Crop wild relative populations of *Beta vulgaris* allow direct mapping of agronomically important genes

Gina G. Capistrano-Gossmann1, D. Ries2, D. Holtgräwe2, A. Minoche3,4, T. Kraft5, S.L.M. Frerichmann1, T. Rosleff Soerensen2, J.C. Dohm6, I. González7, M. Schilhabel1, M. Varrelmann8, H. Tschoep9, H. Uphoff5, K. Schütze10, D. Borchardt10, O. Toerjek10, W. Mechelke10, J.C. Lein10, A.W. Schechert11, L. Frese12, H. Himmelbauer3,6,7, B. Weisshaar2 & F.J. Kopisch-Obuch1,10

Rapid identification of agronomically important genes is of pivotal interest for crop breeding. One source of such genes are crop wild relative (CWR) populations. Here we used a CWR population of <200 wild beets (*B. vulgaris* ssp. *maritima*), sampled in their natural habitat, to identify the sugar beet (*Beta vulgaris* ssp. *vulgaris*) resistance gene *Rz2* with a modified version of mapping-by-sequencing (MBS). For that, we generated a draft genome sequence of the wild beet. Our results show the importance of preserving CWR *in situ* and demonstrate the great potential of CWR for rapid discovery of causal genes relevant for crop improvement. The candidate gene for *Rz2* was identified by MBS and subsequently corroborated via RNA interference (RNAi). *Rz2* encodes a CC-NB-LRR protein. Access to the DNA sequence of *Rz2* opens the path to improvement of resistance towards rhizomania not only by marker-assisted breeding but also by genome editing.

1 Plant Breeding Institute, Kiel University, Am Botanischen Garten 1-9, Kiel 24118, Germany. 2 CeBiTec & Faculty of Biology, Bielefeld University, Universitätsstraße 25, Bielefeld 33615, Germany. 3 Max Planck Institute for Molecular Genetics, Ihnestraße 73, Berlin 14195, Germany. 4 Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney NSW 2010, Australia. 5 Syngenta Seeds AB, Box 302, Landskrona 26123, Sweden. 6 Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU), Muthgasse 18, 1190 Vienna, Austria. 7 Centre for Genomic Regulation (CRG), Carrer del Dr. Aiguader 88, Barcelona 08003, Spain. 8 Department of Phytopathology, Institute of Sugar Beet Research (IfZ), Holtenser Landstraße 77, Göttingen 37079, Germany. 9 SESVanderHave N.V., Industriepark, Tienen 3300, Belgium. 10 KWS SAAT SE, Grimsehlstraße 31, Einbeck 37555, Germany. 11 Strube Research GmbH & Co. KG, Hauptstraße 1, Söllingen 38387, Germany. 12 Federal Research Centre for Cultivated Plants (JKI), Erwin-Baur-Str. 27, Quedlinburg 06484, Germany. Correspondence and requests for materials should be addressed to B.W. (email: bernd.weisshaar@uni-bielefeld.de) or to F.J.K.-O. (email: Friedrich.KopischObuch@kws.com).
Crop wild relatives (CWRs) are an invaluable genetic resource for crop improvement. CWR populations collected in their natural habitat not only provide access to agricultonically important traits but can also serve to directly identify genes underlying these traits. Gene identification has been done so far mainly in three types of populations: germplasm collections, elite breeding materials, and synthetic populations such as biparental or multi-parent advanced generation inter-crossed populations. Compared to those populations, CWR populations have usually undergone many generations of outcrossing and thus display low linkage disequilibrium and high population admixture.

Rhizomania is the most important sugar beet disease next to infection by beet cyst nematode (Heterodera schachtii). It is caused by the beet necrotic yellow vein virus (BNYVV) and reduces the yield of sugar beet by up to 80%. The BNYVV virus is transmitted by the soil-borne plasmodiophoromycete Polymyxa betae. Starting from the first appearance in Italy in 1952, rhizomania has spread to almost all sugar beet-growing areas of the world. The only way to control this soil-borne disease is through the cultivation of rhizomania-resistant varieties that have been made available since the mid-1980s. Resistance breeding has relied mostly on the resistance genes Rz1 and Rz2 as single loci or a combination of both. The resistance conferred by Rz1 has been discovered in sugar beet germplasm, while Rz2 is derived from the wild beet Beta vulgaris ssp. maritima. Alleles that provide resistance show a dominant inheritance at both loci, and Rz2 confers a higher resistance level than Rz1 (ref. 8). Rz1 and Rz2 were mapped at a distance of about 20 cM on chromosome (Chr) 3 of the sugar beet genome. A potentially additional rhizomania-resistance locus has been mapped to sugar beet Chr3 (BvChr3) and is most likely identical to Rz2 as it is derived from the same B. vulgaris ssp. maritima origin in Denmark.

Here we demonstrate the potential of a CWR population for gene-level resolution mapping by identifying the sugar beet (B. vulgaris ssp. vulgaris) rhizomania-resistance locus Rz2 in a wild beet population of B. vulgaris ssp. maritima through a modified version of mapping-by-sequencing (MBS). The results not only confirm, in yet another case, the potential of CWR for crop improvement and the importance of preserving CWR in their natural habitats but also provide access to an agronomically important gene conferring resistance against rhizomania that encodes a CC-NB-LRR protein.

Results

A CWR population to identify a candidate for Rz2 by MBS. Aiming at fine mapping and identification of the Rz2 gene, we sampled seeds from individual plants of the B. vulgaris ssp. maritima CWR population that is growing along the coast of Kalundborg, Denmark (Fig. 1). This population, referred to as Kalundborg population, is known to segregate for Rz2 (ref. 6). Given the allogamous flowering biology of this wild beet species, we assumed thorough population admixture and low linkage disequilibrium in the population, which are both features required for high-resolution mapping.

We crossed 189 different parental wild beets with a rhizomania-susceptible sugar beet line and used the resulting 189 full sib families for a replicated progeny test. This test for rhizomania resistance consisted of four independent greenhouse assays, each including semiquantitative immuno-detection of BNYVV virions. The families showed segregation for rhizomania resistance and allowed inference of the genotype at the Rz2 locus of the respective parental wild beets (homozygous resistant (Rz2/Rz2), heterozygous (Rz2/rz2) or homozygous susceptible (rz2/rz2)). The results were explained by assuming a single locus, showing that Rz2 is the only rhizomania-resistance locus segregating in the Kalundborg population. Referring back to their original location, resistant genotypes appeared to be randomly distributed along the coastal line (Fig. 1, Supplementary Data 1). For most of the markers, the allele frequencies did not deviate significantly from Hardy–Weinberg equilibrium (HWE; Supplementary Table 1) and thus suggest a random mating population.

For MBS, we selected only parental wild beet individuals with high confidence in phenotype classification and used them to create one bulk of only four homozygous-resistant wild beets and one bulk with eight homozygous-susceptible wild beets. The two bulks were each sequenced using the Illumina technology (Supplementary Table 2). For access to a useful reference sequence for read mapping, the Rz2 donor accession WB42 (ref. 6) was sequenced and assembled to draft quality (Supplementary Table 3). The assembly was concatenated according to the sugar beet reference version 1.2 (ref. 14) to yield pseudochromosomes. RefBeet1.2 covers 567 Mbp of the B. vulgaris ssp. vulgaris genome that has an estimated haploid genome size of about 730 Mbp in nine chromosomes. The initial assembly, which was designated WB42-v1, was subsequently corrected manually with regard to the positioning of WB42 scaffolds and contigs to BmChr3 to yield the optimized assembly WB42-v2.

MBS analysis with WB42-v2 resulted in the detection of four closely connected intervals within a genomic region of 0.72 Mbp on BmChr3 (Fig. 2a,b, Supplementary Figs 1 and 2), overlapping with 25 almost continuous scaffolds and contigs of WB42-v2. Four smaller intervals spanning altogether about 0.09 Mbp at three loci were found on three other chromosomes (Supplementary Fig. 1, Supplementary Table 4). In the case of the scaffolds and contigs that cause signals outside of BmChr3, even detailed manual inspection did not allow us to deduce a reliable position in the genome sequence. However, mapping with additional markers indicated that these sequences also belong to BmChr3 but are not closely linked to the resistance locus defined by the phenotype (Supplementary Table 5). This excludes these three loci as possible location for Rz2. Gene prediction was carried out using the WB42-v2 assembly of B. vulgaris ssp. maritima together with RNA-Seq read evidence generated from tissue derived from rhizomania-infected plants related to line WB42. A total of 33,922 genes with expression evidence were predicted, including 60 genes in the target region on BmChr3. DNA and deduced peptide sequences of these 60 genes, and of 14 additional genes from the same region without expression evidence, were compared to the NCBI NR protein database and the current B. vulgaris annotation BeetSet-2 (ref. 16) (Supplementary Data 2). We detected a candidate for Rz2 that (i) encodes a CC-NB-LRR protein, (ii) is absent from BeetSet-2 obtained from the susceptible genotype KWS2320 and (iii) was supported by 100% transcript evidence in the WB42-v2 gene prediction. The Rz2 candidate had been included in the initial B. vulgaris gene prediction with the ID Bv3_jumg as a truncated gene but omitted from BeetSet-2 due to the absence of detectable transcription in the reference genotype KWS2320. Apart from its truncation in the reference sequence genotype, Bv3_jumg was located in a region that displays strong synteny between WB42 and KWS2320. The region was well assembled, co-linear and displayed no indication of read coverage deviation (Supplementary Fig. 3), indicating that no duplicated or paralogous loci from elsewhere in the genome interfered.

Fine mapping of the Rz2 candidate to rhizomania resistance.

A total of 33 single-nucleotide polymorphisms (SNPs) segregating in the Kalundborg population, and distributed over the 0.72-Mbp
target region, were converted to molecular markers (see Supplementary Table 6) and used to genotype the complete panel of the 189 parental wild beets (Supplementary Data 1). By association analysis of the marker results with the corresponding resistance phenotype data, 19 markers were significantly associated with resistance in at least two independent rhizomania-resistance tests (Fig. 2c; Supplementary Table 7). The best association was observed for marker CAU3880, which detected an SNP within the gene Bv3_jumg that encodes the CC-NB-LRR candidate for Rz2. Notably, CAU3880 was also the only marker that displayed association with rhizomania resistance (single marker F-test, P value = 6.1E-4, r² = 0.17) in an independent second CWR population that was sampled in Brighton, France (see Supplementary Data 3). Therefore, we hypothesized that the CC-NB-LRR gene Bv3_jumg containing CAU3880 is the best candidate for Rz2.

To validate this hypothesis, the markers with the highest association in the CWR populations were tested using diversity panels of sugar beet inbred lines representing the germplasm of three breeding companies (Supplementary Table 8). Across this association panel, only markers CAU3880, CAU4188 and CAU3882 showed perfect co-segregation with rhizomania resistance. In addition, detailed analysis of the recombination events detected in the Kalundborg CWR population showed that neighbouring genes upstream of Rz2 were excluded by CAU3881, while genes downstream of Rz2 were excluded by CAU4188. In the Kalundborg CWR population, the marker CAU4188 also excluded an ankyrin repeat encoding gene that is located downstream of Rz2 and included in BeetSet-2 (Bv3_tftt, see Supplementary Data 2). We concluded that Rz2 is the only gene that is fully associated with resistance to rhizomania.

Validation of Rz2 as the rhizomania-resistance gene by RNAi.

To finally corroborate Rz2 (Bv3_jumg) as the rhizomania-resistance gene, an RNAi experiment was performed using the resistant sugar beet genotype 6921_RR. An RNAi construct was designed to inactivate the intact allele of Rz2KD-R, and this construct was transformed into 6921_RR. The transgenic line displays sensitivity to rhizomania to an extent comparable to the susceptible control (Supplementary Fig. 4).

Sequence analysis of the available Rz2 alleles revealed that alleles associated with susceptibility display defects in the CC-NB-LRR gene Bv3_jumg. The susceptible allele from the Kalundborg population (rz2KD,S) contained a premature stop codon, and the allele from the susceptible sugar beet reference genotype KWS2320 contained a transposon insertion (rz2KWS2320; Fig. 2d,e, Supplementary Table 9). We concluded that an intact allelic version of Bv3_jumg represents Rz2 and that susceptibility to rhizomania is caused by the absence of a functional allele of this gene. This conclusion is further supported by the similarity of Rz2 and the encoded CC-NB-LRR protein to other identified plant virus-resistance genes of the same type.

Discussion

The CC-NB-LRR family of the disease-resistance genes, which includes Rz2 on the basis of high amino acid sequence similarity, has been named after the domains that these resistance-conferring receptors typically contain: a nucleotide-binding site (NB) domain, a leucine-rich repeat (LRR), and a coiled-coil (CC) domain. CC-NB-LRR proteins can recognize a wide variety of taxonomically unrelated pathogens, including viruses, bacteria, fungi and even insects. Activation of these genes results in a hypersensitive response characterized by the rapid death of cells in a local region surrounding an infection site, restricting the growth and spread of pathogens to other parts of the plant.

The identification of Rz2 in a CWR population of an allogamous diploid species demonstrates the potential for high-resolution mapping of major genes. The small population size (N<200) drastically reduces the cost for phenotyping and marker assays. In comparison, when working with synthetic populations, 8,283
(ref. 20) and 5,457 (ref. 21) F2 plants were required for identifying sugar beet bolting time genes in a map-based cloning approach. In contrast to multi-parent advanced generation inter-crossed populations that required a minimum of eight crop generations2, it took only one step of seed increase to obtain a CWR-derived mapping population. Potentially, this step could even be skipped if the trait has high single plant heritability, such as bolting20,21 or other monogenic traits22. Although no direct comparison within B. vulgaris is available, the advantage of CWR populations over populations that are derived from germplasm collections or elite material13,23,24 is the high population admixture.

The confirmation of the correct identification of the candidate gene Bv3_jumg as Rz2 might have the complication that potentially existing additional sequence-related genes could be off-targets of the RNAi construct. However, read coverage analyses of the Bv3_jumg region as well as the correlation of

Figure 2 | The Rz2 region identified by MBS. (a) Plot of delta-AFe values for BmChr3. The delta-AFe values, indicated by yellow dots, clearly show a skew in the AFe values of the two pools created from four and eight plants unequivocally identified as homozygous resistant and susceptible, respectively. The blue bar indicates the target region defined by genetic intervals from MBS. (b) Sketch of the 0.72 Mbp target region. Blue bars indicate genetic intervals identified by MBS. (c) Association analysis results after testing 189 families of the Kalundborg population. Black and white circles represent markers associated and not associated with resistance, respectively. Arrows indicate the number of recombinant plants identified. (d) Location of Bv3_jumg identified as Rz2 by recombinant analysis. (e) Structure of the Rz2 gene. Rz2KD-R: allele identified in resistant plants in Kalundborg, rz2KWS2320: allele present in the susceptible sugar beet genotype KWS2320; rz2KD-S: allele identified in susceptible plants in Kalundborg. White box: location of conserved domains in the encoded protein sequence for three main domains of a CC-NB-LRR protein; grey box: transposon insertion location; number and letters: predicted amino acid changes.
the inactivated alleles (rz2) with susceptibility argue strongly against the existence of such genes in the B. vulgaris genome. Considering also the marker CAU3880 that is located within the candidate gene, we are confident that Bv3_jun0g is Rz2.

Our approach relies on naturally occurring variation and is, therefore, independent of the production of a mutant population, unlike the recently published MutRenSeq method53. MutRenSeq combines chemical mutagenesis with targeted exome capture sequencing of resistance gene analogues (RGAs). The two approaches are distinct and complementary for the molecular identification of causal resistance genes. However, our approach also allows access to any other type of genes causal for a clear phenotype.

In the work presented, a CWR natural population was explored by application of MBS technology and allowed direct identification of a major locus of high agronomic importance. This was facilitated by the allogamous flowering biology of B. vulgaris ssp. maritima that led to an accumulation of recombination events facilitated by the allogamous flowering biology of a major locus of high agronomic importance. This was also allows access to any other type of genes causal for a clear phenotype.

In addition, 129 full sib families developed from the Brighton population were phenotyped, as described above, in one test using B-type BYVVV30,31. As controls, 20 plants of an Rz2 line, an Rz1 line and one rhizomania-susceptible line were used in each test. The results are summarized in Supplementary Data 3.

Molecular marker analysis. Leaf samples were taken from each wild beet plant that was grown for production of full sib families. Extracting of genomic DNA was performed using the standard CTAB method52. For fine mapping, a total of 33 SNPs distributed within the target region identified by MBS of 0.72 Mbp in length were cleaved to Konpetitive Allele-Specific PCR (see Supplementary Table 6) or cleaved amplified polymorphic sequence markers and used to genotype the complete panel of 189 families. Konpetitive Allele-Specific PCR genotyping assays were performed in a total volume of 5 μl containing 0.1 ng of genomic DNA according to the manufacturer’s guidelines (LG¢ Genetics LLC, Beverly, MA, USA). Cleaved amplified polymorphic sequence genotyping was used for CAU4188 (BW42 allele: A; alternative allele: C; position in BmChr3: 6,894,501). An amplicon of 457 bp was amplified using the primers 5′-TGGGAAATTTGCGGAGGA- GA-3′ and 5′-GGAGCTTCCAGATGGCCTTGC-3′ by PCR. After PCR, an enzyme digestion using Sphi was performed according to the manufacturer’s guidelines (ThermoFisher Scientific, Southampton, UK). Markers CAU4220, CAU4021 and CAU4022 (Supplementary Table 5) were evaluated by direct Sanger sequencing of amplicons generated by PCR.

For the markers in the 0.72 Mbp target region, deviations from the HWE expectations were estimated using the marker results listed in Supplementary Data 1. The P values for deviation from HWE are listed in Supplementary Table 1.

Genetic association analysis. Genotyping data from SNP markers targeting the 0.72 Mbp region on B. vulgaris ssp. maritima chromosome 3 (BmChr3) were used in the analysis. Association analysis for rhizomania resistance was performed in accordance with a Generalized Linear Model method using the TASSEL software (version 3.0). For assuming an association, an adjusted P value (Bonferroni correction) of <0.0015 was required, and r2 (correlation coefficient r2 = 0.1) was used to evaluate the magnitude of the marker effects. The results are summarized in Supplementary Table 7.

Generation of a BW42 draft genome sequence. A genome sequence of the Rz2 donor line BW42 was generated. DNA from a single plant was sequenced with the Illumina technology. Data generated from one pair-end (PE) library with an insert size of 600 bp and from two mate-pair (MP) libraries with span sizes of 2.5 and 5 kb, respectively, were combined. PE sequencing on an Illumina HiSeq 2000 instrument was performed using a 2 × 100 cycle protocol for the PE library and a 2 × 50 protocol for the MP libraries. Illumina sequencing reads were quality filtered and trimmed prior to assembly. Reads were removed if they did not pass the chastity filter, had at least one uncalled base, contained at least six bases of the Illumina adaptor sequence in both ends or had less than two-thirds of the bases with quality value Q≥30 within the first half of the read. Additionally, low-quality reads were marked as ‘?’ in the quality string (‘B-tails’) were trimmed out, resulting in 390 million high-quality PE read-pairs equivalent to 100-fold coverage of the beet genome. Quality filtering, trimming and redundancy elimination resulted in 30 million MPFs for the 2.5 kb span size library and 35 million MPFs with span size of 5 kb. Sequence read data were submitted to Sequence Read Archive (SRA). The reads were assembled using SOAPdenovo V1.05 (ref. 34) with standard parameters and a kmer-size of 49. Gaps were filled using SOAP GapCloser v1.12. The final assembly size was 531,940,822 bp in 57,361 scaffolds and contigs > 500 bp. The largest scaffold had a length of 899,438 bp, and N50 size was 59,342 bp (see Supplementary Table 3). We refer to this initial assembly as BW42-v0 and to the 57,361 sequences in this assembly as ‘BW42-v0 contigs’.

Concatenation of contigs to yield BW42 pseudochromosomes. For BWs, we generated 125,422 pseudochromosomes from the BW42-v0 draft. Contigs were ordered according to the sugar beet reference genome sequence ReBeeet-1.2 (ref. 14) assuming strong synteny between the two subspecies B. vulgaris ssp. vulgaris and B. vulgaris ssp. maritima. We used BLAST+ (ref. 35) to determine sequence similarity among the BW42-v0 contigs to ReBeeet-1.2. The BW42-v0 contigs were ordered according to their sequence coordinates. To further align the BW42-v0 contigs, we aligned the reference with the dust filter turned off, an e-value cutoff of 1e-50 and a culling limit of 1; (ii) high scoring pairs were combined and the best overall alignment for each BW42-v0 contig was kept; (iii) BW42-v0 contigs were ordered by their mapping position; hits in minus orientation were included as reversed-complement, (iv) the contigs mapping to each chromosome were concatenated to pseudochromosomes with stretches of 50 N’s between contigs. We refer to this initial sorted and concatenated assembly as BW42-v1.

Mapping-by-sequencing. BWs were performed by applying a workflow15 that has been modified to fit the CWR setting, initially using BW42-v1 as a target genome.
sequence and an adapted interval detection (see below). MBS for Rz2 was based on the determination of allele frequencies in two DNA pools that were built from four resistant (R) and susceptible (S) sugarcane lines. Equal amounts of genomic DNA isolated from freeze-dried leaf samples of single plants were pooled according to the resistance phenotype and sequenced using the Illumina technology.

Library preparation and sequencing of pooled DNA. Library preparation for the DNA pools R1/B2444 and S1/B2446 was performed according to the Illumina TruSeq Sample Preparation v2 guide. DNA from each pool was fragmented by ultrasonication. After end repair and A-tailing, individual indexed PE adapters were ligated to the DNA fragments, which allow a multiplexed PE sequencing run. The adapter-ligated fragments were size selected on a 2% low melt agarose gel to a size of 350–650 bp. After enrichment, PCR of fragments of the final library, that carry adapters on both ends, was performed with a yield of 10–15 ng/mL. The average fragment size of each library was determined on a BioAnalyzer High Sensitivity DNA chip. The samples were sequenced on a HiSeq 2000. Cluster generation for a high output run was done on a cBot using the TruSeq PE Cluster Kit v3. and 2 > 100 bp reads were generated using the TruSeq SBS Kit v3. Sequence read data of both pools were submitted to SRA. After completion of the sequencing runs, basecalling, demultiplexing and fastq file generation was performed using the CASAVA-1.8.2 programs. The results are summarized in Supplementary Table 2.

Targeted reordering of the WB42 sequence assembly. The allele frequency values (AFe) determined from sequence data (see below) of the two DNA pools indicated several genetic intervals with highly significant separation between the resistant and susceptible pool. When using WB42-v1 assembly as genome sequence, these intervals were located mainly on BmChr3 but in addition also on four other pseudochromosomes as well as in the unassigned assembly fraction. We concluded that the region of interest was assembled incorrectly and that the WB42-v1 assembly had insufficient quality for MBS and thus set out to develop a WB42 assembly that would be a better mapping resolution. We focussed on BmChr3 for targeted reordering based on data from genetic markers that place Rz2 on BmChr3. We manually inspected scaffolds and contigs from ReBfeet-1.2 that caused localization of WB42-v0 contigs to BmChr3 and analysed WB42-v0 contigs that might be incorrectly localized to other parts of the genome although the correct localization might be BmChr3. We used data from a genotyping-by-sequencing experiment15 to identify true SNPs in regions of BvChr3 that displayed very low variation between the parents of the sugar beet mapping population that was used to build ReBfeet-1.2 (ref. 14) and also data from BeetMap-3 (ref. 36). The regions that displayed very low variation between the parents turned out to contribute significantly to the wrong localization of contigs in WB42-v1. The respective ReBfeet-1.2 scaffolds were addressed by marker-based genetic anchoring. Validated SNPs were genotyped by direct Sanger sequencing and manual SNP calling from the tracelines in the KWS1 population36 and genetically anchored between 62.3 and 68.1 cM on BvChr3 (cM values according to BeetMap-3). Some sections within RefBeet-1.2 scaffolds were evaluated, resulting in support according to BeetMap-3). Some sections within RefBeet-1.2 scaffolds were addressed by marker-based genetic anchoring. Validated SNPs were genotyped by direct Sanger sequencing and manual SNP calling from the tracelines in the KWS1 population36 and genetically anchored between 62.3 and 68.1 cM on BvChr3 (cM values according to BeetMap-3). Some sections within RefBeet-1.2 scaffolds were addressed by marker-based genetic anchoring. Validated SNPs were genotyped by direct Sanger sequencing and manual SNP calling from the tracelines in the KWS1 population36 and genetically anchored between 62.3 and 68.1 cM on BvChr3 (cM values according to BeetMap-3). Some sections within RefBeet-1.2 scaffolds were addressed by marker-based genetic anchoring. Validated SNPs were genotyped by direct Sanger sequencing and manual SNP calling from the tracelines in the KWS1 population36 and genetically anchored between 62.3 and 68.1 cM on BvChr3 (cM values according to BeetMap-3). Some sections within RefBeet-1.2 scaffolds were addressed by marker-based genetic anchoring. Validated SNPs were genotyped by direct Sanger sequencing and manual SNP calling from the tracelines in the KWS1 population36 and genetically anchored between 62.3 and 68.1 cM on BvChr3 (cM values according to BeetMap-3).

Calculation of allele frequencies from sequencing data. All operations from postprocessing and mapping of reads over InDel realignment, base quality score recalibration to variant calling and calculation of AFe values were performed on the compute cluster of the CeBiTeC. The reads of the sequenced pools R1/B2444 and S1/B2446 were processed by adapter trimming, quality filtered by removing reads with stretches of four consecutive bases with a mean quality value < 30 and removal of bases at the read heads and tails with quality values < 30 (ref. 15). The remaining reads were mated to the ordered and concatenate WB42 sequences v1 and v2 using BWA-MEM. SNPs and InDels were called using GATK38. AFe and delta-AFe values were calculated from the proportion of reads for the two states of a given sequence variant within and between pools, respectively15.

Interval identification and prediction of causal variation. For visual inspection of the results, the delta-AFe values of detected variants were plotted along all WB42 ORF regions for the five WB42 pseudochromosomes or for selected regions (see Supplementary Figs 1 and 2). To reduce background noise, only variants with a read coverage between 0.75 and 2.5 times the average coverage of uniquely mapped reads were plotted. Additionally, an established algorithm13 for automatic interval detection was adapted to the CWR scenario. An interval was defined as a genomic region containing a follow-up of variants with AFe values of ≥ 0.1 in the susceptible (S1/B2444) pool. Interruptions of the follow-up were ignored if they consist of at least 1 variation with an AFe > 0.1, flanked on each side by a variant with AFe values of ≤ 0.1. Interval detection was started at seed variants with a delta-AFe value close to 1. The exact, lowest delta-AFe value of valid seeds (X) was calculated depending on the read coverage and tolerance to a small number of non-supportive reads. X = (P – 100 × E/C) with phenotypic difference of the pools P = 0.9, the coverage of both pools combined for the variant C and an estimate for errors introduced by HiSeq sequencing E = 0.01. Adjacent intervals were merged into one if the gap between them was ≤ 0.5 cM. The threshold interval after the merging was < 0.1 bp. The remaining intervals were filled with a mixture of rhizomania-infected soil and sterile sand in 1:1 proportions. Before transplantation, the soil in the containers was thoroughly soaked with nutrient solution11. Root samples of KD_091343 were collected 3, 6, 9, 12, 15 and 20 days after transfer to the infected soil. The infection efficiency of the soil was confirmed by assays of the rhizomania-susceptible line KD_091115, which were grown in parallel.

Total RNA was extracted from the infected root tissue using thepeqGold Plant RNA Kit (PEQLAB) according to the manufacturer’s instructions. After on-column DNase treatment with thepeqGold DNase I Digest Kit, the RNA was quantified and performed with RNA-Seq (2 × 100 bp). RNA-Seq data were mapped to the HiSeq 1500 by using the Illumina RNA Sequencing Kit complemented with reverse transcriptase according to the instructions of the manufacturer. One barcoded library was created for each of the six time points. Raw RNA-Seq reads were processed as described19 and mapped to the WB42-v2 genome sequence (see below). In total, RNA-Seq yielded 177 million reads (535 million reads of the single reads after trimming of 86 bp. Sequence read data of the six libraries were submitted to SRA. For expression detection and gene structure validation, the resulting BAM files were analysed with Readexplorer2 (ref. 40).

WB42 ORF prediction. All scripts were provided by AUGUSTUS35 and used following AUGUSTUS’ instructions. Ab initio gene prediction on the WB42-v2 genome sequence was carried out using AUGUSTUS version 2.7 with the default functional annotation as well as the accession nos. of the best BLASTp hit were submitted to SRA. For expression detection and gene structure validation, the resulting BAM files were analysed with Readexplorer2 (ref. 40). Functional annotation of WB42 genes in intervals. The deduced 74 peptide sequences of the predicted WB42 genes located in the 0.72 Mbp target region on BmChr3 (60 with expression evidence and 14 without) containing the four main MBS intervals (see Supplementary Table 4) were analysed to collect functional annotation information for these genes. Note that the termini of the genetic intervals were determined by the position of variants within WB42-v2 scaffolds, which places these termini within scaffolds. As a result, the sum of the length of the scaffolds that overlap the four genetic intervals was longer than the total length of the four genetic intervals adding up to 0.53 Mbp (see above). In functional annotation, a BLASTp analysis of the WB42-v2 peptide sequences was carried out against the non-redundant protein sequence database (NCBI nr protein, update date: 2016/06/01 containing 88,499,796 sequences, E-value cutoff limit e-10), and functional annotation as well as the accession nos. of the best BLASTp hit were extracted. Second, the peptide sequences were analysed by BLASTp for their best hit in the set of all genes from ReBfeet14, including those without transcript evidence (protein sequences extracted from ReBfeet.unfiltered_genes.1302.gff3, which is available at http://bvisq.nolgen.mpg.de/, E-value cutoff limit e-30). Third, sequences were used as query in a Blast Like Alignment Tool (BLAT) analysis against all peptide sequences deduced from BeetSet-2 (ref. 16). BLAT hits > 50% query coverage are listed in Supplementary Data 2. BLAT hits with query coverage > 80% and an identity of > 85% were considered as an indication for the detection of true homologues of WB42 genes in BeetSet-2. Finally, we searched for known resistance gene domains in the WB42-v2 peptide sequences of the target region by using PROSITE model24 that had been established for sugar beet. A domain was considered as present if it was at least 95% complete with an E-value of ≤ 1e-5. We identified two RGAs; both RGA genes are highlighted in Supplementary Data 2.
One of the two, namely, Bv3_jung, a resistant standard sugar beet genotype (line 6921_RR) was transformed with a DNA construct encoding a double-stranded hairpin RNA resulting in 6921_RNAi. The dsRNA was aimed at post-transcriptional gene silencing of the resistant Bv3_jung allele. In order to provide a suitable DNA construct, a defined target sequence region of the resistant Bv3_jung allele of 434 base pair length (nt 247–680 of the mRNA) was selected, amplified by PCR and cloned both in sense and antisense direction in the vector pZFN, a modified version of pHLRA, which is suitable for the synthesis of hairpin structures. This vector contains a dual CaMV 35S promoter, two multiple cloning sites separated by intronic sequences, an intron from the Arabidopsis thaliana gene AtAAP6 and the nos terminator region. Transformation of sugar beet was performed according to established protocols and by using kanamycin as selection drug. Following a number of selection steps, successful transformation was examined on transgenic shoots via PCR by detection of the nptII gene, the AtAAP6 intron sequence, the two T-DNA border sequences (LB/RB) and the absence of vir. Positive shoots were clonally multiplied in vitro to 30 shoots in each case, rooted and transferred into soil in the greenhouse. Approximately 2 weeks later, the transgenic sugar beet plants (6921_RNAi) were transplanted into a rhizomania assay as described above. As controls, the non-transformed-resistant genotype (6921_RR) that was used for RNAi transformation and the susceptible line D108_ssm were included (Supplementary Fig. 4).

References

1. Brescghello, F. & Sorrells, M. E. Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323–1330 (2006).
2. Huang, B. E. *et al.* MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* 128, 999–1017 (2015).
3. Tamada, T. & Baba, T. Beet necrotic yellow vein virus from rhizomania-affected sugar beet in Japan. *Ann. Phytopathol. Soc. Jpn* 39, 325–332 (1973).
4. McGann, G. R., Grimmer, M. K., Mutasa-Gottgens, E. S. & Stevens, M. Progress towards the understanding and control of sugar beet rhizomania disease. *Mol. Plant Pathol.* 10, 129–141 (2009).
5. Scholten, O. E. & Lange, W. Breeding for resistance to rhizomania in sugar beet. *Euphytica* 112, 219–231 (2002).
6. Scholten, O. E., De Bock, T. S., Klein-Lankhorst, R. M. & Lange, W. Inheritance of resistance to beet necrotic yellow vein virus in *Beta vulgaris* conferred by a second gene for resistance. *Theor. Appl. Genet.* 99, 740–746 (1999).
7. Lewellen, R. T. Registration of 11 sugarbeet germplasm C79 lines with resistance to *P. capsici*. *Crop Sci.* 37, 948–951 (1997).
8. Acosta-Leal, R., Bryan, B. K. & Rush, C. M. Host effect on the genetic diversification of beet necrotic yellow vein virus single-plant populations. *Phytopathology* 100, 1204–1212 (2010).
9. Barzen, E., Mechelke, W., Ritter, E., Seitzer, J. F. & Salamini, F. RFLP markers for sugar beet breeding: chromosomal linkage maps and location of major genes for rhizomania resistance, monogerm and hypocotyl colour. *Plant J.* 2, 601–611 (1992).
10. Scholten, O. E., Klein-Lankhorst, R. M., Esselink, D. G., De Bock, T. S. & Lange, W. Identification and mapping of random amplified polymorphic DNA (RAPD) markers linked to resistance against beet necrotic yellow vein virus (BNYVV) in Beta accessions. *Theor. Appl. Genet.* 94, 123–130 (1997).
11. Gidner, S. *et al.* QTL mapping of BNYVV resistance from the WB41 source in sugar beet. *Genome* 48, 279–285 (2005).
12. Schneeberger, R. Using next-generation sequencing to isolate mutant genes from breeding panels. *BMC Genomics* 16, 184 (2015).
13. de Ronde, D., Butterbach, P. & Kornelink, R. Dominant resistance against plant viruses. *Front. Plant Sci.* 5, 307 (2014).
14. McDowell, J. M. & Woffenden, B. J. Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.* 21, 178–183 (2003).
15. van den Ende, E., Kato, N. & Lawton, M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature* 411, 848–853 (2001).
16. Pin, P. A. *et al.* The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. *Curr. Biol.* 22, 1095–1101 (2012).
17. Dally, N., Xiao, K., Holmgräve, D. & Jung, C. The B2 flowering time locus of beet encodes a zinc finger transcription factor. *Proc. Natl Acad. Sci. USA* 111, 10365–10370 (2014).
18. Francis, S. Sugar-beet powder milderw (Erysiphe betae). *Mol. Plant Pathol.* 3, 119–124 (2002).
19. Stich, B. *et al.* Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). *Theor. Appl. Genet.* 117, 1167–1179 (2008).
20. Wu, R. L. & Zeng, Z. B. Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157, 899–914 (2000).
21. Steurlemong, B. *et al.* Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655 (2016).
22. Vincent, H. *et al.* A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265–275 (2013).
23. Moxted, N., Dulloo, M. E. & Ford-Lloyd, B. V. *Enhancing Crop Genepool Use: Capturing Wild Related and Landrace Diversity for Crop Improvement* (CABI, 2016).
24. Grimmer, M. K., Bean, K. M. R. & Asher, M. J. C. Mapping of five resistance genes to sugar-beet powder mildew using AFLP and anchored SNP markers. *Theor. Appl. Genet.* 115, 67–75 (2007).
25. Whitney, E. D. Identification, distribution, and testing for resistance to beet (*Beta vulgaris*) and sugar (*Beta accessions*). *Genome* 50, 737–746 (2007).
26. Vincent, H. *et al.* Reliable isolation of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655 (2016).
27. Van der Auwera, G. A. *et al.* Rapid generation of high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Gigascience* 1, 1–12 (2012).
28. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl Acad. Sci. USA* 81, 8014–8018 (1984).
29. Koenig, R., Luddecke, P. & Haebeler, A. M. Detection of beet necrotic yellow vein virus strains, variants and mixed infections by examining single-strand conformation polymorphisms of immunocapture RT-PCR products. *J. Gen. Virol.* 76, 2051–2055 (1995).
30. Bravaz, M. *et al.* Restriction fragment length polymorphism analysis of reverse transcription-PCR products reveals the existence of two major strain groups of beet necrotic yellow vein virus, *J. Gen. Virol.* 75, 1835–1842 (1994).
31. Saghai-Marof, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Influence of a 30 bp deletion in the repeat unit on the generation of RFLP patterns. *Mol. Gen. Genet.* 123, 105–113 (1969).
32. Sagha-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl Acad. Sci. USA* 81, 8014–8018 (1984).
33. Sagha-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl Acad. Sci. USA* 81, 8014–8018 (1984).
34. Convertino, C. & Bhandari, A. M. Detection of beet necrotic yellow vein virus strains, variants and mixed infections by examining single-strand conformation polymorphisms of immuno capture RT-PCR products. *J. Gen. Virol.* 76, 2051–2055 (1995).
35. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 11, 1110 (2013).
39. Stracke, R. et al. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (Beta vulgaris). BMC Plant Biol. 14, 249 (2014).
40. Hilker, R. et al. ReadXplorer 2-detailed read mapping analysis and visualization from one single source. Bioinformatics 32, 3702–3708 (2016).
41. Stanke, M., Scho¨ffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7, 62 (2006).
42. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
43. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
44. Kent, W. J. BLAT—the BLAST-like alignment tool. Genome Res. 12, 656–664 (2002).
45. Lindsey, K. & Gallois, P. Transformation of sugarbeet (Beta vulgaris) by Agrobacterium tumefaciens. J. Exp. Bot. 41, 529–536 (1990).

Acknowledgements
This work was supported by the German Ministry of Food and Agriculture (BMEL) grant FKZ 2814505310 to F.J.K.-O. and M.V. and the German Ministry of Education and Science (BMBF) grant FKZ 0315957D (Acronym: NuGGET) to B.W. We thank Monika Bruisch for help with plant sampling, as well as Meike Pfeiler and Prisca Viehöver for excellent wet lab work. In addition, we thank Christian Jung for providing laboratory infrastructure and support.

Author contributions
F.J.K.-O. conceived the study; G.G.C.-G. and F.J.K.-O. designed and analysed the mapping experiments; K.S., O.T., D.B., W.M. and J.C.L. provided the transgenic RNAi results; G.G.C.-G., D.B., D.H., F.J.K.-O. and B.W. wrote the manuscript; D.R., D.H., T.R.S. and B.W. established the pseudochromosomes of WB42 using RefBeet1.2 as backbone, performed the NGS analysis, genome annotation and RNA-Seq analysis; A.W.S., T.K., M.V., H.T., H.U., J.C.L. and W.M. developed the material, did resistance tests and selected elite material for the validation panel; L.F. recommended exploitation of the Kalundborg and Brighton CWR populations; G.G.C.-G., D.H. and S.L.M.F. carried out marker analysis, resistance test analysis, association analysis and sequence analysis of the candidate genes; A.M., J.D., I.G. and H.H. sequenced and assembled the genome of WB42; and M.S. sequenced B2444 and B2446.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: O.T., D.B., W.M. and J.C.L. have filed a patent application that includes the molecular complementation of Rz2 (WO 2014/202044 A1). All other authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Capistrano-Gossmann, G. G. et al. Crop wild relative populations of Beta vulgaris allow direct mapping of agronomically important genes. Nat. Commun. 8, 15708 doi: 10.1038/ncomms15708 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.