

SHORT REPORT

Open Access

Re-sequencing and genetic variation identification of a rice line with ideal plant architecture

Shuangcheng Li^{1,2,3}, Kailong Xie², Wenbo Li², Ting Zou², Yun Ren², Shiquan Wang^{1,2,3}, Qiming Deng^{1,2,3}, Aiping Zheng^{1,2}, Jun Zhu^{1,2}, Huainian Liu¹, Lingxia Wang¹, Peng Ai², Fengyan Gao², Bin Huang², Xuemei Cao² and Ping Li^{1,2,3*}

Abstract

Background: The ideal plant architecture (IPA) includes several important characteristics such as low tiller numbers, few or no unproductive tillers, more grains per panicle, and thick and sturdy stems. We have developed an *indica* restorer line 7302R that displays the IPA phenotype in terms of tiller number, grain number, and stem strength. However, its mechanism had to be clarified. **Findings:** We performed re-sequencing and genome-wide variation analysis of 7302R using the Solexa sequencing technology. With the genomic sequence of the *indica* cultivar 9311 as reference, 307 627 SNPs, 57 372 InDels, and 3 096 SVs were identified in the 7302R genome. The 7302R-specific variations were investigated via the synteny analysis of all the SNPs of 7302R with those of the previous sequenced non-IPA-type lines IR24, MH63, and SH527. Moreover, we found 178 168 7302R-specific SNPs across the whole genome and 30 239 SNPs in the predicted mRNA regions, among which 8 517 were Non-syn CDS. In addition, 263 large-effect SNPs that were expected to affect the integrity of encoded proteins were identified from the 7302R-specific SNPs. SNPs of several important previously cloned rice genes were also identified by aligning the 7302R sequence with other sequence lines. **Conclusions:** Our results provided several candidates account for the IPA phenotype of 7302R. These results therefore lay the groundwork for long-term efforts to uncover important genes and alleles for rice plant architecture construction, also offer useful data resources for future genetic and genomic studies in rice.

Keywords: Rice, IPA, Re-sequencing, SNP, InDel, SV

Findings

Field performances of 7302R

We examined the field performances of 7302R (Figure 1) by comparing several of its yield-related traits with those of IR24, MH63, and SH527, which are considered as core representative restorer lines for hybrid rice. No obvious differences were found in the yield components of 1000 grain weight (Figure 1E) and weight per plant (Figure 1G) between 7302R and other lines. However, a significant decrease in the tiller number per plant

(Figure 1B) and an increase in grain number per main panicle (Figure 1C) were observed. Moreover, the increase in grain number resulted in an apparent increase in the weight per main panicle (Figure 1F). Understandably, the seed setting rate of 7302R (Figure 1D) was decreased, which might have been a negative result of the huge increase in grain number. In addition, the stem of 7302R became stronger than those of the other rice lines. Upon comparison with other currently cultivated rice varieties, the newly developed *indica* restorer line 7302R was found to display the IPA phenotype, particularly in terms of tiller number, grain number, and stem strength.

Genome sequencing and variation identification

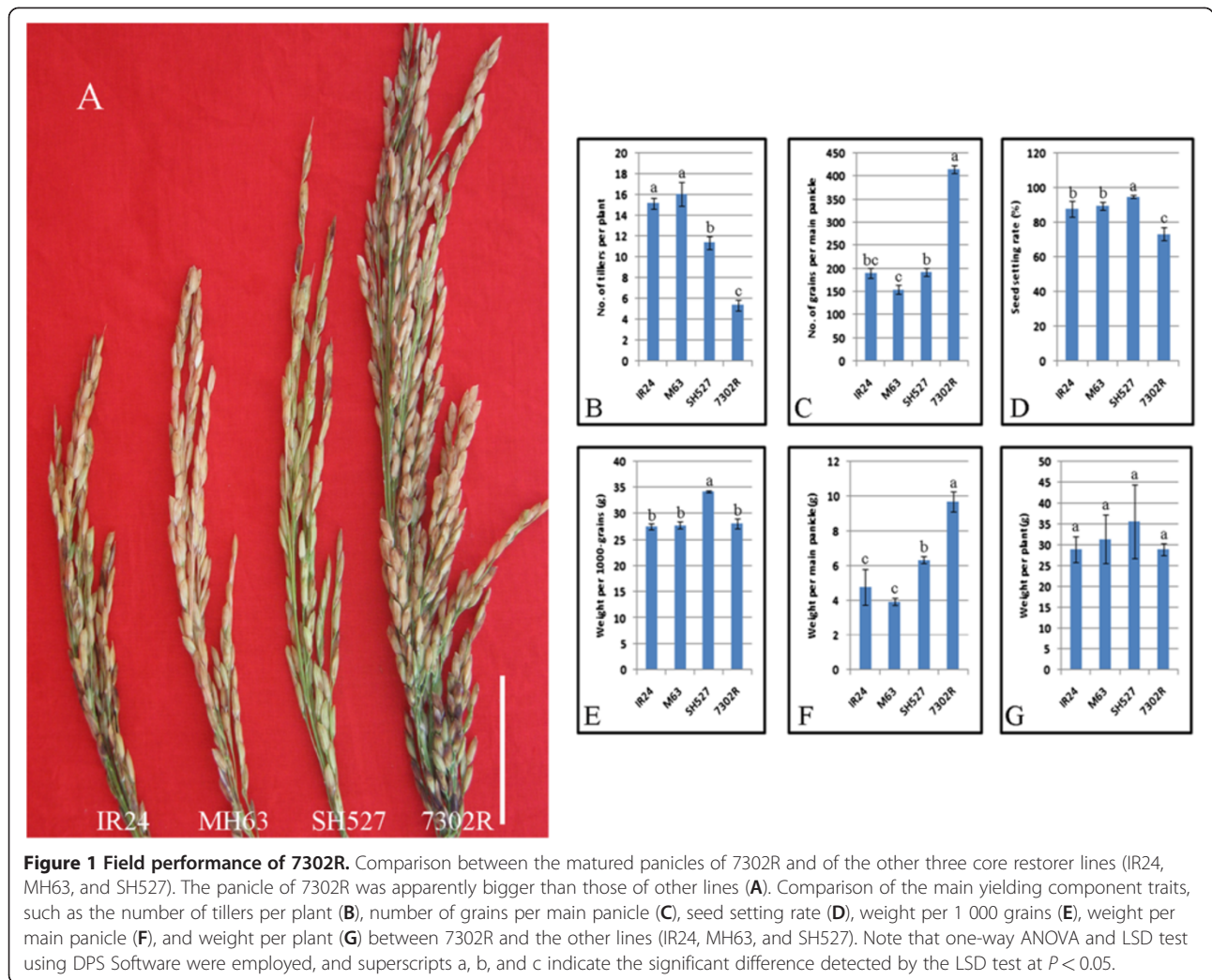
The 7302R genotypes were determined with approximately a 10-fold coverage by genome sequencing

* Correspondence: liping6575@163.com

¹State Key Laboratory of Hybrid Rice, Sichuan Agricultural University, Chengdu 611130, China

²Rice Research Institute, Sichuan Agricultural University, 211 Huimin Road, 611130 Wenjiang, Sichuan, China

Full list of author information is available at the end of the article



using the Solexa sequencing technology. According to the protocol, a DNA library with an average insertion length of 484 bp was constructed and 6.35 G bases were generated. The alignment of reads was used to build consensus genome sequences for 7302R. Moreover, approximately 4.77 G high-quality raw databases were aligned with the reference sequence of cultivar 9311 using the SOAPaligner (Li et al. 2008). An overall effective depth of 13× coverage was achieved (Table 1), and the resulting consensus sequence covered approximately 82.57% of the reference genome.

Genome-wide variations were then examined via SOAPsnp11 and SOAPsv using a conservative quality filter pipeline (Li et al. 2009), and 307 627 SNPs, 57 372

InDels, and 3 096 SVs were yielded from the 7302R genome (Table 2, Additional files 1, 2 and 3). We have previously re-sequenced three important representative restorer lines, namely, IR24, MH63, and SH527, using the same technology (Li et al. 2012). The overall genome diversity among these re-sequenced lines was much lower than that reported for a more diverse population (Huang et al. 2010) because of the inherent relationship between these samples, suggesting a close relationship between the sequenced lines. The relative close relationship was also consistent with the previous result of restorer lines that have narrow genetic backgrounds (Duan et al. 2002). A phylogenetic tree was constructed (Tamura et al. 2007) using several authentic collections

Table 1 Summary of original sequencing data (9311 as the reference)

Sample	Insert size	Bases (G)	Mapped bases (G)	Depth	Coverage (%)	Mismatch rate (%)
7302R	484	6.35	4.77	13.26	82.57	0.65

Table 2 SNP, InDel and SV detected among the 7302R genome (9311 as the reference)

Chr.	SNP	InDel	SV
Chr01	39,674	8,046	483
Chr02	36,901	7,199	231
Chr03	32,756	6,451	327
Chr04	27,159	4,810	282
Chr05	26,229	4,951	236
Chr06	24,092	4,424	225
Chr07	21,525	3,699	118
Chr08	27,935	5,200	421
Chr09	13,531	2,524	114
Chr10	22,382	3,826	253
Chr11	13,777	2,490	193
Chr12	21,666	3,752	213

of SNPs for each sequenced line, and a relatively distant relationship was observed between 7302R and the other restorer lines (Figure 2B).

The frequencies of SNPs, InDels, and SVs for 7302R were plotted at a 100 kb sliding window, with a step size of 50 kb along each chromosome, by comparing them with those of IR24, MH63, and SH527. The SNP/InDel/SV frequency was defined as the corresponding number of SNPs/InDels/SVs divided by the number of nucleotides within the 100 kb interval,

excluding the uncovered nucleotides. Each sample was compared with the corresponding interval to identify regions that showed non-random variation frequencies (Figure 2A). Overall, 47/292, 91/124, and 39/571 SNP, InDel, and SV high/low regions were identified in the 7302R genome (Table 3). The most abundant chromosomes in the SNP high region were chr. 7, chr. 5, chr. 12, and chr. 2. Moreover, the most abundant chromosomes in the InDel high region were chr. 7, chr. 5, chr. 3, and chr. 2. Among these chromosomes, chr. 5 and chr. 2 were shown to be distributed with more SV high regions. These results show that these chromosomes are more abundant in genetic variations. In addition, several chromosomal loci in the SNP high region were also found covered by the InDel and SV high regions (Figure 2A), which might suggest that those regions were the most polymorphic.

Identification and characterization of 7302R-specific SNPs

As the differences between 7302R and those none-IPA type lines may reflect the genetic improvement of the IPA-type rice from the current none-IPA cultivars, an investigation of the 7302R-specific variations was therefore performed using a synteny analysis of all the SNPs of the 7302R compared with those of IR24, MH63, and SH527. We revealed a total of 178 168 7302R-specific SNPs across the whole genome, and the distribution in each chromosome is shown in Figure 3. The chr. 2, chr. 4,

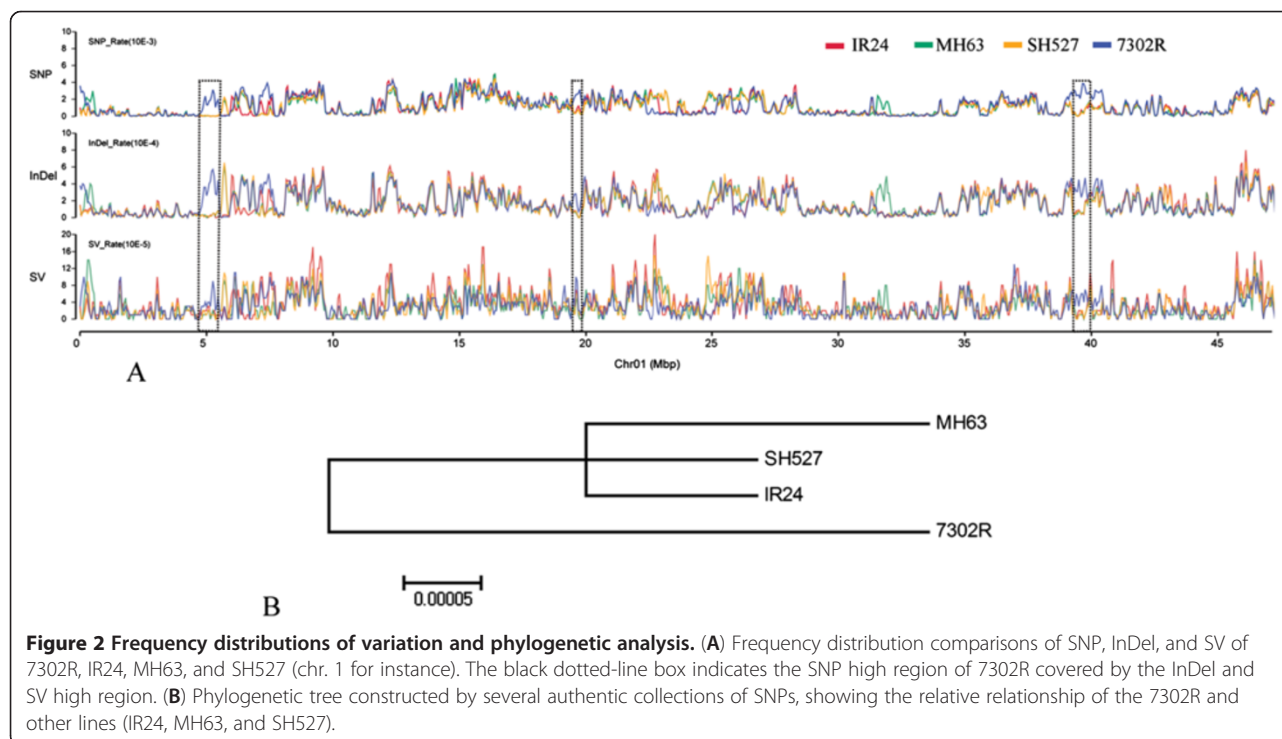


Figure 2 Frequency distributions of variation and phylogenetic analysis. (A) Frequency distribution comparisons of SNP, InDel, and SV of 7302R, IR24, MH63, and SH527 (chr. 1 for instance). The black dotted-line box indicates the SNP high region of 7302R covered by the InDel and SV high region. (B) Phylogenetic tree constructed by several authentic collections of SNPs, showing the relative relationship of the 7302R and other lines (IR24, MH63, and SH527).

Table 3 High and low frequency regions of variation distribution in 7302R genome

Chr.	SNP_high	SNP_low	InDel_high	InDel_low	SV_high	SV_low
Chr01	0	64	3	20	3	72
Chr02	6	13	19	9	8	49
Chr03	0	32	15	16	3	93
Chr04	3	32	2	5	3	71
Chr05	8	16	19	13	10	28
Chr06	5	32	6	11	4	65
Chr07	10	27	13	19	1	60
Chr08	1	14	2	6	2	23
Chr09	4	17	4	1	0	22
Chr10	1	12	4	2	2	17
Chr11	2	23	0	18	0	42
Chr12	7	10	4	4	3	29
Total	47	292	91	124	39	571

and chr. 7 were found to be the three most abundant chromosomes in the 7302R-specific SNPs.

The SNPs in the coding regions were analyzed to further understand the potential functional effects of the 7302R-specific SNPs. A total of 30 239 SNPs were located in the predicted mRNA regions, among which 4 946 were synonymous coding sequences (Syn CDS) and 8 517 were non-synonymous coding sequences (Non-syn CDS) (Table 4).

Two hundred sixty-three large-effect SNPs that were expected to affect the integrity of the encoded proteins were further identified from the 7302R-specific SNPs (Figure 4). These included 187 premature terminations, 27ATG changes, and 49 stop changes. Be accordance with the distribution of 7302R-specific SNPs, Chr. 2, chr. 4, and chr. 7 were also the three most abundant large-effect SNP chromosomes.

Gene Ontology analyses were further conducted for the genes in the 7302R-specific SNPs to explore the gene functions. Investigations showed that the top GOs were protein kinase activity, nucleic acid binding, catalytic activity, protein binding, and DNA binding (Figure 5). Our finding is partially consistent with our previous result of the gene function analysis of variations between restorer lines (Li et al. 2012).

Variation analysis on important rice genes

We investigated the natural variations among ~60 genes, which might explain the phenotypic differences of the sequenced sample. A large number of SNPs (Table 5) were detected both in the DNA sequence and in the coding regions of genes related to disease resistance, such as *Pib* (Wang et al. 1999), *Xa1* (Yoshimura et al. 1998), and *Xa21* (Song et al. 1995). Other disease

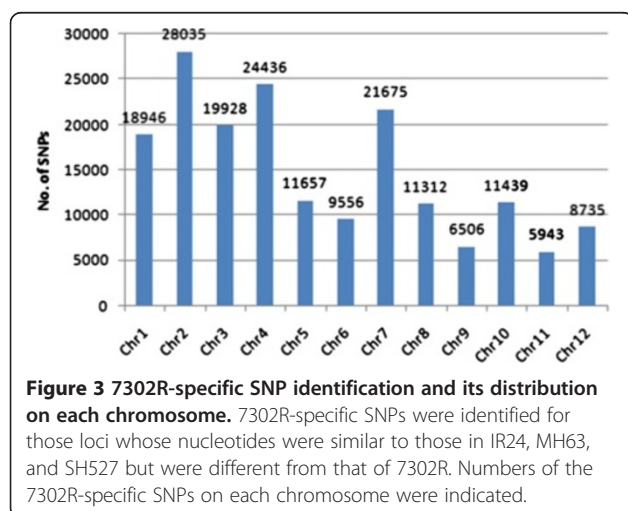
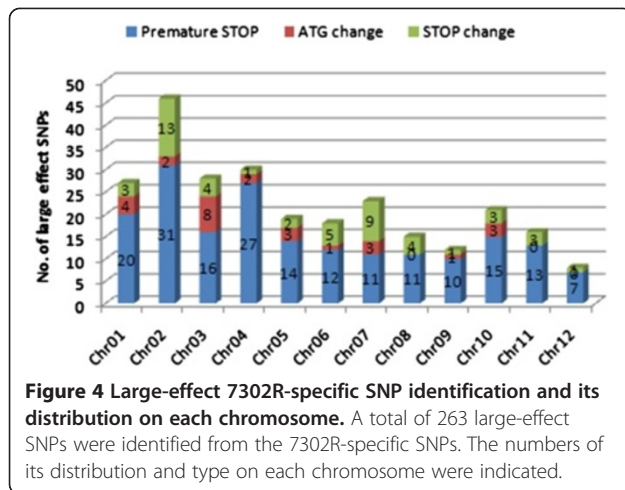
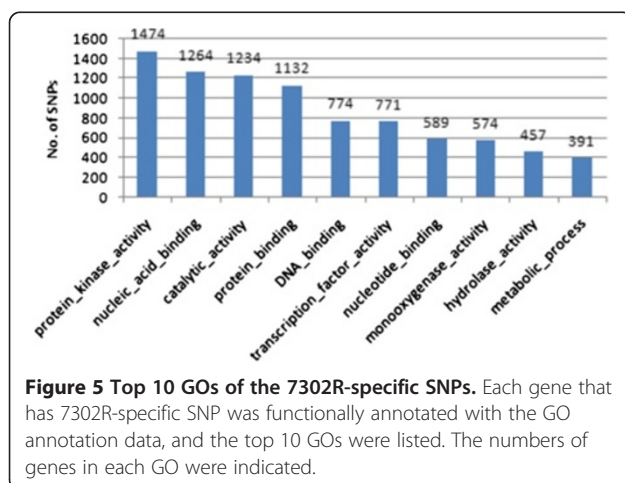


Table 4 Annotations of gene located 7302R-specific SNPs

Chr.	Syn_CDS	Non-syn_CDS	mRNA
Chr01	461	898	3,445
Chr02	733	1,169	4,710
Chr03	536	928	3,978
Chr04	562	1,032	3,131
Chr05	298	558	1,893
Chr06	385	579	2,247
Chr07	702	1,055	3,482
Chr08	265	409	1,637
Chr09	248	413	1,382
Chr10	278	585	1,632
Chr11	266	507	1,406
Chr12	212	384	1,296



resistance genes, such as *Pi9* (Qu et al. 2006), *Xa26* (Sun et al. 2004), *rTGA2.1* (Heather et al. 2005), and *Pi-ta* (Bryan et al. 2000), have at least one SNP in the predicted mRNA region. Our finding is also consistent with the previous result that genes mediating disease resistance in plants are particularly diverse due to pathogen pressure (Lai et al. 2010). However, genes related to rice developmental processes, yield, and quality, such as *FLO4* (Kang et al. 2005), *DEP1* (Huang et al. 2009), *GS3* (Fan et al. 2006; Mao et al. 2010), *EUI1* (Zhu et al. 2006), *Gn1a* (Ashikari et al. 2005) and *qSW5* (Shomura et al. 2008), had rare or no variations in the coding regions although they were found to have several SNPs in the DNA sequence. Especially, no variations were found in the *IPA1* (Jiao et al. 2010) locus and might suggest the IPA phenotype of 7302R was not associated with the recently isolated rice architecture gene. Interestingly, a number of SNPs were found both in the DNA sequence (~15) and in the coding regions (~11) of *Rf1a* (Wang et al. 2006), a possible allelic gene



for *Rf4* (Ahmadikhah et al. Ahmadikhah and Karlov 2006), which is the major restoring gene of the WA-CMS line. These variations may be due to the differences between the restoring abilities of 7302R and other sequenced restorer lines. This observation is consistent with and might well explain the restoring range and ability difference observed in the breeding practice between them. In the same way, the variations in the *SSIIIa* (Fujita et al. 2007) DNA sequence (~19) and in the coding regions (~7) might have been responsible for the grain quality difference.

In the present study, we report variations over the whole genome of a rice cultivar with an IPA phenotype. However, further analysis of more related lines is necessary to better understand the IPA mechanism although useful information have been proposed to account for the IPA phenotype. Several follow-up steps can also be taken to determine candidate genes that may contribute to this phenotype. The large-effect SNPs and known important rice gene-located SNPs should also be strictly

Table 5 SNP Detection of cloned important rice genes

Gene	DNA	mRNA
<i>ALK</i>	1	0
<i>Bph-14</i>	1	0
<i>DWARF10</i>	2	0
<i>DWARF 27</i>	2	0
<i>DEP1</i>	7	0
<i>EUI1</i>	4	0
<i>FLO4</i>	19	0
<i>GIF1</i>	1	0
<i>Gn1a</i>	1	0
<i>GS3</i>	6	0
<i>GW2</i>	2	0
<i>HTD2</i>	1	0
<i>OsGT1</i>	3	0
<i>Pi9</i>	1	1
<i>Pib</i>	25	14
<i>P-id2</i>	7	0
<i>Pi-ta</i>	8	2
<i>qSW5</i>	3	0
<i>Rf1a</i>	15	11
<i>rTGA2.1</i>	1	1
<i>sd1</i>	2	0
<i>OsSSIIIa</i>	19	7
<i>Xa1</i>	5	4
<i>Xa21</i>	11	7
<i>Xa26</i>	3	1
<i>Xa5</i>	9	0

selected and considered for functional verifications. Furthermore, we have developed several genetic populations with 7302R for the dissection and mapping of IPA components. The QTL mapping result and variation distributions are expected to make candidate fixing and further functional confirmation easy. The present study therefore lays the groundwork for long-term efforts to uncover genes and alleles important in rice plant architecture construction, also offers useful data resources for future genetic and genomic studies in rice.

Additional files

Additional file 1: Data files are generally TXT. For Windows user, "Editplus" or "UltraEdit" is recommended as the browser program.

SNPs information of 7302R; Format description(left to right). 1. Chromosome name. 2. Position of locus. 3. Nucleotide at corresponding locus of reference sequence. 4. Genotype of sequencing sample. 5. Quality value. 6. nucleotide with the highest probability(first nucleotide). 7. Quality value of the nucleotide with the highest probability. 8. Number of supported reads that can only be aligned to this locus. 9. Number of all supported reads that can be aligned to this locus. 10. Nucleotide with higher probability. 11. Quality value of nucleotide with higher probability. 12. Number of supported reads that can only be aligned to this locus. 13. Number of all supported reads that can be aligned to this locus. 14. Total number of reads that can be aligned to this locus. 15. Order and quality value. 16. Estimated copy number for this locus. 17. Presence of this locus in the dbSNP database. 1 refers to presence and 0 refers to inexistence. 18. The distance between this locus and another closest SNP.

Additional file 2: InDels information of 7302R; Format description (left to right). 1. Chromosome. 2. Position. 3. Indel type and number. 4. Bases. 5. Strand (+, positive; -, negative; *, both positive and negative). 6. Homozygosis or heterozygosis. 7. Average quality. 8. The number of supported read pair. 9. The number of all crossed read pair.

Additional file 3: SVs information of 7302R; Format description(left to right). 1. Chromosome name. 2. Type of structure variation. 3. Minimal value of start position in cluster. 4. Maximal value of end position in cluster. 5. Estimated start position of this structure variation. 6. Estimated end position of this structure variation. 7. Length of SV. 8. Breakpoint of SV (only for insertion). 9. Unusual matching mode (F refers to align with forward sequence, R refers to align with reverse sequence). 10. Number of paired-end read which support this structure variation.

Abbreviations

IPA: Ideal plant architecture; SNP: Single-nucleotide polymorphisms; InDel: Insertion/deletion polymorphisms; SV: Structural variations; Non-syn CDS: Non-synonymous coding sequences.

Competing interests

The authors declare no potential competing interests.

Authors' contributions

PL Conceived and designed the experiments; SL, KX, WL, TZ, YR, SW, QD, AZ and JZ Performed the experiments and analysis; HL and LW Collected samples and performed the phenotyping; FG, BH, PA and XC Contributed to analysis tools, SL and PL Analyzed the data, SL Wrote the paper, All authors read and approved the final manuscript.

Accession codes

Raw sequence data obtained in our study have been deposited in the NCBI Short Read Archive with accession number SRP006823.

Acknowledgments

This work was supported by the National Basic Research (973) Program of China (2011CB100101), the National Natural Science Foundation of China

(30800084), and the Special Programs for GM Crops Initiative of China (2011ZX08001-001).

Author details

¹State Key Laboratory of Hybrid Rice, Sichuan Agricultural University, Chengdu 611130, China. ²Rice Research Institute, Sichuan Agricultural University, 211 Huimin Road, 611130 Wenjiang, Sichuan, China. ³Key Laboratory of Crop Genetic Resources and Improvement, Ministry of Education, Sichuan Agricultural University, 625014 Ya'an, Sichuan, China.

Received: 14 May 2012 Accepted: 30 June 2012

Published: 24 July 2012

References

- Ahmadihah A, Karlov GI (2006) Molecular mapping of the fertility-restoration gene Rf4 for WA-cytoplasmic male sterility in rice. *Plant Breeding* 125:363–367
- Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. *Science* 309:741–745
- Bryan G, Wu K, Farrall L, Jia Y, Hershey H, McAdams S, Faulk K, Donaldson G, Tarchini R, Valent B (2000) A Single Amino Acid Difference Distinguishes Resistant and Susceptible Alleles of the Rice Blast Resistance Gene Pi-ta. *Plant Cell* 12:2033–2046
- Duan S, Mao J, Zhu Y (2002) Genetic Variation of Main Restorer Lines of Hybrid Rice in China Was Revealed by Microsatellite Markers. *Acta Genetica Sinica* 29:250–254
- Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 112:1164–1171
- Fujita N, Yoshida M, Kondo T, Saito K, Utsumi Y, Tokunaga T, Nishi A, Satoh H, Park J, Jane J, Miyao A, Hirochika H, Nakamura Y (2007) Characterization of SSIIa-Deficient Mutants of Rice: The Function of SSIIa and Pleiotropic Effects by SSIIa Deficiency in the Rice Endosperm. *Plant Physiol* 144:2009–2023
- Heather A, Patrick E, Chern M, Pamela C (2005) Alteration of TGA factor activity in rice results in enhanced tolerance to *Xanthomonas oryzae* pv. *Oryzae* The *Plant Journal* 43:335–347
- Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. *Nat Genet* 41:494–497
- Huang X, Wei X, Sang T, Zhao Q, Feng Q et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* 42:541–544
- Kang H, Park S, Matsuoka M, An G (2005) White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C4-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J* 42:901–911
- Lai J, Li R, Xu X, Jin W, Xu M et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* 42:1027–1030
- Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714
- Li R, Li Y, Fang X, Yang H, Wang J et al (2009) SNP detection for massively parallel whole-genome resequencing. *Genome Res* 19:1124–1132
- Li S, Wang S, Deng Q, Zheng A, Zhu J et al (2012) Identification of Genome-Wide Variations among Three Elite Restorer Lines for Hybrid-Rice. *PLoS One* 7(2): e30952. doi:10.1371/journal.pone.0030952
- Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc Natl Acad Sci* 107:19579–19584
- Qu S, Liu G, Zhou B, Bellizzi M, Zeng L et al (2006) The Broad-Spectrum Blast Resistance Gene Pi9 Encodes an NBS-LRR Protein and is a Member of a Multigene Family in Rice. *Genetics* 105:044891, genetics.105.044891
- Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* 40:1023–1028

- Song W, Wang G, Chen L, Kim H, Pi L et al (1995) A Receptor Kinase-Like Protein Encoded by the Rice Disease Resistance Gene, Xa21. *Science* 270:1804–1806
- Sun X, Cao Y, Yang Z, Xu C, Li X et al (2004) Xa26, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. *Plant J* 37:517–527
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol* 24:1596–1599
- Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L et al (1999) The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J* 19:55–64
- Wang Z, Zou Y, Li X, Zhang Q, Chen L et al (2006) Cytoplasmic Male Sterility of Rice with Boro II Cytoplasm Is Caused by a Cytotoxic Peptide and Is Restored by Two Related PPR Motif Genes via Distinct Modes of mRNA Silencing. *The Plant Cell Online* 18:676–687
- Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z-X et al (1998) Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc Natl Acad Sci* 95:1663–1668
- Zhu Y, Nomura T, Xu Y, Zhang Y, Peng Y et al (2006) ELONGATED UPPERMOST INTERNODE Encodes a Cytochrome P450 Monooxygenase That Epoxidizes Gibberellins in a Novel Deactivation Reaction in Rice. *Plant Cell* 18:442–456

doi:10.1186/1939-8433-5-18

Cite this article as: Li et al.: Re-sequencing and genetic variation identification of a rice line with ideal plant architecture. *Rice* 2012 **5**:18.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
