT
S103	ACGTTGGATGTTCTTGGGGCTGAAGGAGTC	ACGTTGGATGAAAGCTGACATTCGCCATCC	GGGCTGAAGGAGTCGTAGTC
S104	ACGTTGGATGAGATTGATTTGGCCATCCCC	ACGTTGGATGTTCTGTTGAACTTGAAAGCC	CCATATTGAATCATGACAAAGATTTT
S105	ACGTTGGATGCTTGTAGTTTCTATGCTGAC	ACGTTGGATGGCTCATGTGGCAACAAG	AGCTATGCAATGGTACTA
S106	ACGTTGGATGCTTGTGGGACAAAGCCAAATG	ACGTTGGATGATGCGAGACAAACAGGTATGG	TTGTGGACAAAGCCAAATGATTTTCT
S107	ACGTTGGATGATCCCTCCACTCTCAGAAC	ACGTTGGATGTCAAATGCAGTACCTGTCC	CTCCACTCTCAGAACCTCTATTC
S108	ACGTTGGATGTAAGAAGTAAACCTGGGAGC	ACGTTGGATGGGACATTTCTCAATAGC	TTGGACTAACTTTTATGTTTATG
S109	ACGTTGGATGTTCAGCAAGAGCTCTTGTCC	ACGTTGGATGAGAGTCTTGGTAGTTCACTG	AACAATATTAGAAAAAGATCAGG
S110	ACGTTGGATGTCAAATACACACGCATGCAC	ACGTTGGATCGGAGCTCATCTCGGCGG	TGCGCGGCTAGTGTCA
S111	ACGTTGGATGACCTCCCGATCAAAGGCAAG	ACGTTGGATGAAGAAATCAGAGCGGATGAGC	CAAAGCAAGTTGTTTCCG
S112	ACGTTGGATGTCATGCAAGGTGAGATGAGC	ACGTTGGATGTAGTCCCTACTGAGGAAG	GCACCTCAGTACAATGCTT
S113	ACGTTGGATGACCATCGGTACAAAGGTTTC	ACGTTGGATGTAGAGGATGTCCGACATCCTG	TCACAAGGTTTCCCCCGAGCACGTGT
S114	ACGTTGGATGTCCAAAGCCTAAGCCACAAC	ACGTTGGATGATTTGGCTTGGGGTTAGGAC	CCAAAACCTGAACCCAAACCT

Table S3. (Cont'd.).

SNP_ID	1st-PCR		UEP_SEQ
	2nd-PCR	1st-PCR	
S115	ACGTTGGATGATGAAAGTCCCTAGCTAGCTC	ACGTTGGATGTCGAAACCGGATTCGATCCACC	AGTAGCTCACATCGGGAGCAGCCAT
S116	ACGTTGGATGAGCTCTCAAATTTGGTGTGTC	ACGTTGGATGGGTACGGAAACAAGGTAACCTC	TGAACCGAATAATCCACACTGAGG
S117	ACGTTGGATGTTGATGTCGGCTCCACGTA	ACGTTGGATGAAGATTCCCATGGAAAGCC	GTGGACGTGTTGGTGAACCTTC
S118	ACGTTGGATGCCGACAGCATCAGGTAATAT	ACGTTGGATGTGAGTGCAGCAAGCCACAGTAAG	AATATTAATAATAGAGAAGGTGAGT
S119	ACGTTGGATGCTGCGGATTCCTCCGATTTG	ACGTTGGATGGGGCTCTATCTATTTTCGTG	TCCGATTTCTGTTCTTCC
S120	ACGTTGGATGAAAGTATCTCTCTCCTCCAG	ACGTTGGATGCAATATTGGCCACCATCACC	AAGCGTATGAAGGCTATG
S121	ACGTTGGATGAAAGTGGCAAGCTAGTG	ACGTTGGATGCGATCTTCCACATTTGGTC	GTGGGCAAGCTAGTGAGCATGCATA
S122	ACGTTGGATGTTGTTGGCTGTCAGTTG	ACGTTGGATGGAGATTAACAGGCAATGCAG	CTTACCTTAAGACTTGGCTATTTCTT
S123	ACGTTGGATGCTTGCCACTTGTAGAACAG	ACGTTGGATGTAGAAAGCAAAAGATGAGGC	GGCTGGTCAGTTGATATCTGTGCTT
S124	ACGTTGGATGAACCTAGCCAATACTGACG	ACGTTGGATCGGGAAATGACTGTTATGAGC	GCTAGAACAGCAGGCTCTACC
S125	ACGTTGGATGGAATTTATTTACTATGCGCTC	ACGTTGGATGTTGTCAATGGCGTCTTAC	GAAAGAAGAAGAGGAGGTC
S126	ACGTTGGATGCACACCAAAAGACGACAC	ACGTTGGATGAAGACTTGGCAAAAGCATCGG	ATGCGCTCAATATGTACCT
S127	ACGTTGGATGTTGGAGGATCCTTATGGAG	ACGTTGGATGTGTTCTTCTTGTGGAGTGG	AAGAAGCGACACCCCTCTC
S128	ACGTTGGATGGCATATTTGGTGTGAAAAAC	ACGTTGGATGATTGGTGTCTCCAAGTGTG	GAGGATCCTTATGGAGCAGTCAGTT
S129	ACGTTGGATGCAATGGAGGTTCTTCCAAC	ACGTTGGATGCCATCTCTCATCAAGACC	GTGATGAAAAACTATCAAGGC
S130	ACGTTGGATGTGGAGATAGCTTAGTCTCGG	ACGTTGGATGATCACCGCAAAATCTCTCCC	TTCCAACATCTTAATTTTCTTG
S131	ACGTTGGATGGTTCGCCAGCCATGTTCTTG	ACGTTGGATGACGATCGTCTCACTGCAATC	TTATTTTCATGTGTAGGTGACT
S132	ACGTTGGATGTAATAGGAAGTGTTAAGGG	ACGTTGGATGCCAGGAAAAACCATCACCTC	GCCAGCCATGTTCTTGTCTCTCA
S133	ACGTTGGATGTGCAAAAAAGCTAGAGGAG	ACGTTGGATGTCTAAACCTGTGCTGTTGG	GGGTAATTAATAAAGCAAGGATTA
S134	ACGTTGGATGTGGTTGTGCAACTTACAAG	ACGTTGGATGATCAGCGGCAGGAGTAGATG	AAAAAGGCCATGAAACAT
S135	ACGTTGGATGCCTTACTTGAGAGTTCAGC	ACGTTGGATGATGCAACAGAAAGGATCTCGC	AACTTACAAGATACTAAATGCTTGCA
S136	ACGTTGGATGAAAGGATCTGACTAAGGG	ACGTTGGATGATACTCTTCAACTGCTGG	TTGAGAGTTCAGCATTTAACTATT
S137	ACGTTGGATGATCGCCGTCGAAAAAGAA	ACGTTGGATGGAGGCCAACCGCCGGCTG	TAGATACACCAGCAAAAGTACCA
S138	ACGTTGGATGCGATGGAAGTATTTTCTTC	ACGTTGGATGGAAAGAGAACTTTCTCCA	GAAAAAATGACCGCCA
S139	ACGTTGGATGAAGCGAAATCGAGTGGTGC	ACGTTGGATGTGGTTGTGATGTTCCGGAGC	TTTCTCATTTGGTGTGTAAGATG
S140	ACGTTGGATGCACCGAGCAATAATCCTAAC	ACGTTGGATGCCGATTTGGTTGTCTCGAAC	TCCTTATTGTACCCCTCTGGATT
S141	ACGTTGGATGAGTATCAACCCTCCAC	ACGTTGGATGGTAAATATCAGCTTGGTTGC	TCCTAATCTCATGGAAACACTTTC
S142	ACGTTGGATGTGGACATCTTGTGTTTG	ACGTTGGATGGACCAAGAGTCCGATCCAAAG	TGTCGTACAAATCTTTGAGTC
S143	ACGTTGGATGGATGACACGTCATCATCTC	ACGTTGGATGACAAAGAACTTGTGAAGAG	TCCAAATAGTTATGACAAAAGTT
S144	ACGTTGGATGAAACAGTGGATCACAGTAGC	ACGTTGGATGAATTTCTCTCGTACCGAAGC	CAGTCATCATCTCTCTTTTTC
S145	ACGTTGGATGAATGTAAGTGGTACAGCTCC	ACGTTGGATGCCCCAATAACTCCAATGAG	TAGCAGAAAGCTGCTTACTTCGACCAT
S146	ACGTTGGATGCCGGTCTAAGTATATGC	ACGTTGGATGCTGAAAAACCTCTCCCCTAC	TACAGCTCCAAGCATCAGTTCCC
S147	ACGTTGGATGTTCTATCTGGACTACCTCG	ACGTTGGATGCGGATCACTGAAACCGATAGG	TGACATTTTTCTGTTGCATGTT
S148	ACGTTGGATGATGTCAGTCTGCTTCTTC	ACGTTGGATGGTATACCTCGGGAGAAAAG	CTATTTTCATTTCCCCCAA
S149	ACGTTGGATGGTAAACAGAGTTGACAAAATG	ACGTTGGATGGTGGCCAAAAGTAAAGAAAG	CGCTGCTTCTTCTCTCTCTCTC
S150	ACGTTGGATGGCATAAACTTCTCTGCTCTC	ACGTTGGATGTACCTAAAGGTGCAAAAAGT	TTATTAGTGTCTGAATGGATAATC
S151	ACGTTGGATGATCTGCGACAAACGGGCAAG	ACGTTGGATGAGCAAGACCTCCACGTAAAC	CTGGCCATATGCTTTATCTC
S152			GACAAACGGGCAAGCAGGTGCG

Table S3. (Cont'd.).

SNP_ID	2nd-PCR	1st-PCR	UEP_SEQ
S153	ACGTTGGATGGAAGTTACCGGACTGGAC	ACGTTGGATGGATGTGCTCTGATGACTATGC	ATGGTCGAGCGAACTTT
S154	ACGTTGGATGACAGGCATACTACCTCTACC	ACGTTGGATGAGAGCTTAAATCTCCCAAG	TGGATTTTGATGACGGG
S155	ACGTTGGATGAGATAAGCAGCGATGGATGG	ACGTTGGATGTCTCTCTTTGTGACGCGG	TGGCGTTGGCAGTGGCGTTGTCTTC
S156	ACGTTGGATGACGAGAAGTTCAAGGAGGAG	ACGTTGGATGAGTTCTGCAATGGTAAAGCCTC	CGAGAAAGTTCAAGGAGGAGGAGGTCAAC
S157	ACGTTGGATGGTCATAGCCCAATTACCTTGC	ACGTTGGATGGAGTCAATTGAGGATGCACTG	ACCTTGCATGTCTCACG
S158	ACGTTGGATGGTCCGAAACTTCTCGATCAC	ACGTTGGATGGAGACTGAAGGCATTTGGATG	GATCACAAGACCTTCAACCAAATC
S159	ACGTTGGATGTATCCAGCACACCAAAG	ACGTTGGATGGGCTTTTCCAGTCGTA AAC	GGCACACCGAAGTCAAAGT
S160	ACGTTGGATGTACCACTACGCATGTATGCC	ACGTTGGATGTGAAACCAAAATATATCGCC	TATGCCAAGCAAAAAAATATATGCCG
S161	ACGTTGGATGTATCTCCCGTTTTTCAGG	ACGTTGGATGATTGAAGAGGGCAGCAAAGG	CATTCTGGTGTAGCATCACCT
S162	ACGTTGGATGTGCTGCTTTGGCATTGTCG	ACGTTGGATGTGATGGAGATGTTGACGATG	GGCATTGTGCTGCCACCCG
S163	ACGTTGGATGTGCGCTGCCTCTCTGCT	ACGTTGGATGATAGAATGCAGCTAGCGAGG	CGGCCACTGCCCTTACC CGCGGAGTC
S164	ACGTTGGATGGTCCCTTACCACTGCTGATG	ACGTTGGATGTAGATTATAGAACACCAGC	GGAGTTTTGTTAAAAAACCAAAGG
S165	ACGTTGGATGCTCTATCTCACTGCAATTGG	ACGTTGGATGAGTACACGTAGCCATGGATG	TAGCTAGCCAAGTCTCTCA
S166	ACGTTGGATGGCATTCTGGTAAGTTCTTC	ACGTTGGATGTCACTTTTGTGAGTCC	GTCTTACATATATAGTCTTACCT
S167	ACGTTGGATGTACCTCAAGCCTTCTACAC	ACGTTGGATGGTGATCAAACTAAAACCGC	TCCTCCTTTCGGCCTATC
S168	ACGTTGGATGATTTGTGGCTCACCTGCTTC	ACGTTGGATGTCAAAGGTCTCTACCGGTT	CCTGTTCAATTTGATCATGTACTG
S169	ACGTTGGATGGCACTAGAGATGACCAACAC	ACGTTGGATGCAGCAGGGAGA ACTTATTT	GAGATGACCAACAATGGAGAACCT
S170	ACGTTGGATGAATGAGACCTTAGGCTCC	ACGTTGGATGGTCCACCTCTCGTTTTAG	TAGGCTCCAGCCGTACA
S171	ACGTTGGATGCTGATCTTCTGTGGCTTG	ACGTTGGATGTGGCATCGTCAATAAC	ACCAACAACATGGTACTC
S172	ACGTTGGATGTCTGTGGGCTGCATACTG	ACGTTGGATGCAGTCTCTCTTGTGCAATAC	TTTGGCTTTTCTGAAGCGT
S173	ACGTTGGATGGCTTGAATATGTGTGGAGC	ACGTTGGATGCCAGTTAAACCCCAACAAC	TGTGGAGCAGAATAGATAGTT
S174	ACGTTGGATGAATGGTCCCAGCCTTCCCAG	ACGTTGGATGTTAACTCCAGAACCTGGTG	TTCCAGACGTCATCGTATTC
S175	ACGTTGGATGATGCTTCCGCCATCGCTC	ACGTTGGATGAATCGGAGGTACGGCAAAG	CCCCCTCTACCTCGCCT
S176	ACGTTGGATGTTCTTGGGCAGACTGGA AAC	ACGTTGGATGAGAGATTGGAGAAGAGCCAC	TACTTGGGGAGGTCACCTT
S177	ACGTTGGATGCCCTTGTGCAATTTGCCATC	ACGTTGGATGGCAGATGCTATGACACAGAG	TTTTGCCATCTCTATGTAGAAATC
S178	ACGTTGGATGATTGCCCTCCTCTCCTGCT	ACGTTGGATGATCGAGGGAGAGGGTGATG	CCGCCACATCCCTTCCCTA
S179	ACGTTGGATGCTCTGTGAAGAGTTGTGGG	ACGTTGGATGTATGGTTTTAGCGGAAGAG	TATCATTGGTAAAGTCCATCCTCC
S180	ACGTTGGATGAGAAAACCTAGAGGGCAAC	ACGTTGGATGAAGCAGCAGCGTGAGGATTG	AGGCCAACAGGAGGGAGTC
S181	ACGTTGGATGCCATCTTTCTTTGCTTGTG	ACGTTGGATGGTTGATGCGTCTTTTGTAGC	TAATTTCAATTTCTAGTGCCTCGT
S182	ACGTTGGATGGTCAAGCCATTCTGTTTTTC	ACGTTGGATGCCAAGTGTCAATCATGAGC	CTTCTTTTCTGTGTGAATGAA
S183	ACGTTGGATGCAGTTTCATGATCGACAGGG	ACGTTGGATGACATATACITTCGTTCAAAG	CGACAGGGGATTTCTACTG
S184	ACGTTGGATGAGTTGCAGCTTGTATGCACC	ACGTTGGATGATCACTCCGTTTAAAGCTCCG	ACTCCATCTTTTGTAGATCTT
S185	ACGTTGGATGAGCCCAATGCCATAGTATCC	ACGTTGGATGGACCTTGGCCATATTGTACGC	CCCAAAGGAAGCTTCTT
S186	ACGTTGGATGCCTCTCATCTGCACAAATTC	ACGTTGGATGATCTGCAAGCAAGCACATCC	CACAAATCTGCAGACGCTC
S187	ACGTTGGATGTGTTGAGCTGAACGATGAC	ACGTTGGATGCTGATACTAGTGAATGAGC	CTTACATTTTTATAGTACTCCT
S188	ACGTTGGATGCACCGGTTTAGCTTCACTC	ACGTTGGATGAAAACCTAGATCAATGACACC	CAAGGAAAGGCTTGAAGCGTTT
S189	ACGTTGGATGAACACAGCAAGGAGGAGTG	ACGTTGGATGCTAAACTCCTTTGCTCTTGG	CCCAAATCTATGCAATGGA
S190	ACGTTGGATGCTATTGAGGCCAATCTTGGC	ACGTTGGATGGAGAACTCCAGGTATACACTTG	GAAACTAGCTCGTCAATGAGG

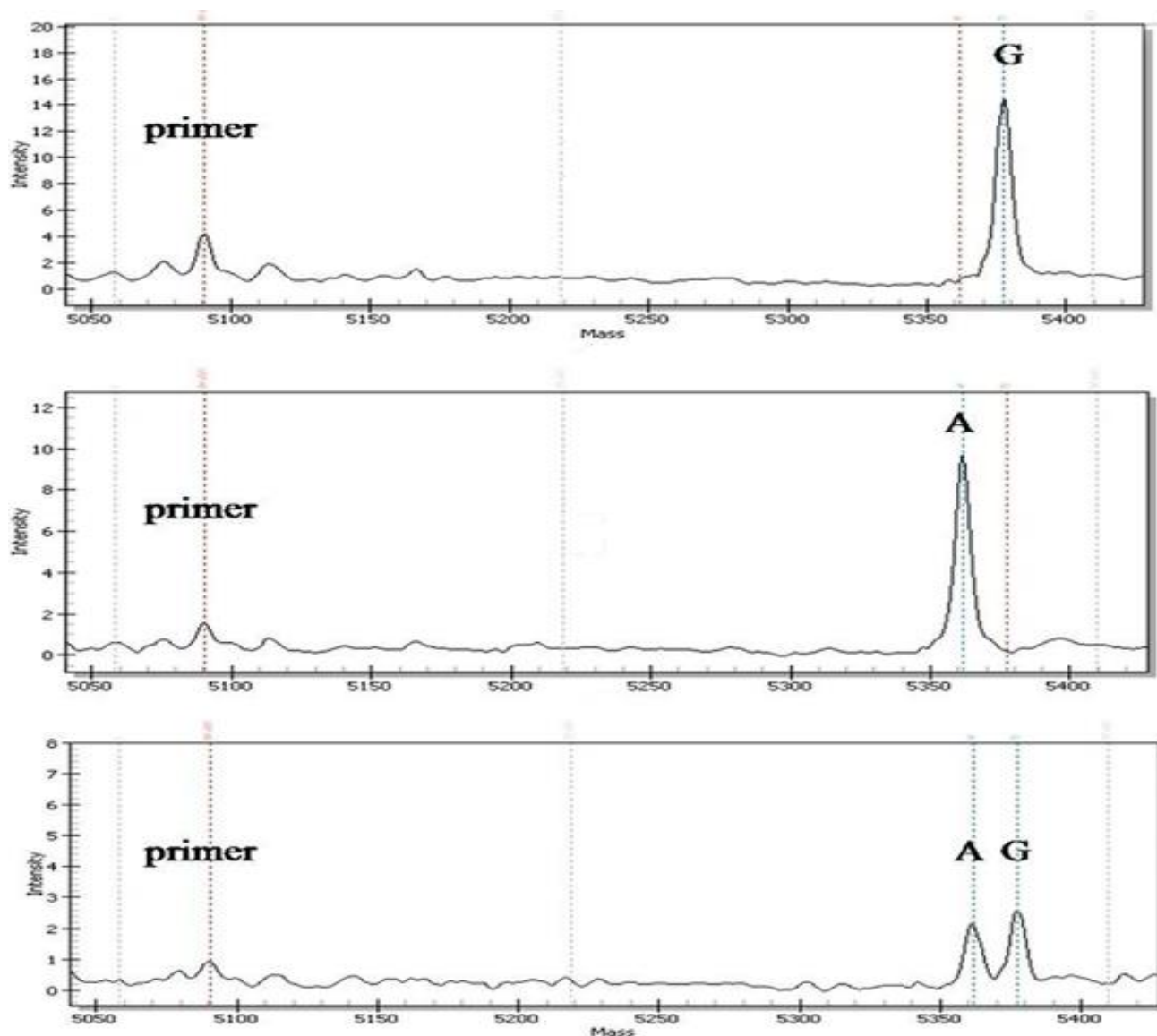


Fig. 2. Mass spectrum representing probe primer extension peaks and allele peaks 'G', 'A,' and 'AG', observed in the genotyping of 109 *Oryza sativa* L. ssp. samples at 190 single nucleotide polymorphism (SNP) loci.

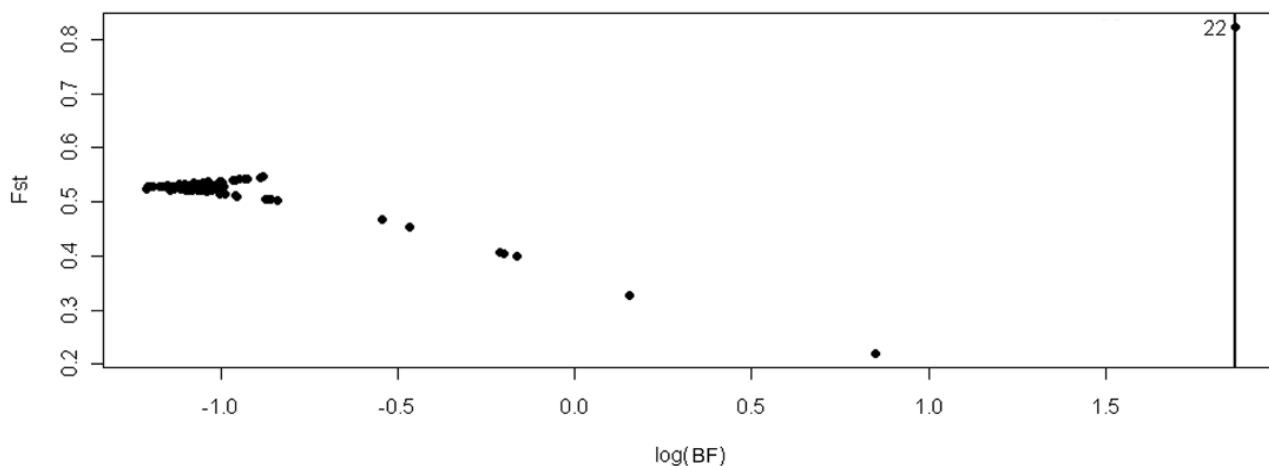


Fig. 3. Identification of outlier loci presumably subject to selection based on a Bayesian genomic scan. Each point corresponds to a single nucleotide polymorphism (SNP) marker. SNP22 was consistently identified as an outlier through 10 independent iterations.

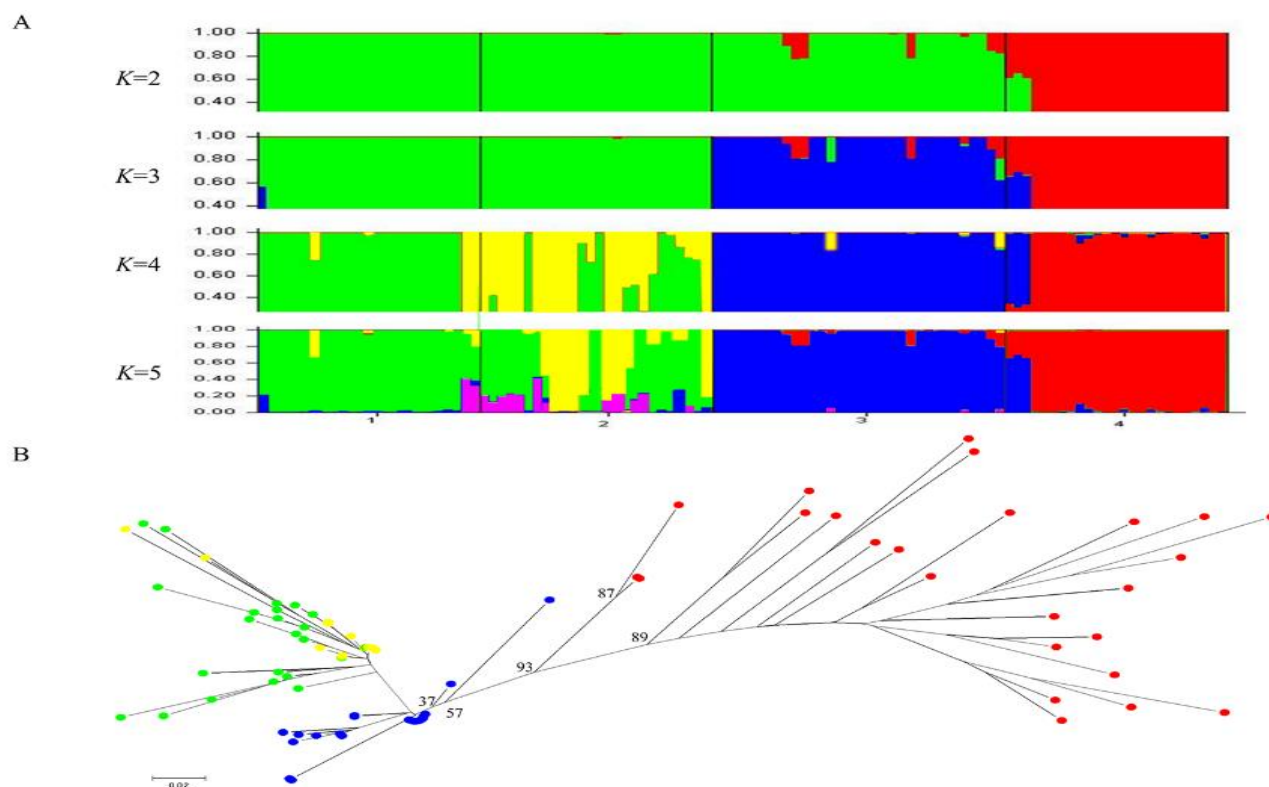


Fig. 4. The population structure and Neighbor-joining consensus tree of the sampled groups with *indica* groups
A, Population structure analysis plots for $k = 2-5$ obtained using the Structure 2.3.1 software based on 189 SNPs in *Oryza sativa* L. ssp. The four populations analyzed are displayed as follows (from left): exotic *japonica*, historical *japonica*, modern *japonica*, and *indica*.
B, Neighbor-joining consensus tree were built from 189 functional SNPs. Numbers at the main branch nodes represent bootstrap values. Exotic *japonica*, historical *japonica*, modern phase *japonica*, and *indica* are indicated in green, yellow, blue, and red, respectively.

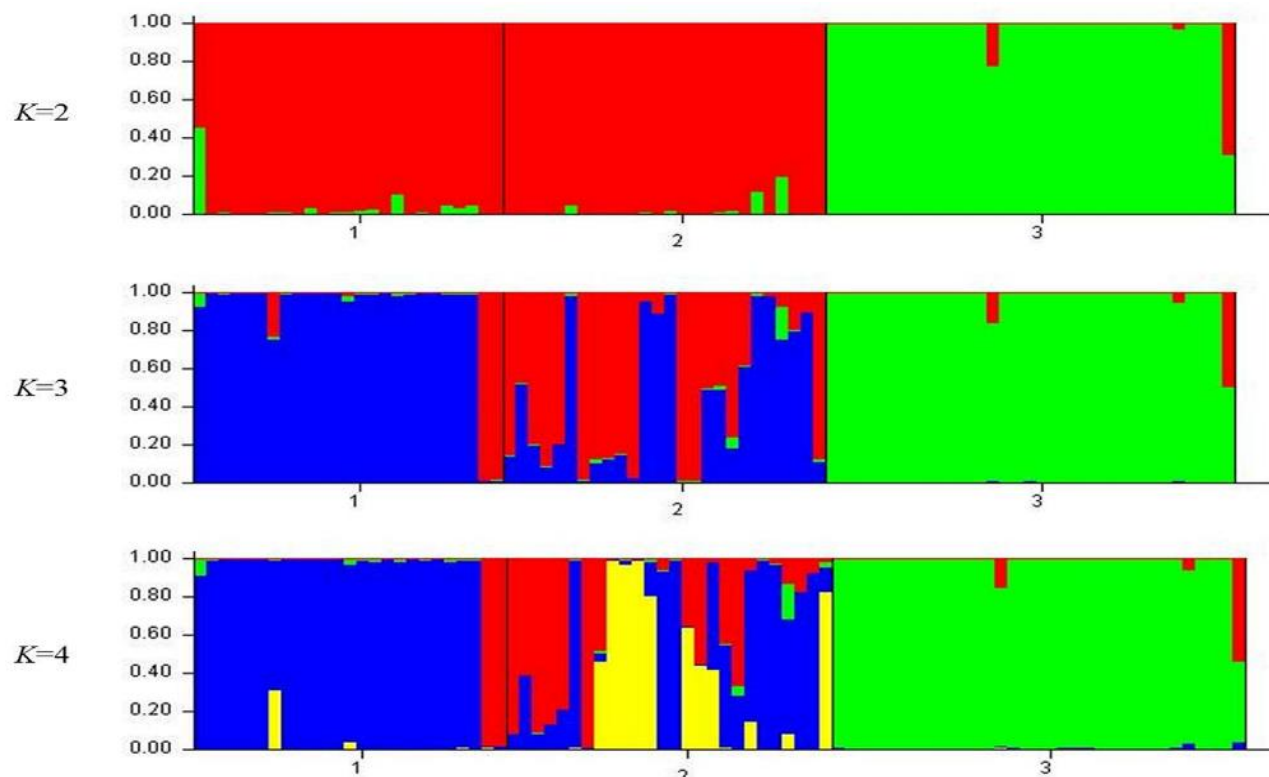


Fig. 5. The population structure of the sampled groups without *indica* groups
 Population structure analysis plots for $k = 2-4$ obtained using the Structure 2.3.1 software based on 189 SNPs in *Oryza sativa* L. ssp. The three populations analyzed are displayed as follows (from left): exotic *japonica*, historical *japonica*, and modern *japonica*.

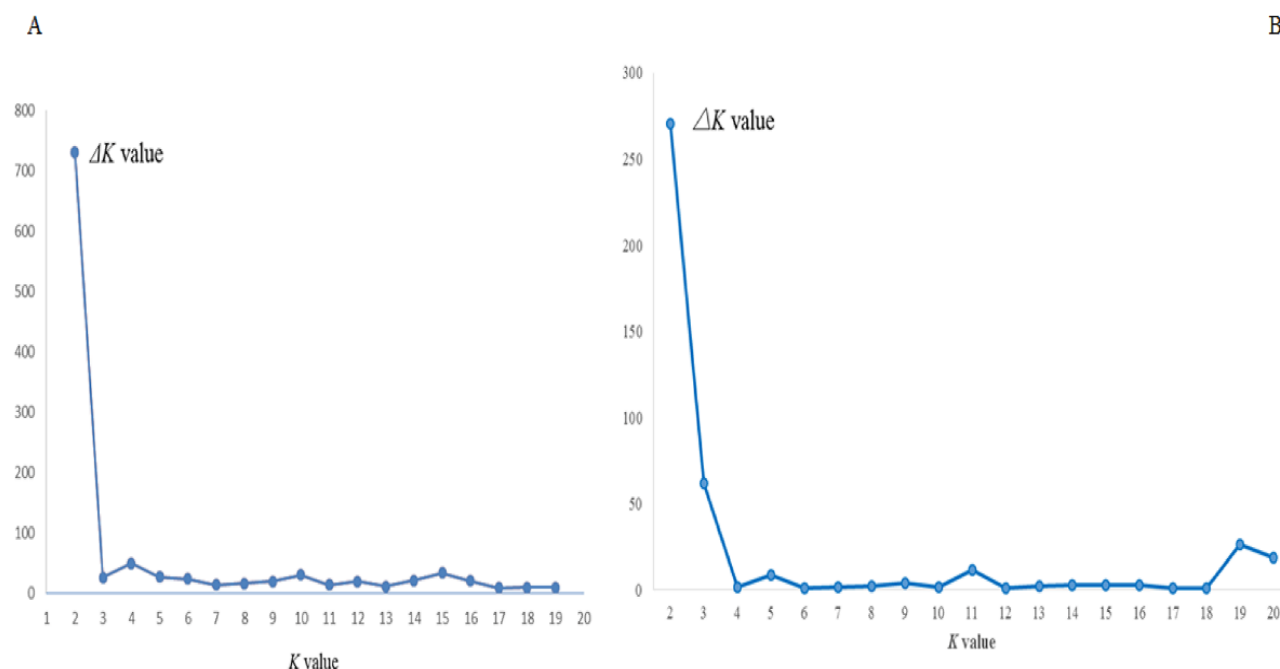


Fig. 6. The structure consisting of two groups ($k=2$) best fit the clustering results by Evanno's ΔK method.

Discussion

The introduction of *indica* lineage into the northeastern Chinese *japonica* rice has significantly improved agronomic performance of rice varieties, however the genetic diversity in these populations doesn't seem to have been enriched as a result. This is in contrast to a previous report that use of *indica* rice as a breeding resource can enrich the genetic diversity of modern varieties analyzed by SSR markers (Huang *et al.*, 2010). We speculate that the introgression from *indica* was not sufficiently detectable using these SNP markers. Breeders usually focus on one or a few desirable traits, such that a few core parental alleles are isolated from accumulated genetic variation, thus breeding parents frequently have a narrow genetic background. Although hybridization has generated new haplotype polymorphisms across the modern Japanese *japonica* genome, haplotype diversity has decreased overall as a consequence of artificial selection (Yamamoto *et al.*, 2010; Yonemaru *et al.*, 2012). In general, modern varieties with good eating quality in Japan have been bred successfully through hybridization and selection, and its population is still pure *temperate japonica* (Nagasaki *et al.*, 2010). Although ecological conditions in *japonica*-producing areas of northeastern China and Japan are similar, different breeding strategies may have quickly led the rice genome to different paths diverged. In contrast, the development and formation of modern *japonica* in northeastern China has been in less than 50 years. The genetic bottleneck created by selection and ecological stress has caused its novel genomic and phylogenetic status. In the present study, the neo-diversity arisen from the introgression of other germplasm resources may introduce valuable genetic variation for improved varieties in northeastern China. Our study has opened a new door to exploration of this field.

Acknowledgement

This work was supported by the National Key R&D Program of China (2017YFD0300700), the National Natural Science Foundation of China (31371587 and 31430062), the Program for Changjiang Scholars and Innovative Research Team in University (IRT13079), and the Cultivation Plan for Youth Agricultural Science and Technology Innovative Talents of Liaoning Province (2014046).

References

- Ali, N., Farhatullah, N.U. Khan, M.A. Rabbani, I. Hussain, S. Ali, S.A. Khan and M.Q. Kakar. 2017. Genetic diversity of *Brassica Raps L.* indigenous landraces based on cluster and principal component analyses. *Pak. J. Bot.*, 49: 1891-1901.
- Arai-Kichise, Y., Y. Shiwa, H. Nagasaki, K. Ebana, H. Yoshikawa, M. Yano and K. Wakasa. 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *japonica* rice by whole-genome sequencing. *Plant & Cell Physiol.*, 52: 274-282.
- Caicedo, A.L., S.H. Williamson, R.D. Hernandez, A. Boyko, A.F. Alon, T.L. York, N.R. Polato, K.M. Olsen, R. Nielsen and S.R. McCouch. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics*, 3: e163.
- Cheng, C., R. Motohashi, S. Tsuchimoto, Y. Fukuta, H. Ohtsubo and E. Ohtsubo. 2003. Polyphyletic origin of cultivated rice: based on the interspersed pattern of SINEs. *Mol. Biol. & Evol.*, 20: 67-75.
- Ebana, K., J.I. Yonemaru, S. Fukuoka, H. Iwata, H. Kanamori, N. Namiki, H. Nagasaki and M. Yano. 2010. Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza sativa L.*). *Breed. Sci.*, 60: 390-397.
- Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.*, 14: 2611-2620.

- Foll, M. and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180: 977-993.
- Gabriel, S., L. Ziaugra and D. Tabbaa. 2009. SNP genotyping using the Sequenom MassARRAYiPLEX platform. *Curr. Prot. Human Gen.*, 2.12: 11-12.12. 16.
- Garris, A.J., T.H. Tai, J. Coburn, S. Kresovich and S. McCouch. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 169: 1631-1638.
- He, Z., W. Zhai, H. Wen, T. Tang, Y. Wang, X. Lu, A.J. Greenberg, R.R. Hudson, C.I. Wu and S. Shi. 2011. Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genetics*, 7: e1002100.
- Huang, P., J. Molina, J.M. Flowers, S. Rubinstein, S.A. Jackson, M.D. Purugganan and B.A. Schaal. 2012. Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. *Mol. Ecol.*, 21: 4593-4604.
- Huang, X., N. Kurata, X. Wei, Z.X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu and W. Li. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490: 497-501.
- Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu and Z. Zhang. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42: 961-967.
- Jeffreys, H. 1961. Theory of probability. In: Oxford University Press, Oxford UK.
- Li, C., A. Zhou and T. Sang. 2006. Rice domestication by reducing shattering. *Science*, 311: 1936-1939.
- Librado, P. and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25: 1451-1452.
- Londo, J.P., Y.C. Chiang, K.H. Hung, T.Y. Chiang and B.A. Schaal. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences of USA*, 103: 9578-9583.
- Lu, B.R., X. Cai and J. Xin. 2009. Efficientindica and japonica rice identification based on the InDel molecular method: Its implication in rice breeding and evolutionary research. *Prog. In Nat. Sci.*, 19: 1241-1252.
- Mao, T., X. Li, Z.J. Xu, J.Y. Wang, L. Tang, H. Xu, S.K. Jiang, Z. Zhang and Z.Y. Li. 2018. Development of functional markers for quantitative trait locus underlying rice quality components. *Pak. J. Bot.*, 50: 645-653.
- Molina, J., M. Sikora, N. Garud, J.M. Flowers, S. Rubinstein, A. Reynolds, P. Huang, S. Jackson, B.A. Schaal and C.D. Bustamante. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of USA*, 108: 8351-8356.
- Nagasaki, H., K. Ebana, T. Shibaya, J.I. Yonemaru and M. Yano. 2010. Core single-nucleotide polymorphisms—a tool for genetic analysis of the Japanese rice population. *Breed. Sci.*, 60: 648-655.
- Oka, H. 1988. Origin of cultivated rice. Japan Scientific Societies Press. Tokyo, Japan.
- Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.
- Sang, T. and S. Ge. 2007. Genetics and phylogenetics of rice domestication. *Curr. Opin. Gen. & Develop.*, 17: 533-538.
- Sun, J., D. Liu, J.Y. Wang, D.R. Ma, L. Tang, H. Gao, Z.J. Xu and W.F. Chen. 2012. The contribution of intersubspecific hybridization to the breeding of super-high-yielding japonica rice in northeast China. *Theor. & App. Gen.*, 125: 1149-1157.
- Sun, J., Q. Qian, D.R. Ma, Z.J. Xu, D. Liu, H.B. Du and W.F. Chen. 2013. Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New Phytol.*, 197: 290-299.
- Sweeney, M.T., M.J. Thomson, Y.G. Cho, Y.J. Park, S.H. Williamson, C.D. Bustamante and S.R. McCouch. 2007. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics*, 3: e133.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. & Evol.*, 30: 2725-2729.
- Wei, X., W.H. Qiao, Y.T. Chen, R.S. Wang, L.R. Cao, W.X. Zhang, N.N. Yuan, Z.C. Li, H.L. Zeng and Q.W. Yang. 2012. Domestication and geographic origin of *Oryza sativa* in China: insights from multilocus analysis of nucleotide variation of *O. sativa* and *O. rufipogon*. *Mol. Ecol.*, 21: 5073-5087.
- Xu, X., X. Liu, S. Ge, J.D. Jensen, F. Hu, X. Li, Y. Dong, R.N. Gutenkunst, L. Fang and L. Huang. 2012. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnol.*, 30: 105-111.
- Yamamoto, T., H. Nagasaki, J. I. Yonemaru, K. Ebana, M. Nakajima, T. Shibaya and M. Yano. 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genom.*, 11: 267.
- Yonemaru, J.I., T. Yamamoto, K. Ebana, E. Yamamoto, H. Nagasaki, T. Shibaya and M. Yano. 2012. Genome-wide haplotype changes produced by artificial selection during modern rice breeding in Japan. *PLoS One*, 7: e32982.
- Zhang, L.B., Q. Zhu, Z.Q. Wu, J. Ross-Ibarra, B. S. Gaut, S. Ge and T. Sang. 2009. Selection on grain shattering genes and rates of rice domestication. *New Phytol.*, 184: 708-720.
- Zhao, X., L. Yang, Y. Zheng, Z. Xu and W. Wu. 2009. Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). *J. Gen. & Genom.*, 36: 435-442.
- Zhu, B.F., L. Si, Z. Wang, Y. Zhou, J. Zhu, Y. Shangguan, D. Lu, D. Fan, C. Li, H. Lin, Q. Qian, T. Sang, B. Zhou, Y. Minobe and B. Han. 2011. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.*, 155: 1301-1311.