nature biotechnology

Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture

Burkhard Steuernagel[1,2,7], Sambasivam K Periyannan[3,7], Inmaculada Hernández-Pinzón[1], Kamil Witek[1], Matthew N Rouse[4], Guotai Yu[2], Asyraf Hatta[2,5], Mick Ayliffe[3], Harbans Bariana[6], Jonathan D G Jones[1], Evans S Lagudah[3] & Brande B H Wulff[1,2]

© 2016 Nature America, Inc. All rights reserved.

Wild relatives of domesticated crop species harbor multiple, diverse, disease resistance (R) genes that could be used to engineer sustainable disease control. However, breeding R genes into crop lines often requires long breeding timelines of 5–15 years to break linkage between R genes and deleterious alleles (linkage drag). Further, when R genes are bred one at a time into crop lines, the protection that they confer is often overcome within a few seasons by pathogen evolution[1]. If several cloned R genes were available, it would be possible to pyramid R genes[2] in a crop, which might provide more durable resistance[1]. We describe a three-step method (MutRenSeq)-that combines chemical mutagenesis with exome capture and sequencing for rapid R gene cloning. We applied MutRenSeq to clone stem rust resistance genes *Sr22* and *Sr45* from hexaploid bread wheat. MutRenSeq can be applied to other commercially relevant crops and their relatives, including, for example, pea, bean, barley, oat, rye, rice and maize.

Plant diseases can devastate crop yields and pose a threat to global food security. R genes offer an economical and environmentally responsible solution to control plant disease, and cloning of these genes would enable durable R gene deployment strategies. Many R genes are present in gene families, with members in close physical proximity, such that dissection of the locus by recombination is not practical. Functional dissection based on recombination is further confounded by the extreme sequence diversity and R gene copy number variation often present between different haplotypes[3,4]. In addition, many plant genomes carry large chromosomal regions that impair positional cloning due to suppressed recombination[5]. Therefore, complementary approaches that are not reliant on positional cloning are required. Most R genes encode proteins with nucleotide binding and leucine-rich repeats (NLRs)[2]. A typical plant genome contains hundreds of NLR-encoding genes, many of which reside in complex clusters of linked paralogs[6]. R gene enrichment sequencing (RenSeq) of this specific gene class involves capturing fragments from a genomic or cDNA library using biotinylated RNA oligonucleotides designed to be complementary to the NLR-encoding genes of a reference genome[7,8]. RenSeq was used to identify trait-linked single-nucleotide polymorphisms in NLRs in a potato population that was segregating for disease resistance[7]. However, extensive sequence diversity among parental R gene families prevented the identification of the individual R genes responsible for resistance.

To identify R genes that mediate resistance, we instead used RenSeq to compare the R gene complement of ethyl methane sulfonate (EMS)-derived, loss-of-resistance mutants with wild-type progenitors. This modified version of RenSeq, dubbed "MutRenSeq," enabled the rapid identification of genes responsible for resistance without any positional fine mapping (**Fig. 1**). Obtaining loss-of-function mutants is straightforward since R gene suppressor screens typically recover mutations in the R gene (~90% of mutants) rather than second-site suppressors (**Supplementary Table 1**). In this report, we isolate two wheat stem rust resistance genes, *Sr22* and *Sr45*, that mediate resistance to the stem rust pathogen *Puccinia graminis* f. sp *tritici*. MutRenSeq will enable the rapid isolation of functional R genes from plant species amenable to mutational genomics and is particularly applicable to organisms with large genomes for which whole genome sequencing of multiple individuals is neither straightforward nor cost-effective[9].

First, we tested MutRenSeq using genomic DNA from six EMS-derived bread wheat plants containing mutations (four point mutants and two deletions) in the previously cloned wheat stem rust resistance gene, *Sr33* (ref. 10; **Supplementary Table 2**). We designed a cereal NLR bait library containing 60,000 120-mer RNA probes (**Supplementary Data**) with ≥95% identity to predicted NLR genes present in the Triticeae species barley (*Hordeum vulgare*), hexaploid bread wheat (*Triticum aestivum*), tetraploid pasta wheat (*T. durum*), red wild einkorn (*T. urartu*), domesticated einkorn (*T. monococcum*), and three goatgrass species (*Aegilops tauschii*, *Ae. sharonensis*, and *Ae. speltoides*). We prepared barcoded short insert (500–700 bp) libraries from genomic DNA of the *Sr33* wild type and each of the six mutants and performed NLR capture. Quantitative PCR on the enriched libraries indicated a 500- to 1,000-fold increase in NLRs relative to other genes. We pooled the enriched libraries and sequenced them using Illumina short-read sequencing-by-synthesis technology

[1]The Sainsbury Laboratory, Norwich, UK. [2]John Innes Centre, Norwich, UK. [3]Commonwealth Scientific and Industrial Research Organization (CSIRO) Agriculture, Canberra, ACT, Australia. [4]USDA-ARS Cereal Disease Laboratory and Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota, USA. [5]Department of Agriculture Technology, Universiti Putra Malaysia, Serdang, Malaysia. [6]University of Sydney, Plant Breeding Institute, Cobbitty, NSW, Australia. [7]These authors contributed equally to this work. Correspondence should be addressed to B.W. (brande.wulff@jic.ac.uk) or E.S.L. (evans.lagudah@csiro.au).

Received 19 August 2015; accepted 16 March 2016; published online 25 April 2016; doi:10.1038/nbt.3543

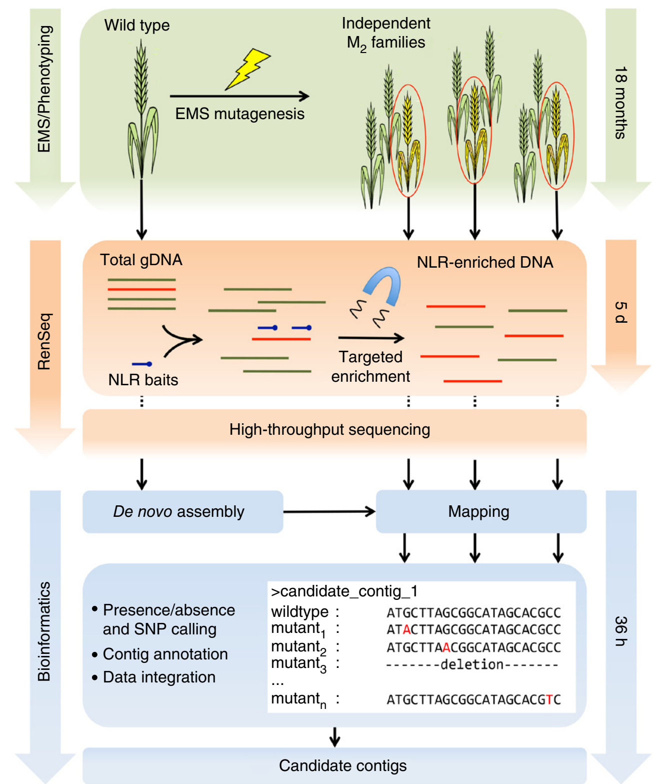

Figure 1 Mutational genomics strategy for resistance gene cloning. Step 1 (green): EMS mutagenesis of resistant plant, creation of independent M_2 families and screening for susceptible mutants (highlighted in yellow). Step 2 (orange): target enrichment using a *Triticeae* NLR–specific bait library and sequencing of the wild-type and susceptible mutants (indicated by arrows). Step 3 (blue): data analysis and candidate calling. A *de novo* assembly of the enriched sequences of the resistant wild type is used as a reference for mapping. Subsequent SNV and presence/absence calls are integrated and scored.

(**Supplementary Table 3**; EBI study number PRJEB10070) and performed a *de novo* assembly of the *Sr33* wild-type sequence obtained from 2.9 Gb of 250-bp paired-end sequence data. This resulted in 8,235 genomic contigs (14.5 Mb) associated with NLR–containing regions (**Supplementary Table 4**). We identified three contigs that spanned 98% of the coding region of *Sr33* (**Supplementary Fig. 1**). We next compared the reads from different mutants to the wild-type assembly and searched for NLR–associated contigs containing mutations (single-nucleotide variants (SNVs) or deletions). The number of NLR mutations ranged from 39 to 142 per mutant (**Supplementary Table 5**). Thirty-one NLR contigs (different from *Sr33*) were identified, which carried independent mutations in two mutant lines, two contigs carried mutations in three lines, whereas two contigs contained mutations in four lines (**Supplementary Table 6**). These latter two contigs were both from the *Sr33* gene and identified the previously characterized *Sr33* point mutations and deletion mutations (which spanned both contigs), verifying the efficacy of this method to identify causative mutations in a single R gene in hexaploid wheat.

We next used MutRenSeq to clone the stem rust resistance gene *Sr22*, which was introgressed into wheat chromosome 7A from the diploid A-genome relatives (*T. boeoticum* and *T. monococcum*)[11,12]. In cultivar Schomburgk, *Sr22* confers resistance to commercially important races of the stem rust pathogen, including the Ug99 race group, which threatens wheat production in Africa. *Sr22* is one of the few R genes that is effective against Yemeni and Ethiopian stem-rust isolates[13] (**Fig. 2a** and **Supplementary Table 7**). However, deployment of *Sr22* has been hampered owing to poor agronomic performance associated with the *Sr22*-introgression conferred by linked gene alleles[14] (linkage drag). Further, efforts to clone *Sr22* in wheat with standard map-based approaches were unsuccessful owing to suppressed recombination in the *Sr22* region (**Supplementary Fig. 2**). We carried out an *Sr22* EMS suppressor screen using Schomburgk seeds and identified six independent susceptible mutants from 1,300 M_2 families (**Fig. 2a**). We sequenced the genomic NLR complement of Schomburgk (wild-type *Sr22*) and the six mutants using Illumina short-read sequencing (**Supplementary Table 3**; EBI study number PRJEB10099), and compared the mutant NLR complements to wild type. The number of mutations ranged from 44 to 84, and we identified 23 contigs that were mutated in two mutants, three contigs that were mutated in three mutants, and a single 3,408-bp contig, that contained independent mutations in five of the six mutants (**Fig. 2b** and **Supplementary Table 8**). This contig had homology (detected using BLAST) with the C-terminal region of an *Ae. tauschii* NLR homolog. We used the 5′ end of the *Ae. tauschii* NLR to search the *Sr22* wild-type assembly and identified a contig that carried an EMS-induced mutation in the N-terminal region of the same gene in the remaining mutant (**Fig. 2b**). We were able to physically join the two contigs using PCR of genomic and cDNA templates to obtain the full-length sequence of the predicted open reading frame of *Sr22* (**Fig. 2b** and **Supplementary Fig. 3a**). We also confirmed the presence of mutations by PCR and Sanger sequencing of each mutant DNA (NCBI study number SRP070803). All six mutations are GC to AT transitions that cause nonsense (two) or missense (four) mutations (**Fig. 2b**). To further verify *Sr22* cloning, we used the sequence to generate a PCR molecular marker, which co-segregated with *Sr22* in 2,300 gametes (**Supplementary Fig. 2**). Finally, we screened by PCR and sequencing accessions of *T. boeoticum* and its domesticated form *T. monococcum*, which have been postulated to carry *Sr22* (ref. 15), for alleles of *Sr22*, and obtained highly homologous sequences (>96% in coding region) **Supplementary Figs. 4–6**). In a *T. monococcum* mapping population, the *Sr22* homolog co-segregated with stem-rust resistance in 2,300 gametes and mapped to the orthologous location defined in the hexaploid wheat Schomburgk (**Supplementary Figs. 2** and **7**).

The 250-bp paired-end sequencing reads were unable to bridge a gap created by a 2,920-bp intron located between exon 1 and 2 of *Sr22* (**Fig. 2b**). The presence of large introns in many R genes presents a limitation in our pipeline in cases where few mutants are available or where the mutations are on either side of an intron, or when no sequenced homologs are available that can be used to define contigs belonging to the same gene. However, this limitation could be overcome if the resistant parent is sequenced using a long-read sequencing technology such as PacBio[16].

Instead of using long-read sequencing, we sequenced the Schomburgk (*Sr22*) leaf transcriptome and used the cDNA-derived reads to join contigs. Anchoring the transcriptome reads to the *Sr22* 3′ contig enabled the identification of two candidate 5′ contigs (**Supplementary Fig. 8**). Only one of these 5′ contigs matched the mutated sequence present in the mutant line described above and this contig was nearly identical to the *Ae. tauschii Sr22* homolog. Therefore, transcriptome sequencing allowed us to identify the entire coding sequence of *Sr22* without relying on the existence of a close homolog in a public database. We aligned the leaf transcriptome reads, as well as 5′ and 3′ RACE products, to the genomic sequence to determine the *Sr22* exon structure (**Supplementary Fig. 3**). The gene

© 2016 Nature America, Inc. All rights reserved.
npg

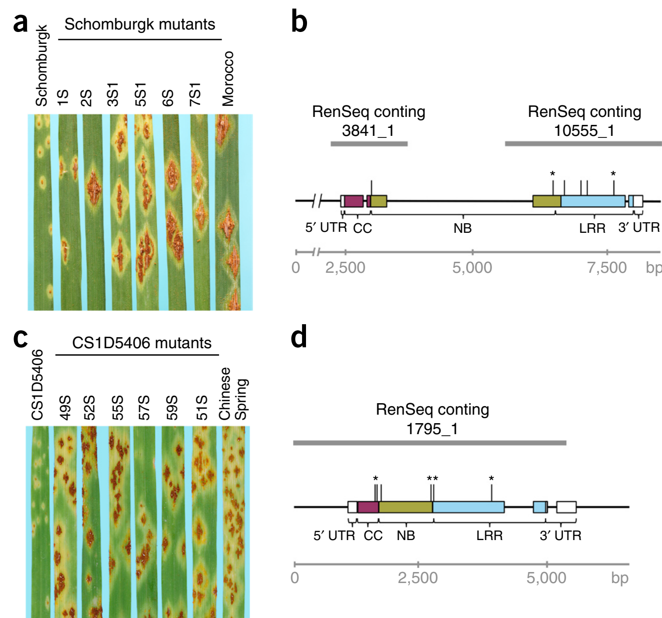

Figure 2 Cloning of *Sr22* and *Sr45* by sequencing EMS-induced susceptible mutants. (**a**,**c**) Stem rust infection phenotype of Schomburgk (*Sr22*), CS1D5406 (*Sr45*), mutants, and comparison to susceptible cultivars Morocco and Chinese Spring. Seedlings at two full-leaf stages were inoculated with stem rust urediniospores, as described[10]. Lines showing higher infection types when compared to the parental lines (Schomburgk and CS1D5406) were selected as susceptible mutant lines (**a**,**c**). Inactivation of *Sr22* and *Sr45* resistance in the mutants was confirmed in homozygous M_3 and M_4 progeny. (**b**,**d**) Schematic diagrams of *Sr22* (**b**) and *Sr45* (**d**) loci showing intron-exon boundaries, protein domain structure (colored boxes), 5′ and 3′ UTR (white boxes), and location of missense and nonsense (*) mutations. Contigs obtained by RenSeq are shown in gray. The *Sr22* and *Sr45* contigs were extended by local RenSeq and/or genome walking, and the full-length *Sr22* and *Sr45* sequences were confirmed by PCR amplification from genomic DNA and sequencing.

has four exons and three introns and spans 5,918 bp from the translation initiation to termination codons with a coding sequence of 2,826 bp and 5′ and 3′ UTRs of 75 bp and 335 bp (**Fig. 2b**). The predicted protein of 941 amino acids contains domains with homology to a coiled coil, a nucleotide binding site, and leucine-rich repeats (**Fig. 2b** and **Supplementary Fig. 9**).

To obtain the promoter region of *Sr22*, we used a modified RenSeq approach, (local RenSeq). This approach uses *in vitro*-transcribed biotinylated RNA probes (>500 nt) targeting the 5′ and 3′ ends of assembled contigs to enrich for long (>2 kb) fragments from genomic DNA libraries. Using this approach we cloned 2.5 kb of the promoter region (**Fig. 2b**). In addition, we cloned 1.6 kb of terminator region using genome walking (**Fig. 2b**). We confirmed the physical continuity of these upstream and downstream sequences with *Sr22* by (i) sequencing long-range PCR amplicons derived from mutant templates and recovering the diagnostic SNVs, and (ii) by obtaining an error-free, full-length, high-fidelity PCR-derived clone of *Sr22* (**Supplementary Fig. 3a**). In total, we cloned 8,957 bp of contiguous sequence spanning *Sr22* without needing to use a large-insert genomic BAC or fosmid library. We transformed the stem-rust susceptible cultivar Fielder with the *Sr22* clone and obtained six independent transgenic lines. We grew the transgenic Fielder lines in an automated growth cabinet and inoculated the transgenic plants at the third-leaf developmental stage with an Australian stem rust race (#98-1,2,3,5,6) which is prevalent in wheat fields in Australia and virulent on Fielder. All the transgenic lines were resistant to wheat stem rust with an infection phenotype similar to that of Schomburgk *Sr22* (**Fig. 3**). These data provide further evidence that we have cloned a functional, single *Sr22* gene.

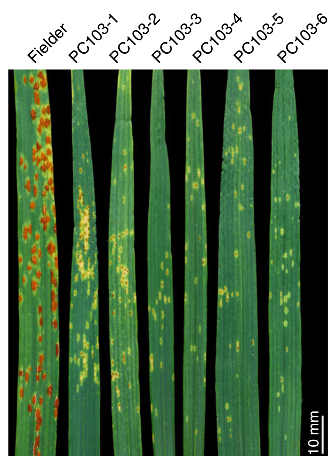

Figure 3 Stem rust infection phenotypes of Fielder and transgenic wheat lines carrying the *Sr22* gene derived from Schomburgk. PC103 (1–6) represent independent transgenic lines carrying the *Sr22* gene construct. These lines all show a typical *Sr22* stem rust resistance phenotype when challenged with the avirulent Australian stem rust strain 98-1,2,3,5 and 6.

Next, we used MutRenSeq to clone the stem-rust resistance gene *Sr45*. This gene was introgressed into wheat chromosome 1D from the D-genome progenitor *Ae. tauschii*[14]. It confers resistance to stem-rust pathogen races from Africa, India, and Australia, but virulence has been reported in Canada[17–19]. We performed an *Sr45* EMS suppressor screen and identified six susceptible mutants (**Fig. 2c**). Comparison of the RenSeq profiles (EBI study number PRJEB10112) of these mutants with the wild-type parent CS1D5406 (ref. 17) (carrying *Sr45*) revealed 28 contigs with two mutations in independent mutants, two contigs with three mutations in independent mutants, and a single 5,266-bp contig with independent point mutations in all six mutants comprising four nonsense and two missense mutations (**Fig. 2d** and **Supplementary Table 9**). We developed a PCR molecular marker from this contig and showed that it co-segregates with stem-rust resistance in a high-resolution mapping population (2,300 gametes; **Supplementary Fig. 2b**). Based on motif analysis[20] and cDNA sequencing we predict that the *Sr45* candidate contig encodes a CC-NLR protein of 1,230 amino acids. The gene spans 5,822 bp and includes a 226-bp 5′ UTR, 3,693 bp of coding sequence, two introns of 395 and 113 bp, and a 590-bp 3′ UTR (**Fig. 2d**). Additionally, 1,508 bp of *Sr45* downstream sequence that includes the entire 3′ UTR (identified through 3′ RACE) were identified through genome walking. We confirmed the physical juxtaposition of these sequences by long-range PCR and sequencing (**Supplementary Fig. 3b**). Based on the number of NLR contigs and the observed mutation rates in these contigs in the Sr45 mutant lines (**Supplementary Tables 4** and **5**), we calculated that the probability of finding the same NLR contig mutated in all six *Sr45* mutants by chance was 1 in 148,559 (**Supplementary Table 10**). We cannot *sensu stricto* rule out the possibility that another linked gene, which we did not identify in our analysis, is actually *Sr45*.

R genes are often present in clusters of related paralogs[6]. Using BLAST sequence searches we estimated the number of homologs of *Sr22* and *Sr45* in Schomburgk and CS1D5406, respectively.

© 2016 Nature America, Inc. All rights reserved.

Sr22 belongs to a small gene family with three homologs, whereas *Sr45* belongs to a larger family with 8–12 homologs (**Supplementary Fig. 10**). Therefore, MutRenSeq can identify genes belonging to multigene families.

Together with *Sr33*, *Sr35*, and *Sr50*, *Sr22* and *Sr45* are two of-five major dominant *Sr* genes cloned so far in wheat[10,21]. All five genes confer resistance to the Ug99 race group of *P. graminis* f. sp. *tritici*, whereas *Sr22, Sr33*, and *Sr50* confer broad-spectrum resistance to multiple pathogen races. Approximately 60 *Sr* genes have been genetically identified in wheat, several of which also provide broad-spectrum resistance at all plant developmental stages, for example, *Sr26*, *Sr32*, *Sr39*, *Sr40,* and *Sr47*. MutRenSeq could be used to clone these genes rapidly. Pyramiding cloned *Sr* genes at a single transgene locus is predicted by modelling to enhance the durability of resistance[1]. The physical co-location of genes at the same transgene locus would ensure co-segregation, enabling facile tracking in breeding programs and avoiding single genes again being deployed against the pathogen.

In conclusion, we report the use of mutational genomics (MutRenSeq) for cloning two *Sr* genes from the large (17×10^9 bp), hexaploid wheat genome. MutRenSeq is fast (<24 months), cheap, independent of fine mapping, and the generation of a physical contig across the map interval, and easily scalable, allowing the rescue of R genes from wheat-alien introgressions that are not currently being used in agriculture owing to linkage drag. This approach can be applied to most crops or their wild relatives, and will allow the cloning of R genes that could be used in multi-R gene pyramids, a strategy that promises more durable disease resistance in crops.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. *Sr22* and *Sr45* loci are accessible through EBI: LN883743 and LN883757. Short read raw data are available from EBI, study numbers: PRJEB10070 (*Sr33*), PRJEB10099 (*Sr22*), and PRJEB10112 (*Sr45*). *Sr22* alleles are accessible through EBI: LN883744–LN883756. Reads from resequencing of causal mutations in mutant lines are available from NCBI: SRP070803. The programs and scripts used in this analysis are available as **Supplementary Code** and have been published on Github (https://github.com/steuernb/MutantHunter/). The bait library is available as **Supplementary Data** (https://github.com/steuernb/MutantHunter/). All primer sequences are available in **Supplementary Table 11** and infection type scores are available in **Supplementary Table 12**.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was supported by funds from the Gatsby Charitable Foundation, UK; Two Blades Foundation, USA; Biotechnology and Biological Sciences Research Council, UK; Borlaug Global Rust Initiative (BGRI) Durable Rust Resistance in Wheat (DRRW) Project (administered by Cornell University with a grant from the Bill & Melinda Gates Foundation and the UK Department for International Development); USDA-ARS National Plant Disease Recovery System; Grains Research and Development Corporation, Australia; and a fellowship to A.H. from Universiti Putra Malaysia (UPM), Malaysia. We are grateful to colleagues in The Sainsbury Laboratory and the Two Blades Foundation for helpful discussions. This research was supported in part by the NBI Computing infrastructure for Science (CiS) group and Dan MacLean's group by providing computational infrastructure.

AUTHOR CONTRIBUTIONS

B.S., S.K.P., I.H.-P., K.W., M.N.R., G.Y., A.H., M.A., and H.B. performed experiments. B.S., S.K.P., E.S.L. and B.B.H.W. wrote the manuscript. B.S., S.K.P., K.W., J.D.G.J., E.S.L., and B.B.H.W. contributed to the design of the study.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. McDonald, B.A. & Linde, C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379 (2002).
2. Dangl, J.L., Horvath, D.M. & Staskawicz, B.J. Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751 (2013).
3. Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E. & Michelmore, R.W. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**, 2870–2894 (2004).
4. Smith, S.M., Pryor, A.J. & Hulbert, S.H. Allelic and haplotypic diversity at the rp1 rust resistance locus of maize. *Genetics* **167**, 1939–1947 (2004).
5. Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J. & Anderson, L.K. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat. Rev. Genet.* **8**, 77–84 (2007).
6. Hulbert, S.H., Webb, C.A., Smith, S.M. & Sun, Q. Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39**, 285–312 (2001).
7. Jupe, F. *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the *NB-LRR* gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**, 530–544 (2013).
8. Andolfo, G. *et al.* Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol.* **14**, 120 (2014).
9. Henry, I.M. *et al.* Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and Next-Generation Sequencing. *Plant Cell* **26**, 1382–1397 (2014).
10. Periyannan, S. *et al.* The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**, 786–788 (2013).
11. Gerechter-Amitai, Z.K., Wahl, I., Vardi, A. & Zohary, D. Transfer of stem rust seedling resistance from wild diploid einkorn to tetraploid durum wheat by means of a triploid hybrid bridge. *Euphytica* **20**, 281–285 (1971).
12. The, T.T. Chromosome location of genes conditioning stem rust resistance transferred from diploid to hexaploid wheat. *Nat. New Biol.* **241**, 256 (1973).
13. Olivera Firpo, P. *et al.* Phenotypic and genotypic characterization of race TKTTF of *Puccinia graminis* f. sp. *tritici* that caused a wheat stem rust epidemic in southern Ethiopia in 2013/14. *Phytopathology* **105**, 917–928 (2015).
14. The, T.T. *et al.* in *7th International Wheat Genetics Symposium* (eds. Miller, T.E. & Koebner, R.M.D.) 901–909 (Bath Press, Bath, UK, 1988).
15. Rouse, M.N. & Jin, Y. Stem rust resistance in A-Genome diploid relatives of wheat. *Plant Dis.* **95**, 941–944 (2011).
16. Witek, K. *et al.* Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* http://dx.doi.org/10.1038/nbt.3540 (2016).
17. Kerber, E.R. & Dyck, P.L. Inheritance in hexaploid wheat of leaf rust resistance and other characters derived from *Aegilops squarrosa. Can. J. Genet. Cytol.* **11**, 639–647 (1969).
18. Marais, G.F., Potgieter, G.F., Roux, H.S. & le Roux, J. An assessment of the variation for stem rust resistance in the progeny of a cross involving the *Triticum* species *aestivum, turgidum* and *tauschii. S. Afr. J. Plant Soil* **11**, 15–19 (1994).
19. Periyannan, S. *et al.* Identification of a robust molecular marker for the detection of the stem rust resistance gene *Sr45* in common wheat. *Theor. Appl. Genet.* **127**, 947–955 (2014).
20. Steuernagel, B., Jupe, F., Witek, K., Jones, J.D. & Wulff, B.B. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics* **31**, 1665–1667 (2015).
21. Saintenac, C. *et al.* Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science* **341**, 783–786 (2013).

© 2016 Nature America, Inc. All rights reserved.

npg

ONLINE METHODS

Re-identification of *Sr33*. We used previously published EMS mutants (**Supplementary Table 2**)[10] to test the MutRenSeq method. Two contigs (contig_3037_2 and contig_7846_1) were identified using our RenSeq/EMS pipeline. Both contigs were supported by two SNV mutants and two deletion mutants (**Supplementary Fig. 1**). To further test the coverage of the *Sr33* locus as obtained by RenSeq, we aligned all assembled contigs to the locus, which revealed a third contig (contig_3037_1).

EMS mutagenesis and screening for *Sr22* and *Sr45* mutants. We performed EMS mutagenesis as described by Periyannan *et al.*[10] on seeds of wheat cultivar Schomburgk carrying *Sr22*. In a kill-curve analysis, 0.3% EMS was identified as the optimum dose to cause 50% seedling mortality and reduced growth. We then treated 3,000 seeds with this concentration and advanced the plants to the M_2 generation. We harvested 1,300 single heads (M_2 families) and screened these for response to Australian stem rust race 40-1,2,3,4,5,6,7 (Plant Breeding Institute culture number 383). We identified six susceptible mutants (1S, 2S, 3S1, 5S1, 6S, and 7S1) derived from different M_2 families that were confirmed in the M_3 generation (**Fig. 2a**). A similar experiment was also performed to create stem rust–susceptible mutants from seeds of CS1D5406 carrying *Sr45*. A total of 680 single heads (M_2 families) were harvested from the cultivation of 2,000 seeds of CS1D5406 treated with 0.3% EMS. In subsequent screening for stem rust resistance, six lines (49S1, 52S1, 55S1, 57S1, 59S1, and 61S1) were identified as susceptible mutants and were reconfirmed in the M_3 generation (**Fig. 2c**).

Design of enrichment library. We designed 60,000 NLR baits of 120 nucleotides using a proprietary script from MYcroarray (http://www.mycroarray.com/). Source sequences were derived from publicly available gene annotations[22,23] and additional genomic data of *Aegilops sharonensis* (EBI|PRJEB5333). In the case of *Ae. sharonensis* we used RNA-Seq data (EBI|PRJEB5340) to *de novo* predict gene models according to Mayer *et al.*[23]. We included all genes encoding an NB-ARC (Pfam|PF00931) domain being aware that this may include additional non-NLR gene families. Predicted protein sequences were screened for NB-ARC domains using pfam_scan[24]. All predicted exons from a selected gene were added to the set of source sequences. Exons shorter than 120 bp were elongated by 60 bp, both 5′ and 3′. Resulting sequences were further refined by (i) screening for repetitive elements using RepeatMasker (http://www.repeatmasker.org/) and the Triticeae Repeat library TREP (http://wheat.pw.usda.gov/ITMI/Repeats/), (ii) clustering using CD-HIT[25] (identity threshold 95%) to remove redundant sequences, and (iii) sequences were ensured to have no reverse complementary motifs. This latter step was achieved by aligning all sequences against all sequences using BLASTn[26] and reverse complementing conflicting sequences.

DNA extraction and library preparation. We extracted total genomic DNA from leaf tissue using the method described by Lagudah *et al.*[27]. DNA quantification was performed using a NanoDrop spectrophotometer (Thermo Scientific) and the Quant-iT PicoGreen dsDNA assay (Life Technologies). DNA samples were normalized to 3 μg and sheared to an average length of 500 bp in a Covaris S2 Focused-ultrasonicator. A small aliquot was assayed by gel electrophoresis and additional size selection was carried out using Agencourt AMPure XP beads (Beckman Coulter Genomics). Samples were end-repaired followed by 3′dA addition using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs). Illumina sequencing adapters were ligated onto the ends and following purification with AMPure XP beads, the DNA was PCR-amplified (six cycles) using indexed PCR primers (NEBNext Multiplex Oligos for Illumina, New England BioLabs) and the Illumina PE1.0 PCR primer. After purification using AMPure XP beads, quality assays were performed with a Bioanalyzer DNA 1000 chip (Agilent) and the PicoGreen assay to determine the average fragment sizes and concentrations.

Target enrichment and sequencing of Illumina short insert DNA libraries. DNA libraries were enriched according to the MYbaits (MYcroarray) protocol using MYbaits reagents. Briefly, 500 ng of the prepared libraries were hybridized in hybridization buffer (10× SSPE, 10× Denhardt's solution, 10 mM EDTA, 0.2% SDS) to the biotinylated RNA baits for 40 h at 65 °C. After hybridization, bound DNA was recovered using magnetic streptavidin-coated beads as follows: the hybridization mix was added to Dynabeads MyOne Streptavidin C1 (Invitrogen, Life Technologies) that had been washed three times and resuspended in binding buffer (1 M NaCl; 10 mM Tris-HCl, pH 7.5; 1 mM EDTA). After 30 min at RT, the beads were pulled down and washed once at RT for 15 min with 1× SSC/0.1% SDS, followed by three 10 min washes at 65 °C with 0.1× SSC/0.1% SDS. Captured DNA was eluted with 0.1 M NaOH and neutralized with 1 M Tris-HCl, pH 7. Libraries were purified with AMPure XP beads and PCR amplified (14–16 cycles) using Q5 High Fidelity DNA Polymerase (New England BioLabs) and Illumina P5 and P7 primers (**Supplementary Table 11**). The enriched libraries were paired-end sequenced on the Illumina MiSeq or HiSeq platforms at TGAC (**Supplementary Table 3**). These raw data are available from EBI, study numbers: PRJEB10070(*Sr33*), PRJEB10099(*Sr22*), PRJEB10112(*Sr45*).

NLR read assembly and comparison. Primary data from wild type was *de novo* assembled using CLC assembly cell (http://www.clcbio.com/products/clc-assembly-cell/) and standard parameters. Raw data of each mutant and wild type was aligned to the wild-type assembly using BWA[28]. The resulting SAM file was filtered for reads mapping as a proper pair using SAMtools[29] and parameter –f 2. The result was converted to mpileup format using SAMtools mpileup and parameters –BQ0. *De novo* assembled wild-type contigs were aligned to source sequences of the bait library using BLASTn[26]. Only sequences of the contigs having local alignments to the source sequences were considered for further analysis. Using the mpileup format, potentially mutated nucleotide positions were identified. A position was considered for further analysis if the local coverage derived from the mapping of wild-type raw data against wild-type assembly was at least tenfold and the alternative allele frequency of a mutant was at least 10%. The latter step is thought to be extremely sensitive and capable of identifying a large number of false-positive positions that are filtered out in a subsequent step. As it is highly unlikely that two independently mutated plants have mutations at the same position, every position that was found for more than one mutant was filtered out. Regions with an average coverage less than 10% of the median overall coverage were considered as a deletion mutation for a line. Resulting candidate contigs were ranked by the number of SNV or deletion mutations within subsequences with a local alignment to the bait source sequences.

Local target enrichment (for *Sr22*). The genomic DNA library for local RenSeq was constructed essentially as described above for standard RenSeq, with minor modifications. Four μg of genomic DNA was sheared using gTUBEs with the manufacturer's settings for 5 kb and a library was constructed from 2 μg of DNA using NEBNext Ultra DNA Library Prep Kit for Illumina (P5-P7 adaptors) with extended PCR amplification time (5 min) to allow amplification of longer DNA fragments. All QC steps were performed as described above for standard RenSeq libraries.

To generate biotinylated RNA probes complementary to the 5′ end of *Sr22*, we PCR-amplified from *Sr22* genomic DNA the first 600 nt corresponding to the 5′ candidate contig (contig_10555_1) and cloned it into the pGEM-T Easy Vector (Promega) under the transcriptional control of the T7 promoter. To generate transcripts *in vitro*, we linearized plasmid clones with PstI (NEB) and synthesized biotinylated RNA using the MAXIscript T7 Transcription Kit (Life Technologies) with 60% UTP and 40% Biotin-16-UTP (Cambio). Biotinylated RNA was purified and concentrated using RNA Clean & Concentrator-5 kit (Zymo Research) and quantified with a Nanodrop spectrophotometer.

The synthesized biotinylated 5′ probe and the long-insert library were mixed in a 1:1 ratio and hybridized as described above for standard RenSeq. To further enrich for 5′ sequences, post-enrichment PCR was performed with P5 forward primer and a reverse primer specific to the 5′ region of the candidate contig (position 20-54 nucleotides) with the same conditions used for library construction. PCR products were run on a 1% agarose gel and fragments of 2–4 kb were excised, purified using a Nucleospin Extract Kit (Macherey-Nagel), and cloned into the Zero Blunt PCR Cloning vector (Life Technologies). After transformation we screened 30 bacterial colonies using P5 and an *Sr22*-specific primer and obtained six positive colonies. The cloned fragment lengths varied from 1 to 2.5 kb. The six inserts were Sanger sequenced and all were found to have identical overlapping sequences, with the longest cloned fragment being

© 2016 Nature America, Inc. All rights reserved.

2,545 bp. To confirm that the cloned fragment was derived from the 5′ region of *Sr22*, we PCR-amplified a 2.5-kb fragment that included 1.6 kb of 5′ sequence and the first 800 bp of the *Sr22* ORF, that is, the region encompassing the 2S mutation (using primers KW_sr22_p_seq_F1 & KW_sr22_p_seq_R5 –; **Supplementary Table 11**). PCR was performed on genomic DNA from wild type and the 2S mutant and produced a single product for each template that was directly Sanger sequenced. Sequencing of the wild-type PCR product confirmed the sequence of the *Sr22* 5′ region whereas the sequence of the PCR product from S2 encoded the expected mutation.

Extending RenSeq contigs by BLAST search, whole transcriptome sequencing, and long-range PCR. *BLAST search*[26]. The *Sr22* 5′ candidate contig (contig_10555_1) was aligned to NCBI nonredundant protein sequences using BLASTx. The highest similarity observed was to the 3′ end of a disease resistance gene analog (RGA) from *Ae. tauschii* (gi|475585845|gb|EMT20152.1). We then used the 5′ end of the *Ae. tauschii* RGA that extended beyond contig_10555_1 and searched the *Sr22* wild-type assembly using tBLASTn. The most similar contig identified, contig_3841_1, was the only sequence with homology to the 5′ region of the *Ae. tauschii* RGA. This same region of sequence contained a nonsynonymous mutation in the *Sr22* mutant 2S.

RNA-Seq. Paired-end reads from whole transcriptome sequencing of Schomburgk (carrying *Sr22*) were quality trimmed using Sickle (https://github.com/najoshi/sickle). Trimmed reads were aligned to the 3′ contig of *Sr22* (contig_10555_1) using BLASTn[26]. All reads with a local alignment (≥99% identity) and their mates were then aligned to the other contigs of the *Sr22* assembly. All contigs were then scored by the number of times they were the most similar alignment by a read.

For validation of the entire *Sr22* coding sequence transcriptome sequences were mapped to the final predicted *Sr22* locus using Tophat2 (ref. 30). The alignment was manually inspected using Savant[31] (**Supplementary Fig. 8**).

Long-range PCR. The full-length sequence of the *Sr22* open reading frame was further confirmed by PCR amplification using the GoTaq Long PCR Master Mix (Promega) and the primer pair S22F14 and S22R4 (**Supplementary Table 11**). PCR products were cloned into the pCR-XL-TOPO vector (Life Technologies) and subsequently sequenced.

Genetic mapping of *Sr22* and *Sr45*. We generated a mapping population by crossing Schomburgk, the hexaploid wheat line carrying *Sr22* derived from *Triticum boeoticum* accession G-21 (ref. 32), to the stem rust–susceptible cultivar Westonia. We screened 1,150 F_2 seeds with the closely linked flanking Simple Sequence Repeat (SSR) markers cfa2123 and cfa2019 (ref. 33) as per the method of *Kota et al.*[34], and identified 180 recombinants. Seeds of the selected recombinants were advanced to the F_3 generation and tested for rust response using the method and rust pathogen isolates described in Periyannan *et al.*[35]. The DNAs of 180 recombinant lines were subsequently screened with cssu22 (ref. 35), the chromosome 7A–specific wheat expressed sequence tags (wESTs) BE443521, BG274853, BE498985, BG262287, BF201318, and BG604641 (ref. 36), and the markers AT7D7092 and AT7D7094 derived from the homoeologous chromosome 7D region[37]. Markers BE498985 and BG274853 were found to be distal and separated from *Sr22* by one and three recombination events, respectively, whereas markers BE443521 and AT7D7094 were proximal and separated from *Sr22* by eight and two recombination events, respectively. The remaining markers, cssu22, BG262287, BF201318, and BG604641, were found to co-segregate with *Sr22* (**Supplementary Fig. 2**).

In addition to the hexaploid mapping family, mapping of the *Sr22* locus was verified using 1,150 F_2 plants derived from the cross between the *T. monococcum* accession PI190945 (carrying *Sr22*)[38] and the stem rust susceptible accession PI272557 (**Supplementary Fig. 7**). We used markers BE498985 and AT7D7094 to identify recombinants. The presence or absence of *Sr22* in recombinants was determined by screening with *P. graminis* f. sp. *tritici* isolate 06ND76C (race QFCSC) that is avirulent on wheat lines with *Sr22* (ref. 37). None of the co-segregating markers from the Schomburgk x Westonia population were found to be polymorphic between PI190945 and PI272557.

Mapping of the identified *Sr22* gene using the above hexaploid and diploid populations were performed using sequence-specific primer pairs (S22F2/S22R3 specific to *Sr22* in Schomburgk, S22F2/S22R10 specific to *Sr22* in PI190945, S22F2/S22R51 specific to *sr22* in Westonia and S22F2/S22R22 specific to *sr22* in PI272557; **Supplementary Table 11**). Using multiple sequence alignment, sequences diagnostic for the resistant *Sr22* haplotypes were identified and were used in the development of a corresponding gene-specific PCR marker (S22GMF and S22GMR) that can be used in marker-assisted selection of *Sr22* resistance (**Supplementary Figs. 5** and **6a**).

For mapping *Sr45*, the high-resolution mapping population (CS1D5406 × Chinese Spring) described in Periyannan *et al.*[19] and primer pair S45F1/S45R1 (**Supplementary Table 11** and **Supplementary Fig. 6b**) specific to the *Sr45* gene sequence were used.

Genome walk and identification of 5′ and 3′ regions of *Sr22* and *Sr45*. To identify the 5′ and 3′ regions of *Sr22* and *Sr45* genome walking was performed using the APAgene GOLD Genome Walking Kit (Bio S&T Inc., Canada). Primers S22R13, S22R12, and S22R11 were used to amplify regions 5′ of the *Sr22* ATG start codon whereas the primers S22F35, S22F36, and S22F37 were used to amplify sequences 3′ of the gene ORF. Similarly, the 3′ region of the *Sr45* gene sequence was amplified using primers S45F11, S45F12 and S45F13.

5′ and 3′ RACE. To identify the terminal sequences of *Sr22* and *Sr45* cDNAs, 5′ and 3′ RACE (rapid amplification of cDNA ends) reactions were performed using the method described in Krattinger *et al.*[39]. The 5′ RACE reaction was performed using primers S22R13 (for *Sr22*) and S45G1 (for *Sr45*) which are specific for the 5′ region of each gene, respectively, whereas the 3′ RACE reaction was performed using S22F13 (for *Sr22*) and S45G3 (for *Sr45*) primers from the 3′ genic region. To confirm intron locations, we amplified cDNAs using S22F24 and S22R39 (for *Sr22* from *T. boeoticum*), S22F24 and S22R37 (for *Sr22* from *T. monococcum*) and S45F7 and S45R1 (for *Sr45*) primer pairs (**Supplementary Table 11**).

Haplotype analysis. In addition to the *Sr22* gene introgressed into Schomburgk from *T. boeoticum*, a second source of stem rust resistance, derived from *T. monococcum*, was introgressed into chromosome 7A in other hexaploid wheat backgrounds. This second introgression is believed to also encode *Sr22* on the basis of identical resistance specificities to stem rust races. The hexaploid wheat W3534 (ref. 32) is archetypical of the *T. monoccoccum*-derived *Sr22* and in a subsequent extensive analysis of *T. monococcum* collections, several accessions including PI190945 were postulated to also carry *Sr22*. PI190945 was confirmed to possess a single dominant gene that segregated independently of *T. monococcum* stem rust resistance genes other than *Sr22* (ref. 38). In addition, variants of the *Sr22* gene sequence present in the parents of the above mapping families were identified from 32 *T. monococcum* accessions, 16 *T. boeoticum* accessions (from ICARDA), DV92 and W3534 (ref. 32) described in **Supplementary Table 12**. *Sr22*-related sequences were amplified using two overlapping primer pairs: S22F25 and S22R16, and S22F2 and S22R36. For Westonia, the primer pair S22F29 and S22R51 was used instead of S22F25 and S22R16. We obtained the corresponding *Sr22* sequence from PI190945 and it co-segregated with the resistance phenotype located on chromosome 7A in the diploid mapping population (PI190945 × PI272557) and is orthologous with the *Sr22* gene location in Schomburgk (**Supplementary Fig. 2**). W3534 and PI190945 were shown to possess an identical *Sr22*-like gene sequence. However, they differed from the reference *Sr22* gene sequence in Schomburgk by 33 nonsynonymous nucleotide substitutions (**Supplementary Fig. 5**), suggesting allelic variation at the *Sr22* locus for these genotypes. Additional sequences from *T. monococcum* accessions, PI289605 and PI330550 (**Supplementary Table 12**) that showed similar resistance responses to *P. graminis* f. sp. *tritici* races as PI190945, were identical with each other but differed from PI190945 and W3534 at seven predicted amino acid positions. Similar predicted amino acid changes were also found between the sequences of resistant *T. boeoticum* accessions (IG44857 and IG44921) and Schomburgk whereas the sequence from a third *T. boeoticum* accession, IG44855, was nearly identical to that of Schomburgk except for a predicted two amino acid deletion at the N-terminal region (**Supplementary Fig. 5**). Unexpectedly, the predicted amino acid sequence of *T. monococcum* accession PI573523, which has a low rust infection phenotype, was closely related to the susceptible *T. boeoticum* accession IG44878 suggesting the possibility of resistance from a locus other than *Sr22* in this line. Overall, a number of predicted amino acids

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/nbt.3543

were found to provide distinctions between the diploid A genome taxa, however, the sequence of one stem rust–susceptible accession from *T. monococcum* was grouped with the *T. boeoticum*–resistant genotypes indicating the possibility of an intermediate taxa or a misclassification (**Supplementary Fig. 4**).

***Sr22* and *Sr45* copy number analysis.** We performed an *in silico* analysis to estimate the number of NLR sequences present in the RenSeq assemblies that show homology to the *Sr22* and *Sr45* coding regions. The alignment was performed with BLASTn[26] using default parameters (including: -task megablast -evalue 10). Each sequence with homology to the coding region of either of these genes is shown in **Supplementary Figure 10**, relative to genomic sequences of *Sr22* and *Sr45*.

Transformation of *Sr22* into wheat line Fielder. A genomic fragment of 7.815 kb that encodes the *Sr22* gene and 5′ (1.292 kp) and 3′ (0.605 kb) regulatory regions was PCR amplified from Schomburgk using primers S22F14 and S22R4. The amplified fragment was cloned into the NotI site of the binary vector pVecBARII, a derivative of pWBvec8 (ref. 40) in which the 35S hygromycin gene has been replaced with a 35S BAR selectable marker gene. The *Sr22* gene was introduced into wheat cultivar Fielder using the *Agrobacterium*-transformation protocol described by Ishida *et al.* 2014 (ref. 41) and phosphinothricin as a selective agent. Six independent primary transgenic plants carrying the *Sr22* gene as well as sibling lines without the *Sr22* transgene were recovered and grown in an automated growth cabinet set with day and night temperature of 23 °C, 16 h light and 8 h dark conditions. The plants were inoculated with the Australian *Pgt* race 98-1,2,3,5, and 6, which is virulent on Fielder at the fully developed third-leaf stage. After 24 h of incubation in a closed transparent plastic box under high humidity, the plants were restored to the original growth conditions and observed for rust development. Scoring for rust infection were done after 14 d after inoculation where all six transgenic plants showed rust resistance phenotypes consistent with the *Sr22* infection type whereas Fielder control plants showed full susceptibility (**Fig. 3**).

22. International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788 (2014).
23. Mayer, K.F. *et al.* A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).
24. Finn, R.D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
25. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
26. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
27. Lagudah, E.S., Appels, R., Brown, A.H.D. & McNeil, D. The molecular-genetic analysis of *Triticum tauschii*, the D-genome donor to hexaploid wheat. *Genome* **34**, 375–386 (1991).
28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
29. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
30. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
31. Fiume, M. *et al.* Savant Genome Browser 2: visualization and analysis for population-scale genomics. *Nucleic Acids Res.* **40**, W615–W621 (2012).
32. Paull, J.G., Pallotta, M.A., Langridge, P. & The, T.T. RFLP markers associated with *Sr22* and recombination between chromosome 7A of bread wheat and the diploid species *Triticum boeoticum*. *Theor. Appl. Genet.* **89**, 1039–1045 (1994).
33. Khan, R.R. *et al.* Molecular mapping of stem and leaf rust resistance in wheat. *Theor. Appl. Genet.* **111**, 846–850 (2005).
34. Kota, R., Spielmeyer, W., McIntosh, R.A. & Lagudah, E.S. Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr2* from pseudo-black chaff in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **112**, 492–499 (2006).
35. Periyannan, S.K., Bansal, U.K., Bariana, H.S., Pumphrey, M. & Lagudah, E.S. A robust molecular marker for the detection of shortened introgressed segment carrying the stem rust resistance gene *Sr22* in common wheat. *Theor. Appl. Genet.* **122**, 1–7 (2011).
36. Akhunov, E.D. *et al.* Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* **11**, 702 (2010).
37. Luo, M.C. *et al.* A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc. Natl. Acad. Sci. USA* **110**, 7940–7945 (2013).
38. Rouse, M.N. & Jin, Y. Genetics of resistance to race TTKSK of *Puccinia graminis* f. sp. *tritici* in *Triticum monococcum*. *Phytopathology* **101**, 1418–1423 (2011).
39. Krattinger, S.G. *et al.* A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**, 1360–1363 (2009).
40. Wang, M.B., Li, Z.Y., Matthews, P.R. & Upadhyaya, N.M. Improved vectors for Agrobacterium tumefaciens-mediated transformation of monocot plants. *Acta Hortic.* **461**, 401–408 (1998).
41. Ishida, Y., Tsunashima, M., Hiei, Y. & Komari, T. Wheat (Triticum aestivum L.) transformation using immature embryos. *Methods Mol. Biol.* **1223**, 189–198 (2015).

© 2016 Nature America, Inc. All rights reserved.

npg