

1 A fully phased interspecific grapevine rootstock genome
2 sequence representing *V. riparia* and *V. cinerea* and allele-
3 aware annotation of the phylloxera resistance locus *Rdv1*

4
5
6 Bianca Frommer¹, Ludger Hausmann², Daniela Holtgräwe¹, Prisca Viehöver¹,
7 Bruno Hüttel³, Richard Reinhardt³, Reinhard Töpfer², Bernd Weisshaar¹

8
9 ¹ Bielefeld University, Chair of Genetics and Genomics of Plants, Faculty of Biology
10 & Center for Biotechnology (CeBiTec), Bielefeld, Germany

11 ² Julius Kuehn Institute (JKI), Institute for Grapevine Breeding Geilweilerhof, Sie-
12 beldingen, Germany

13 ³ Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Re-
14 search, Cologne, Germany

15
16 Email addresses:

17 BF: frommer@CeBiTec.Uni-Bielefeld.DE / ORCID: 0000-0002-5792-0102

18 LH: ludger.hausmann@julius-kuehn.de / ORCID: 0000-0002-4046-9626

19 DH: dholtgra@CeBiTec.Uni-Bielefeld.DE / ORCID: 0000-0002-1062-4576

20 PV: viehoeve@CeBiTec.Uni-Bielefeld.DE

21 BH: huettel@mpipz.mpg.de / ORCID: 0000-0001-7165-1714

22 RR: reinhardt@mpipz.mpg.de / ORCID: 0000-0001-9376-2132

23 RT: reinhard.toepfer@julius-kuehn.de / ORCID: 0000-0003-1569-2495

24 BW: bernd.weisshaar@Uni-Bielefeld.DE / ORCID: 0000-0002-7635-3473
25

26 Abstract

27 The phylloxera resistant rootstock cultivar ‘Börner’ is an interspecific hybrid derived
 28 from *Vitis riparia* and *V. cinerea* and a valuable resource for *Vitis* disease resistances.
 29 We created a fully phased, high-quality ‘Börner’ genome sequence named BoeRC
 30 using long PacBio reads. Comprehensive gene annotation of both ‘Börner’ haplo-
 31 types, designated BoeRip and BoeCin, was applied to describe the phylloxera resis-
 32 tance locus *Rdv1*. Using a mapping population derived from a susceptible *V. vinifera*
 33 breeding line and ‘Börner’, the *Rdv1* locus was further delimited. *Rdv1*, which is de-
 34 rived from *V. cinerea* and included in the haplotype BoeCin, was compared with se-
 35 quences of phylloxera-susceptible and phylloxera-tolerant cultivars. Between flanking
 36 regions that display high synteny, we detected and precisely characterized a diverse
 37 sequence region that covers between 202 to 403 kbp in different haplotypes. In
 38 BoeCin, five putative disease resistance genes were identified that represent likely
 39 candidates for conferring resistance to phylloxera.

40 Keywords

41 Assembly, PacBio, SMRT sequencing, grapevine, rootstock, genome, Canu, annota-
 42 tion, resistance, resistance gene analogs, phylloxera, *Rdv1*
 43

44 Background

45 *Vitis vinifera* Linné subspecies *vinifera* is the most cultivated *Vitis* species worldwide.
 46 It is economically highly important and appreciated for wine production, table grapes,
 47 dry fruits like raisins, and other products. The species is endemic to Europe, and do-
 48 mestication and development of cultivars has a very long history ¹. *V. vinifera* is
 49 highly susceptible to many diseases and pests including downy and powdery mildew
 50 as well as to the root aphid phylloxera (*Daktulosphaira vitifoliae* Fitch) ². In the 19th
 51 century, phylloxera caused a crisis of unprecedented dimension for viticulture and
 52 questioned the survival of viticulture in general. *Vitis vinifera* plants died quickly after
 53 phylloxera attack of the roots and huge cultivation areas were destroyed. Only the
 54 grafting of susceptible *V. vinifera* cultivars as scions to phylloxera-tolerant rootstocks
 55 derived from American wild *Vitis* species or their interspecific hybrids rescued viti-
 56 culture ^{3, 4}. The phylloxera plague remains as a serious case even if it is controlled at
 57 the moment.

58 The selection of a good rootstock that provides resistances or tolerances to the almost
 59 ubiquitous phylloxera is crucial for nowadays viticulture. One newly developed root-
 60 stock variety is the cultivar ‘Börner’, an interspecific F1 hybrid derived from a cross
 61 of the American *Vitis* genotypes *Vitis riparia* GM183 and *Vitis cinerea* Arnold.
 62 ‘Börner’ inherited full resistance to phylloxera on roots from *V. cinerea* and develops
 63 no nodosities or tuberosities ⁵ upon infestation with phylloxera (Fig. 1). *V. cinerea* is
 64 an American wild species known to often provide high resistance at the root and very
 65 high resistance at the leaf against phylloxera ⁶. A quantitative trait locus (QTL), lo-
 66 cated on chromosome (chr) 13, was identified to provide phylloxera resistance ⁶ and is
 67 referred to as *Rdv1* (resistance to *D. vitifoliae*) ⁷. In addition, ‘Börner’ inherited posi-
 68 tive viticultural traits including tolerance to drought, resistance to black rot caused by
 69 the fungus *Guignardia bidwellii* ⁸, and resistance to downy mildew caused by *Plas-*
 70 *mopara viticola* ^{9, 10}.

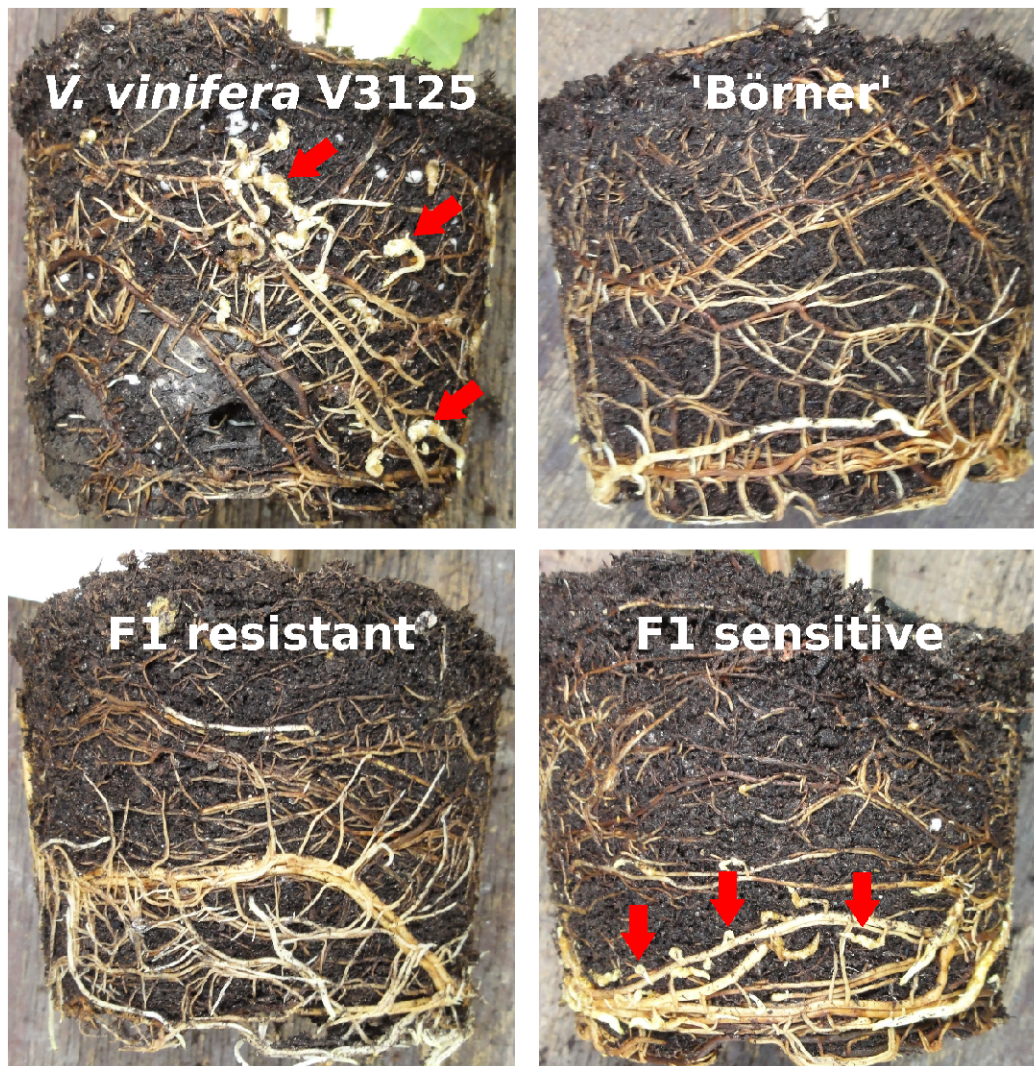


Fig. 1: Root balls of susceptible and resistant grapevine plants attacked by phylloxera. Some phylloxera-induced nodosities on the roots are marked with red arrows. The pictures show roots of the grapevines indicated, the two F1 plants selected with the phenotypes as indicated are derived from the mapping population V3125 x 'Börner' ⁶.

Disease resistance, usually conferred through resistance genes, makes plants capable of surviving attacks from a broad variety of pathogens and pests ¹¹. So called resistance gene analogues (RGAs) are identified by the protein domains they encode and structural features which are similar to validated resistance genes ¹². Based on the protein domain structure that also indicates the localisation and/or site of action within the cell, they are broadly classified as nucleotide-binding site and leucine rich repeat domain receptor (NLRs) or pattern-recognition receptors (PRRs) ¹¹. Usually, PRRs

act at the cell surface because of a transmembrane (TM) domain, while NLRs cover a nucleotide-binding site (NBS) domain and usually act intracellularly.

Vitis genomes cover a haploid set of 19 chromosomes, and the haploid genome size varies around 500 Mbp^{13, 14}. The inbred cultivar ‘PN40024’ has been sequenced^{15, 16} and the resulting assembly serves as a reference genome sequence (referred to as PN40024). However, since *Vitis* species and cultivars like almost all perennial plants are highly heterozygous and genetically diverse^{10, 17, 18}, there is a need for more *Vitis* genome sequences. Several additional *Vitis* genome sequences have become available which were generated from long reads, including those of Cabernet Sauvignon^{19, 20}, Carménère²¹, Chardonnay^{14, 22}, *Muscadinia rotundifolia* cultivar Trayshed²³ and *V. riparia* Gloire de Montpellier (referred to here as VitRGM)²⁴. These assemblies each cover a primary and an alternate pseudo-haplotype of which the alternate pseudo-haplotype is usually of reduced contiguity. The cultivar *V. riparia* Gloire de Montpellier is tolerant against root infestation by phylloxera²⁵, but to the best of our knowledge it is unknown if this tolerance is linked to the *Rdv1* locus or to another region of the genome.

Long read DNA sequencing technology, like "Single Molecule, Real Time" (SMRT) sequencing²⁶ provided by Pacific Biosciences (PacBio), is one option to generate high quality genome sequence assemblies. The long reads more likely span problematic genomic regions and thus pave the way for more contiguous assemblies²⁷. Also, information of the haplotype and phase differences are more completely contained in a single read. This fact, together with the transmission of exactly one haplotype from two parents to a single offspring according to Mendel's laws, is exploited by the "trio binning" approach for generating fully phased genome sequence assemblies²⁸.

To access and resolve the *Rdv1* locus, we set out to generate a haplotype-resolved genome sequence of ‘Börner’ with the goal to overcome complications caused by high heterozygosity and the complexity of resistance gene clusters. Using SMRT sequencing to generate long reads and additional Illumina short read data from both parents of ‘Börner’, we assembled both haplotype sequences of ‘Börner’ at chromosome level. The two truly phased haplotypes of the diploid interspecific hybrid ‘Börner’ represent the genome sequences of the two species *V. riparia* and *V. cinerea*. Structural and functional gene annotation was performed for both haplotype sequences, and RGAs were studied. The new sequence and annotation data as well as

117 additional mapping results were used to dissect the *Rdv1* locus of ‘Börner’ at the gene
118 level.

119 Results

120 Sequencing data of ‘Börner’ and its parents

121 To assemble the ‘Börner’ genome sequence, ~66 Gbp of raw SMRT sequencing data
122 comprising 7,328,737 subreads with an average read length of 9,016 bp and an N50
123 length of 12,963 bp were generated. The expected diploid genome size (2n) of
124 ‘Börner’ is 2x500 Mbp; thus the calculated coverage with long reads for a genome
125 sequence with merged haplotypes would be 132 times, or more than 60-fold coverage
126 for each individual haplotype. To allow k-mer-based binning of the long reads accord-
127 ing to the two haplotypes that ‘Börner’ inherited from its parents, the parental geno-
128 types were sequenced with Illumina technology to about 140-fold coverage.

129 A ‘Börner’ genome sequence assembly in two phases

130 To generate a ‘Börner’ genome sequence with two separated haplotypes, the trio-
131 binning approach was applied²⁸. The two haplotype-specific ‘Börner’ read subsets or
132 bins, designated Vrip and Vcin, contain approximately 50 % of all reads and bases.
133 These were assembled with Canu (see Methods). Statistics of the two resulting haplo-
134 type assemblies are shown in Table 1. Both assemblies reached the expected genome
135 size of ~500 Mbp. The BoeRip and BoeCin haplotype assemblies represent "1n dou-
136 ble haploid" genome sequences of the two species *V. riparia* and *V. cinerea*, respec-
137 tively.

138 **Table 1 Assembly statistics for the raw and scaffolded haplotype assemblies**
139 **BoeRip and BoeCin of ‘Börner’.**

	Scaffold Level ^a		Total	
	BoeRip	BoeCin	Contigs ^b	Scaffolds
Sequences	971	767	1,843	1,738
Size (bp)	495,882,484	501,563,116	997,019,452	997,445,600
Largest seq. (bp)	14,614,105	16,784,012	16,784,012	16,784,012
N50 (Mbp)	5.29	5.34	4.79 ^c	5.29
N90 (Mbp)	0.72	0.51	0.49 ^c	0.55

140 ^ascaffolded haplotype assemblies

141 ^braw haplotype assemblies, numbers include contigs shorter than 10 kbp

142 ^csee Supplementary File 1 Fig. S1 for details

Pseudochromosome construction and validation by genetic map

The assignment of scaffolds to pseudochromosomes was mainly based on the PN40024 reference sequence and was achieved by using RBHs (see Methods). For BoeRip, 446 scaffolds accounting for 95.21 % of all bases were assigned to pseudochromosomes. For BoeCin, 320 scaffolds holding 93.89 % of all bases were assigned. Since the reference contains a pseudochromosome chr00 for unassigned sequences, this has been created as well and is referred to as "chrUn". In addition, some scaffolds find no clear corresponding sequence in the reference and result in a pseudochromosome "chrNh" (for no hit, see Fig. 2A). As a result, the haplotype assemblies each consist of 21 pseudochromosome sequences representing the 19 true chromosomes and two artificial sequences.

To validate the pseudochromosomes, 314 SSR markers that show unequivocal chromosome mappings on the haplotype assemblies were evaluated relative to their position on PN40024. A total of 309 markers confirm the pseudochromosome assignment. Conflicting markers were checked manually and corrected if possible (see Methods). The completeness of the assemblies was evaluated by detection of plant core genes with BUSCO (Fig. 2B). For both BoeRip and BoeCin, about 96.5 % of the conserved single copy gene set were detected. This indicated that both haplotype assemblies are more complete than the assemblies PN40024 and VitRGM (Supplementary File 2 Table S2).

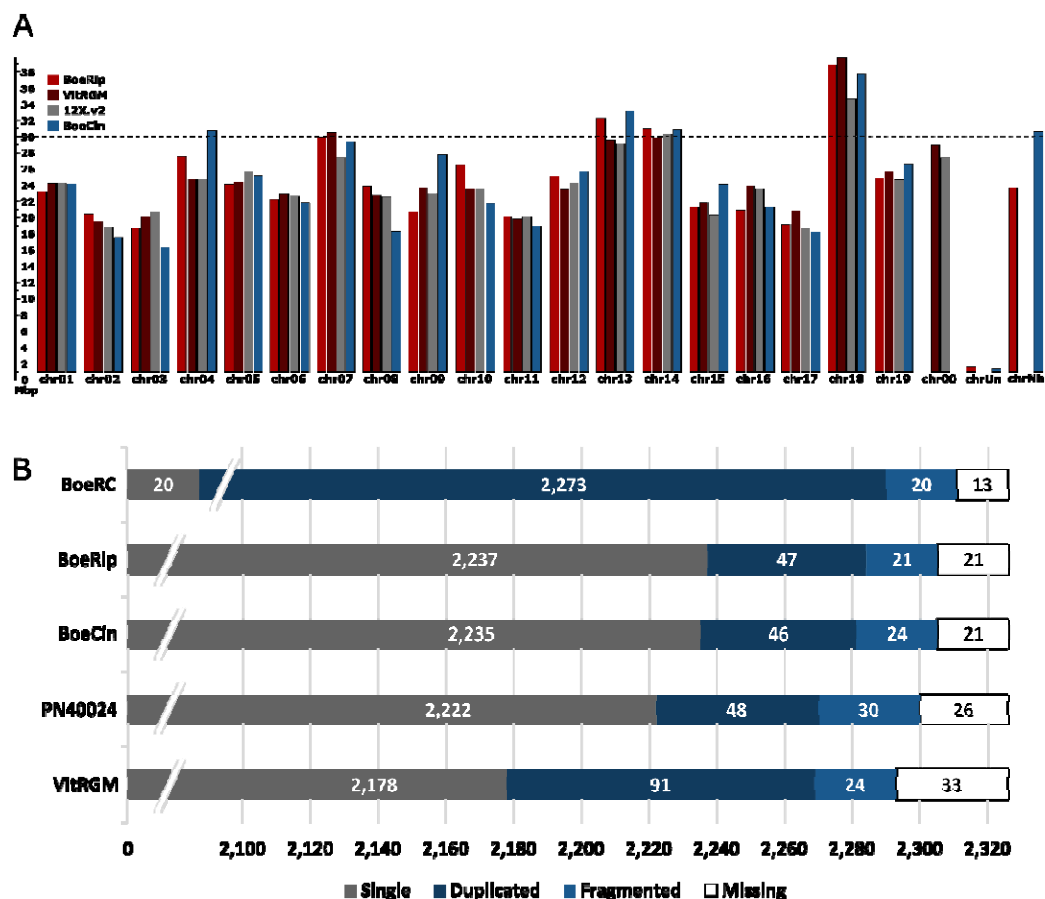


Fig. 2 Comparison of the ‘Börner’ genome sequence assembly with other *Vitis* assemblies. (A) Pseudochromosome lengths comparison of both ‘Börner’ haplotypes, PN40024 and VitRGM. The bars represent the pseudochromosomes of the different assemblies. Red, BoeRip; dark red, VitRGM; grey, PN40024; blue, BoeCin. ChrUn holds sequences that were assigned to chr00 of PN40024 or VitRGM, and chrNh collects sequences with no assignment (Supplementary File 2 Table S1). (B) Plant core gene content (2,326 eudicots genes in reference set) of the assemblies in comparison to the PN40024 and VitRGM genome assemblies. Note that the bar graph is truncated at the left and focusses on only the duplicated, fragmented and missing BUSCO genes. The track labelled BoeRC represents the whole ‘Börner’ assembly (combined BoeRip and BoeCin sequences; merged results from the two haplotypes).

Evaluation of assembly quality and phasing

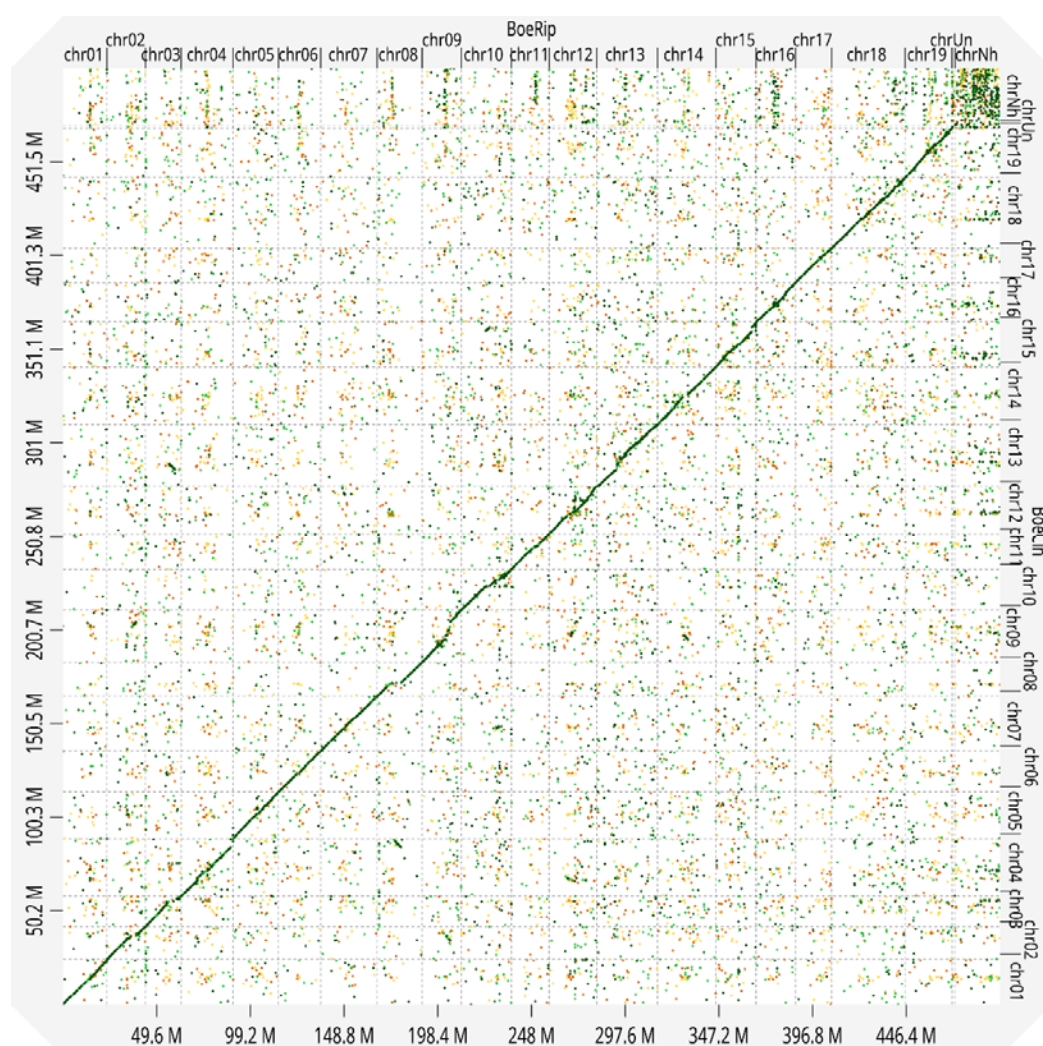
For validation of the phasing, bacterial artificial chromosome (BAC) sequences (Supplementary File 2 Table S3) with known haplotype/parental origin were mapped on BoeRip and BoeCin (Supplementary File 2 Table S4). The BACs were selected to cover regions on chr01 and chr14 and were sequenced with a read coverage of about 1,000-fold (Supplementary File 2 Table S5). BAC contigs of the same haplotype map

181 with almost no sequence difference, while BAC contigs derived from the other haplo-
 182 type display a wide range of mismatches and InDels (Supplementary File 1 Fig. S2).
 183 On both haplotypes, ~3.6 Mbp (chr01) and ~5.5 Mbp (chr14) were covered by BAC
 184 contigs of both phases and support a correct phasing over 9 Mbp.
 185 In addition, the assembly quality evaluation tool Merqury (phasing assessment for
 186 genome sequence assemblies) was used to assess the quality of both haplotype assem-
 187 blies. The overall base quality value (QV, consensus quality value, representing the
 188 log-scaled probability of error for consensus base calls) of 37.42 found for BoeRC
 189 indicates a very high level of correctness (see Supplementary File 2 Table S2 for
 190 comparisons to other assemblies). It should be noted that Merqury requires Illumina
 191 data for the complete trio (both parents and F1 as available for ‘Börner’ and its ances-
 192 tors) to calculate the complete set of quality values.
 193 Also the phasing accuracy was evaluated with Merqury (Supplementary File 1 Fig.
 194 S3). Based on haplotype-specific k-mers (referred to as hap-mers in Mercury), 2,759
 195 phase blocks with an N50 of 2.25 Mbp were calculated for BoeRip and 1,852 blocks
 196 with an N50 of 2.42 Mbp for BoeCin. Almost all bases were covered by a block and
 197 the switch error rate was estimated to be less than 0.05 % (Supplementary File 2 Ta-
 198 ble S6; Supplementary File 1 Fig. S4). Overall, the quality of the ‘Börner’ genome
 199 sequence as well as its separation into haplotypes derived from *V. riparia* GM183
 200 (BoeRip) and *V. cinerea* Arnold (BoeCin) is considered to be high.

201 **Sequence similarities and variations between BoeRip, BoeCin, PN40024 and** 202 **VitRGM**

203 All-versus-all alignments between the pseudochromosomes of the two ‘Börner’ haplo-
 204 types, and between each haplotype and PN40024 as well as VitRGM, were computed
 205 and visualised as dot plots (Fig. 3, Supplementary File 1 Fig. S5 to S8). All homolo-
 206 gous pseudochromosomes show strong synteny, yet some rearrangements and smaller
 207 and larger gaps indicating insertions, deletions and/or missing sequence were de-
 208 tected.
 209 On average 90.20 % of all bases with an identity of 95.52 % were aligned between the
 210 pseudochromosomes of the two haplotypes except of chrUn and chrNh (see Methods).
 211 The SNP and InDel frequency for the protein coding fraction was 1/1,001 bp and 1/34
 212 bp for the non-coding fraction of the genome sequence. When aligning with

213 PN40024, on average 88.59 % and 87.51 % of all bases were aligned with an identity
 214 of 94.82 % (BoeRip) and 94.94 % (BoeCin) over all pseudochromosomes, respec-
 215 tively. The alignments resulted in a SNP and InDel frequency of 1/932 bp (BoeRip)
 216 and 1/935 bp (BoeCin) for the coding regions and of 1/32 bp for the non-coding re-
 217 gions.
 218 An alignment of BoeRip and BoeCin with VitRGM resulted in 93.22 % (BoeRip) and
 219 88.45 % (BoeCin) aligned bases with an identity of 98.53 % (BoeRip) and 95.29 %
 220 (BoeCin). The SNP and InDel frequency was 1/3,128 bp (BoeRip) and 1/1,019 bp
 221 (BoeCin) for coding regions and 1/94 bp (BoeRip) and 1/35 bp (BoeCin) for non-
 222 coding regions.



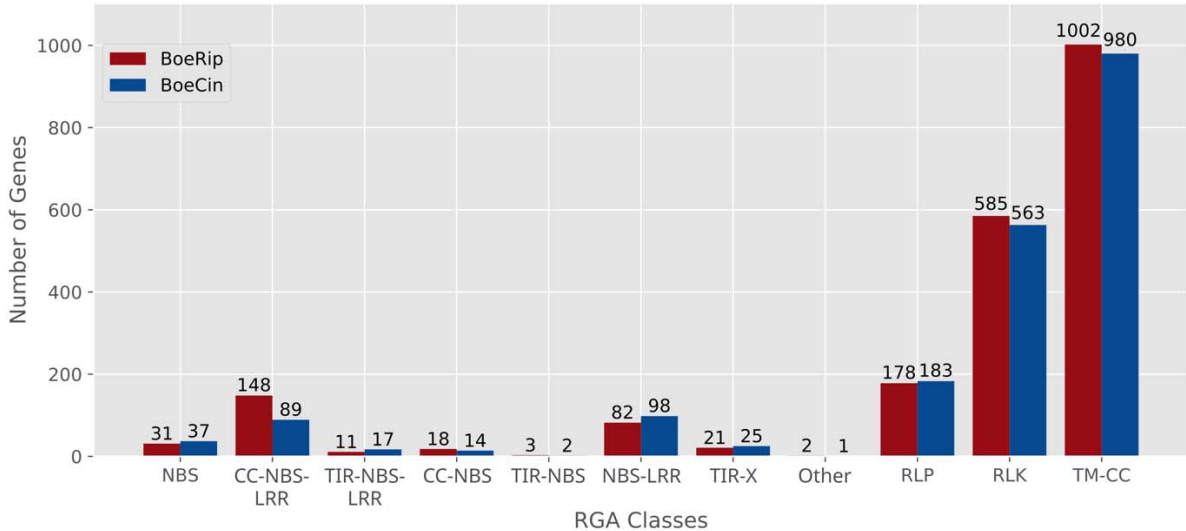
223
 224 **Fig. 3 All-versus-all dot plot of both ‘Börner’ haplotypes.** The graphic shows the
 225 dot plot between concatenated pseudochromosomes of BoeRip and BoeCin.

226 **Gene annotation**

227 In total, ~274 Mbp (55.24 %, BoeRip) and ~278 Mbp (55.46 %, BoeCin) repetitive
228 sequences were identified with the MAKER pipeline (Supplementary File 2 Table
229 S7). The final gene annotation comprises 27,738 and 27,995 protein-coding genes
230 (see Data availability for links to the genome sequence annotation of the *V. cinerea*
231 and the *V. riparia* haplotypes) and 525 and 419 tRNA genes for BoeRip and BoeCin,
232 respectively (Supplementary File 2 Table S7). Of the protein-coding genes, 19,500
233 genes of each haplotype were captured in RBHs. Based on RBHs, BoeRip shares
234 17,922 genes with VCost.v3 (PN40024) and 18,928 genes with VitRGM; BoeCin
235 shares 17,837 genes with VCost.v3 (PN40024) and 18,126 genes with VitRGM.

236 **Resistance genes**

237 The annotated genome sequence of ‘Börner’ was used to identify RGAs based on
238 their typical protein domain structure. Resistance gene annotation with RGAugury
239 revealed 2,081 RGAs for BoeRip and 2,009 RGAs for BoeCin (Fig. 4). Of these,
240 1,509 were classified as RBHs between BoeRip and BoeCin.



242 **Fig. 4 Distribution of genes to RGA classes based on characteristic protein**
243 **domains.** The red bars display the number of RGA genes of BoeRip and the blue bars
244 the RGA genes of BoeCin. RGAs were classified according to the domains encoded
245 by the predicted genes. Designations were adapted according to ²⁹. NBS, Nucleotide
246 Binding Site; CC-NBS-LRR with CC for Coiled-Coil and LRR for Leucine Rich
247 Repeat; TIR-NBS-LRR with TIR for Toll/Interleukin-1 Receptor like; TIR-X with X
248 for unknown domain; Other including TIR-CC-NBS-LRR or TIR-NBS-LRR-CC-TM

249 with TM for TransMembrane region; RLP, Receptor Like Protein; RLK, Receptor
250 Like Kinase.

251 **Candidate genes for *Rdv1***

252 At the sequence level, the region between GF13-01 and GF13-11 that contains *Rdv1*
253 displays a highly diverged structure between BoeRip, BoeCin, PN40024 and VitRGM
254 (Fig. 5A, Supplementary File 1 Fig. S9, S10).

255 The interval flanked by these genetic markers includes 2,978,547 bp with 179 genes
256 for BoeRip and 2,538,570 bp with 173 genes for BoeCin, according to the BoeRC
257 assembly and its annotation, as well as 2,796,801 bp with 273 genes for PN40024 and
258 5,854,243 bp with 450 genes for VitRGM according to published data. The interval
259 between GF13-01 and GF13-11 in VitRGM, which has been generated with a FAL-
260 CON-Unzip and which is representing a pseudo-haplotype, is twice as large as in the
261 other assemblies.

262 To further delimitate *Rdv1*, new genetic markers were designed and used to establish
263 a local map. F1 individuals with different flanking haplotypes were selected from the
264 mapping population V3125 x 'Börner', genotyped and tested for phylloxera resis-
265 tance. Since marker assay design tuned out to be difficult for gene-level resolution,
266 genotyping-by-sequencing (GBS) was applied to five crucial F1 genotypes. Mapping
267 of SSR and GBS markers as well as the presence or absence of the resistance encoded
268 by BoeCin led to delimitation of *Rdv1* to a region between the markers GBS-04 and
269 GBS-11 (Fig. 5B-D, Supplementary File 2 Table S8). The GBS markers were used to
270 precisely locate recombination sites that were identified between SSR markers. Since
271 the exact positions of the GBS markers are unknown for PN40024 and VitRGM, the
272 markers GF13-04.2 and GF13-26, located very close to the markers GBS-04 and
273 GBS-11, were used to specify the interval size.

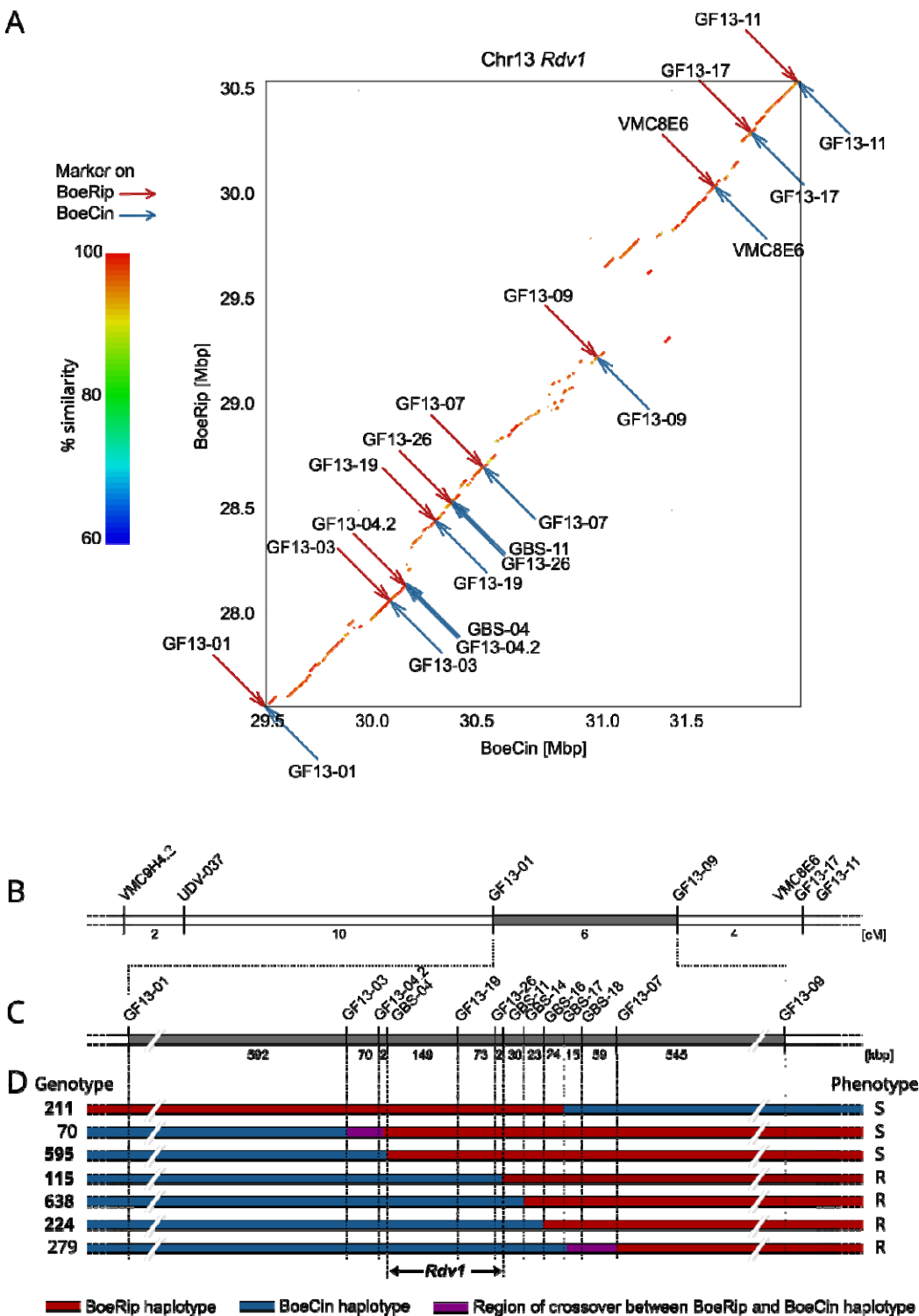


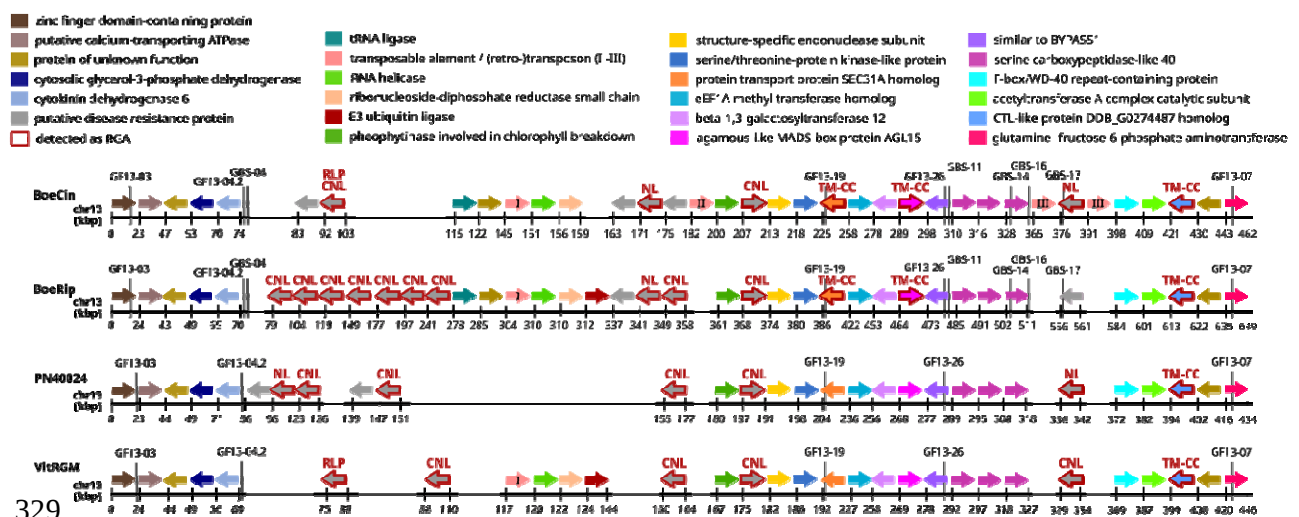
Fig. 5 Integration of genetic map and haplotype-specific physical map of the *Rdv1* region on pseudochromosome 13. (A) Dot plot of the *Rdv1* region between BoeRip and BoeCin. Genetic markers mapping on BoeRip are represented as red arrows and markers mapping on BoeCin as blue arrows. The *Rdv1* locus is located

279 between GBS-04 and GBS-11. **(B-D)** Fine mapping of the phylloxera resistance locus
 280 *Rdv1*. **(B)** Genomic region of the genetically mapped *Rdv1* locus on chromosome 13
 281 of ‘Börner’. The grey bar represents the interval between the delimiting SSR markers
 282 GF13-01 and GF13-09⁶, the numbers indicate the genetic distance between adjacent
 283 markers in cM. **(C)** Physical position of the SSR and GBS markers used for fine
 284 mapping of this region. The bar depicts the enlarged region taken from (B). Numbers
 285 refer to the physical distance of adjacent markers in kbp based on BoeCin (haplotype
 286 conferring resistance). **(D)** Local map of seven relevant F1 individuals from the
 287 V3125 x ‘Börner’ population revealed after haplotype-specific genotyping, only the
 288 ‘Börner’ haplotypes are shown. F1 genotypes analysed by GBS are highlighted in
 289 bold. The phenotypes are indicated on the right side: R, resistant; S, susceptible. Red
 290 and blue bars indicate the BoeRip and BoeCin haplotypes, respectively. Regions with
 291 unlocalised recombination sites are shown in purple.

292 Based on the gene annotation data generated for the assemblies, the RGA studies and
 293 intensive manual curation (see Methods), the genetically delimited *Rdv1* locus was
 294 evaluated in detail at the gene level (Fig. 6). The locus has a size of 223,738 bp with
 295 20 genes and eight potential resistance genes in BoeCin, a size of 402,698 bp with 25
 296 genes and 13 resistance genes in BoeRip, a size of 202,484 bp with 15 genes and
 297 seven resistance genes in PN40024 and 208,610 bp with 16 genes and four resistance
 298 genes (all detected as RGAs) in VitRGM. Complete gene structures were counted as
 299 "normal" genes even if the gene belongs to a TE. The *Rdv1* locus is covered by a sin-
 300 gle contig in both ‘Börner’ haplotypes.

301 A resistance gene cluster that displays a very divergent structure and different gene
 302 content among the three analysed haplotypes and the pseudo-haplotype VitRGM is
 303 present between the cytokinin dehydrogenase encoding gene (dark blue in Fig. 6, an-
 304 notation ID *BoeCin13g18380*, tagged by GF13-04.2/GBS-04) and the BYPASS1-
 305 related gene (purple in Fig. 6, annotation ID *BoeCin13g18400*, tagged by GF13-
 306 26/GBS-11). The haplotype conferring dominant resistance (BoeCin) contains eight
 307 potential resistance genes (Supplementary File 2 Table S9). Three of them, located in
 308 the southern part of the *Rdv1* locus (*BoeCin13g18393*, CC-NBS-LRR class;
 309 *BoeCin13g18396*, TM-CC class; and *BoeCin13g18399*, TM-CC class), are located in
 310 a syntenic region that contains similar alleles of the genes detected in all four haplo-
 311 types (see Discussion). In the following, the distinction between true haplotype se-
 312 quences (PN40024, BoeRip and BoeCin) and the pseudo-haplotype VitRGM, that
 313 may contain merged sequences from the two phases of *V. riparia* Gloire de Montpel-

314 lier, is neglected. Due to presence of these three genes in susceptible genotypes with
315 highly similar alleles, they were considered not to be relevant for phyloxera resis-
316 tance. The remaining five potential resistance genes (*BoeCin13g18381*;
317 *BoeCin13g18382*, CC-NBS-LRR/RLP; *BoeCin13g18388*; *BoeCin13g18389*, NBS-
318 LRR; *BoeCin13g18390*) can be considered as candidate genes for causing phyloxera
319 resistance. *BoeCin13g18381*, *BoeCin13g18382* and *BoeCin13g18389* are putative
320 disease resistance genes encoding RPP13-like proteins. The other two
321 (*BoeCin13g18388*, *BoeCin13g18390*) are similar to the *Arabidopsis thaliana* resis-
322 tance gene *AtLRRAC1*. To visualise the similarities between the protein-coding genes,
323 especially the putative disease resistance genes among and within the haplotypes, a
324 similarity matrix was calculated between *BoeCin* and itself as well the three other
325 haplotypes (Supplementary File 2 Table S10 to S13). The similarities detected con-
326 firm the relations mentioned above, including the relatedness of, for example, the
327 similarity of the putative disease resistance proteins encoded by *BoeCin13g18388* and
328 *BoeCin13g18390*.



330 **Fig. 6 Genes of *BoeCin*, *BoeRip*, *PN40024* and *VitRGM* at and surrounding the**
331 ***Rdv1* locus.** The genes (alleles) were clustered according to similarity and are shown
332 in one colour for each orthogroup. The grey box represents the *Rdv1* locus delineated
333 by the genetic markers GF13-04.2/GBS-04 and GF13-26/GBS-11; additional markers
334 are included upstream and downstream of the *Rdv1* locus. The kbp values on the axis
335 were adapted such that the start of the first gene is bp zero (relative coordinates). At
336 the top, the colour code for each group of related genes/alleles and the corresponding
337 functional annotation is given. Grey coloured genes are potential resistance genes
338 according to their functional annotation; genes with a red edging were identified as
339 RGA (see text). Above each RGA, the domain type classification of the encoded

protein is mentioned. See legend to Fig. 4 for acronyms, CC-NBS-LRR abbreviated as CNL, NBS-LRR as NL. Three genes qualified as TE genes of different types, these are marked with roman numbers; I: mutator-like element (MULE), II: similar to transposon TX1 protein, III: similar to retrovirus-related polymerase polyprotein.

Discussion

Final ‘Börner’ genome sequence assembly

For the fully phased genome assembly, the sequence reads were binned into parental subsets prior to assembly. The binning process produced two main read sets, both containing approximately half of all reads. An additional small read set with relatively short reads that were probably derived from homozygous regions was also generated. The binning process provided an excellent basis for generating high quality, haplotype-separated genome assemblies, with sufficient sequence coverage. Two fully phased genome assemblies of the grapevine cultivar ‘Börner’ were computed. About 90 % of all bases were included in contigs equally or larger than 530 kbp (BoeRip) and 450 kbp (BoeCin). In contrast to other published grapevine genome assemblies, most of them generated with FALCON and FALCON-Unzip, the genome assembly size was not over- or underestimated by Canu and thus reaches for both haplotypes approximately 500 Mbp (Supplementary File 2 Table S2). As noted before^{14, 19, 24}, FALCON-Unzip sometimes overestimates the primary contigs and underestimates the haplotigs because it orders some alternative sequences of heterozygous regions to the primary contigs. Regarding genome assembly, TrioCanu/Canu^{28, 30} turned out to be a very good choice for generating the genome sequence assembly BoeRC of the interspecific grapevine cultivar ‘Börner’. The N50 value of both ‘Börner’ haplotype assemblies outreaches most of the other *Vitis* genome assemblies. Comparing the N50, QV and overall k-mer completeness with some available chromosome level *Vitis* assemblies, BoeRC has with distance the highest N50 value, the highest quality values (estimated accuracy of 99.9998 %) and the highest degree of completeness (98.97 %) (Supplementary File 2 Table S2). To further demonstrate that the ‘Börner’ haplotype assemblies are fairly complete, the assemblies were investigated for the plant core genes. As in other high-quality genome sequence assemblies of grapevine like that of Cabernet Sauvignon²⁰, more than 98 % of complete core genes (BUSCOs) were identified.

372 **Phasing**

373 The haplotype assemblies can be considered as fully and correctly phased. Through
374 mapping of phased BAC sequences derived from parts of chr01 and chr14 of
375 ‘Börner’, it was shown that several Mbp of chr01 and chr14 are truly phased. The
376 haplotype-specific BAC sequences coincide with the correct haplotype and share
377 plenty of mismatches with the other haplotype. Also, the mapping validated correct-
378 ness of contiguity and chromosome assignment in those regions.

379 Additionally, the phasing was analysed through k-mer analysis with Merqury. The k-
380 mer completeness of BoeRC over both haplotypes was rated to 99.04 % (BoeRip) and
381 99.51 % (BoeCin), while haplotype precision was estimated to 99.07 % (BoeRip) and
382 99.92 % (BoeCin) (Supplementary File 2 Table S2). Thus, the haplotype assemblies
383 can be considered as completely and correctly phased. The purity of the haplotypes
384 was also expressed through the very low switch error rate of 0.046 % (BoeRip) and
385 0.035 % (BoeCin) and the high block N50 value of 2.25 Mbp (BoeRip) and 2.42 Mbp
386 (BoeCin). None of the scaffolds showed a notable amount of k-mers from the other
387 haplotype (Supplementary File 1 Fig. S2).

388 There would potentially be room for improvement of the assemblies BoeRip and
389 BoeCin by methods like chromosome conformation capture sequencing (Hi-C). Even
390 if there are indications that the organization of chromosomes into territories within the
391 nucleus of diploid organisms could allow to extract data supporting correct haplotype
392 separation³¹, we did not follow this route and anticipate that the quality of the BoeRC
393 assembly is sufficient for almost any application with relevance to viticulture.

394 **Similarity of ‘Börner’ haplotypes and comparison with PN40024 and VitRGM**

395 The assemblies BoeRip and BoeCin show overall similarity to each other and to
396 PN40024. Even if the haplotypes among themselves have almost the same SNP fre-
397 quency as with PN40024, more precisely 1 SNP in 33 bp compared to 1 SNP in 31
398 bp, more bases were aligned between the haplotypes with a higher identity than with
399 PN40024. The difference in SNP frequency becomes more specific when comparing
400 the SNP frequency of the coding regions. Here, BoeRip and BoeCin show less SNPs
401 (1/1,001 bp) than with PN40024 (1/932 bp BoeRip & 1/935 bp BoeCin). Because of
402 the high heterozygosity between haplotypes of grapevine³², it is not surprising that
403 the SNP frequency between ‘Börners’ haplotypes is almost the same as to another

404 *Vitis* species. Reported SNP frequencies for grapevines range from 1 SNP per 64 bp³³
 405 to 1 SNP per 200 bp¹⁷. The SNP frequency between *V. riparia* and *V. vinifera* was
 406 determined to 1 SNP per 78 bp³⁴. Generally, it seems that SNP frequencies between
 407 and within *Vitis* species are not that different. However, the SNP frequency between
 408 BoeRip and VitRGM, both representing *V. riparia* genome sequences, was signifi-
 409 cantly lower than between BoeCin and VitRGM, for coding as well as non-coding
 410 regions. This may indicate that *V. riparia* GM183 and *V. riparia* Gloire de Montpel-
 411 lier were derived from related populations, or that the species *V. riparia* generally
 412 covers less variation.

413 High sequence identities and a good chromosome assignment were revealed through
 414 dot plots between homologous pseudochromosomes of the haplotypes of BoeRC and
 415 PN40024 or VitRGM. The "point clouds" often observed in the dot plots between
 416 homologous pseudochromosomes can be assumed to pinpoint the location of the cen-
 417 tromere as centromeric regions are often highly repetitive and thus hard to sequence
 418 and assemble. The gaps detected in comparison with PN40024 may represent missing
 419 bases in the PN40024 genome sequence assembly. Almost all identified rearrange-
 420 ments between the haplotypes of BoeRC and PN40024 were found between both hap-
 421 lotypes and PN40024, and most of them are inversions. This calls for an improvement
 422 of the 'PN40024' reference genome sequence based on long read sequence data.

423 **Rdv1 locus**

424 The *Rdv1* locus has been genetically mapped to a region of 224 kbp in size in the
 425 BoeCin haplotype of the 'Börner' genome. The genome regions orthologous to the
 426 *Rdv1* locus from BoeCin and different *Vitis* species are flanked on both sides by large
 427 syntenic regions. The high variability and divergence that is known for resistance
 428 gene clusters³⁵ was also detected within the *Rdv1* locus.

429 In syntenic regions, alleles may vary significantly in lengths between haplotypes, still
 430 they encode the same protein even among different *Vitis* species. For example, the
 431 PN40024 gene *Vitvi13g01613* which encodes a putative glycerol-3-phosphate dehy-
 432 drogenase (dark blue gene in Fig. 6, northern part of the locus), is 3,600 bp longer
 433 than its BoeCin ortholog. The 466 amino acid long protein sequence is identical be-
 434 tween all haplotypes analysed except one amino acid substitution each in PN40024
 435 and BoeRip. Increased gene length of *Vitvi13g01613* in PN40024 is caused by a lar-

ger third intron. Several TE fragments and especially one larger LTR/Copia-like TE explain the difference in the length of intron three.

In addition to SNPs, quite some presence/absence variation (PAV) contributes to the divergence of the *Rdv1* region in *Vitis*. The various PAVs were, at least in several cases, explained by detection of TEs (either DNA transposons or retrotransposons) in the haplotypes (Fig. 6, Supplementary File 1 Fig. S11). For example, two intact LTR/Copia TEs and two intact DNA/hAT TEs were detected in the *Rdv1* region of BoeRip in the context of the tandem array of seven RGAs of the CC-NBS-LRR (CNL) type that is only present in BoeRip. Another genic PAV is the E3 ubiquitin ligase coding gene (*BoeRip13g18552*) that was detected in BoeRip and VitRGM which might be specific for *V. riparia*, although this hypothesis needs to be tested with more data from *V. riparia*. The resistance gene *BoeCin13g18388* encodes a LRR domain and is about 2,400 bp longer in BoeCin than in BoeRip (*BoeRip13g18553*). The reason is again an intron in the 5'UTR region of BoeCin which includes an LTR/Gypsy TE. This TE insertion causes a shifted translation start and the protein of the BoeCin allele is 49 amino acids longer at the N-terminus than the protein encoded by the BoeRip allele. In case of the MULE transposon (TE with similarity to mutator-like elements, marked with I in Fig. 6), a gene coding for a transposase protein was identified in BoeCin, BoeRip and VitRGM. There are hints from the TE annotation (Supplementary File 2 Table S7) that the MULE transposon might contain two ORFs in opposite orientation which is characteristic for some MULEs. However, the terminal inverted repeats were not identified. Because of the great diversity within the same and between transposon superfamilies, generally and also with regard to the structure and length of MULEs^{36, 37}, further work is needed to precisely identify all TEs and TE fragments that contribute to PAVs at this and other loci.

The *Rdv1* region is characterized by one extended resistance gene cluster. Orthologous regions in the other studied haplotypes derived from different *Vitis* species display very diverse cluster structures with respect to length in kbp and number of resistance genes. In the BoeCin haplotype that confers the dominant resistance to phylloxera and which segregates in offspring of 'Börner', the cluster has a length of about 350 kbp and includes 10 potential resistance genes. The cluster extends at the southern end beyond the genetically delimited *Rdv1* locus. In the BoeRip haplotype that

468 confers recessive susceptibility to phylloxera, the cluster includes 14 potential resis-
469 tance genes.

470 In both ‘Börner’ haplotypes as well as in VitRGM, the cluster shows an insertion of a
471 conserved array of four to six genes compared to PN40024 representing *V. vinifera* in
472 the comparison. The inserted genes (or the genes deleted in PN40024) encode a tRNA
473 ligase, a protein of unknown function, a mutator-like element (MULE) DNA transpo-
474 son, a DEAD-box ATP-dependent RNA helicase, a ribonucleoside-diphosphate re-
475 ductase small chain and a ubiquitin ligase. These genes display no features of resis-
476 tance genes and are found with high identity values in the susceptible haplotype
477 BoeRip. They are, therefore, no candidate genes for *Rdv1*.

478 Within the southern part of the locus, RGAugury detected two RGAs with TM-CC
479 domains in the encoded proteins that are potential false positive results (orange and
480 pink in Fig. 6, with red edging in BoeRip and BoeCin due to the RGAugury output).
481 Based on the functional annotation (Supplementary File 2 Table S9), the two or-
482 thogroups code for homologs of SEC31A (a component of the coat protein complex II
483 (COPII) involved in the formation of transport vesicles) and homologs of AGL15
484 (agamous-like MADS-box factor potentially involved in control of development). The
485 genes in the syntenic block that starts at its northern end with the homolog of *SLX1*
486 (encoding a subunit of a structure-specific endonuclease complex involved in process-
487 ing diverse DNA damage intermediates, light ochre in Fig. 6) and which extends to
488 the south beyond the delimiting markers GF13-04.2/GBS-04, are also very unlikely to
489 be candidates for *Rdv1*.

490 To the north, the synteny extends for two more genes and terminates at a TE gene
491 annotated as TX1-like non-LTR retrotransposon (*BoeCin13g18391*, marked II in Fig.
492 6) that is only detected in BoeCin. Southern to this TX1-like TE, a syntenic or-
493 thogroup (including *BoeCin13g18393* in BoeCin) encodes a CC-NBS-LRR (CNL)
494 resistance protein related to the wild potato resistance protein RGA4. Since this gene
495 is represented by closely related alleles in BoeCin and the haplotypes conferring sus-
496 ceptibility (amino acid sequence identity: BoeRip 99.9 %, PN40024 97.1 %), it is not
497 considered to be a candidate for *Rdv1*.

498 Based on this evaluation, the most promising genes mediating phylloxera root resis-
499 tance are the five resistance genes northern and southern to the array of four to six
500 genes conserved in BoeCin, BoeRip and VitRGM (*BoeCin13g18381*;

501 *BoeCin13g18382*, CNL/RLP; *BoeCin13g18388*; *BoeCin13g18389*, NL;
502 *BoeCin13g18390*). Ideally, an F1 genotype from the V3125 x 'Börner' cross with a
503 recombination site between the *BoeCin* and *BoeRip* haplotypes should have been
504 identified. However, such an event was not detected. It is possible that the high diver-
505 gence and sequence dissimilarity throughout the *Rdv1* locus causes a suppression or at
506 least reduction of recombination events at the locus, which could explain this failure.
507 A reduction of recombination frequency was, for example, also described for the
508 *Rpv3.1* locus that confers resistance to *P. viticola* in *Vitis*³⁸.
509 The five remaining resistance genes can be divided into two types with respect to
510 functional annotation. The resistance genes *BoeCin13g18388* and *BoeCin13g18390*
511 were not detected as RGAs by RGAugury, but encode proteins similar to the *A.*
512 *thaliana* resistance gene *AtLRRAC1* (*At3g14460*). *AtLRRAC1* has been reported to be
513 involved in the defence against the biotrophic fungus *Golovinomyces orontii* as well
514 as the hemi-biotrophic bacteria *Pseudomonas syringae*, and might be relevant for sig-
515 nalling via cAMP-dependent defence pathways³⁹. The three other resistance genes,
516 namely *BoeCin13g18381*, *BoeCin13g18382* CC-NBS-LRR/RLP and
517 *BoeCin13g18389* NBS-LRR (NL), encode proteins similar to *RPPL1* (Recognition of
518 *Peronospora Parasitica* 13-Like 1). *AtRPPL1* (*At3g14470*) received its annotation
519 from *AtRPP13* (*At3g46530*), a gene that confers resistance to the biotrophic oomycete
520 *Peronospora parasitica* and which encodes an NBS-LRR protein^{40, 41}. In *A. thaliana*,
521 *AtLRRAC1* and *AtRPPL1* form a cluster of two neighbouring genes, a feature that is
522 conserved in *Vitis* (Fig. 6). It will be interesting to figure out if the conserved co-
523 localisation of two different resistance genes has functional implications for resistance
524 to phylloxera in *Vitis* species.
525 In *BoeCin* and *BoeRip*, one of the resistance genes is classified as encoding an NBS-
526 LRR (NL) protein. Despite an identical NBS domain with only very few substitutions,
527 the LRR region varies greatly between the two alleles (*BoeCin13g18389* and
528 *BoeRip13g18554*) and is shorter in *BoeRip* because of an early stop codon. Also for
529 the other four candidate resistance genes, differences between the alleles in *BoeCin*,
530 *BoeRip* and PN40024 were detected. For example, the protein encoded by
531 *BoeCin13g18382* was detected as RGA based on CC-NBS-LRR domains in addition
532 to features of an RLP. However, the genes detected at syntenic positions (alleles) of
533 *BoeRip* and PN40024 lack these features either in part or fully.

Since the PN40024 sequence is derived from the most susceptible line in our comparison, the protein sequences of the five *Rdv1* candidate resistance genes of BoeCin were compared with all PN40024 protein sequences. No identical PN40024 protein sequence or proteins with identical resistance domain sequences were identified. Thus, the five BoeCin resistance genes are promising candidate genes for conferring resistance to phylloxera.

Conclusions

The fully phase-separated genome sequence assembly BoeRC from the phylloxera-resistant rootstock ‘Börner’, and its structural and functional gene annotation, build a cornerstone for the investigation of loci from ‘Börner’ that are linked to various valuable traits. Here, the focus was on phylloxera resistance. The sequences of the haplotypes BoeCin conferring resistance to phylloxera and BoeRip conferring susceptibility allowed to precisely map recombination sites by GBS in crucial genotypes of a mapping population of V3125 x ‘Börner’. Subsequently, detailed examination and comparative genomics of the gene content of the *Rdv1* locus allowed to delimit the locus to a region of 123 kb in BoeCin haplotype with five resistance genes as potential candidates involved in mediating phylloxera resistance at the root. The resources generated will allow to study additional traits and have the potential to support resistance breeding for the benefit of viticulture.

Methods

Reference data sets

In this study, the ‘PN40024’ genome sequence¹⁶ assembly 12X.v2 (urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences) and its annotation, the VCost.v3 gene annotation (<https://urgi.versailles.inra.fr/Species/Vitis/Annotations>)⁴² as well as the *V. riparia* Gloire de Montpellier genome sequence assembly VitRGM and its annotation (https://ncbi.nlm.nih.gov/assembly/GCF_004353265.1)²⁴ were used for comparisons.

Plant material and DNA and RNA extraction

‘Börner’ is listed in the *Vitis* International Variety Catalogue with the variety number VIVC 1499, its parents *V. riparia* GM183 and *V. cinerea* Arnold with variety numbers VIVC 4686 and VIVC 13645, respectively. All three cultivars do not belong to an endangered species and were obtained and are grown at the Institute for Grapevine Breeding Geilweilerhof at Siebeldingen (JKI Siebeldingen) in accordance with German legislation.

Dormant wood cuttings of the interspecific rootstock variety ‘Börner’ were raised in a growth chamber on soil at 18°-24°C (64°F-75°F), 70 % relative humidity and 16 h light and 8 h darkness. Prior to harvest the plants were cultivated in the dark for 72 hours. Small, young leaves were harvested and stored on ice until use. Extraction of high molecular DNA for SMRT sequencing was performed starting with 1.5 g fresh weight and was carried out with the NucleoSpin Plant II Maxi kit for DNA (Macherey and Nagel, Düren; Germany) according to the recommendations of the manufacturer for plants. DNA integrity, quality and quantity was checked using a TapeStation (Agilent) and was found to peak at a fragment size higher than 48.5 kbp.

Leaf material of ‘Börners’ parents was harvested from cuttings grown in the green house, frozen in liquid nitrogen and stored at -80°C. The DNA extraction was carried out with the DNeasy Plant Maxi Kit (Qiagen, Hilden; Germany) according to the instructions of the manufacturer. Tissue samples of leaves and tendrils of ‘Börner’ were collected from grown in the field. Total RNA was isolated using the Spectrum Plant RNA-Kit (Sigma, Taufkirchen; Germany).

Plant material for the local map of *Rdv1* was selected from the previously described mapping population V3125 x ‘Börner’^{6, 43} and another set of 310 F1 genotypes obtained after repeating the cross in 2006. To search for recombinant F1 lines, SSR markers associated with *Rdv1* (Supplementary File 2 Table S14) were used for genotyping as described⁴³. For GBS, young leaf material was harvested from selected individual F1 genotypes as well as from the maternal genotype V3125 and used for DNA extraction. Genomic DNA was prepared using an established CTAB-based protocol⁴⁴. Subsequently, the DNA obtained was treated with RNase and quantified using PicoGreen.

Library construction for ‘Börner’ genome and transcriptome sequencing

PacBio Sequel libraries were prepared according to the SMRTbell Template Prep Kit 1.0 and then size-selected with BluePippin for a target insert size of greater than 10 kb. Seventeen 1Mv3 SMRT cells were run on a Sequel I sequencer using the Binding Kit 1.0 and the Sequencing Chemistry version 1.0 (all from PacBio) for six hours per cell. The output BAM files of the sequencing step were loaded into SMRT Link 5.0.1 according to the recommendation of the manufacturer (PacBio Reference Guide 2018) and converted to FASTA format for downstream processing. The length distribution over all reads was calculated with the assembler Canu³⁰ and is provided in Supplementary File 1 Fig. S12.

The RNA samples from leaves and tendrils of ‘Börner’ were paired-end sequenced using Illumina technology on a HiSeq-2000 essentially as described⁴⁵. The new RNA-Seq read data (ENA/SRA study accession no. PRJEB46079) were processed together with data submitted before⁴⁵. These already existing RNA-Seq reads of ‘Börner’ (PRJEB34983) cover leaves (ERR3894001), senescent leaves (winter leaves, ERR3895010), inflorescences (ERR3894002), tendrils (ERR3894003) and roots (ERR3895007). Read data were trimmed with Trimmomatic-0.38⁴⁶ allowing reads larger than 80 nt (new data) and 90 nt (published data) to be kept. The trimming statistics of the RNA-Seq data is provided in Supplementary File 2 Table S15.

Illumina sequencing of the parental genomes of ‘Börner’

Library preparation for the *V. riparia* GM183 and *V. cinerea* Arnold samples was performed according to the Illumina TruSeq DNA Sample Preparation v2 Guide. Genomic DNA (1500 ng each) was fragmented by nebulisation. After end repair and A-tailing, individual adaptors were ligated to the fragments for PE sequencing. The adaptor ligated fragments were purified on a 2 % low melt agarose gel and size selected by excising a band ranging from 500-800 bp. After enrichment PCR of DNA fragments that carry adaptors on both ends, the final libraries were quantified by using the Quant-iT PicoGreen dsDNA assay on a FLUOstar Optima Platerreader (BMG Labtech) and qualified on a BioAnalyzer High Sensitivity DNA Chip (Agilent). The *V. riparia* GM183 library was sequenced on a MiSeq PE run (2 x 250 nt) and on two lanes of a 2 x 100 PE run on a HiSeq-1500 in high output mode. The *V. cinerea* Arnold library was sequenced in one 2 x 150 bp PE run on a HiSeq-1500 in rapid mode.

624 The read data were submitted to ENA (ENA/SRA study accession no. PRJEB45595).
625 All genomic short read data were quality trimmed with Trimmomatic-0.36⁴⁶ allowing
626 reads equal or longer 80 nt to be kept. The trimming statistics is provided in Supple-
627 mentary File 2 Table S16.

628 **Sequencing of pools of BACs from a BAC library of ‘Börner’**

629 For the creation of phased sequences of ‘Börner’, 8 pools of BAC constructs contain-
630 ing a total of 440 mapped BACs from two loci (one on chr01, the other on chr14)
631 were selected from a ‘Börner’ BAC library¹⁰. The BACs were selected and distrib-
632 uted to the pools according to the positions of BAC end sequences on the correspond-
633 ing regions of PN40024. Mapping to the reference assembly was carried out with the
634 CLC Genomics Workbench v8.0 toolkit (<https://www.qiagenbioinformatics.com/>).
635 Pools were arranged by excluding overlap of BAC inserts to allow assembly of indi-
636 vidual BAC insert sequences. The BAC Pools1-4 were sequenced on a 454 Life Sci-
637 ences Genome Sequencer GS FLX and PE (2 x 250 nt) on the Illumina MiSeq plat-
638 form. The BAC Pools5-8 were sequenced on the MiSeq platform only. The read data
639 were submitted to ENA (ENA/SRA study accession no. PRJEB46081). The MiSeq
640 raw data were quality trimmed with Trimmomatic-0.32 using ‘ILLUMINACLIP:
641 2:40:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36’⁴⁶. The
642 trimming statistics for the BAC pools is provided in Supplementary File 2 Table S5.

643 **Library construction and sequencing for GBS**

644 The barcoded libraries were prepared using Illumina library preparation kits using
645 TrueSeq technology and sequenced on a NextSeq500. Sequencing was performed in
646 150 nt single-end (SE) modus and the read data were submitted to ENA. The
647 ENA/SRA study accession no. is PRJEB53997. Information on genotypes, sequenc-
648 ing strategy, data volume and run IDs are summarized in Supplementary File 2 Table
649 S17.

650 **Phenotyping of phylloxera resistance**

651 The test system used to assess phylloxera resistance of selected F1 individuals was
652 essentially described previously^{6, 47}. The phylloxera population used for the artificial
653 inoculation of potted plants was derived from leaves of naturally infested rootstock
654 cultivars from the germplasm repository at JKI Siebeldingen. The phylloxera resis-

655 tance tests started at the beginning of the vegetation season when enough infested leaf
656 material was available as inoculum. Leaf pieces with 8 to 12 (3-5 mm) galls contain-
657 ing mature eggs were inserted in the soil of rooted cuttings of the F1 individuals. The
658 inoculation procedure was repeated after 3 to 4 weeks. Approximately 4 weeks after
659 the second inoculation the plants were rated as susceptible if nodosities were visible,
660 or resistant if the roots developed normally. Each F1 genotype was tested in triplicate
661 and the assays were repeated at least twice.

662 **Diploid genome sequence assembly**

663 To compute a phased genome assembly, a binning and assembly approach was used.
664 The binning prior to assembly was performed with the tool TrioCanu of the assembler
665 Canu v1.7^{28, 30}. TrioCanu uses short reads of the parents and long reads of the F1 as
666 input. It bins the long reads of the F1 into parental subsets based on k-mer compari-
667 sons. The binning step was performed on a compute cluster using TrioCanu default
668 parameters (k-mer size of 20). Binning resulted in one read subset for each parent and
669 an additional small subset with reads that could not be assigned to either of the two
670 parental haplotypes, referred to as UAR subset. In the following, the parental subsets
671 are referred to as Vrip and Vcin read subset according to ‘Börners’ parents and the
672 two different haplotypes combined into ‘Börner’, respectively. The UAR subset con-
673 tained only 0.13 % of all bases with an average read length of 1.5 kbp.

674 After binning, the read subsets were used to compute the two haplotype assemblies
675 BoeRip and BoeCin of ‘Börner’ with the assembler Canu v1.7³⁰. To compute the
676 BoeRip haplotype assembly, the Vrip and the UAR subsets were used in order to in-
677 clude identical and/or homozygous regions. Likewise, the BoeCin haplotype assembly
678 was generated from the Vcin read subset and the UAR subset. Therefore, the basis of
679 the haplotype assemblies, BoeRip and BoeCin, are approximately 32 to 34 Gbp read
680 data with an estimated genome coverage of more than 60-fold each. For both sepa-
681 rately computed haplotype assemblies, Canu was run on a compute cluster utilizing
682 the parameters ‘genomeSize=550m’, ‘corMhapSensitivity=normal’, ‘correctedError-
683 Rate=0.065’, ‘canuIteration=1’ and ‘stopOnReadQuality=false’. During the assembly
684 processes about 2.5 % of the corresponding input data (bp) remained unassembled. A
685 search with blastn of the Basic Local Alignment Search Tool (BLAST) package+
686 v2.8.1^{48, 49} and the grapevine chloroplast⁵⁰ and mitochondrial sequences⁵¹ was car-

ried out on the haplotype assemblies and detected contigs were removed. Finally, the haplotype assemblies were polished twice with *arrow* v2.2.2 (smrtlink-release_5.1.0.26412) and the corresponding read subsets. Assembly statistics were computed with QUAST v4.6.3⁵².

691 Scaffolding with ‘Börner’ BAC end sequences

To validate the quality of the two genome sequence assemblies, and also to scaffold a few smaller contigs, high-quality paired BAC end sequences from the ‘Börner’ genome (ENA/GenBank accession numbers KG622866 - KG692309¹⁰) were used. The 69,444 BAC end sequences cover 39,360,203 bp, have an average sequence length of 566 bp, and represent an eight-fold coverage of the ‘Börner’ genome based on average BAC insert length. First, the 62,498 BAC end sequences that represented pairs from one BAC were mapped to BoeRip and BoeCin with HISAT2 v2.1.0⁵³ using the parameters ‘--maxins 2000000000’, ‘--secondary’, ‘--no-softclip’ and ‘--no-spliced-alignment’. Alignments with a mapping quality >10 were filtered out and alignment statistics were collected with SAMtools v1.8⁵⁴. After mapping and quality filtering, the BAC end sequences were used to scaffold the haplotype assemblies with SSPACE (standard) v3.0⁵⁵. SSPACE was run with the parameters ‘-x 0’, ‘-m 50’, ‘-k 3’, ‘-n 500’ and ‘-T 10’ and by using BWA-SW⁵⁶ for mapping. Contig extension was disabled (‘-x 0’) as the BAC end sequences are not phased and thus would eventually contaminate the haplotypes with a sequence from the other haplotype. However, a test run with enabled extension showed that the BAC end sequences were anyway not suitable for gap filling.

709 Scaffold to pseudochromosome assignment

To build pseudochromosomes (the nucleotide sequence representation of chromosomes), an *ad hoc* gene prediction was generated with AUGUSTUS v3.3⁵⁷. Prior to gene prediction, the RNA-Seq reads from ‘Börner’ (see above) were processed to create a *de novo* transcriptome assembly with Trinity v2.8.5⁵⁸ with default settings. Hint files with information about exon and intron positions for the haplotype assemblies were created through mapping all trimmed RNA-Seq reads and the transcriptome assembly on the haplotype assemblies. The RNA-Seq reads were mapped with HISAT2 v2.1.0⁵⁹ and soft-clipping disabled (‘--no-softclip’). The transcriptome sequences

were mapped with BLAT v36x2⁶⁰ and the parameters ‘-stepSize=5’ and ‘-minIdentity=93’. Hints were generated according to the AUGUSTUS documentation. Additionally, AUGUSTUS was trained with the CRIBI v2.1 gene annotation of PN40024¹⁷ to retrieve parameter sets for grapevine. Finally, genes were predicted on both haplotype assemblies BoeRip and BoeCin with the generated *Vitis* parameter set and the hint files.

The protein sequence of the primary transcript variant was extracted from the PN40024 CRIBI v2.1 annotation and from the *ad hoc* annotation of BoeRip and BoeCin. Reciprocal best BLAST hits (RBHs) between the protein sets of each of both haplotypes and the filtered proteins of PN40024 were computed with blastp utilizing an e-value threshold of 0.0001. For RBHs, two rounds of BLAST searches were performed. Here, one round was run with the proteins of one haplotype as input query and one with the PN40024 proteins as input query. Only RBHs with a percentage of identical matches and a query coverage per subject ≥ 80 % for at least one direction (e.g. PN40024 protein sequences as input query) were kept. Based on RBHs encoded, the scaffolds were assigned to the pseudochromosomes. Scaffolds with less than 10 RBHs in total and scaffolds that contain a significant number of RBHs linking a different pseudochromosome (30 % distance to the 2nd rank in terms of RBH numbers required) were filtered out for later assignment or moved to the chromosomally unassigned part of the respective assembly.

Additionally, reciprocal hits between the scaffold sequences and the pseudochromosome sequences of PN40024 were computed with blastn. The e-value, identity and query coverage filters remained the same as above. The nucleotide RBHs were iteratively computed and scaffolds assigned to pseudochromosomal positions. If the classifications based on protein and nucleotide level contradicted, the protein RBH classification was preferred.

To further refine the pseudochromosomes of the haplotypes, an additional assignment based on VitRGM was performed. Also, reciprocal hits between the scaffold sequences of one haplotype and the pseudochromosomes of the other haplotype and *vice versa* were calculated as described above. After manual construction of pseudochromosomes for the BoeRip and BoeCin phases, the refinement with the corresponding other haplotype was repeated until no further scaffolds could be assigned. The pseudochromosomes were constructed with gaps of 100 bp length between the scaffolds

and with 10 kbp as minimal scaffold length. Due to the length filter, 61 (BoeRip) and 22 (BoeCin) scaffolds were discarded.

The orientation and order of the scaffolds on the pseudochromosomes was verified with dot plots and manually adapted if necessary. Thus, DNAdiff v1.3 of the MUMmer package v4.0.0beta2⁶¹ was run pairwise on homologous pseudochromosomes of BoeRip or BoeCin and PN40024 with default settings and the resulting 1-to-1 alignments were visualized with mummerplot v3.5.

The pseudochromosomes were validated with 340 simple sequence repeat (SSR) genetic markers from⁴³ and the 4 SSR markers ATP3⁶², Gf13_11a, VMC8E6⁶ and Gf14-42⁸ (Supplementary File 2 Table S14). Primer sequences were mapped to the pseudochromosomes with primersearch of the EMBOSS v6.6.0.0 package⁶³ and with a blastn search. A total of 314 markers were assigned to unequivocal sequence positions, the remaining markers do not map, show non-evaluable multi-mappings or map with too many mismatches (Supplementary File 2 Table S18). Several markers did map only to one of the haplotypes (16 to only BoeRip and 30 to only BoeCin, see Supplementary File 2 Table S19). Emerging disagreements between the marker position on PN40024 and the BoeRip and BoeCin phases were further investigated through all-versus-all dot plots computed with the webtool D-Genies v1.2.0⁶⁴ and by using long read mapping to detect sequence positions that are not or only very weakly supported by continuously mapping reads. The haplotype specific read subsets together with the unassigned read subset were mapped with minimap2⁶⁵ v2.17 ('-ax map-pb --secondary=no') to the corresponding haplotype assembly and the coverage values per base were calculated with SAMtools. The haplotype assemblies are both covered at about 54x (Supplementary File 1 Fig. S13). The back-mapping results of the read subsets were consulted to reveal miss-assemblies if indicated by a genetic marker. A sequence location was further investigated if five reads end in a 10 bp region and if the read coverage five bp around the region drops to $\leq$ five.

Conflicting marker VVMD28 maps on chr03 of PN40024 and BoeCin, yet on chr13 of BoeRip. However, eight markers allocate the sequence to chr13 of BoeRip and neither a significant breakpoint was found in the read mappings nor in the all-versus-all dot plot (Supplementary File 1 Fig. S3). VCHR16B maps on chr16 of PN40024 and BoeCin, but on chr03 of BoeRip. Here, one SSR marker assigns the sequence to chr03 and no significant breakpoint was found in the read mapping or in the all-

versus-all dot plot. The three SSR marker VVS4, VMC5G6.1 and UDV-126 map on chr08 of PN40024 and BoeRip, but on chr04 of BoeCin. The corresponding sequence was assigned to chr04 according to 376 protein RBHs, because only 141 protein RBHs assign it to chr08. Moreover, three SSR markers allocate the sequence to chr04, too. Through investigating the all-versus-all dot plot, an approximately four Mbp large alignment between chr04 and chr08 was found. However, since no position for a split was detected in the read mappings, the sequence remained on chr04 (Supplementary File 1 Fig. S6). GF10-06, VMC3D7 and UDV-073 mapped wrongly on chr02 of BoeRip instead of chr10; in this case the investigation of the respective contig resulted in a split of the sequence and a corrected assignment of the contig fragment in question. Due to 348 protein RBHs the sequence was initially assigned to chr02. The sequence showed the second most protein RBHs (171) with chr10. On marker side, 13 markers assign the sequence to chr02 and only three to chr10. However, the all-versus-all dot plot showed a large alignment between chr02 of BoeRip and chr10 of PN40024 (Supplementary File 1 Fig. S14), and the read mappings indicated a breakpoint position with a coverage drop to ≤ 5 . Thus, the 12.3 Mbp sequence was split into a 7.7 Mbp large sequence that was assigned to chr10 and a 4.6 Mbp large sequence that remained at chr02. Another case was detected in the all-versus-all dot plots between chr13 of BoeCin and chr03 of PN40024 (Supplementary File 1 Fig. S6). The corresponding sequence was assigned to chr13 through 485 protein RBHs. Only 36 protein RBHs would allocate this sequence to chr03. Furthermore, 10 SSR marker support the assignment to chr13 and none to chr03. No coverage drop was found in the read mappings, the sequence remained on chr13 of BoeCin. Finally, only five genetic markers remained unresolved (VVMD28, VCHR16B, VVS4, VMC5G6.1, UDV-126).

To estimate the completeness of the assemblies, the plant core genes were localized with the program Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.2 utilizing the database ‘eudicots_odb10’ (2,326 genes)^{66, 67}. For comparison, the BUSCOs of PN40024 and of VitRGM were also determined with identical parameters. The pseudochromosome lengths were visualized with cvit v1.2.1⁶⁸.

814 Validation of phasing

815 To validate the phasing of the ‘Börner’ haplotypes, the BAC sequence data of pool1-4
 816 were *de novo* assembled with Newbler v2.6. The data of pool5-8 were assembled with
 817 CLC Genomics Workbench 8.0 using a word size of 30. Remaining vector sequences
 818 were removed with vectorstrip of the EMBOSS package v6.2. Assembled sequences
 819 shorter than 500 bp were discarded. The short reads of the parents *V. riparia* GM183
 820 and *V. cinerea* Arnold were mapped with CLC’s map reads to reference algorithm to
 821 the assembled BAC sequences (linear gap cost, length & similarity fraction 1.0).
 822 Based on these mapping results regarding read coverage and percent of covered bases,
 823 the BAC sequences were assigned to either *V. riparia* GM183 or *V. cinerea* Arnold.
 824 BAC sequences with no read coverage were discarded. The assembled and phase-
 825 separated BAC contig sequences are available at
 826 <https://doi.org/10.4119/unibi/2962639>.
 827 The ‘Börner’ BAC sequences with known haplotype allocation were mapped against
 828 the BoeRip and BoeCin haplotype assemblies with minimap2 v2.17⁶⁵, sorted with
 829 SAMtools and the mappings visualized with the Integrative Genomics Viewer (IGV)
 830 v2.5.2^{69, 70}. The number of base calls of the aligned BAC sequences on each haplo-
 831 type were determined with SAMtools’ mpileup. Continuously and correct aligned
 832 BAC sequences show high base calls and low numbers of SNP/InDels (small nucleo-
 833 tide polymorphism, insertion-deletion polymorphism) to one of the two haplotypes
 834 and the opposite result with the alternative haplotype.

835 k-mer based reference-free assembly evaluation with Mercury

836 Mercury v1.3⁷¹ was used to evaluate the diploid BoeRC assembly. First, the best k-
 837 mer size was determined with Mercurys best_k.sh script for the expected 500 Mbp
 838 haploid genome size resulting in 19-mers. Consequently, 19-mer databases were
 839 computed for the *V. riparia* GM183 and *V. cinerea* Arnold Illumina reads as well as
 840 for existing Illumina reads from ‘Börner’¹⁰. With these, hap-mer (haplotype-specific
 841 k-mers as defined by Mercury⁷¹) databases were generated and the haplotype assem-
 842 blies evaluated using the scaffolds of both haplotypes. For building of phased blocks,
 843 the parameter ‘num_switch 10’ and ‘short_range 20,000’ were applied. For compari-
 844 son, Mercury was also run on PN40024, on VitRGM, on both Cabernet Sauvignon
 845 haplotype assemblies v1.1 and on both *M. rotundifolia* cultivar ‘Trayshed’ haplotype

846 assemblies v2.0²³. The 19-mer profiles were computed from PN40024 WGS reads
847 SRR5627797, from VitRGM WGS reads SRR8379638, from Cabernet Sauvignon
848 WGS reads SRR3346861 and SRR3346862 and from *M. rotundifolia* WGS reads
849 SRR6729333. Prior to analyses, the reads were trimmed with Trimmomatic-v0.39
850 allowing a minimum read length of 80 or 90 (PN40024 reads).

851 **Gene annotation**

852 The ‘Börner’ haplotypes were annotated *ab initio* with MAKER v3.01.03 following
853 the MAKER-P pipeline^{72, 73}. As input data, haplotype-specific repeat libraries were
854 created according to the MAKER advanced repeat library protocol. The monocotyle-
855 dons repeat library of RepBase (RepeatMaskerEdition-20181026,
856 model_org=monocotyledons)⁷⁴, the transposable element (TE) sequences from
857 MAKER, the ‘Börner’ *de novo* transcriptome assembly described above, *Vitis* protein
858 sequences (NCBI, Protein DB “Vitis”[Organism]), plant protein sequences (UniProt,
859 release 2020_02, "Viridiplantae [33090]" AND reviewed:yes) and *Vitis* full-length
860 cDNAs (NCBI, Nucleotide DB “Vitis” [Organism] AND complete cds[Title]) were
861 used. The gene predictors SNAP v2006-07-28 (three rounds)⁷⁵ and GeneMark v3.60
862⁷⁶ were trained on both, BoeRip and BoeCin. AUGUSTUS v3.2.3, tRNAscan-SE
863 v1.3.1 and EvidenceModeler v1.1.1⁷⁷, SNAP and GeneMark together were ran
864 through MAKER to compute the final gene annotation. The option ‘split_hit’ was set
865 to 20,000 and the previously generated *Vitis* parameter set was adjusted for AUGUS-
866 TUS.

867 The gene models from BoeRip and BoeCin were refined with PASA v2.4.1⁷⁸. As
868 evidence data, the *de novo* and reference-guided ‘Börner’ transcriptome assemblies,
869 the *Vitis* full-length cDNAs and *Vitis* EST data (NCBI, txid3603[Organism:exp] AND
870 is_est[filter]) were used. The parameter file is available as Supplementary File 3 Data
871 S1. The refinement with PASA was iteratively applied three times on the gene mod-
872 els. The reference-guided assembly was computed with Trinity v2.10.0 using the
873 ‘Börner’ RNA-Seq data aligned with HISAT2 v2.2.0. Only alignments with MQ >10
874 were given to Trinity and ‘--genome_guided_max_intron’ was set to 20,000.

875 After structural gene annotation, a functional gene annotation was added through a
876 BLAST search (package+ v2.10.0) against the UniProt/Swiss-Prot database (release
877 2020_02), through domain identification with InterProScan5 v5.42-78.0⁷⁹ and the

878 PFAM database v32.0. Gene models with Annotation Edit Distance (AED) > 0.5 were
879 filtered from the final gene model set. Also, predicted genes expected to encode a
880 polypeptide with less than 50 amino acids and no functional annotation were dis-
881 carded. The annotation data for BoeRC are available at
882 (<https://doi.org/10.4119/unibi/2962793>). For comparison, functional annotation was
883 also carried out for the PN40024 VCost.v3 structural gene annotation with identical
884 resources.

885 **Prediction of resistance genes**

886 Resistance gene analogs (RGAs) were predicted for both haplotypes with the pipeline
887 RGAugury v2.1.7²⁹. RGAugury employed ncoil⁸⁰, PfamScan v1.6⁸¹, InterProScan5
888 v5.45-80.0⁷⁹ and Phobius-1.01⁸² on the protein sequences to predict known resis-
889 tance domains. The initial filtering against the RGAdb was disabled since the database
890 has not been updated since 2016, unknown RGAs not included in the database were of
891 interest and the identification of (resistance) domains for none RGAs was enabled.
892 For InterProScan5 domain search through RGAugury, the databases Gene3D-4.2.0⁸³,
893 Pfam-33.1⁸⁴, SMART-7.1⁸⁵ and SUPERFAMILY-1.75⁸⁶ were utilized.
894 The RGA annotations of the two haplotypes were correlated through determination of
895 RBHs based on protein sequences. Only RBHs with an identity and coverage ≥ 90
896 were considered.

897 **Assembly wide variant detection**

898 The SNP and InDel positions of the SNP output from the DNAdiff analysis (see
899 above) were compared with the positions of the coding and non-coding gene regions
900 and classified accordingly.

901 **Genetic mapping**

902 The annotated genes and nucleotide sequences of the phylloxera resistance QTL on
903 chr13, referred to as *Rdv1*, of BoeRip, BoeCin, PN40024 and VitRGM were com-
904 pared. Before this study, the QTL *Rdv1* was delimited with the associated genetic
905 markers GF13-01 and GF13-12⁶; the marker positions were identified based on the
906 respective marker assay primer sequences with primersearch (EMBOSS package⁶³).
907 To narrow down the location of *Rdv1*, additional genetic fine mapping was carried out
908 with a mapping population of in total 498 F1 genotypes (see above) and newly de-

signed sequence-tagged SSR (STS) markers. The additional markers were derived from PN40024¹⁶ (Supplementary File 2 Table S14) and used to screen the F1 genotypes from the mapping population.

Five relevant genotypes with marker data indicating recombination events within the initial *Rdv1* QTL were subjected to GBS to locate the recombination points between the two haplotypes of 'Börner'. V3125 was included as a control for comparative read mapping. The SE reads were quality trimmed using Trimmomatic-0.39⁴⁶ with the following parameters: ILLUMINACLIP: 4:30:15 LEADING:30 TRAILING:30 SLIDINGWINDOW:4:15 MINLEN:60. Trimmed reads of each of the six studied genotypes as well as reads from 'Börner'¹⁰ were separately mapped to BoeCin and BoeRip using the HISAT2 version 2.2.0⁵⁹ with the following parameters: --no-softclip --no-spliced-alignment. The obtained SAM files were converted to sorted BAM files as well as indexed using SAMtools. Variant calling against BoeRC and estimation of genomic recombination sites was supported by QIAGEN CLC Genomics Workbench 22.0 (QIAGEN, Aarhus, Denmark). SNPs detected in the *Rdv1* region on chr13 for each genotype were filtered with respect to read coverage and frequency. Homozygous variants were counted if > 90 % of the reads support the variation, for heterozygous variants a frequency between 25 % and 75 % was required. An example case (GBS-04) for the assessment of GBS markers is shown in Supplementary File 1 Fig. S15. Recombination sites were determined by monitoring the switch from either the BoeCin haplotype to the BoeRip haplotype or *vice versa* between a set of at least two variant positions.

Allele sequence comparison and TE detection

The genes (alleles, orthologs) of the *Rdv1* region of BoeRip, BoeCin, PN40024 and VitRGM between the markers GF13-03 and GF13-07 were placed into groups with OrthoFinder v2.3.11⁸⁷ using the longest protein sequence annotated for the splicing variants of each gene. Missing alleles (or, alternatively phrased, missing predictions of orthologous genes at syntenic positions) were investigated and re-annotated by aligning the protein sequences of the other cultivars against the genome sequences with exonerate v2.4.0⁸⁸. Only alignments with at least 70 % coverage of the protein sequence were considered. To obtain gene annotations for remaining poorly annotated sequences despite existing RNA-Seq support in the genomic region studied,

941 BRAKER was run with strictly filtered RNA-Seq mappings (only uniquely mapped
942 PE reads on both ‘Börner’ haplotypes offered as combined target, both reads of a pair
943 must map); adding untranslated regions (UTRs) and prediction of alternative tran-
944 scripts from RNA-Seq data was enabled. Since the focus here is on phylloxera resis-
945 tance at the root, an additional AUGUSTUS species model was trained on BoeRC
946 with the uniquely mapped root RNA-Seq reads, the above described *Vitis* protein and
947 plant protein sequences and eudicot protein sequences from OrthoDB v10.1⁸⁹. The
948 results were manually evaluated and restored gene models of BoeRip and BoeCin
949 were refined two times with PASA. Through investigation of the RNA-Seq mappings
950 from all available tissues to BoeCin, the gene predictions in the *Rdv1* region of all
951 four sequence regions studied were manually screened for low supported gene mod-
952 els. Transcripts encoding for protein sequences with less than 50 amino acids, tagged
953 as ‘protein of unknown function’ (PUF) and with no ortholog were removed. RGA
954 classes were determined for the manually curated genes/alleles with RGAugury as
955 described above. TEs were identified with the Extensive *de novo* TE Annotator
956 (EDTA)⁹⁰. Finally, the results were manually collected in allelic representations of
957 the genes of the *Rdv1* region of BoeRip, BoeCin, PN40024 and VitRGM. An ex-
958 tended version of the gene overview at the *Rdv1* locus that includes TEs has also been
959 prepared (Supplementary File 1 Fig. S11).
960

961 **Competing interests**

962 The authors declare that they have no competing interests.

963 **Authors' contributions**

964 Sampling and phenotyping were done at the JKI (Julius Kuehn Institute, Institute for
965 Grapevine Breeding Geilweilerhof, Siebeldingen, Germany). Sequencing and data
966 analysis were performed at Bielefeld University, Faculty of Biology & Center for Bio-
967 technology (CeBiTec) and MPIPZ (Cologne). The design of the experiments was set
968 up by LH, DH and BF. Genotype selection, sampling and DNA isolations were car-
969 ried out by LH and DH. PV, BH and RR accomplished library preparation and se-
970 quencing. LH performed the SSR analysis. BW supervised the work at Bielefeld Uni-
971 versity. RT supervised the work at JKI. RT and BW acquired project funding and
972 wrote the project proposal. All bioinformatic data analyses, creation of figures, tables
973 and writing of the initial draft of the manuscript were performed by BF and DH. BF
974 and BW finalised the manuscript. All authors have read and approved the final manu-
975 script.

976 **Acknowledgements**

977 The authors like to thank all members of the Genetics and Genomics of Plants team at
978 Bielefeld University as well as the members of the Julius Kuehn Institute for Grape-
979 vine Breeding at Geilweilerhof.

980 The project was supported by funds from the Federal Ministry of Food and Agricul-
981 ture (BMEL), based on a decision by the Parliament of the Federal Republic of Ger-
982 many via the Federal Office for Agriculture and Food (BLE) under the innovation
983 support program (project acronym MureViU, 28-1-82.066-15), as well as by the EU
984 COST Action INTEGRAPE (CA 17111). We acknowledge support for the article
985 processing charge by the Deutsche Forschungsgemeinschaft and the Open Access
986 Publication Fund of Bielefeld University.

987 This work was supported by the BMBF-funded de.NBI Cloud within the German
988 Network for Bioinformatics Infrastructure (de.NBI).

989 Data availability

990 The SMRT sequencing reads of 'Börner' (ERR6182799), the WGS reads of 'Börners'
 991 parents *V. riparia* GM18 (ERR6182967, ERR6183009) and *V. cinerea* Arnold
 992 (ERR6183041) as well as the 'Börner' haplotype assemblies (ERS6637871) were
 993 deposited in ENA/GenBank/DDJ under project no. PRJEB45595. The haplotype as-
 994 semblies BoeRip and BoeCin at chromosome level with structural and functional an-
 995 notation are available at (<https://doi.org/10.4119/unibi/2962793>). The RNA-Seq data
 996 of 'Börner' leaf (ERR6182801) and tendril (ERR6182801) were deposited in
 997 ENA/GenBank/DDJ under project no. PRJEB46079, the raw sequencing reads of
 998 'Börner' BACs (ERR6183016-ERR6183021, ERR6183023-ERR6183025,
 999 ERR6183030-ERR6183032) under project no. PRJEB46081. The assembled and
 1000 phase-separated BAC contigs are available at (<https://doi.org/10.4119/unibi/2962639>).
 1001 DNA sequence read data from the F1 genotypes of the mapping population and the
 1002 maternal parent V3125 generated for GBS were deposited in ENA/GenBank/DDJ
 1003 under project no. PRJEB53997.

1004 Supplementary information

1005 Supplementary File 1 Fig. S1: Cumulative length of both 'Börner' haplotype assem-
 1006 blies
 1007 Supplementary File 1 Fig. S2: Sequence region with BAC sequences of both haplo-
 1008 types
 1009 Supplementary File 1 Fig. S3: Spectra-copy number and assembly spectrum plot of
 1010 the 'Börner' genome sequence assembly
 1011 Supplementary File 1 Fig. S4: Analysis of phasing with Mercury
 1012 Supplementary File 1 Fig. S5: All-versus-all dot plot between pseudochromosomes of
 1013 the *V. riparia* haplotype of 'Börner' and the *V. vinifera* reference sequence.
 1014 Supplementary File 1 Fig. S6: All-versus-all dot plot between pseudochromosomes of
 1015 the *V. cinerea* haplotype of 'Börner' and the *V. vinifera* reference sequence
 1016 Supplementary File 1 Fig. S7: All-versus-all dot plot between pseudochromosomes of
 1017 the *V. riparia* haplotype of 'Börner' and those of *V. riparia* Gloire de Montpellier
 1018 Supplementary File 1 Fig. S8: All-versus-all dot plot between pseudochromosomes of
 1019 the *V. cinerea* haplotype of 'Börner' and those of *V. riparia* Gloire de Montpellier
 1020 Supplementary File 1 Fig. S9: Dot plot of the extended *Rdv1* sequence region on
 1021 pseudochromosome 13 between BoeRip and PN40024
 1022 Supplementary File 1 Fig. S10: Dot plot of *Rdv1* on pseudochromosome 13 between
 1023 BoeCin and PN40024

1024 Supplementary File 1 Fig. S11: Extended version of Fig. 6, "Genes of BoeCin,
1025 BoeRip, PN40024 and VitRGM at and surrounding the *Rdv1* locus", transposable
1026 elements/TEs included
1027 Supplementary File 1 Fig. S12: Length distribution calculated over all subreads
1028 Supplementary File 1 Fig. S13: Coverage plot of the 'Börner' haplotype assemblies
1029 Supplementary File 1 Fig. S14: Display of initial assembly analysis of chr02 of
1030 BoeRip
1031 Supplementary File 1 Fig. S15: Example for determination of allelic constitution in
1032 selected F1 genotypes
1033
1034 Supplementary File 2 Table S1: Pseudochromosome sizes of the haplotype assemblies
1035 BoeRip and BoeCin
1036 Supplementary File 2 Table S2: Quality statistics of *Vitis* genome assemblies
1037 Supplementary File 2 Table S3: Assembly statistics of the BAC contigs
1038 Supplementary File 2 Table S4: Mapping statistics of the BAC contigs on the BoeRip
1039 and BoeCin haplotype assembly
1040 Supplementary File 2 Table S5: Statistics of the trimmed BAC sequences
1041 Supplementary File 2 Table S6: Haplotype blocks of BoeRip and BoeCin
1042 Supplementary File 2 Table S7: Repeat content of the haplotypes
1043 Supplementary File 2 Table S8: GBS marker sites for the parents 'Börner' and V3125
1044 and selected F1 offspring
1045 Supplementary File 2 Table S9: BoeCin, BoeRip, PN40024 and VitRGM genes in
1046 *Rdv1* region
1047 Supplementary File 2 Table S10: Amino acid sequence similarity matrix BoeCin to
1048 BoeCin deduced from genes in *Rdv1* region
1049 Supplementary File 2 Table S11: Amino acid sequence similarity matrix BoeCin to
1050 BoeRip deduced from genes in *Rdv1* region
1051 Supplementary File 2 Table S12: Amino acid sequence similarity matrix BoeCin to
1052 PN40024 deduced from genes in *Rdv1* region
1053 Supplementary File 2 Table S13: Amino acid sequence similarity matrix BoeCin to
1054 VitRGM deduced from genes in *Rdv1* region
1055 Supplementary File 2 Table S14: Primer sequences for SSR marker
1056 Supplementary File 2 Table S15: RNA-Seq read trimming statistics
1057 Supplementary File 2 Table S16: Trimming statistics of *V. riparia* and *V. cinerea* ge-
1058 nomic reads
1059 Supplementary File 2 Table S17: Description of samples applied to GBS
1060 Supplementary File 2 Table S18: Mapping positions of the genetic SSR marker on the
1061 PN40024 genome assembly and on the 'Börner' haplotype assemblies BoeRip and
1062 BoeCin
1063 Supplementary File 2 Table S19: Distribution of the 344 genetic marker used for
1064 pseudochromosome validation over BoeRip and BoeCin.
1065
1066 Supplementary File 3 Data S1: Parameter file for PASA gene model refinement

1067

References

1. Ramos-Madrigal J, *et al.* Palaeogenomic insights into the origins of French grapevine diversity. *Nature Plants* **5**, 595-603 (2019).
2. Vezzulli S, *et al.* Genomic Designing for Biotic Stress Resistant Grapevine. In: *Genomic Designing for Biotic Stress Resistant Fruit Crops* (ed Kole C). Springer Nature Switzerland AG (2022).
3. Granett J, Walker MA, Kocsis L, Omer AD. Biology and management of grape phylloxera. *Annual Review of Entomology* **46**, 387-412 (2001).
4. Tello J, Mammerler R, Cajic M, Forneck A. Major Outbreaks in the Nineteenth Century Shaped Grape Phylloxera Contemporary Genetic Structure in Europe. *Scientific Reports* **9**, 17540 (2019).
5. Forneck A, Kleinmann S, Blaich R, Anvari SF. Histochemistry and anatomy of phylloxera (*Daktulosphaira vitifoliae*) nodosities on young roots of grapevine (*Vitis* spp). *Vitis* **41**, 93-97 (2002).
6. Zhang J, Hausmann L, Eibach R, Welter LJ, Töpfer R, Zyprian EM. A framework map from grapevine V3125 (*Vitis vinifera* 'Schiava grossa' x 'Riesling') x rootstock cultivar 'Börner' (*Vitis riparia* x *Vitis cinerea*) to localize genetic determinants of phylloxera root resistance. *Theoretical and Applied Genetics* **119**, 1039-1051 (2009).
7. Töpfer R, Hausmann L, Harst M, Maul E, Zyprian E, Eibach R. New horizons for grapevine breeding. In: *Fruit, vegetable and cereal science and biotechnology*,). Global Science Books (2011).
8. Rex F, Fechter I, Hausmann L, Töpfer R. QTL mapping of black rot (*Guignardia bidwellii*) resistance in the grapevine rootstock 'Börner' (*V. riparia* Gm183 × *V. cinerea* Arnold). *Theoretical and Applied Genetics* **127**, 1667-1677 (2014).
9. Ochssner I, Hausmann L, Töpfer R. Rpv14, a new genetic source for *Plasmopara viticola* resistance conferred by *Vitis cinerea*. *Vitis* **55**, 79-81 (2016).
10. Holtgräwe D, *et al.* A Partially Phase-Separated Genome Sequence Assembly of the *Vitis* Rootstock 'Börner' (*Vitis riparia* x *Vitis cinerea*) and Its Exploitation for Marker Development and Targeted Mapping. *Frontiers in Plant Science* **11**, 156 (2020).
11. Han GZ. Origin and evolution of the plant immune system. *New Phytologist* **222**, 70-83 (2019).
12. Sekhwal MK, Li P, Lam I, Wang X, Cloutier S, You FM. Disease Resistance Gene Analogs (RGAs) in Plants. *International Journal of Molecular Sciences* **16**, 19248-19290 (2015).
13. Lodhi MA, Reisch BI. Nuclear DNA content of *Vitis* species, cultivars, and other genera of the Vitaceae. *Theoretical and Applied Genetics* **90**, 11-16 (1995).
14. Zhou Y, *et al.* The population genetics of structural variants in grapevine domestication. *Nature Plants* **5**, 965-979 (2019).

15. Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R. Genome Sequence Initiatives. In: *Genetics, Genomics and Breeding of Grapes* (eds Adam-Blondon A-F, Martinez-Zapater JM, Kole C) (2011).
16. Jaillon O, *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463-467 (2007).
17. Vitulo N, *et al.* A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biology* **14**, 99 (2014).
18. Riaz S, *et al.* Genetic diversity analysis of cultivated and wild grapevine (*Vitis vinifera* L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biology* **18**, 137 (2018).
19. Chin CS, *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* **13**, 1050-1054 (2016).
20. Massonnet M, *et al.* The genetic basis of sex determination in grapes. *Nature Communications* **11**, 2902 (2020).
21. Minio A, Massonnet M, Figueroa-Balderas R, Castro A, Cantu D. Diploid Genome Assembly of the Wine Grape Carmenere. *G3 Genes Genomes Genetics* **9**, 1331-1337 (2019).
22. Roach MJ, *et al.* Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genetics* **14**, e1007807 (2018).
23. Cochetel N, Minio A, Massonnet M, Vondras AM, Figueroa-Balderas R, Cantu D. Diploid chromosome-scale assembly of the *Muscadinia rotundifolia* genome supports chromosome fusion and disease resistance gene expansion during *Vitis* and *Muscadinia* divergence. *G3 Genes Genomes Genetics* **11**, jkab033 (2021).
24. Girollet N, Rubio B, Lopez-Roques C, Valiere S, Ollat N, Bert PF. De novo phased assembly of the *Vitis riparia* grape genome. *Scientific Data* **6**, 127 (2019).
25. Ollat N, *et al.* Rootstocks as a component of adaptation to environment. In: *Grapevine in a changing environment: a molecular and ecophysiological perspective* (eds Geros H, Chaves MM, Gil HM, Delrot S). John Wiley and Sons (2016).
26. Eid J, *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009).
27. Michael TP, VanBuren R. Building near-complete plant genomes. *Current Opinion in Plant Biology* **54**, 26-33 (2020).
28. Koren S, *et al.* De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* **36**, 1174-1182 (2018).
29. Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* **17**, 852 (2016).

30. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**, 722-736 (2017).
31. Xu Z, Dixon JR. Genome reconstruction and haplotype phasing using chromosome conformation capture methodologies. *Briefings in Functional Genomics* **19**, 139-150 (2020).
32. Sweetman C, Wong DC, Ford CM, Drew DP. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics* **13**, 691 (2012).
33. Lijavetzky D, Cabezas JA, Ibanez A, Rodriguez V, Martinez-Zapater JM. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* **8**, 424 (2007).
34. Salmaso M, *et al.* Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. *Molecular Breeding* **14**, 385-395 (2004).
35. Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* **39**, 285-312 (2001).
36. Dupeyron M, Singh KS, Bass C, Hayward A. Evolution of Mutator transposable elements across eukaryotic diversity. *Mobile DNA* **10**, 12 (2019).
37. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics* **9**, 411-414 (2008).
38. Folia S, *et al.* Gene duplication and transposition of mobile elements drive evolution of the Rpv3 resistance locus in grapevine. *The Plant Journal* **101**, 529-542 (2020).
39. Bianchet C, *et al.* An *Arabidopsis thaliana* leucine-rich repeat protein harbors an adenyl cyclase catalytic center and affects responses to pathogens. *Journal of Plant Physiology* **232**, 12-22 (2019).
40. Bittner-Eddy PD, Beynon JL. The *Arabidopsis* downy mildew resistance gene, RPP13-Nd, functions independently of NDR1 and EDS1 and does not require the accumulation of salicylic acid. *Molecular Plant Microbe Interactions* **14**, 416-421 (2001).
41. Allen RL, *et al.* Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957-1960 (2004).
42. Canaguier A, *et al.* A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* **14**, 56-62 (2017).
43. Fechter I, *et al.* QTL analysis of flowering time and ripening traits suggests an impact of a genomic region on linkage group 1 in *Vitis*. *Theoretical and Applied Genetics* **127**, 1857-1872 (2014).
44. Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B. An *Arabidopsis thaliana* T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics. *Plant Molecular Biology* **53**, 247-259 (2003).

45. Frommer B, *et al.* Genome Sequences of Both Organelles of the Grapevine Rootstock Cultivar 'Börner'. *Microbiology Resource Announcement* **9**, e01471-01419 (2020).
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
47. Forneck A, Walker MA, Blaich R, Yvon M, Leclant F. Interaction of phylloxera (*Daktulosphaira vitifoliae* Fitch) with grape (*Vitis* spp.) in simple isolation chambers. *American Journal of Enology and Viticulture* **52**, 28-34 (2001).
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
49. Camacho C, *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
50. Jansen RK, *et al.* Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evolutionary Biology* **6**, 32 (2006).
51. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* **26**, 99-110 (2009).
52. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* **32**, 3321-3323 (2016).
53. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-360 (2015).
54. Li H, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
55. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579 (2011).
56. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
57. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215-ii225 (2003).
58. Grabberr MG, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).
59. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-915 (2019).
60. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Research* **12**, 656-664 (2002).

61. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology* **14**, e1005944 (2018).
62. Fechter I, *et al.* Candidate genes within a 143 kb region of the flower sex locus in Vitis. *Molecular Genetics and Genomics* **287**, 247-259 (2012).
63. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277 (2000).
64. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
65. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
66. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
67. Waterhouse RM, *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* **35**, 543-548 (2018).
68. Cannon EK, Cannon SB. Chromosome visualization tool: a whole genome viewer. *International Journal of Plant Genomics* **2011**, 373875 (2011).
69. Robinson JT, *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26 (2011).
70. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-192 (2013).
71. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).
72. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, (2011).
73. Campbell MS, *et al.* MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiology* **164**, 513-524 (2014).
74. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
75. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
76. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research* **2014**, e119 (2014).

77. Haas BJ, *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7 (2008).
78. Haas BJ, *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654-5666 (2003).
79. Jones P, *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
80. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164 (1991).
81. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* **8**, 298 (2007).
82. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**, 1027-1036 (2004).
83. Sillitoe I, *et al.* CATH: expanding the horizons of structure-based functional annotations for genome sequences. *Nucleic Acids Research* **47**, D280-D284 (2019).
84. Mistry J, *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419 (2021).
85. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* **46**, D493-D496 (2018).
86. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology* **313**, 903-919 (2001).
87. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
88. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
89. Kriventseva EV, *et al.* OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* **47**, D807-D811 (2019).
90. Ou S, *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).