

Methods

MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with *de novo* assembly of gap regions identifies the rice blast resistance gene *Pii*

Hiroki Takagi^{1,2}, Aiko Uemura¹, Hiroki Yaegashi¹, Muluneh Tamiru¹, Akira Abe¹, Chikako Mitsuoka¹, Hiroe Utsushi¹, Satoshi Natsume^{1,2}, Hiroyuki Kanzaki¹, Hideo Matsumura³, Hiromasa Saitoh¹, Kentaro Yoshida^{1,4}, Liliana M. Cano⁴, Sophien Kamoun⁴ and Ryohei Terauchi¹

¹Iwate Biotechnology Research Center, Kitakami, Iwate, 024-0003, Japan; ²United Graduate School of Iwate University, Morioka, Iwate, 020-8550, Japan; ³Shinshu University, Ueda, Nagano, 386-8567, Japan; ⁴The Sainsbury Laboratory, Norwich Research Park, Norwich, UK

Author for correspondence:
Ryohei Terauchi
Tel: +81 197 68 2911
Email: terauchi@ibrc.or.jp

Received: 8 March 2013
Accepted: 13 May 2013

New Phytologist (2013)
doi: 10.1111/nph.12369

Key words: *de novo* assembly, MutMap, next-generation sequencing, R-gene, rice, whole-genome sequencing.

Introduction

Developments in DNA sequencing technology have enabled whole-genome sequencing (WGS) of plant species to become routine. One of the applications of WGS is the identification of causative mutations responsible for mutant phenotypes of interest. For this purpose, a series of methods have been reported including SHOREmap (Schneeberger *et al.*, 2009), NGM (Austin *et al.*, 2011) and MutMap (Abe *et al.*, 2012).

In MutMap (Abe *et al.*, 2012), a recessive mutant with a phenotype of interest is crossed to the parental line used for the mutagenesis and DNA from multiple individuals of mutant F2 progeny are bulk-sequenced. The resulting short reads are aligned to the 'reference sequence' constructed for the parental line, and the result of this alignment is used to infer the genomic location of the causal mutation responsible for the phenotype. The prerequisite for MutMap is that the genomic fragment spanning the causative mutation is present in the parental 'reference sequence'. However, this is not guaranteed, as the parental line 'reference sequence' is constructed based on the 'reference genome' of a representative cultivar/line of the species, for example the cv Nipponbare in rice

Summary

- Next-generation sequencing allows the identification of mutations responsible for mutant phenotypes by whole-genome resequencing and alignment to a reference genome. However, when the resequenced cultivar/line displays significant structural variation from the reference genome, mutations in the genome regions missing from the reference (gaps) cannot be identified by simple alignment.
- Here we report on a method called 'MutMap-Gap', which involves delineating a candidate region harboring a mutation of interest using the recently reported MutMap method, followed by *de novo* assembly, alignment, and identification of the mutation within genome gaps.
- We applied MutMap-Gap to isolate the blast resistant gene *Pii* from the rice cv Hitomebore using mutant lines that have lost *Pii* function.
- MutMap-Gap should prove useful for cloning genes that exhibit significant structural variations such as disease resistance genes of the nucleotide-binding site-leucine rich repeat (NBS-LRR) class.

(*Oryza sativa*) and the Col ecotype in *Arabidopsis thaliana*. This is done by resequencing the parental line genome and replacing nucleotides of the 'reference genome' with those of the parental line at all the single nucleotide polymorphism (SNP) positions detected between the two lines. Consequently, MutMap cannot identify mutations located within parental line-specific genomic regions (gaps) that are missing when compared with the 'reference genome'. To identify mutations in such gap regions, we have introduced MutMap-Gap, a combination of MutMap and targeted *de novo* assembly of genomic gap regions.

Materials and Methods

Constructing the Hitomebore 'reference sequence' based on the Nipponbare 'reference genome'

To generate the northern Japanese rice (*Oryza sativa*) cv Hitomebore parental line 'reference sequence', we aligned 12.25 Gb of Hitomebore wild-type (WT) parental line sequence reads obtained by Illumina (San Diego, CA, USA) sequencing to the Nipponbare reference genome (build five genome sequence;

<http://rapdblegacy.dna.affrc.go.jp/download/index.html>) by BWA (Li & Durbin, 2009) as described in Abe *et al.* (2012). The Hitomebore 'reference sequence' was constructed by replacing Nipponbare nucleotides with those of Hitomebore at the 124 968 SNP positions that were identified between the two cultivars.

WGS of bulked DNA from F2 progeny and MutMap analysis

DNA samples were prepared by bulking DNA extracted from leaves of mutant F2 individuals as described previously (Abe *et al.*, 2012). The library for the Illumina GAIIX sequencer was prepared from 5 µg DNA samples. Short reads were aligned to the Hitomebore 'reference sequence' using BWA software (Li & Durbin, 2009). Alignment files were converted to SAM/BAM files using SAMtools (Li *et al.*, 2009), and the aligned short reads were filtered by Coval (S. Kosugi *et al.*, unpublished; <http://sourceforge.net/projects/coval105/?source=directory>) to improve SNP calling accuracy. The SNP index was calculated as described previously (Abe *et al.*, 2012). Sliding window analysis was applied with a 4 Mb window size and a 10 kb increment. The details are given in the Supporting Information, Fig. S1.

De novo assembly

A mate-pair library of the Hitomebore WT genome for contig scaffolding was prepared using the Mate Pair Library Prep Kit (Illumina) and *de novo* assembly was performed with CLC (<http://www.clcbio.com>) software. The generated contigs were used to generate scaffolds using mate-pair sequence reads of Hitomebore WT DNA with SSPACE (Boetzer *et al.*, 2011).

Mutant screen

Using a total of 3033 Hitomebore mutant lines generated by ethylmethanesulfonate (EMS) mutagenesis, we carried out a spray inoculation test with the rice blast fungus *Magnaporthe oryzae* isolate TH68-126 (race 033.1) carrying *AVR-Pii* (Yoshida *et al.*, 2009) and incompatible with Hitomebore (with *Pii*).

Results

The principle of MutMap-Gap

MutMap-Gap is a WGS-based method developed for the identification of the causative nucleotide change for a given mutant phenotype in a genomic region that is missing from the reference genome. MutMap-Gap combines the previously reported MutMap method (Abe *et al.*, 2012) with targeted gap filling by *de novo* assembly. This is followed by identification of the causative mutation responsible for the phenotype in the assembled gap region as described in Figs 1 and 2 using rice as an example.

First, mutant lines are developed by mutagenesis of the parental cv 'P' with a chemical mutagen such as EMS. Given that cv P

is different from the cv Ref for which an accurate genome sequence is available, we first need to generate a 'reference sequence' of the P genome. For this purpose, we sequence the cv P and align the resulting reads to the Ref genome, which is the publicly available reference sequence. Following this step, nucleotides of the Ref genome are replaced with those from the P genome at all SNP positions identified between the two cultivars (Fig. 1a). Although the majority of sequence reads obtained for P are expected to be aligned to the Ref genome, the short reads derived from a P-specific genomic region cannot be aligned, and thus are collected as unmapped reads.

Assume that we are interested in a mutant line 'M' generated in the cv P background, and that the causal mutation for the phenotype under consideration resides in a P-specific genomic region (Fig. 1b). As the genomic region containing the causal mutation is not represented in the P 'reference sequence', simply aligning the 'M' sequence reads to the P 'reference sequence' cannot identify it. However, we can recover such a P-specific genomic region by *de novo* assembly. To this end, we first delineate the approximate position of the causative mutation by MutMap. Briefly, 'M' is backcrossed to P to generate F2 progeny, and DNA from *c.* 20 mutant F2 individuals are bulked and subjected to whole-genome resequencing. The resulting short reads are aligned to the cv P 'reference sequence' (Fig. 1b) to identify SNPs. For each SNP position, the SNP index, the ratio of mutant-type short reads to the total short reads covering a particular SNP position, is calculated and SNP index plots are generated to show the relationship between the SNP index and chromosomal position graphically. The SNPs in the candidate region close to the causal mutation are expected to have a higher SNP index (SNP index \sim 1), whereas those in unlinked regions show a SNP index of *c.* 0.5. Finding a peak in the SNP index allows the identification of the approximate genomic interval harboring the causal mutation. However, as our candidate interval is located within the gap region, it is not possible to identify the causal mutation by MutMap alone (Fig. 1c).

To target a mutation located in the gap region, we apply *de novo* assembly to reconstruct the P 'reference sequence' within the target interval delineated by the initial MutMap step (Fig. 2). For *de novo* assembly, we utilize two types of sequence reads derived from the cv P: sequence reads aligned to the target interval region as delineated by MutMap; and sequence reads that could not be aligned (unmapped) to the Ref genome (Fig. 2a). Using these two types of reads, we perform *de novo* assembly using CLC (<http://www.clcbio.com>), and SSPACE (Boetzer *et al.*, 2011) software to recover scaffolds presumably located in the target interval (Fig. 2a).

Finally, the short reads derived from the bulked DNA of F2 plants showing a mutant phenotype are aligned to the combined sequences of the scaffolds produced by the *de novo* assembly and P 'reference sequence' (Fig. 2b). This procedure allows the identification of SNPs residing within the newly generated scaffolds, for which the SNP index is calculated. The SNPs showing an SNP index of 1 are the likely candidates for the causative mutation for the mutant phenotype. The novelty of MutMap-Gap resides in the targeted *de novo* assembly of the genomic region corresponding to the gap in the Ref genome sequence, as

delineated by MutMap, and identification of the causal mutation in the assembled sequence by SNP index analysis.

MutMap-Gap identifies rice *Pii*, a blast resistance gene

As a proof of principle, we applied MutMap-Gap in the identification of the rice blast (*M. oryzae*) resistance (R-) gene *Pii*. Rice blast is a destructive and widespread disease caused by an ascomycete pathogen

M. oryzae and accounts for significant yield losses worldwide. For efficient marker-assisted breeding of rice blast resistance, the identification of R genes is important. *Pii* confers resistance to rice against the blast pathogen harboring the corresponding *AVR-Pii* gene. The complete genome sequence of the rice ssp. *japonica* cv Nipponbare was published in 2005 (International Rice Genome Sequencing Project, 2005). However, Nipponbare is known to lack *Pii*, indicating that this cultivar cannot be directly used for cloning of the *Pii* gene.

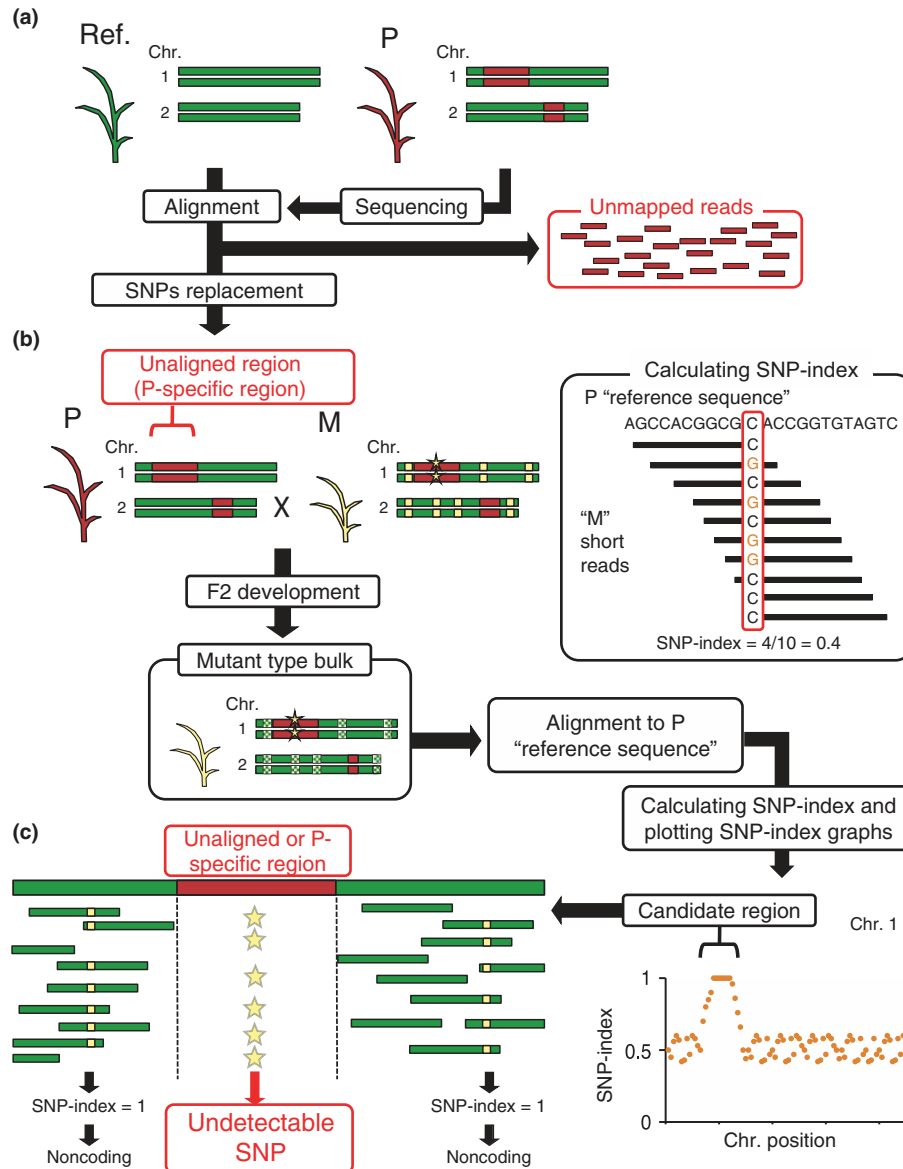


Fig. 1 MutMap applied to a rice (*Oryza sativa*) cultivar differentiated from the one used for generating the reference genome cannot identify mutations in regions missing from the reference. (a) Generation of a 'reference sequence' for the cv P used for mutagenesis. P was sequenced by an Illumina sequencer and the resulting reads were aligned to the publicly available reference genome of the rice cv Nipponbare (indicated in green), which is designated Ref. Although the majority of the P short reads are aligned to the Ref genome, the reads derived from P-specific genomic regions (represented in red) cannot be aligned, and are thus left as unmapped reads. Following the alignment and identification of the single nucleotide polymorphisms (SNPs) between cvs P and Ref, the nucleotides of Ref are replaced with those of P at all SNP positions to construct a P 'reference sequence' (b) Cv P is mutagenized by ethylmethanesulfonate (EMS) to generate a mutant 'M' showing a phenotype of interest, for example, semi-dwarfism. The mutations resulting from EMS treatment are indicated by yellow squares, and a yellow star indicates the causative mutation responsible for the M phenotype. For MutMap analysis, the plant M is backcrossed to P to generate F2 progeny that segregate for the wild-type and mutant phenotypes. DNA from 20 or more mutant F2 individuals are bulked in equal proportion and subjected to whole-genome sequencing (WGS), followed by alignment to the 'reference sequence' of P. (c) An SNP index is calculated for all SNPs, and an SNP index plot is generated to facilitate delineation of the target genomic region harboring the causal mutation. If the causal mutation is located within a P-specific genomic region, MutMap analysis alone cannot identify it. Chr., chromosome.

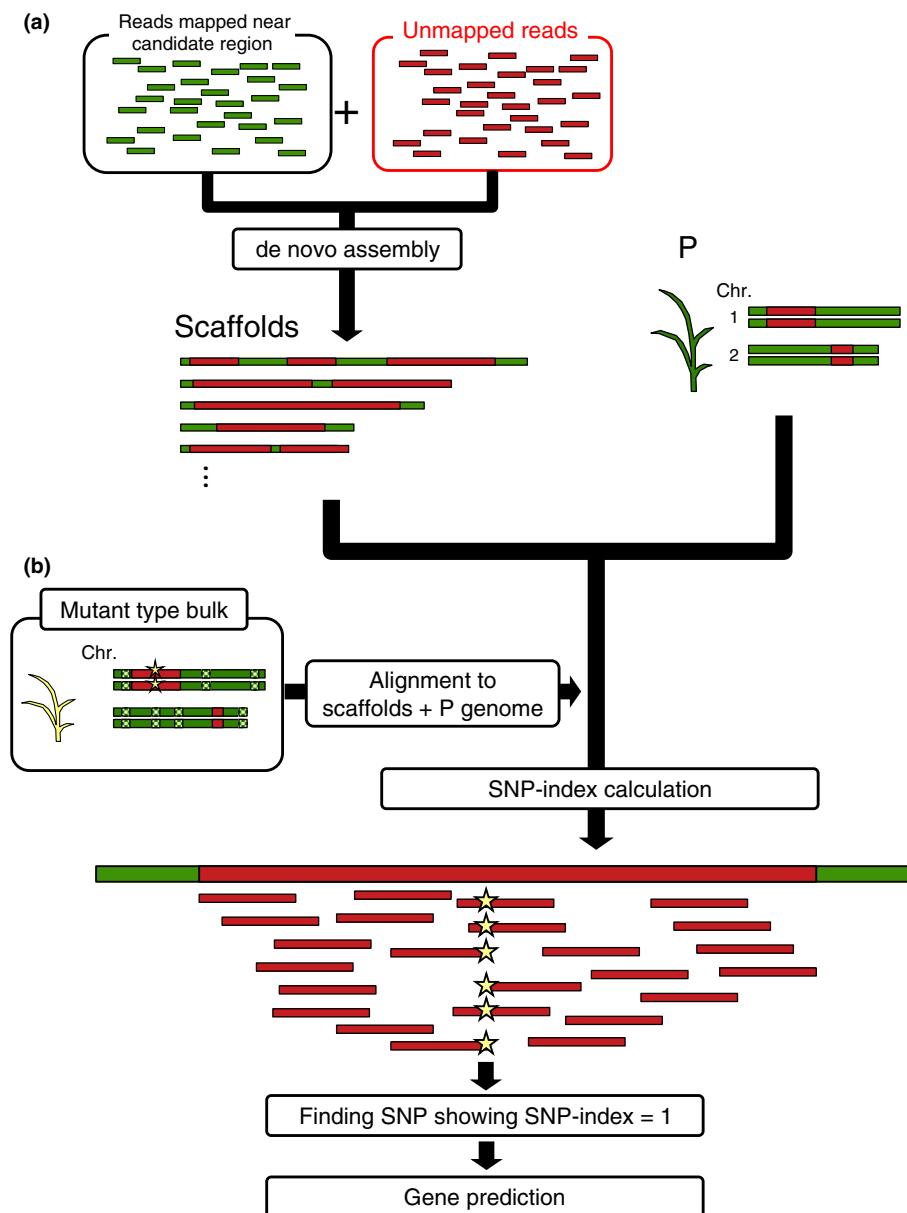


Fig. 2 MutMap-Gap fills the gap within a target genome region delineated by MutMap using *de novo* assembly. (a) Using a combination of the unmapped reads collected in the previous step (Fig. 1a) and the short reads aligned to the target region (Fig. 1c), *de novo* assembly is performed to reconstruct the sequence of the target interval. The scaffolds generated by the *de novo* assembly are combined with the P 'reference sequence' and serve as a 'P + scaffolds' reference sequence for alignment purposes in the following step. (b) Short reads derived from the bulk DNA of mutant F2 progeny are aligned to the 'P + scaffolds' reference. The scaffolds with single nucleotide polymorphisms (SNPs) showing an SNP index of 1 are likely to contain the causal nucleotide change responsible for the mutant phenotype. Chr., chromosome.

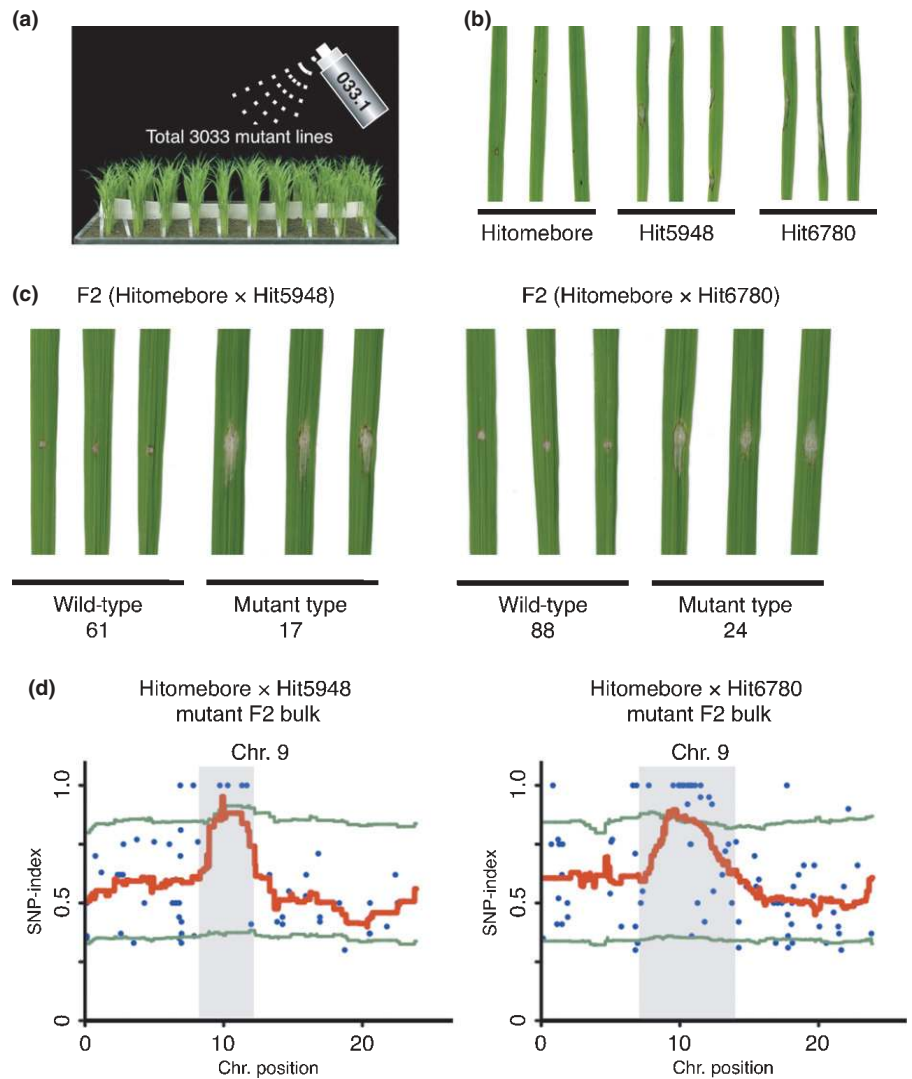
To isolate *Pii*, we used the 'Hitomebore' cultivar containing the *Pii* gene. Whole-genome resequencing of Hitomebore allowed us to construct a Hitomebore 'reference sequence' based on the Nipponbare genome sequence by replacing Nipponbare nucleotides with those of Hitomebore at all the SNP positions (124 968 positions) identified between these two cultivars (see the Materials and Methods section). Of the 389 Mb Nipponbare genome, *c.* 358 Mb (92%) was covered by the 10.71 Gb Hitomebore sequence reads generated, corresponding on average to $\times 27.5$ coverage (Table S1). We found short reads amounting to a total of 251 Mb that were unmapped to the Nipponbare reference genome sequence. These short reads are probably derived from Hitomebore-specific genomic regions that are not represented in the Nipponbare genome.

For identification of Hitomebore mutants that had lost *Pii* function, we carried out a spray inoculation test using an incompatible blast race 033.1 expressing *AVR-Pii* (Yoshida *et al.*, 2009)

on a total of 3033 EMS-mutagenized Hitomebore lines (Fig. 3a). We identified two independent *Pii*-deficient candidate mutants, Hit5948 and Hit6780, which showed susceptible phenotypes following inoculation with race 033.1 (Fig. 3b). For mapping the causal mutations by MutMap, we independently crossed the two mutants to Hitomebore WT and generated two sets of F2 progeny. The Hit5948 \times WT F2 progeny segregated 61 WT to 17 mutant phenotypes, whereas the Hit6780 \times WT F2 progeny segregated 88 WT to 24 mutant phenotypes. In both cases, the segregation conformed to a 3:1 ratio ($\chi^2 = 0.43$, ns for Hit5948 \times WT F2; $\chi^2 = 0.76$, ns for Hit6780 \times WT F2), indicating that the phenotypes of the two mutants were caused by single recessive mutations (Fig. 3c).

For MutMap analysis, we bulked the DNA of the mutant F2 progeny (17 individuals for Hit5984 and 24 individuals for Hit6780) and undertook WGS. We carried out 75 bp paired-end sequencing and obtained 2.45 and 2.87 Gbp sequence reads for

Fig. 3 MutMap reveals genomic regions harboring candidate mutations of Hit5984 and Hit6780, two rice (*Oryza sativa*) mutants that have lost *Pii* resistance. (a) A total of 3033 rice cv Hitomebore ethylmethanesulfonate (EMS) mutant lines were screened for their resistance response by inoculation with *Magnaporthe oryzae* isolate 033.1, which contains *AVR-Pii* avirulence genes. (b) The cv Hitomebore contains the *Pii* R gene, and thus shows resistance to *M. oryzae* isolate 033.1, whereas the two mutants, Hit5948 and Hit6780, are susceptible, suggesting they have lost *Pii* resistance. (c) The segregation of two progeny sets generated by crossing the two *Pii* mutants, Hit5948 and Hit6780, back to the Hitomebore wild-type for resistant (wild-type) and susceptible (mutant type) phenotypes confirmed a 3 : 1 ratio, suggesting the phenotype is governed by a single recessive gene. The resistance response is assessed by punch or spot inoculation of *M. oryzae* isolate 033.1. (d) Single nucleotide polymorphism (SNP) index plots of Hit5948 and Hit6780 for chromosome 9 generated by a combined analysis of MutMap and MutMap-Gap. Blue dots indicate SNP index values at a given SNP position. Red lines represent the sliding window average SNP index values of the 4 Mb interval with 10 kb increments. Green lines show the 95% confidence limit of the SNP index value under the null hypothesis of SNP index = 0.5. Chromosomal regions shaded gray indicate the genomic regions presumably harboring the causal mutations. Chr., chromosome.



DNA samples from Hit5948 and Hit6780 mutant F2 bulks, respectively (Table S1). The sequence reads were aligned to the Hitomebore ‘reference sequence’ and SNPs were identified. For all SNP positions, the SNP index was calculated and graphs relating SNP position and SNP index were generated for further analysis (Figs 3d, S2).

MutMap applied to Hit5948 revealed an SNP index peak in the genomic interval from 7.88 to 11.98 Mb on chromosome 9. Of the four SNPs with an SNP index of 1 identified in the candidate region, SNP-10290916 was localized in the second exon of the gene *Os09t0327600-01* predicted in the Nipponbare genome (Table S2). This SNP represented a nonsense mutation, causing an amino acid change from Trp (TGG) to a stop codon (TGA) at the 226th-amino-acid residue. *Os09t0327600-01* encodes a protein with a nucleotide binding site and leucine rich repeat (NBS-LRR) domain, both of which are conserved in plant resistance genes (Zhou *et al.*, 2004; Jones & Dangl, 2006). Scrutiny of *Os09t0327600-01* from Nipponbare showed that this gene encodes a truncated R-protein that is likely to be nonfunctional. We therefore hypothesize that the *Os09t0327600-01* homolog in

Hitomebore functions as *Pii* and that Nipponbare lacks a functional *Pii*.

A similar analysis applied to Hit6780 using MutMap identified a candidate genomic region that is probably harboring the causative mutation in the interval from 7.18 to 13.05 Mb on chromosome 9, an identical region to that in which the causal mutation of Hit5948 was mapped. Although 10 SNPs with an SNP index of 1 were identified in the region, none represented nonsynonymous mutations and no SNP was detected in *Os09t0327600-01*, the candidate gene for Hit5948 (Table S3). We hypothesized that the causative mutation of Hit6780 is located in the Hitomebore-specific genomic region (Fig. 4a).

To identify the causative mutation of Hit6780, we applied MutMap-Gap analysis. We retrieved all of the Hitomebore sequence reads (4849 550 reads) mapped to the 7.18–13.05 Mb region on chromosome 9 and combined them with the Hitomebore sequence reads (3358 005 reads) unmapped to the Nipponbare genome. These combined short reads were used for *de novo* assembly with CLC software (<http://www.clcbio.com>) and generated a total of 2239 contigs that were over 1 kb in size. For

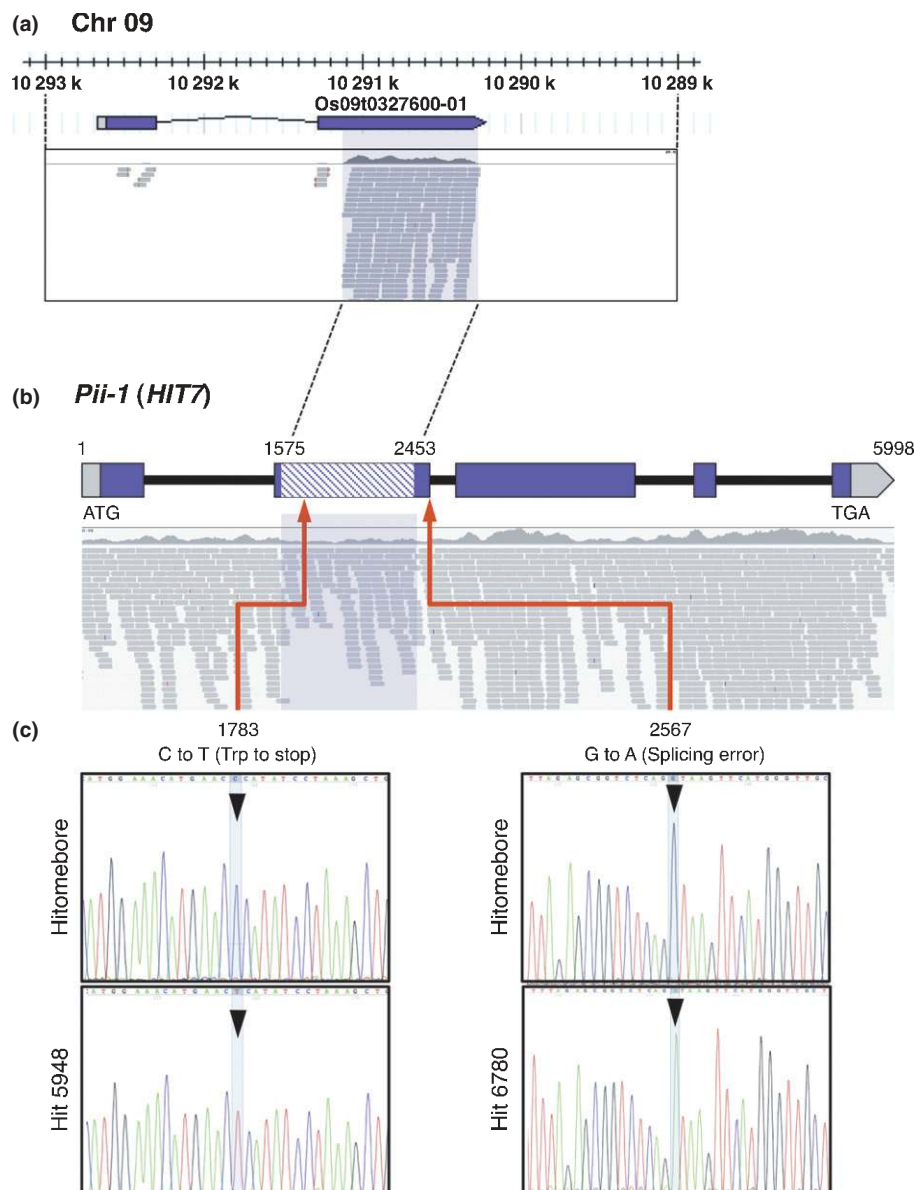


Fig. 4 MutMap-Gap identifies the causal mutation of Hit6780. (a) A snapshot from the Integrative Genomics Viewer (IGV) genome viewer showing the 10 289 000–10 293 000 region of chromosome 9 in the Nipponbare reference genome with the predicted *Os09t0327600-01* gene aligned with short reads derived from the Hitomebore wild-type (WT). The short reads cover only the predicted exon regions of *Os09t0327600-01*. (b) A snapshot from the IGV showing the predicted *HIT7* gene region spanning scaffold 7 generated by *de novo* assembly. The predicted gene structure is shown by boxes (exons and untranslated regions (UTRs)) and lines (introns). The alignment was made with short reads generated from the Hitomebore WT. The hatched area in the coding region of *HIT7* (second exon) shows > 95% similarity with the *Os09t0327600-01* region in (a). Red arrows indicate the positions of the candidate mutations in Hit5948 (left; position 1783) and Hit6780 (right; position 2567). The genomic position of Hit5984 mutation is conserved between Hitomebore and Nipponbare, whereas the position of Hit6780 is not conserved. (c) Confirmation of the candidate single nucleotide polymorphisms (SNPs) by Sanger sequencing. DNA sequencing peak chromatograms of genomic DNA in the regions around the mutations for Hit5948 (left) and Hit6780 (right) are compared with the Hitomebore WT. Arrows indicate the mutated nucleotides.

scaffolding of the contigs with SSPACE (Boetzer *et al.*, 2011) software, *c.* 46 989 929 mate-pair short reads were used, generating *c.* 852 scaffolds with a minimum size of 1 kb (Fig. S3). The scaffolds were then combined with the Hitomebore ‘reference sequence’ and used as a reference for aligning the short reads derived from bulked DNA of Hit6780 F2 plants, as would be done in MutMap analysis. Of the 852 scaffolds generated, only two harbored SNPs with an SNP index of 1. Gene prediction by GENSCAN (<http://genes.mit.edu/GENSCAN.html>) using the two scaffolds revealed that one of the SNPs corresponded to an intergenic region, and the other, in scaffold no. 7 (length = 63 355 bp), was localized within the splicing junction of a gene we tentatively named *HIT7* (Fig. S4). This mutation is predicted to cause mis-spliced mRNA and to introduce a premature stop codon (Fig. S4). The *HIT7* gene contains an NBS-LRR domain, suggesting that this SNP is the likely causal mutation of Hit6780.

We compared the DNA sequence of *HIT7* with that of *Os09t0327600-01* and found a high similarity (nucleotide identity 97.3%) in the region where the candidate SNP of Hit5948 was detected (Fig. 4). Accordingly, we concluded that the causative mutations of Hit5948 and Hit6780 are located within the same gene. Conventional MutMap analysis could detect the Hit5984 mutation, which was localized in the region where sequence similarity was very high between *HIT7* of Hitomebore and *Os09t0327600-01* of Nipponbare, but not the causal SNP of Hit6780. The Nipponbare *Os09t0327600-01* gene seems to have lost R-gene function by truncation of the region corresponding to the C-terminus of the protein (Fig. 4).

To confirm whether the susceptible phenotype of Hit5948 and Hit6780 was caused by mutations in the same gene, we carried out an allelism test by crossing Hit5948 with Hit6780. As expected, the phenotypes of F1 plants heterozygous for the Hit5948 and Hit6780 mutations (Fig. 5a) were susceptible to the

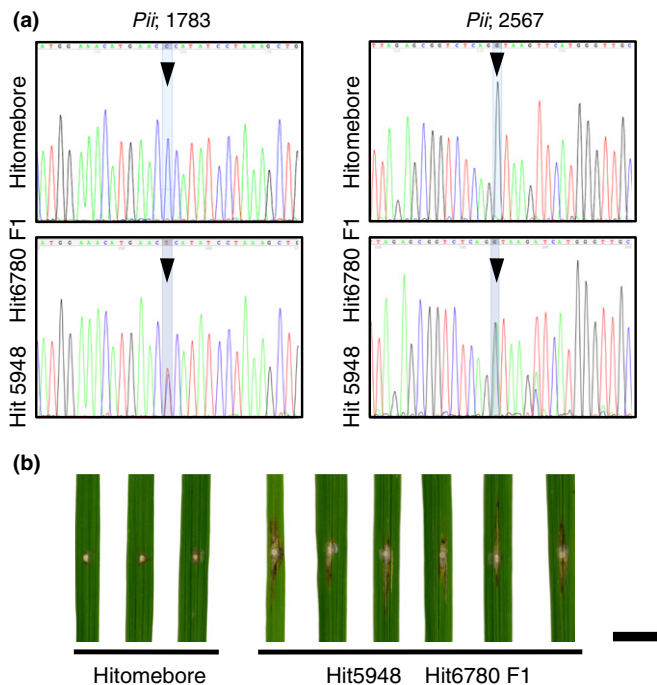


Fig. 5 Allelism test for the Hit5984 and Hit6780 mutations. (a) F1 plants were obtained from a cross between Hit5984 and Hit6780. Genomic DNA was extracted from F1 plants, and two genomic regions spanning the mutations in Hit5984 (position 1783) and Hit6780 (position 2567) were amplified and sequenced. As expected, F1 plants showed heterozygous peaks of T/C and G/A at the 1783th and 2567th positions, respectively. (b) Results of a punch inoculation test for Hitomebore wild-type (WT) and F1 plants with an incompatible *Magnaporthe oryzae* race (033.1). The F1 plants remained susceptible to race 033.1, suggesting that Hit5984 and Hit6780 have defects in the same gene. Bar, 1 cm.

M. oryzae isolate TH68-126 (race 033.1; Fig. 5b), confirming that the phenotypes of Hit5984 and Hit6780 mutants are caused by defects in the same gene, *HIT7*. We further tested the association between the phenotype in the presence or absence of *Pii* and cleaved amplified polymorphic sequence (CAPS) marker polymorphism discriminating *HIT7* and *O509t0327600-01* alleles in a total of 30 rice cultivars. A complete association was observed between the *Pii* phenotypes and the CAPS patterns, supporting the identification of the *Pii* gene as *HIT7* (Fig. S5). We accordingly renamed *HIT7* as *Pii*.

The *Pii* gene is composed of five exons, and the 3078 bp coding sequence spanning the start and stop codons encodes a 1025-amino-acid protein predicted as a putative NBS-LRR type R protein (Fig. S6), which is typical of the majority of disease resistance genes in plants (Belkhadir *et al.*, 2004; Zhou *et al.*, 2004; McHale *et al.*, 2006).

Discussion

Successful identification of a causative nucleotide change responsible for a mutant phenotype by WGS depends on the availability of a good reference genome sequence. If the resequenced mutant belongs to a cultivar/line that is structurally different from the one used for constructing the reference genome, it is possible that

the causative mutation may reside in the genomic region specific to the resequenced cultivar, but absent from reference genome. In the past, such regions have been addressed by generating a bacterial artificial chromosome (BAC) library of the cultivar/line used for the mutant screen, followed by screening of the BAC clones by DNA probes tightly linked to the target region and by sequencing of these clones. This procedure remains laborious and time-consuming. MutMap-Gap, as presented here, circumvents this process, and offers a fast, alternative approach to obtaining sequence information for a genomic region of interest that is missing from the reference genome.

For effective marker-assisted breeding of crop resistance against pathogens, the identification of R genes is important. R genes, which mostly encode the NBS-LRR class of proteins, have a unique mode of evolution and represent a highly divergent group in plants (Clark *et al.*, 2007). The successful identification of the rice blast resistant *Pii* gene from Hitomebore demonstrates the utility of MutMap-Gap for R-gene isolation. How large the genetic distance between the line used for the 'reference genome' and the parental line containing the mutation could be for MutMap-Gap to remain effective awaits future evaluation. We envisage that this method, among others, will broaden the application of WGS in isolating novel plant genes by quick forward genetics approaches.

Acknowledgements

This study was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN), the Ministry of Agriculture, Forestry, and Fisheries of Japan, and the Ministry of Education, Cultures, Sports and Technology, Japan (Grant-in-Aid for Scientific Research on Innovative Areas 23113009 to H.S. and R.T.) and by JSPS KAKENHI (grant no. 24248004 to R.T.). We thank Shigeru Kuroda for general support and Matthew J. Terry for comments on the manuscript.

References

- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M *et al.* 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology* 30: 174–178.
- Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW *et al.* 2011. Next-generation mapping of Arabidopsis genes. *Plant Journal* 67: 715–725.
- Belkhadir Y, Subramaniam R, Dangl JL. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Current Opinion in Plant Biology* 7: 391–399.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579.
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA *et al.* 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317: 338–342.
- International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444: 323–329.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25: 2078–2079.
- McHale L, Tan X, Koehl P, Wichelmore RW. 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biology* 7: 212.
- Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JE, Weigel D, Andersen SU. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* 8: 500–501.
- Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J *et al.* 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21: 1573–1591.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402–405.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Flow chart for MutMap-Gap analysis.

Fig. S2 SNP index plots for the 12 rice chromosomes corresponding to bulk sequencing and MutMap analysis of Hit5948 and Hit6780.

Fig. S3 Summary of *de novo* assembly.

Fig. S4 Confirmation of the splicing error caused by the Hit6780 mutation.

Fig. S5 The association between the *Pii* phenotype and CAPS marker polymorphism.

Fig. S6 The *Pii* gene structure.

Table S1 Summary of Illumina GAIIX sequencing and alignment data

Table S2 SNPs with an SNP index of 1 within the Hit5984 candidate region

Table S3 SNPs with an SNP index of 1 within the Hit6780 candidate region

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.