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

Bianca Frommer[1], Ludger Hausmann[2], Daniela Holtgräwe[1], Prisca Viehöver[1], Bruno Hüttel[3], Richard Reinhardt[3], Reinhard Töpfer[2], Bernd Weisshaar[1]

[1] Bielefeld University, Chair of Genetics and Genomics of Plants, Faculty of Biology & Center for Biotechnology (CeBiTec), Bielefeld, Germany
[2] Julius Kuehn Institute, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany
[3] Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Email addresses:
BF: frommer@CeBiTec.Uni-Bielefeld.DE / ORCID: 0000-0002-5792-0102
LH: ludger.hausmann@julius-kuehn.de / ORCID: 0000-0002-4046-9626
DH: dholtgra@CeBiTec.Uni-Bielefeld.DE / ORCID: 0000-0002-1062-4576
PV: viehoeve@CeBiTec.Uni-Bielefeld.DE
BH: huettel@mpipz.mpg.de / ORCID: 0000-0001-7165-1714
RR: reinhardt@mpipz.mpg.de
RT: reinhard.toepfer@julius-kuehn.de / ORCID: 0000-0003-1569-2495
BW: bernd.weisshaar@Uni-Bielefeld.DE / ORCID: 0000-0002-7635-3473

Abstract

The phylloxera resistant rootstock cultivar ‘Börner’ is an interspecies hybrid derived from *Vitis riparia* and *V. cinerea* and a valuable resource for *Vitis* disease resistances. We created a fully phased, high-quality ‘Börner’ genome sequence named BoeRC using long PacBio reads. Comprehensive gene annotation of both ‘Börner’ haplotypes, designated BoeRip and BoeCin, was applied to describe the phylloxera resistance locus *Rdv1*. Using a mapping population derived from a sensitive *V. vinifera* cultivar and ’Börner’, the *Rdv1* locus was further delimited. *Rdv1*, which is derived from *V. cinerea* and included in the haplotype BoeCin, was compared with sequences of phylloxera susceptible and phylloxera tolerant cultivars. Between flanking regions that display high synteny, we detected and precisely characterized a diverse sequence region that covers between 202 to 403 kbp in different haplotypes. In BoeCin, five putative disease resistance genes were identified that represent likely candidates for conferring resistance to phylloxera.

Keywords

Assembly, PacBio, SMRT sequencing, grapevine, rootstock, genome, Canu, annotation, resistance, resistance gene analogs, phylloxera, *Rdv1*

Background

Vitis vinifera Linné subspecies *vinifera* is the most cultivated *Vitis* species worldwide. It is economically highly important and appreciated for wine production, table grapes, dry fruits like raisins, and other products. The species is endemic to Europe, and domestication and development of cultivars and has a very long history [1]. *Vitis vinifera* is highly susceptible to many diseases and pests including downy and powdery mildew as well as to the root aphid phylloxera (*Daktulosphaira vitifoliae* Fitch) [2]. In the 19th century, phylloxera caused a crisis of unprecedented dimension for viticulture and questioned the survival of viticulture in general. *V. vinifera* plants died quickly after phylloxera attack of the roots and huge cultivation areas were destroyed. Only the grafting of susceptible *V. vinifera* cultivars as scions to phylloxera-tolerant rootstocks derived from American wild *Vitis* species or their interspecific hybrids rescued viticulture [3, 4]. The phylloxera plague remains as a serious case even if it is controlled at the moment. The selection of a good rootstock that provides resistances or tolerances to the almost ubiquitous phylloxera is crucial for nowadays viticulture. One newly developed rootstock variety is the cultivar 'Börner', an interspecific F1 hybrid derived from a cross of the American *Vitis* genotypes *Vitis riparia* GM183 and *Vitis cinerea* 'Arnold'. 'Börner' inherited full resistance to phylloxera on roots from *V. cinerea* and develops no nodosities or tuberosities [5] upon infestation with phylloxera (Fig. 1). *V. cinerea* is an American wild species known to provide full resistance at the root and very high resistance at the leaf against phylloxera [6]. A quantitative trait locus (QTL), located on chromosome (chr) 13, was identified to provide phylloxera resistance [6] and is referred to as *Rdv1* (resistance to *Daktulosphaira vitifoliae*) [7]. In addition, 'Börner' inherited positive viticultural traits including tolerance to frost, resistance to black rot caused by the fungus *Guignardia bidwellii* [8], and resistance against downy mildew caused by *Plasmopara viticola* [9, 10].

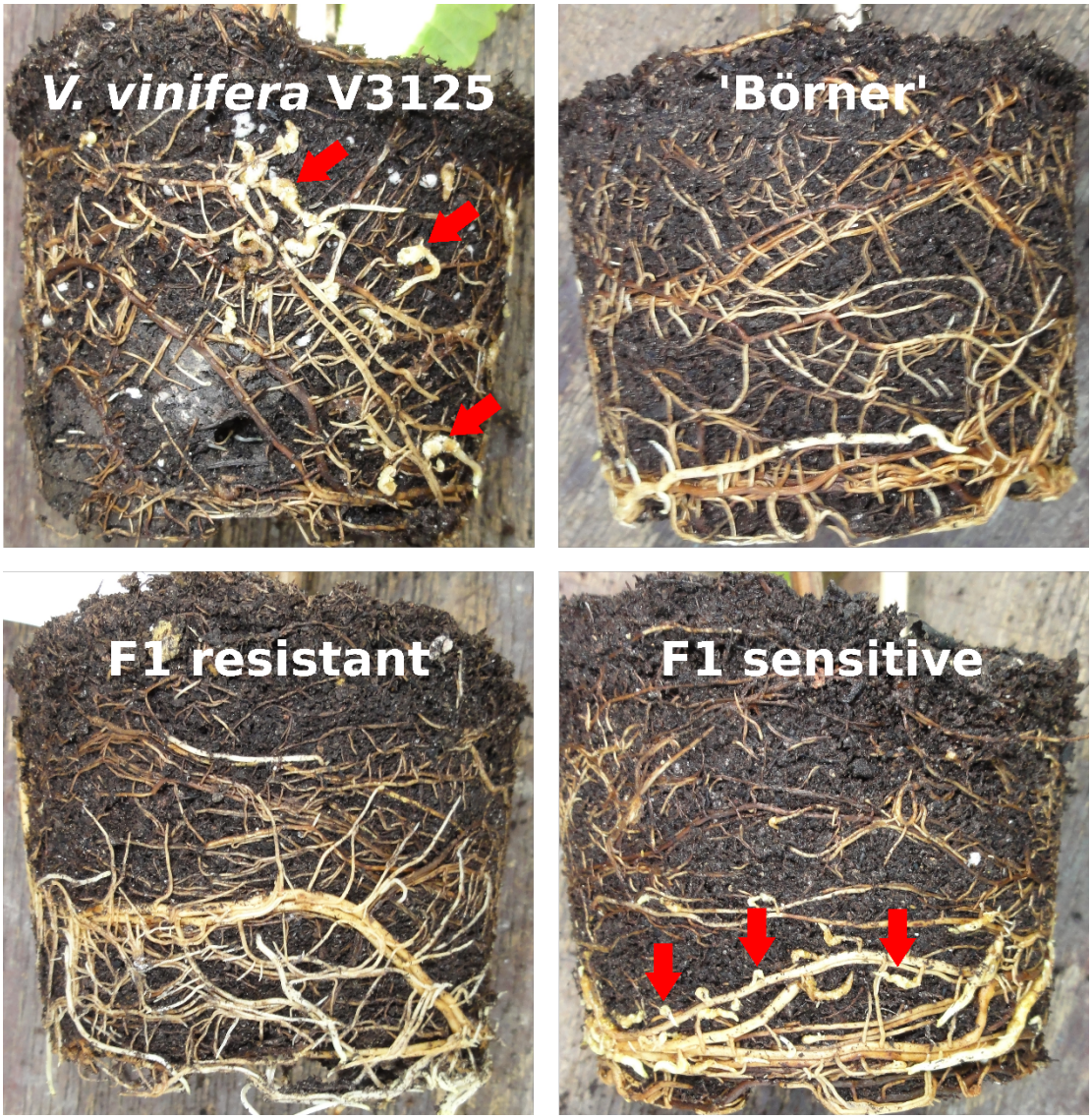


Fig. 1: Root balls of sensitive and resistant grapevine plants attacked by phylloxera. Some phylloxera-induced nodosities are marked on the roots with red arrows. The pictures show roots of the grapevines indicated, the two F1 plants selected with the phenotypes as indicated are derived from the mapping 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 their protein domains and structural features which are similar to validated resistance genes [12]. Based on their domain structure that also indicates their localisation and/or site of action, they are broadly classified as nucleotide-binding site and leucine rich repeat domain receptor (NLRs) or pattern-recog-

84 nition receptors (PRRs) [11]. Usually, PRRs act at the cell surface because of a transmem-
85 brane (TM) domain, while NLRs cover a nucleotide-binding site (NBS) domain and
86 usually act intracellularly.
87 *Vitis* genomes cover a haploid set of 19 chromosomes, and the haploid genome size
88 varies around 500 Mbp [13, 14]. The inbreed cultivar 'PN40024' has been sequenced [15, 16]
89 and the resulting assembly serves as a reference genome sequence (referred to as
90 PN40024). However, since *Vitis* species and cultivars like almost all perennial plants
91 are highly heterozygous and genetically divers [10, 17, 18], there is a need for more *Vitis*
92 genome sequences. Several additional *Vitis* genome sequences have become available
93 which were generated from long reads, including those of Cabernet Sauvignon [19, 20],
94 Carménère [21], Chardonnay [14, 22], *Muscadinia rotundifolia* cultivar Trayshed [23] and *V.*
95 *riparia* Gloire de Montpellier (referred to here as VitRGM) [24]. These assemblies each
96 cover a primary and an alternate pseudo-haplotype of which the alternate pseudo-hap-
97 lotype is usually of reduced contiguity. The cultivar *V. riparia* Gloire de Montpellier is
98 tolerant against root infection by phylloxera [25], but to the best of our knowledge it is
99 unknown if this tolerance is linked to the *Rdv1* locus or to another region of the genome.
100 Long read DNA sequencing technology, like "Single Molecule, Real Time" (SMRT)
101 sequencing [26] provided by Pacific Biosciences (PacBio), is one option to generate high
102 quality genome sequence assemblies. The long reads more likely span problematic ge-
103 nomic regions and thus pave the way for more contiguous assemblies [27]. Also, infor-
104 mation of the haplotype and phase differences are more completely contained in a single
105 read. This fact, together with the transmission of exactly one haplotype from two par-
106 ents to a single offspring according to Mendel's laws, is exploited by the "trio binning"
107 approach for generating fully phased genome sequence assemblies [28].
108 To access and resolve the *Rdv1* locus, we set out to generate a haplotype-resolved ge-
109 nome sequence of 'Börner' with the goal to overcome complications caused by high
110 heterozygosity and the complexity of resistance gene clusters. Using SMRT sequencing
111 to generate long reads and additional Illumina short read data from both parents of
112 'Börner', we assembled both haplotype sequences of 'Börner' at chromosome level.
113 The two truly phased haplotypes of the diploid interspecific hybrid 'Börner' represent
114 the genome sequences of the two species *V. riparia* and *V. cinerea*. Structural and func-
115 tional gene annotation was performed for both haplotype sequences, and resistance gene

116 analogues (RGAs) were studied. The new sequence and annotation data as well as ad-
117 ditional mapping results were used to dissect the *Rdv1* locus of 'Börner' at the gene
118 level.
119
120

121 # Material and Methods

122 **Reference data sets**

123 In this study, the 'PN40024' genome sequence [16] assembly 12X.v2 (urgi.ver-
124 sailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences) and its annotation,
125 the VCost.v3 gene annotation (https://urgi.versailles.inra.fr/Species/Vitis/Annotations)
126 [29] as well as the *V. riparia* Gloire de Montpellier genome sequence assembly VitRGM
127 and its annotation (https://ncbi.nlm.nih.gov/assembly/GCF_004353265.1) [24] were used
128 for comparisons.
129

130 **Plant material and DNA and RNA extraction**

131 'Börner' is listed in the *Vitis* International Variety Catalogue with the variety number
132 *V*IVC 1499, its parents *V. riparia* GM183 and *V. cinerea* Arnold with variety numbers
133 *V*IVC 4686 and *V*IVC 13645, respectively. All three cultivars do not belong to an en-
134 dangered species and were obtained and are grown at the Institute for Grapevine Breed-
135 ing Geilweilerhof at Siebeldingen in accordance with German legislation.
136 Cuttings of the interspecific rootstock variety 'Börner' were raised in a growth chamber
137 on soil at 18°-24°C (64°F-75°F), 70% relative humidity and 16h light and 8h darkness.
138 Prior to harvest the plants were cultivated in the dark for 72 hours. Small, young leaves
139 were harvested and stored on ice until use. Extraction of high molecular DNA for
140 SMRT sequencing was performed starting with 1.5 g fresh weight and was carried out
141 with the NucleoSpin Plant II Maxi kit for DNA (Macherey and Nagel, Düren; Germany)
142 according to the recommendations of the manufacturer for plants. DNA integrity, qual-
143 ity and quantity was checked using a TapeStation (Agilent) and was found to peak at a
144 fragment size higher than 48.5 kbp.
145 Leaf material of 'Börner's parents was harvested from cuttings grown in the green
146 house, frozen in liquid nitrogen and stored at -80°C. The DNA extraction was carried
147 out with the DNeasy Plant Maxi Kit (Qiagen, Hilden; Germany) according to the in-
148 structions of the manufacturer. Tissue samples of leaves and tendrils of 'Börner' were
149 collected from outdoor grown plants. Total RNA was isolated using the Spectrum Plant
150 RNA-Kit (Sigma, Taufkirchen; Germany).

151 Plant material for genotyping-by-sequencing (GBS) was harvested from selected gen-
152 otypes of the mapping population V3125 x 'Börner'. Young leaf material from each
153 individual progeny plant well as from the maternal plant V3125 was used for DNA
154 extraction. Genomic DNA was prepared using an established CTAB-based protocol [30].
155 Subsequently, the DNA obtained was treated with RNAse and quantified using
156 PicoGreen.
157

158 **Library construction for 'Börner' genome and transcriptome sequencing**

159 PacBio Sequel libraries were prepared according to the SMRTbell Template Prep Kit
160 1.0 and then size-selected with BluePippin for a target insert size of greater than 10 kb.
161 Seventeen 1Mv3 SMRT cells were run on a Sequel I sequencer using the Binding Kit
162 1.0 and the Sequencing Chemistry version 1.0 (all from PacBio) for six hours per cell.
163 The output BAM files of the sequencing step were loaded into SMRT Link 5.0.1 ac-
164 cording to the recommendation of the manufacturer (PacBio Reference Guide 2018)
165 and converted to FASTA format for downstream processing. The length distribution
166 over all reads was calculated with the assembler Canu [31] and is provided in Supplemen-
167 tary File 1 Fig. S1.
168 The RNA samples from leaves and tendrils of 'Börner' were paired-end sequenced us-
169 ing Illumina technology on a HiSeq-2000 essentially as described [32]. The new RNA-
170 Seq read data (ENA/SRA study accession no. PRJEB46079) were processed together
171 with data submitted before [32]. These already existing RNA-Seq reads of 'Börner'
172 (PRJEB34983) cover leaves (ERR3894001), winter leaves (ERR3895010), inflores-
173 cences (ERR3894002), tendrils (ERR3894003) and roots (ERR3895007). Read data
174 were trimmed with Trimmomatic-0.38 [33] allowing reads larger than 80 nt (new data)
175 and 90 nt (published data) to be kept. The trimming statistics of the RNA-Seq data is
176 provided in Supplementary File 2 Table S1.
177

178 **Illumina sequencing of the parental genomes of 'Börner'**

179 Library preparation for the *V. riparia* GM183 and *V. cinerea* Arnold samples was per-
180 formed according to the Illumina TruSeq DNA Sample Preparation v2 Guide. Genomic
181 DNA (1500 ng each) was fragmented by nebulisation. After end repair and A-tailing,

182 individual adaptors were ligated to the fragments for PE sequencing. The adaptor li-
183 gated fragments were purified on a 2% low melt agarose gel and size selected by excis-
184 ing a band ranging from 500-800 bp. After enrichment PCR of DNA fragments that
185 carry adaptors on both ends, the final libraries were quantified by using the Quant-iT
186 PicoGreen dsDNA assay on a FLUOstar Optima Platereader (BMG Labtech) and qual-
187 ified on a BioAnalyzer High Sensitivity DNA Chip (Agilent). The *V. riparia* GM183
188 library was sequenced on a MiSeq 2 x 250 nt PE run and on two lanes of a 2 x 100 PE
189 run on a HiSeq-1500 in high output mode. The *V. cinerea* Arnold library was sequenced
190 in one 2 x 150 bp PE run on a HiSeq-1500 in rapid mode. The read data were submitted
191 to ENA (ENA/SRA study accession no. PRJEB45595). All genomic short read data
192 were quality trimmed with Trimmomatic-0.36 [33] allowing reads equal or longer 80 nt
193 to be kept. The trimming statistics is provided in Supplementary File 2 Table S2.
194

195 **Sequencing of pools of BACs from a BAC library of 'Börner'**

196 For the creation of phased sequences of 'Börner', 8 pools of bacterial artificial chromo-
197 some (BAC) constructs containing a total of 440 mapped BACs from two loci (one on
198 chr01, the other on chr14) were selected from a 'Börner' BAC library [10]. The BACs
199 were selected and distributed to the pools according to the positions of BAC end se-
200 quences on the corresponding regions of PN40024. Mapping to the reference assembly
201 was carried out with the CLC Genomics Workbench v8.0 toolkit (https://www.qi-
202 agenbioinformatics.com/). Pools were arranged by excluding overlap of BAC inserts to
203 allow assembly of individual BAC insert sequences. The BAC Pools1-4 were se-
204 quenced on a 454 Life Sciences Genome Sequencer GS FLX and PE (2 x 250 nt) on
205 the Illumina MiSeq platform. The BAC Pools5-8 were sequenced on the MiSeq plat-
206 form only. The read data were submitted to ENA (ENA/SRA study accession no.
207 PRJEB46081). The MiSeq raw data were quality trimmed with Trimmomatic-0.32 us-
208 ing 'ILLUMINACLIP: 2:40:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
209 MINLEN:36' [33]. The trimming statistics for the BAC pools is provided in Supplemen-
210 tary File 2 Table S3.
211

212 **Library construction and sequencing for GBS**

213 The barcoded libraries were prepared using Illumina library preparation kits using
214 TrueSeq technology and sequenced on an NextSeq500. Sequencing was performed in
215 150 nt single-end (SE) modus and the read data were submitted to ENA. The ENA/SRA
216 study accession no. is PRJEB53997. Information on genotypes, sequencing strategy,
217 data volume and run IDs are summarized in Supplementary File 2 Table S4.
218

219 **Phenotyping of phylloxera resistance**

220 The test system used to assess phylloxera resistance of the was essentially described
221 previously [34]. The phylloxera population was derived from naturally infested cultivars.
222 The phylloxera resistance tests started at the beginning of the vegetation season when
223 enough infested leaf material was available as inoculum. Leaf pieces with 8 to 12 galls
224 containing mature eggs were inserted in the soil of rooted cuttings of the F1 individuals.
225 The inoculation procedure was repeated after 3 to 4 weeks. Approximately 4 weeks
226 after the second inoculation the plants were rated as sensitive if nodosities were visible
227 or resistant if the roots developed normally. Each F1 genotype was tested in triplicate
228 and the assays were repeated at least twice.
229

230 **Diploid genome sequence assembly**

231 To compute a phased genome assembly, a binning and assembly approach was used.
232 The binning prior to assembly was performed with the tool TrioCanu of the assembler
233 Canu v1.7 [28, 31]. TrioCanu uses short reads of the parents and long reads of the F1 as
234 input. It bins the long reads of the F1 into parental subsets based on k-mer comparisons.
235 The binning step was performed on a compute cluster using TrioCanu default parame-
236 ters (k-mer size of 20). Binning resulted in one read subset for each parent and an ad-
237 ditional small subset with reads that could not be assigned to either of the two parental
238 haplotypes. In the following, the parental subsets are referred to as Vrip and Vcin read
239 subset according to 'Börner's parents and the two different haplotypes combined into
240 'Börner', respectively. The read subset with unassigned reads is referred to as UAR
241 subset and contained only 0.13% of all bases with an average read length of 1.5 kbp.
242 After binning, the read subsets were used to compute the two haplotype assemblies
243 BoeRip and BoeCin of 'Börner' with the assembler Canu v1.7 [31]. To compute the

244 BoeRip haplotype assembly, the Vrip read subset and the UAR subset were used in
245 order to include identical and/or homozygous regions. Likewise, the BoeCin haplotype
246 assembly was generated from the Vcin read subset and the UAR subset. Therefore, the
247 basis of the haplotype assemblies, BoeRip and BoeCin, are approximately 32 to 34 Gbp
248 read data with an estimated genome coverage of more than 60-fold each. For both sep-
249 arately computed haplotype assemblies, Canu was run on a compute cluster utilizing
250 the parameters 'genomeSize=550m', 'corMhapSensitivity=normal', 'correctedError-
251 Rate=0.065, 'canuIteration=1' and 'stopOnReadQuality=false'. During the assembly
252 processes around 2.5% of the corresponding input data (bp) remained unassembled. A
253 search with blastn of the Basic Local Alignment Search Tool (BLAST) package+ v2.8.1
254 [35, 36] and the grapevine chloroplast [37] and mitochondrial sequences [38] was carried out on
255 the haplotype assemblies and detected contigs were removed. Finally, the haplotype
256 assemblies were polished twice with *arrow* v2.2.2 (smrtlink-release_5.1.0.26412) and
257 the corresponding read subsets. Assembly statistics were computed with QUAST v4.6.3
258 [39].
259

260 **Scaffolding with 'Börner' BAC end sequences**

261 To validate the quality of the two genome sequence assemblies, and also to scaffold a
262 few smaller contigs, high-quality paired BAC end sequences from the 'Börner' genome
263 (ENA/GenBank accession numbers KG622866 - KG692309 [10]) were used. The 69,444
264 BAC end sequences cover 39,360,203 bp, have an average sequence length of 566 bp,
265 and represent an eight-fold coverage of the 'Börner' genome based on average BAC
266 insert length. First, the 62,498 BAC end sequences that represented pairs from one BAC
267 were mapped to BoeRip and BoeCin with HISAT2 v2.1.0 [40] using the parameters '--
268 maxins 2000000000', '--secondary', '--no-softclip' and '--no-spliced-alignment'.
269 Alignments with a mapping quality >10 were filtered out and alignment statistics were
270 collected with SAMtools v1.8 [41]. After mapping and quality filtering, the BAC end
271 sequences were used to scaffold the haplotype assemblies with SSPACE (standard)
272 v3.0 [42]. SSPACE was run with the parameters '-x 0', '-m 50', '-k 3', '-n 500' and '-T
273 10' and by using BWA-SW [43] for mapping. Contig extension was disabled ('-x 0') as

274 the BAC end sequences are not phased and thus would eventually contaminate the hap-
275 lotypes with a sequence from the other haplotype. However, a test run with enabled
276 extension showed that the BAC end sequences were anyway not suitable for gap filling.
277

278 **Scaffold to pseudochromosome assignment**

279 To build pseudochromosomes (the nucleotide sequence representation of chromo-
280 somes), an *ad hoc* gene prediction was generated with AUGUSTUS v3.3 [44]. Prior to
281 gene prediction, the RNA-Seq reads from 'Börner' (see above) were processed to create
282 a *de novo* transcriptome assembly with Trinity v2.8.5 [45] with default settings. Hint files
283 with information about exon and intron positions for the haplotype assemblies were
284 created through mapping all trimmed RNA-Seq reads and the transcriptome assembly
285 on the haplotype assemblies. The RNA-Seq reads were mapped with HISAT2 v2.1.0 [46]
286 and soft-clipping disabled ('--no-softclip'). The transcriptome sequences were mapped
287 with BLAT v36x2 [47] and the parameters '-stepSize=5' and '-minIdentity=93'. Hints
288 were generated according to the AUGUSTUS documentation. Additionally, AUGUS-
289 TUS was trained with the CRIBI v2.1 gene annotation of PN40024 [17] to retrieve pa-
290 rameter sets for grapevine. Finally, genes were predicted on both haplotype assemblies
291 BoeRip and BoeCin with the generated *Vitis* parameter set and the hint files.
292 The protein sequence of the primary transcript variant was extracted from the PN40024
293 CRIBI v2.1 annotation and from the *ad hoc* annotation of BoeRip and BoeCin. Recip-
294 rocal best BLAST hits (RBHs) between the protein sets of each of both haplotypes and
295 the filtered proteins of PN40024 were computed with blastp utilizing an e-value thresh-
296 old of 0.0001. For RBHs, two rounds of BLAST searches were performed. Here, one
297 round was run with the proteins of one haplotype as input query and one with the
298 PN40024 proteins as input query. Only RBHs with a percentage of identical matches
299 and a query coverage per subject ≥80% for at least one direction (e.g. PN40024 protein
300 sequences as input query) were kept. Based on RBHs encoded, the scaffolds were as-
301 signed to the pseudochromosomes. Scaffolds with less than 10 RBHs in total and scaf-
302 folds that contain a significant number of RBHs linking a different pseudochromosome
303 (30% distance to the 2nd rank in terms of RBH numbers required) were filtered out for
304 later assignment or moved to the chromosomally unassigned part of the respective as-
305 sembly.

306 Additionally, reciprocal hits between the scaffold sequences and the pseudochromo-
307 some sequences of PN40024 were computed with blastn. The e-value, identity and
308 query coverage filters remained the same as above. The nucleotide RBHs were itera-
309 tively computed and scaffolds assigned to pseudochromosomal positions. If the classi-
310 fications based on protein and nucleotide level contradicted, the protein RBH classifi-
311 cation was preferred.
312 To further refine the pseudochromosomes of the haplotypes, an additional assignment
313 based on VitRGM was performed. Also, reciprocal hits between the scaffold sequences
314 of one haplotype and the pseudochromosomes of the other haplotype and *vice versa*
315 were calculated as described above. After manual construction of pseudochromosomes
316 for the BoeRip and BoeCin phases, the refinement with the corresponding other haplo-
317 type was repeated until no further scaffolds could be assigned. The pseudochromo-
318 somes were constructed with gaps of 100 bp length between the scaffolds and with 10
319 kbp as minimal scaffold length. Due to the length filter, 61 (BoeRip) and 22 (BoeCin)
320 scaffolds were discarded.
321 The orientation and order of the scaffolds on the pseudochromosomes was verified with
322 dotplots and manually adapted if necessary. Thus, DNAdiff v1.3 of the MUMmer pack-
323 age v4.0.0beta2 [48] was run pairwise on homologues pseudochromosomes of BoeRip or
324 BoeCin and PN40024 with default settings and the resulting 1-to-1 alignments were
325 visualized with mummerplot v3.5.
326 The pseudochromosomes were validated with 340 simple sequence repeat (SSR) ge-
327 netic markers from [49] and the 4 SSR markers ATP3 [50], Gf13_11a, VMC8E6 [6] and Gf14-
328 42 [8] (Supplementary File 2 Table S5). Primer sequences were mapped to the pseudo-
329 chromosomes with primersearch of the EMBOSS v6.6.0.0 package [51] and with a blastn
330 search. A total of 314 markers were assigned to unequivocal sequence positions, the
331 remaining markers do not map, show non-evaluable multi-mappings or map with too
332 many mismatches (Supplementary File 2 Table S6). Several markers did map only to
333 one of the haplotypes (16 to only BoeRip and 30 to only BoeCin, see Supplementary
334 File 2 Table S7). Emerging disagreements between the marker position on PN40024
335 and the BoeRip and BoeCin phases were further investigated through all-versus-all dot-
336 plots computed with the webtool D-Genies v1.2.0 [52] and by using long read mapping
337 to detect sequence positions that are not or only very weakly supported by continuously
338 mapping reads. The haplotype specific read subsets together with the unassigned read

339 subset were mapped with minimap2 [53] v2.17 ('-ax map-pb --secondary=no') to the cor-
340 responding haplotype assembly and the coverage values per base were calculated with
341 SAMtools. The haplotype assemblies are both covered around 54x (Supplementary File
342 1 Fig. S2). The back-mapping results of the read subsets were consulted to reveal miss-
343 assemblies if indicated by a genetic marker. A sequence location was further investi-
344 gated if five reads end in a 10 bp region and if the read coverage five bp around the
345 region drops to ≤ five.
346 Conflicting marker VVMD28 maps on chr03 of PN40024 and BoeCin, yet on chr13 of
347 BoeRip. However, eight markers allocate the sequence to chr13 of BoeRip and neither
348 a significant breakpoint was found in the read mappings nor in the all-versus-all dotplot
349 (Supplementary File 1 Fig. S3). VCHR16B maps on chr16 of PN40024 and BoeCin,
350 but on chr03 of BoeRip. Here, one SSR marker assigns the sequence to chr03 and no
351 significant breakpoint was found in the read mapping or in the all-versus-all dot plot.
352 The three SSR marker VVS4, VMC5G6.1 and UDV-126 map on chr08 of PN40024
353 and BoeRip, but on chr04 of BoeCin. The corresponding sequence was assigned to
354 chr04 according to 376 protein RBHs, because only 141 protein RBHs assign it to
355 chr08. Moreover, three SSR markers allocate the sequence to chr04, too. Through in-
356 vestigating the all-versus-all dot plot, an approximately four Mbp large alignment be-
357 tween chr04 and chr08 was found. However, since no position for a split was detected
358 in the read mappings, the sequence remained on chr04 (Supplementary File 1 Fig. S4).
359 GF10-06, VMC3D7 and UDV-073 mapped wrongly on chr02 of BoeRip instead of
360 chr10; in this case the investigation of the respective contig resulted in a split of the
361 sequence and an corrected assignment of the contig fragment in question. Due to 348
362 protein RBHs the sequence was initially assigned to chr02. The sequence showed the
363 second most protein RBHs (171) with chr10. On marker side, 13 markers assign the
364 sequence to chr02 and only three to chr10. However, the all-versus-all dotplot showed
365 a large alignment between chr02 of BoeRip and chr10 of PN40024 (Supplementary File
366 1 Fig. S5), and the read mappings indicated a breakpoint position with a coverage drop
367 to ≤ 5. Thus, the 12.3 Mbp sequence was split into a 7.7 Mbp large sequence that was
368 assigned to chr10 and a 4.6 Mbp large sequence that remained at chr02. Another case
369 was detected in the all-versus-all dot plots between chr13 of BoeCin and chr03 of
370 PN40024 (Supplementary File 1 Fig. S4). The corresponding sequence was assigned to
371 chr13 through 485 protein RBHs. Only 36 protein RBHs would allocate this sequence

372 to chr03. Furthermore, 10 SSR marker support the assignment to chr13 and none to
373 chr03. No coverage drop was found in the read mappings, the sequence remained on
374 chr13 of BoeCin. Finally, only five genetic markers remained unresolved (VVMD28,
375 VCHR16B, VVS4, VMC5G6.1, UDV-126).
376 To estimate the completeness of the assemblies, the plant core genes were localized
377 with the program Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.2
378 utilizing the database 'eudicots_odb10' (2,326 genes) [54, 55]. For comparison, the BUS-
379 COs of PN40024 and of VitRGM were also determined with identical parameters. The
380 pseudochromosome lengths were visualized with cvit v1.2.1 [56].
381

382 **Validation of phasing**

383 To validate the phasing of the 'Börner' haplotypes, the BAC sequence data of pool1-4
384 were *de novo* assembled with Newbler v2.6. The data of pool5-8 were assembled with
385 CLC Genomics Workbench 8.0 using a word size of 30. Remaining vector sequences
386 were removed with vectorstrip of the EMBOSS package v6.2. Assembled sequences
387 shorter than 500 bp were discarded. The short reads of the parents *V. riparia* GM183
388 and *V. cinerea* Arnold were mapped with CLCs' map reads to reference algorithm to
389 the assembled BAC sequences (linear gap cost, length & similarity fraction 1.0). Based
390 on these mapping results regarding read coverage and percent of covered bases, the
391 BAC sequences were assigned to either *V. riparia* GM183 or *V. cinerea* Arnold. BAC
392 sequences with no read coverage were discarded. The assembled and phase-separated
393 BAC contig ssequences are available at https://doi.org/10.4119/unibi/2962639.
394 The 'Börner' BAC sequences with known haplotype allocation were mapped against
395 the BoeRip and BoeCin haplotype assemblies with minimap2 v2.17 [53], sorted with
396 SAMtools and the mappings visualized with the Integrative Genomics Viewer (IGV)
397 v2.5.2 [57, 58]. The number of base calls of the aligned BAC sequences on each haplotype
398 were determined with SAMtools' mpileup. Continuously and correct aligned BAC se-
399 quences show high base calls and low numbers of SNP/InDels (small nucleotide poly-
400 morphism, insertion-deletion polymorphism) to one of the two haplotypes and the op-
401 posite result with the alternative haplotype.
402

403 **k-mer based reference-free assembly evaluation with Mercury**

404 Merqury v1.3 [59] was used to evaluate the diploid BoeRC assembly. First, the best k-
405 mer size was determined with Merqurys best_k.sh script for the expected 500 Mbp hap-
406 loid genome size resulting in 19-mers. Consequently, 19-mer databases were computed
407 for the *V. riparia* GM183 and *V. cinerea* Arnold Illumina reads as well as for existing
408 Illumina reads from 'Börner' [10]. With these, hap-mer (haplotype-specific k-mers as de-
409 fined by Mercury [59]) databases were generated and the haplotype assemblies evaluated
410 using the scaffolds of both haplotypes. For building of phased blocks, the parameter
411 'num_switch 10' and 'short_range 20,000' were applied. For comparison, Merqury was
412 also run on PN40024, on VitRGM, on the Cabernet Sauvignon haplotype assemblies
413 v1.1 and on the *Muscadinia rotundifolia* cultivar 'Trayshed' haplotype assemblies v2.0
414 [23]. The 19-mer profiles were computed from PN40024 WGS reads SRR5627797, from
415 VitRGM WGS reads SRR8379638, from Cabernet Sauvignon WGS reads
416 SRR3346861 and SRR3346862 and from *Muscadinia rotundifolia* WGS reads
417 SRR6729333. Prior to analyses, the reads were trimmed with Trimmomatic-v0.39 al-
418 lowing a minimum read length of 80 or 90 (PN40024 reads).
419

420 **Gene annotation**

421 The 'Börner' haplotypes were annotated *ab initio* with MAKER v3.01.03 following the
422 MAKER-P pipeline [60, 61]. As input data, haplotype specific repeat libraries were created
423 according to the MAKER advanced repeat library protocol. The monocotyledons repeat
424 library of RepBase (RepeatMaskerEdition-20181026, model_org=monocotyledons) [62],
425 the transposable element (TE) sequences from MAKER, the 'Börner' *de novo* transcrip-
426 tome assembly described above, *Vitis* protein sequences (NCBI, Protein DB "Vitis"[Or-
427 ganism]), plant protein sequences (UniProt, release 2020_02, "Viridiplantae [33090]"
428 AND reviewed:yes) and *Vitis* full-length cDNAs (NCBI, Nucleotide DB "Vitis" [Or-
429 ganism] AND complete cds[Title]) were used. The gene predictors SNAP v2006-07-28
430 (three rounds) [63] and GeneMark v3.60 were trained on both, BoeRip and BoeCin. AU-
431 GUSTUS v3.2.3, tRNAscan-SE v1.3.1 and EvidenceModeler v1.1.1 [64], SNAP and
432 GeneMark together were ran through MAKER to compute the final gene annotation.
433 The option 'split_hit' was set to 20,000 and the previously generated *Vitis* parameter
434 set was adjusted for AUGUSTUS.

435 The gene models from BoeRip and BoeCin were refined with PASA v2.4.1 [65]. As evi-
436 dence data, the *de novo* and reference-guided 'Börner' transcriptome assemblies, the
437 *Vitis* full-length cDNAs and *Vitis* EST data (NCBI, txid3603[Organism:exp] AND
438 is_est[filter]) were used. The parameter file is attached as Supplementary File 3 Data
439 S1. The refinement with PASA was iteratively applied three times on the gene models.
440 The reference-guided assembly was computed with Trinity v2.10.0 using the 'Börner'
441 RNA-Seq data aligned with HISAT2 v2.2.0. Only alignments with MQ >10 were given
442 to Trinity and '--genome_guided_max_intron' was set to 20,000.
443 After structural gene annotation, a functional gene annotation was added through a
444 BLAST search (package+ v2.10.0) against the UniProt/Swiss-Prot database (release
445 2020_02), through domain identification with InterProScan5 v5.42-78.0 [66] and the
446 PFAM database v32.0. Gene models with Annotation Edit Distance (AED) > 0.5 were
447 filtered from the final gene model set. Also, predicted genes expected to encode a pol-
448 ypeptide with less than 50 amino acids and no functional annotation were discarded.
449 The annotation data for BoeRC are available at
450 (https://doi.org/10.4119/unibi/2962793). For comparison, functional annotation was
451 also carried out for the PN40024 VCost.v3 structural gene annotation with identical
452 resources.
453

454 **Prediction of resistance genes**

455 Resistance gene analogs (RGAs) were predicted for both haplotypes with the pipeline
456 RGAugury v2.1.7 [67]. RGAugury employed ncoil [68], PfamScan v1.6 [69], InterProScan5
457 v5.45-80.0 [66] and Phobius-1.01 [70] on the protein sequences to predict known resistance
458 domains. The initial filtering against the RGAdb was disabled as the database was not
459 updated since 2016, unknown RGAs not included in the database were of interest and
460 the identification of (resistance) domains for none RGAs was enabled. For Inter-
461 ProScan5 domain search through RGAugury, the databases Gene3D-4.2.0 [71], Pfam-
462 33.1 [72], SMART-7.1 [73] and SUPERFAMILY-1.75 [74] were utilized.
463 The RGA annotations of the two haplotypes were correlated through determination of
464 RBHs based on protein sequences. Only RBHs with an identity and coverage ≥ 90 were
465 considered.
466

467 **Assembly wide variant detection**

468 The SNP and InDel positions of the SNP output from the DNAdiff analysis (see above)
469 were compared with the positions of the coding and non-coding gene regions and clas-
470 sified accordingly.
471

472 **Genetic mapping**

473 The annotated genes and nucleotide sequences of the phylloxera resistance QTL on
474 chr13, referred to as *Rdv1*, of BoeRip, BoeCin, PN40024 and VitRGM were compared.
475 Before this study, the QTL *Rdv1* was delimited with the associated genetic markers
476 GF13-01 and GF13-12 [6]; the marker positions were identified based on the respective
477 marker assay primer sequences with primersearch (EMBOSS package [51]).
478 To narrow down the location of *Rdv1*, additional genetic fine mapping was carried out
479 with the mapping population of 188 individuals that resulted from a cross of *V. vinifera*
480 V3125 ('Schiava Grossa' x 'Riesling') x 'Börner' [6] and newly designed sequence
481 tagged SSR (STS) marker. The additional markers were derived from PN40024 [16] (Sup-
482 plementary File 2 Table S5) and used to screen the F1 genotypes from the mapping
483 population.
484 Five relevant genotypes with marker data indicating recombination events within the
485 initial *Rdv1* QTL were subjected to GBS to locate the recombination points between
486 the two haplotypes of 'Börner'. V3125 was included as a control for comparative read
487 mapping. The SE reads were quality trimmed using Trimmomatic-0.39[33] with the fol-
488 lowing parameters: ILLUMINACLIP: 4:30:15 LEADING:30 TRAILING:30 SLID-
489 INGWINDOW:4:15 MINLEN:60. Trimmed reads of each of the six studied genotypes
490 as well as reads from 'Börner' [10] were separately mapped to BoeCin and BoeRip using
491 the HISAT2 version 2.2.0 [46] with the following parameters: --no-softclip –no-spliced-
492 alignment. The obtained SAM files were converted to sorted BAM files as well as in-
493 dexed using SAMtools. Variant calling against BoeRC and estimation of genomic re-
494 combination sites was supported by QIAGEN CLC Genomics Workbench 22.0 (QI-
495 AGEN, Aarhus, Denmark). SNPs detected in the *Rdv1* region on chr13 for each geno-
496 type were filtered with respect to read coverage and frequency. Homozygous variants
497 were counted if > 90% of the reads support the variation, for heterozygous variants a
498 frequency between 25% and 75% was required. An example case (GBS-04) for the

499 assessment of GBS markers is shown in Supplementary File 1 Fig. S6. Recombination
500 sites were determined by monitoring the switch from either the BoeCin hapotype to the
501 BoeRip hapotype or vice-versa between a set of at least two variant positions.
502

503 **Allele sequence comparison and TE detection**

504 The genes (alleles, orthologs) of the *Rdv1* region of BoeRip, BoeCin, PN40024 and
505 VitRGM between the markers GF13-03 and GF13-07 were placed into groups with
506 OrthoFinder v2.3.11 [75] using the longest protein sequence annotated for the splicing
507 variants of each gene. Missing alleles (or, alternatively phrased, missing predictions of
508 orthologous genes at syntenic positions) were investigated and re-annotated by aligning
509 the protein sequences of the other cultivars against the genome sequences with exoner-
510 ate v2.4.0 [76]. Only alignments with at least 70% coverage of the protein sequence were
511 considered. To obtain gene annotations for remaining poorly annotated sequences de-
512 spite existing RNA-Seq support in the genomic region studied, BRAKER was run with
513 strictly filtered RNA-Seq mappings (only uniquely mapped PE reads on both 'Börner'
514 haplotypes offered as combined target, both reads of a pair must map); adding UTR
515 regions and prediction of alternative transcripts from RNA-Seq data was enabled. Since
516 the focus here is on phylloxera resistance at the root, an additional AUGUSTUS species
517 model was trained on BoeRC with the uniquely mapped root RNA-Seq reads, the above
518 described *Vitis* protein and plant protein sequences and eudicot protein sequences from
519 OrthoDB v10.1 [77]. The results were manually evaluated and restored gene models of
520 BoeRip and BoeCin were refined two times with PASA. Through investigation of the
521 RNA-Seq mappings from all available tissues to BoeCin, the gene predictions in the
522 *Rdv1* region of all four sequence regions studied were manually screened for low sup-
523 ported gene models. Transcripts encoding for protein sequences with less than 50 amino
524 acids, tagged as 'protein of unknown function' (PUF) and with no ortholog were re-
525 moved. RGA classes were determined for the manually curated genes/alleles with
526 RGAugury as described above. TEs were identified with the Extensive *de novo* TE
527 Annotator (EDTA) [78]. Finally, the results were manually collected in allelic represen-
528 tations of the genes of the *Rdv1* region of BoeRip, BoeCin, PN40024 and VitRGM. An
529 extended version of the gene overview at the *Rdv1* locus that includes TEs has also been
530 prepared (Supplementary File 1 Fig. S7).

Results

Sequencing data of ‘Börner’ and its parents

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

A ‘Börner’ genome sequence assembly in two phases

To generate a ‘Börner’ genome sequence with two separated haplotypes, the trio-binning approach was applied [28]. The two haplotype-specific ‘Börner’ read subsets or bins, designated Vrip and Vcin, contain approximately 50% of all reads and bases. These were assembled with Canu (see Methods). Statistics of the two resulting haplotype assemblies are shown in Table 1. Both assemblies reached the expected genome size of ~500 Mbp. The BoeRip and BoeCin haplotype assemblies represent "1n double haploid" genome sequences of the two species *V. riparia* and *V. cinerea*, respectively.

552 **Table 1** Assembly statistics for the raw and scaffolded haplotype assemblies BoeRip
553 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

554 [a]scaffolded haplotype assemblies
555 [b]raw haplotype assemblies, numbers include contigs shorter than 10 kbp
556 [c]see Supplementary File 1 Fig. S8 for details
557

558 **Pseudochromosome construction and validation by genetic map**

559 The assignment of scaffolds to pseudochromosomes was mainly based on the PN40024
560 reference sequence and was achieved by using RBHs (see Methods). For BoeRip, 446
561 scaffolds accounting for 95.21% of all bases were assigned to pseudochromosomes. For
562 BoeCin, 320 scaffolds holding 93.89 % of all bases were assigned. Since the reference
563 contains a pseudochromosome chr00 for unassigned sequences, this has been created
564 as well and is referred to as "chrUn". In addition, some scaffolds find no clear corre-
565 sponding sequence in the reference and result in a pseudochromosome "chrNh" (for no
566 hit, see Fig. 2A). As a result, the haplotype assemblies each consist of 21 pseudochro-
567 mosome sequences representing the 19 true chromosomes and two artificial sequences.
568 To validate the pseudochromosomes, 314 SSR markers that show unequivocal chromo-
569 some mappings on the haplotype assemblies were evaluated relative to their position
570 on PN40024. A total of 309 markers confirm the pseudochromosome assignment. Con-
571 flicting markers were checked manually and corrected if possible (see Methods).
572 The completeness of the assemblies was evaluated by detection of plant core genes with
573 BUSCO (Fig. 2B). For both BoeRip and BoeCin, about 96.5% of the conserved single
574 copy gene set were detected. This indicated that both haplotype assemblies are more
575 complete than the assemblies PN40024 and VitRGM.
576

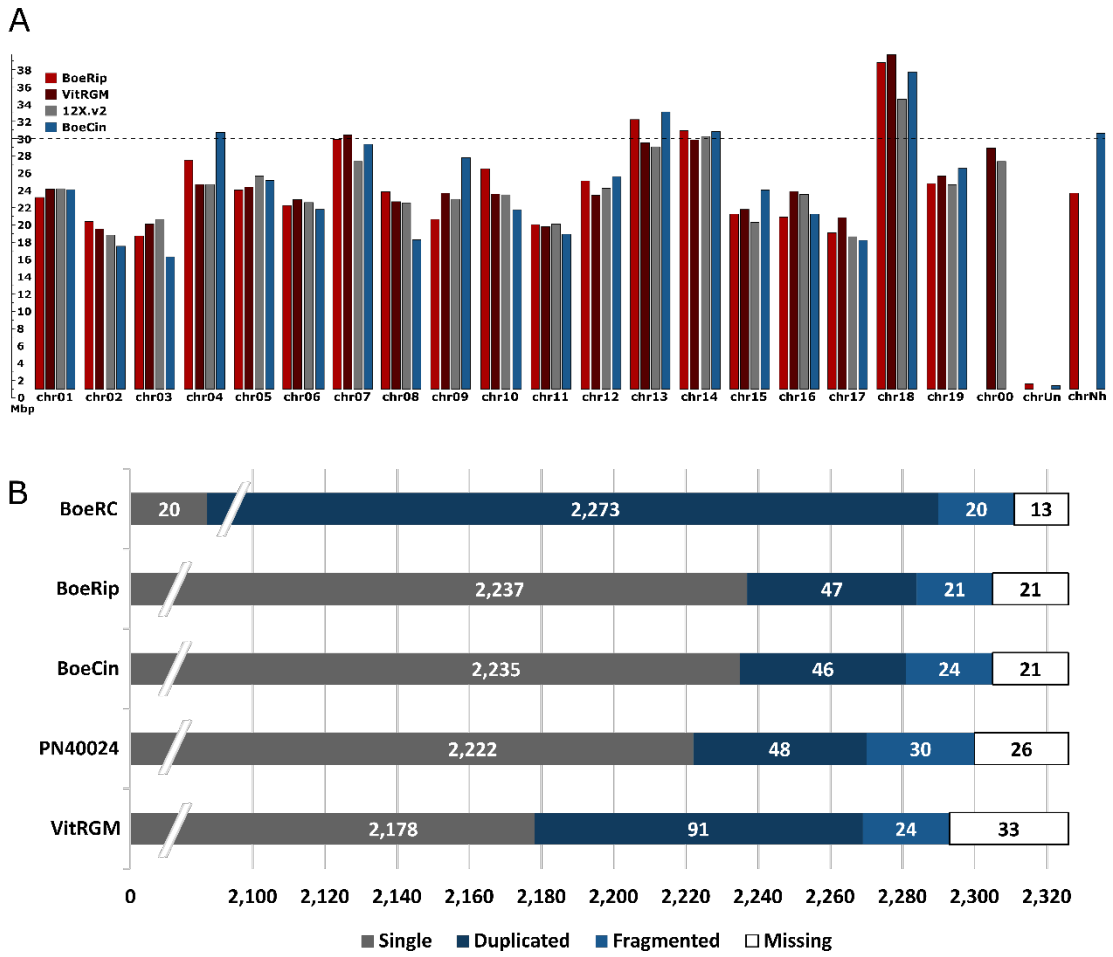


Fig. 2 Comparison of the 'Börner' genome sequence assembly with other *Vitis* assemblies. (A) Pseudochromosome lengths comparison of both 'Börner' haplotypes, PN40024 and VitRGM. The bars represent the pseudochromosomes of different assemblies. Red, BoeRip; dark red, VitRGM; grey, PN40024; blue, BoeCin. ChrUn holds sequences that were assigned to chr00 of PN40024 or VitRGM, and chrNh collects sequences with no assignment (Supplementary File 2 Table S8). **(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; results were summed up).

Evaluation of assembly quality and phasing

For validation of the phasing, BAC sequences (Supplementary File 2 Table S9) with known haplotype/parental origin were mapped on BoeRip and BoeCin (Supplementary File 2 Table S10). 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

595 S3). BAC contigs of the same haplotype map with almost no sequence difference, while
596 BAC contigs derived from the other haplotype display a wide range of mismatches and
597 InDels (Supplementary File 1 Fig. S9). On both haplotypes, ~3.6 Mbp (chr01) and ~5.5
598 Mbp (chr14) were covered by BAC contigs of both phases and support a correct phasing
599 over 9 Mbp.
600 In addition, the assembly quality evaluation tool Merqury (phasing assessment for ge-
601 nome sequence assemblies) was used to assess the quality of both haplotype assemblies.
602 The overall base quality value (QV, consensus quality value, representing the log-
603 scaled probability of error for consensus base calls) of 37.42 found for BoeRC indicates
604 a very high level of correctness (see Supplementary File 2 Table S11 for comparisons
605 to other assemblies). It should be noted that Merqury requires Illumina data for the
606 complete trio (both parents and F1 as available for 'Börner' and its ancestors) to calcu-
607 late the complete set of quality values.
608 Also the phasing accuracy was evaluated with Merqury (Supplementary File 1 Fig.
609 S10). Based on haplotype-specific k-mers (referred to as hap-mers in Mercury), 2,759
610 phase blocks with an N50 of 2.25 Mbp were calculated for BoeRip and 1,852 blocks
611 with an N50 of 2.42 Mbp for BoeCin. Almost all bases were covered by a block and
612 the switch error rate was estimated to be less than 0.05% (Supplementary File 2 Table
613 S12; Supplementary File 1 Fig. S11). Overall, the quality of the 'Börner' genome se-
614 quence as well as its separation into haplotypes derived from *V. riparia* GM183
615 (BoeRip) and *V. cinerea* Arnold (BoeCin) is considered to be high.
616

617 **Sequence similarities and variations between BoeRip, BoeCin, PN40024 and**
618 **VitRGM**

619 All-versus-all alignments between the pseudochromosomes of the two 'Börner' haplo-
620 types, and between each haplotype and PN40024 as well as VitRGM, were computed
621 and visualised as dot plots (Fig. 3, Supplementary File 1 Fig. S3, S4, S12, S13). All
622 homologous pseudochromosomes show strong synteny, still, some rearrangements and
623 smaller and larger gaps indicating insertions, deletions and/or missing sequence were
624 detected.

On average 90.20% of all bases with an identity of 95.52% were aligned between the pseudochromosomes of the two haplotypes exclusive of chrUn and chrNh (see Methods). The SNP and InDel frequency for the protein coding fraction was 1/1,001 bp and 1/34 bp for the non-coding fraction of the genome sequence. When aligning with PN40024, on average 88.59% and 87.51% of all bases were aligned with an identity of 94.82% (BoeRip) and 94.94% (BoeCin) over all pseudochromosomes, respectively. The alignments resulted in a SNP and InDel frequency of 1/932 bp (BoeRip) and 1/935 bp (BoeCin) for the coding regions and of 1/32 bp for the non-coding regions.

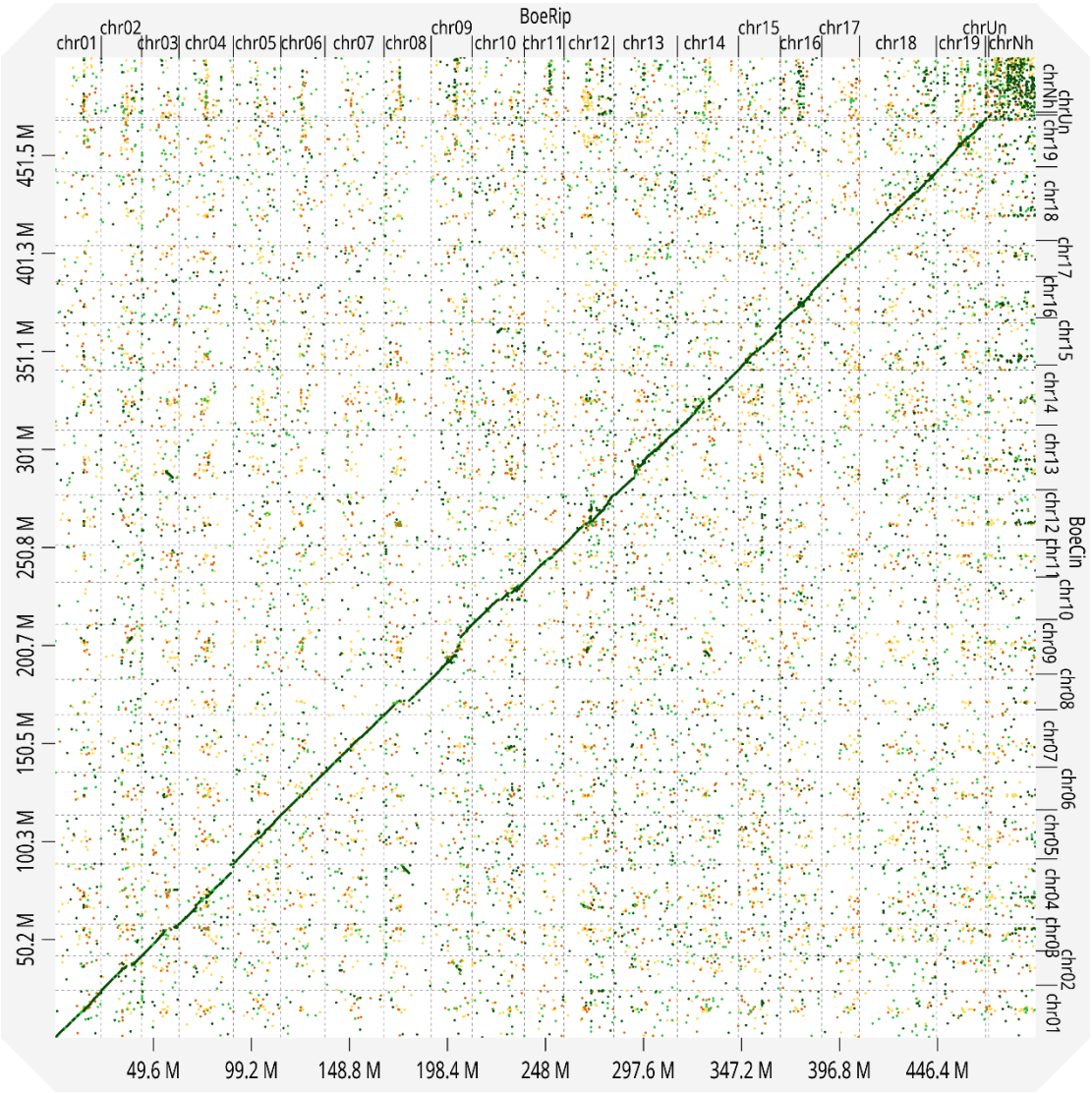


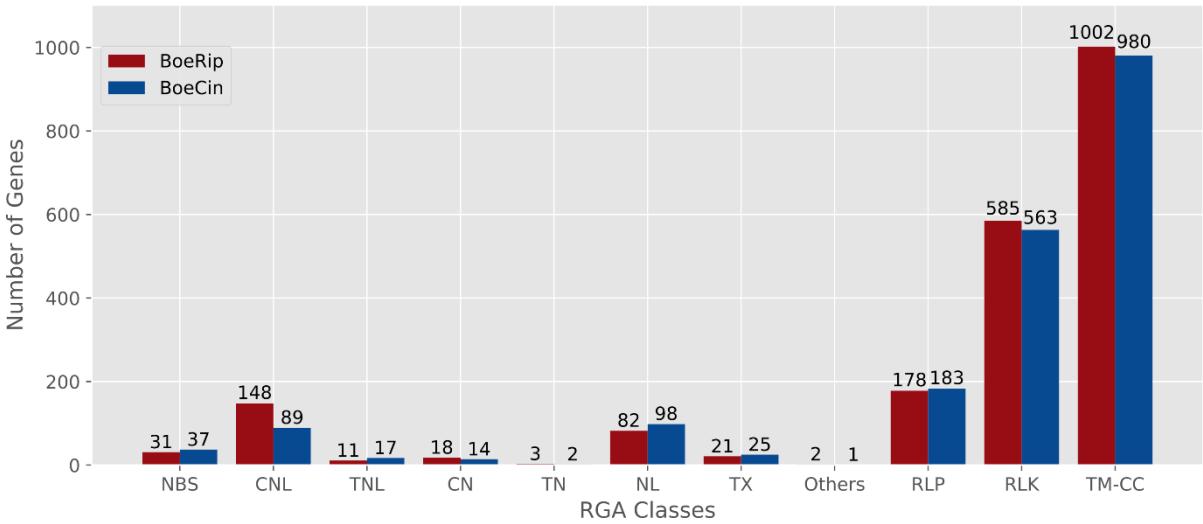
Fig. 3 All-versus-all dot plot of both 'Börner' haplotypes. The graphic shows the dot plot between concatenated pseudochromosomes of BoeRip and BoeCin.

638 **Gene annotation**

639 In total, ~274 Mbp (55.24%, BoeRip) and ~278 Mbp (55.46%, BoeCin) repetitive se-
640 quences were identified with the MAKER pipeline (Supplementary File 2 Table S13).
641 The final gene annotation comprises 27,738 and 27,995 protein-coding genes and 672
642 and 701 tRNAs for BoeRip and BoeCin, respectively. Of the protein-coding genes,
643 19,500 genes of each haplotype were captured in RBHs. Based on RBHs, BoeRip shares
644 17,922 genes with VCost.v3 (PN40024) and 18,928 genes with VitRGM; BoeCin
645 shares 17,837 genes with VCost.v3 (PN40024) and 18,126 genes with VitRGM.
646

647 **Resistance genes**

648 The annotated genome sequence of 'Börner' was used to identify RGAs based on their
649 typical protein domain structure. Resistance gene annotation with RGAugury revealed
650 2,072 RGAs for BoeRip and 2,007 RGAs for BoeCin (Fig. 4). Of these, 1,509 were
651 classified as RBHs between BoeRip and BoeCin.
652



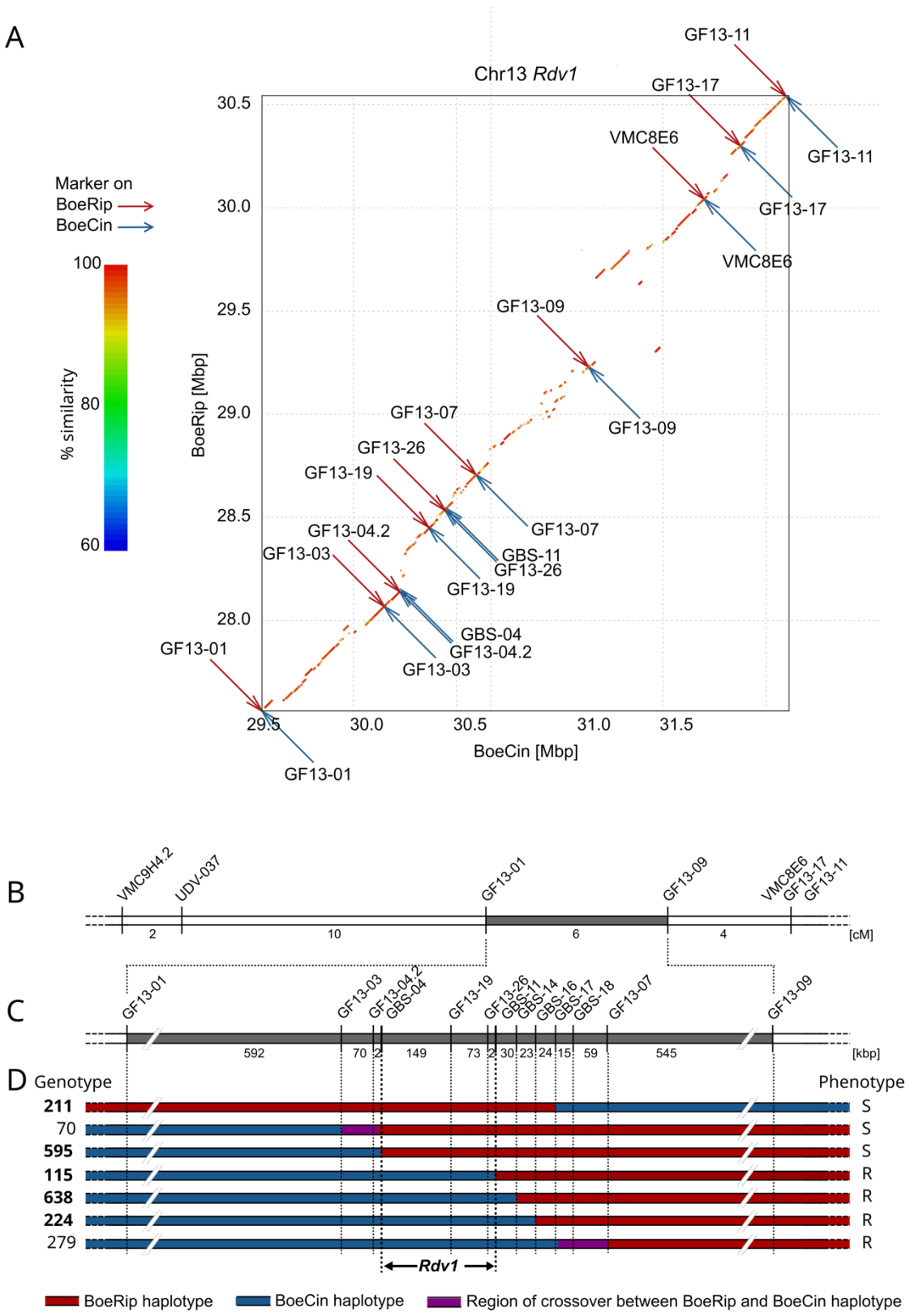
653

654 **Fig. 4 Distribution of genes to RGA classes based on characteristic protein**
655 **domains.** The red bars display the number of RGA genes of BoeRip and the blue bars
656 the RGA genes of BoeCin. RGAs were classified according to the domains encoded by
657 the predicted genes. Designations are according to [67]. NBS, Nucleotide Binding Site
658 (abbreviated with N in other acronyms); CNL, CC-NBS-LRR with CC for Coiled-Coil
659 and LRR for Leucine Rich Repeat; TNL, TIR-NBS-LRR with TIR for Toll/Interleukin-
660 1 Receptor; CN, CC-NBS; TN, TIR-NBS; NL, NBS-LRR; TX, TIR-unknown domain;

661 Others, CC-TIR; RLP, Receptor Like Protein; RLK, Receptor Like Kinase; TM-CC,
662 TM for Transmembrane region.
663

664 **Candidate genes for *Rdv1***

665 At the sequence level, the region between GF13-01 and GF13-11 that contains *Rdv1*
666 displays a highly diverged structure between BoeRip, BoeCin, PN40024 and VitRGM
667 (Fig. 5A, Supplementary File 1 Fig. S14, S15).
668 The interval flanked by these genetic markers includes 2,978,547 bp with 179 genes for
669 BoeRip and 2,538,570 bp with 173 genes for BoeCin, according to the BoeRC assembly
670 and its annotation, as well as 2,796,801 bp with 273 genes for PN40024 and 5,854,243
671 bp with 450 genes for VitRGM according to published data. The interval in VitRGM,
672 which has been generated with a FALCON-Unzip represents a pseudo-haplotype, be-
673 tween GF13-01 and GF13-11 is twice as large as in the other assemblies.
674 To further delimitate *Rdv1*, new genetic markers were designed and used to establish a
675 local map. F1 individuals with different flanking haplotypes were selected from the
676 mapping population V3125 x 'Börner', genotyped and tested for phylloxera resistance.
677 Since marker assay design tuned out to be difficult for gene-level resolution, GBS was
678 applied to five crucial F1 genotypes. Mapping of SSR and GBS markers as well as the
679 presence or absence of the resistance encoded by BoeCin led to delimitation of *Rdv1* to
680 a region between the markers GBS-04 and GBS-11 (Fig. 5B-D, Supplementary File 2
681 Table S14). The GBS markers were used to precisely locate recombination sites that
682 were identified between SSR markers. Since the exact positions of the GBS markers
683 are unknown for PN40024 and VitRGM, the markers GF13-04.2 and GF13-26, located
684 very close to the markers GBS-04 and GBS-11, were used to specify the interval size.
685



686

687

688 **Fig. 5 Integration of genetic map and haplotype-specific physical map of the *Rdv1***
689 **region on pseudochromosome 13. (A)** Dot plot of the *Rdv1* region between BoeRip

690 and BoeCin. Genetic markers mapping on BoeRip are represented as red arrows and
691 markers mapping on BoeCin as blue arrows. The *Rdv1* locus is located between GBS-
692 04 and GBS-11. (**B-D)** Fine mapping of the phylloxera resistance locus *Rdv1*. **(B)**
693 Genomic region of the genetically mapped *Rdv1* locus on chromosome 13 of 'Börner'.
694 The grey bar represents the interval between the delimiting SSR markers GF13-01 and
695 GF13-09 [6], the numbers indicate the genetic distance between adjacent markers in cM.
696 **(C)** Physical position of the SSR and GBS markers used for fine mapping of this region.
697 The bar depicts the enlarged region taken from (B). Numbers refer to the physical
698 distance of adjacent markers in kbp based on BoeCin (haplotype conferring resistance).
699 **(D)** Local map of seven relevant F1 individuals from the V3125 x 'Börner' population
700 revealed after haplotype-specific genotyping, only the 'Börner' haplotypes are shown.
701 F1 genotypes analysed by GBS are highlighted in bold. The phenotypes are indicated
702 on the right side: R, resistant; S, susceptible. Red and blue bars indicate the BoeRip and
703 BoeCin haplotypes, respectively. Regions with unlocalised recombination sites are
704 shown in purple.
705
706 Based on the gene annotation data generated for the assemblies, the RGA studies and
707 intensive manual curation (see Methods), the genetically delimited *Rdv1* locus was
708 evaluated in detail at the gene level (Fig. 6). The locus has a size of 402,698 bp with 25
709 genes and 13 resistance genes in BoeRip, a size of 223,738 bp with 20 genes and eigth
710 resistance genes in BoeCin, a size of 202,484 bp with 15 genes and seven resistance
711 genes in PN40024 and 208,610 bp with 16 genes and four resistance genes (all detected
712 as RGAs) in VitRGM. Gene structures that belong to TEs are counted as "normal"
713 genes. The *Rdv1* locus is covered by a single contig in both 'Börner' haplotypes.
714 A resistance gene cluster that displays a very divergent structure and different gene
715 content among the three analysed haplotypes and the pseudo-haplotype VitRGM is pre-
716 sent between the dehydrogenase encoding gene (dark bue in Fig. 6, annotation ID *Boe-*
717 *Cin13g18380*, tagged by GF13-04.2/GBS-04) and the ligase encoding gene (turquoise
718 in Fig. 6, annotation ID *BoeCin13g18383*, tagged by GF13-26/GBS-11). The haplotype
719 conferring dominant resistance (BoeCin) contains eight potential resistance genes (Sup-
720 plementary File 2 Table S15). Three of them, located in the southern part of the *Rdv1*
721 locus (*BoeCin13g18393*, CNL class; *BoeCin13g18396*, TM-CC class; and *Boe-*
722 *Cin13g18399*, TM-CC class), are located in a synthenic region that contains similar
723 alleles of the genes detected in all four haplotypes (see Discussion). In the following,
724 the distiction between true haplotype sequences (PN40024, BoeRip and BoeCin) and
725 the pseudo-haplotype VitRGM, that may contain merged sequences from the two

phases of *V. riparia* Gloire de Montpellier, is neglected. Due to presence of these genes in sensitive genotypes, they were considered not to be relevant for phylloxera resistance. The remaining five potential resistance genes (*BoeCin13g18381*; *BoeCin13g18382*, CNL/RLP; *BoeCin13g18388*; *BoeCin13g18389,* NL; *BoeCin13g18390*) can be considered as candidate genes for causing phylloxera resistance. *BoeCin13g18381, BoeCin13g18382* and *BoeCin13g18389* are putative disease resistance genes encoding RPP13-like proteins. The other two (*BoeCin13g18388*, *BoeCin13g18390*) are similar to the *Arabidopsis thaliana* resistance gene *AtLRRAC1*.

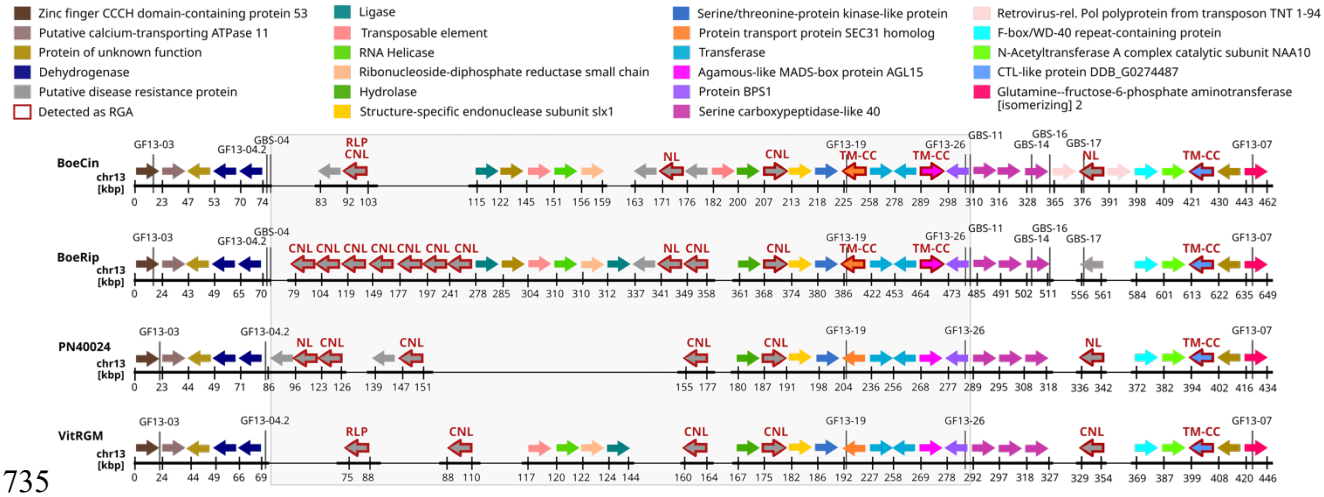


Fig. 6 Genes of BoeCin, BoeRip, PN40024 and VitRGM at and surrounding the *Rdv1* locus. The genes (alleles) were clustered according to similarity and are shown in one colour for each orthogroup. The grey box represents the *Rdv1* locus delineated by the genetic markers GF13-04.2/GBS-04 and GF13-26/GBS-11; additional markers are included upstream and downstream of the *Rdv1* locus. The kbp values on the axis were adapted such that the start of the first gene is bp zero (relative coordinates). At the top, the colour code for each group of related genes/alleles and the corresponding functional annotation is given. Grey coloured genes are potential resistance genes according to their functional annotation; genes with a red edging were identified as RGA (see text). Above each RGA, the domain type classification of the encoded protein is mentioned. See legend to Fig.4 for acronyms.

Discussion

Final 'Börner' genome sequence assembly

For the fully phased genome assembly, the 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. With this basis, 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-Unizp, the genome assembly size was not over- or underestimated by Canu and thus reaches for both haplotypes approximately 500 Mbp (Supplementary File 2 Table S11). 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, 31] 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 S11). 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 Sauvignon [20], more than 98% of complete core genes (BUSCOs) were identified.

Phasing

The haplotype assemblies can be considered as fully and correctly phased. Through mapping of phased BAC sequences derived from parts of chr01 and chr14 of 'Börner',

779 it was shown that several Mbp of chr01 and chr14 are truly phased. The haplotype spe-
780 cific BAC sequences coincide with the correct haplotype and share plenty of mis-
781 matches with the other haplotype. Also, the mapping validated correctness of contiguity
782 and chromosome assignment in those regions.
783 Additionally, the phasing was analysed through k-mer analysis with Merqury. The k-
784 mer completeness of BoeRC over both haplotypes was rated to 99.04% (BoeRip) and
785 99.51% (BoeCin), while haplotype precision was estimated to 99.07% (BoeRip) and
786 99.92% (BoeCin) (Supplementary File 2 Table S11). Thus, the haplotype assemblies
787 can be considered as completely and correctly phased. The purity of the haplotypes was
788 also expressed through the very low switch error rate of 0.046% (BoeRip) and 0.035%
789 (BoeCin) and the high block N50 value of 2.25 Mbp (BoeRip) and 2.42 Mbp (BoeCin).
790 None of the scaffolds showed a notable amount of k-mers from the other haplotype
791 (Supplementary File 1 Fig. S11).
792 There would potentially be room for improvement of the assemblies BoeRip and Boe-
793 Cin by methods like chromosome conformation capture sequencing (Hi-C). Even if
794 there are indications that the organization of chromosomes into territories within the
795 nucleus of diploid organisms could allow to extract data supporting correct haplotype
796 separation [79], we did not follow this route and anticipate that the quality of the BoeRC
797 assembly is sufficient for almost any application with relevance to viticulture.
798

799 **Similarity of 'Börner' haplotypes and comparison with PN40024 and VitRGM**

800 The assemblies BoeRip and BoeCin show overall similarity to each other and to
801 PN40024. Even if the haplotypes display with each other almost the same SNP fre-
802 quency as with PN40024, more precisely 1 SNP in 33 bp compared to 1 SNP in 31 bp,
803 more bases were aligned between the haplotypes with a higher identity than with
804 PN40024. The difference in SNP frequency becomes more specific when comparing
805 the SNP frequency of the coding regions. Here, BoeRip and BoeCin show less SNPs
806 (1/1,001 bp) than with PN40024 (1/932 bp BoeRip & 1/935 bp BoeCin). However,
807 because of the high heterozygosity between haplotypes of grapevine [80], it is not surpris-
808 ing that the SNP frequency between 'Börners' haplotypes is almost the same as to an-
809 other *Vitis* species. Reported SNP frequencies for grapevines range from 1 SNP per 64
810 bp [81] to 1 SNP per 200 bp [17]. The SNP frequency between *V. riparia* and *V. vinifera*

811 was deterimined to 1 SNP per 78 bp [82]. Generally, it seems that SNP frequencies be-
812 tween and within *Vitis* species are not that different.
813 High sequence identities and a good chromosome assignment were revealed trough dot
814 plots between homologous pseudochromosomes of the haplotypes of BoeRC and
815 PN40024 or VitRGM. The "point clouds" often observed in the dot plots between ho-
816 mologous pseudochromosomes can be assumed to pinpoint the location of the centro-
817 mere as centromeric regions are often highly repetitive and thus hard to sequence and
818 assemble. The gaps detected in comparison with PN40024 may represent missing bases
819 in the PN40024 genome sequence assembly. Almost all identified rearrangements be-
820 tween the haplotypes of BoeRC and PN40024 were found between both haplotypes and
821 PN40024, and most of them are inversions. This calls for an improvement of the
822 'PN40024' reference genome sequence based on long read sequence data.
823

824 ***Rdv1* locus**

825 The *Rdv1* locus has been genetically mapped to a region of 224 kbp in size in the Boe-
826 Cin haplotype of the 'Börner' genome. The genome regions orthologous to the *Rdv1*
827 locus from BoeCin and different *Vitis* species are flanked on both sides by large syn-
828 thenic regions. The high variability and divergence that is known for resistance gene
829 clusters [83] was also detected within the *Rdv1* locus.
830 In synthenic regions, alleles may vary significantly in lengths between haplotypes, still
831 they encode the same protein even among different *Vitis* species. For example, the
832 PN40024 gene *Vitvi13g01613* which encodes a putative glycerol-3-phosphate dehydro-
833 genase (dark blue gene in Fig. 6, northern part of the locus), is 3,600 bp longer than its
834 BoeCin ortholog. The 466 amino acid long protein sequence is identical between all
835 haplotypes analysed except one amino acid substitution each in PN40024 and BoeRip.
836 Increased gene length of *Vitvi13g01613* in PN40024 is caused by a larger third intron.
837 Several TE fragments and especially one larger LTR/Copia-like TE explain the differ-
838 ence in the length of intron three.
839 In addition to SNPs, quite some presence/absence variation (PAV) contributes to the
840 divergence of the *Rdv1* region in *Vitis*. The various PAVs were, at least in several cases,
841 explained by detection of TEs (either DNA transposons or retrotransposons) in the hap-
842 lotypes (Supplementary File 1 Fig. S7). For example, two intact LTR/Copia TEs and

843 two intact DNA/hAT TEs were detected in the *Rdv1* region of BoeRip but no intact
844 TE(s) of these classes in the *Rdv1* region of BoeCin. In the *Rdv1* region of VitRGM,
845 one intact DNA/DTC TE was found but no intact TE of these class in the *Rdv1* region
846 of in PN40024.
847 The resistance gene *BoeCin13g18388* encodes a LRR domain and is around 2,400 bp
848 longer in BoeCin than in BoeRip (*BoeRip13g18553*). The reason is again a larger intron
849 in the 3'UTR region of BoeCin with an LTR/Gypsy TE covering the entire leader in-
850 tron. This TE insertion causes a shifted translation start and the protein of the BoeCin
851 allele is 49 amino acids longer at the N-terminus than the protein encoded by the
852 BoeRip allele.
853 In case of the MULE transposon (orthogroup with similarity to mutator-like elements),
854 coding potential a transposase protein was identified in BoeCin, BoeRip and VitRGM.
855 There are hints from the TE annotation (Supplementary File 2 Table S13) that the
856 MULE transposon might contain two ORFs in opposite orientation which is character-
857 istic for some MULEs. However, the terminal inverted repeats were not identified. Be-
858 cause of the great diversity within in the same and between transposon superfamilies,
859 generally and also with regard to the structure and length of MULEs [84, 85], further work
860 is needed to precisely identify all TEs and TE fragments that contribute to PAVs at this
861 and other loci.
862 The *Rdv1* locus itself is characterized by one extended resistance gene cluster. Orthol-
863 ogous regions in the other studied haplotypes derived from different *Vitis* species dis-
864 play very divers cluster structures with respect to length in kbp and number of resistance
865 genes. In the BoeCin haplotype that is conferring the dominant resistance to phylloxera
866 and which segregates in offspring of 'Börner', the cluster has a length of about 350 kbp
867 and includes 10 resistance genes. The cluster extends at the southern end beyond the
868 genetically delimited *Rdv1* locus. In the BoeRip haplotype that confers recessive sus-
869 ceptibility to phylloxera, the cluster includes 14 resistance genes.
870 In both 'Börner' haplotypes as well as in VitRGM, the cluster shows an insertion of a
871 conserved array of four to six genes compared to PN40024. The inserted genes (or the
872 genes deleted in PN40024) encode a tRNA ligase, a protein of unknown function, a
873 mutator-like element (MULE) DNA transposon, a DEAD-box ATP-dependent RNA
874 helicase, a ribonucleoside-diphosphate reductase small chain and for a ubiquitin-ligase.
875 These genes display no features of resistance genes and are found with high identity

876 values in the sensitive haplotype BoeRip. They are, therefore, no candidate genes for
877 *Rdv1*.
878 Within the southern part of the locus, RGAugury detected two RGAs with TM-CC do-
879 mains in the encoded proteins that are potential false positive results. The two genes
880 (orange and pink in Fig. 6, with red edging in BoeRip and BoeCin due to the RGAugury
881 output). Based on the functional annotation (Supplementary File 2 Table S15), the two
882 orthogroups code for homologs of SEC31A (a component of the coat protein complex
883 II (COPII) involved in the formation of transport vesicles) and homologs of AGL15
884 (agamous-like MADS-box factor potentially involved in control of development). The
885 genes in the syntenic block that starts at its northern end with the homolog of *SLX1*
886 (encoding a subunit of a structure-specific endonuclease complex involved in pro-
887 cessing diverse DNA damage intermediates, light ochre in Fig. 6) and which extends to
888 the south beyond the delimiting markers GF13-04.2/GBS-04, are also very unlikely to
889 be candidates for *Rdv1*.
890 To the north, the synteny extends for two more genes and terminates at a TX1-like non-
891 LTR retrotransposon (*BoeCin13g18391*) that is only detected in BoeCin. Southern to
892 this TX1-like TE, a syntenic orthogroup (including *BoeCin13g18393* in BoeCin) en-
893 codes a CC-NBS-LRR (CNL) resistance protein related to the wild potato resistance
894 protein RGA4. Since this gene is represented by closely related alleles in BoesCin and
895 the haplotypes conferring susceptiblity (amino acid sequence identity: BoeRip 99.9 %,
896 PN40024 97.1 %), it is not considered to be a candidate for *Rdv1*.
897 Based on this evaluation, the most promising genes mediating phylloxera root re-
898 sistance are the fife resistance genes northern and southern to the array of four to six
899 genes conserved in BoeCin, BoeRip and VitRGM (*BoeCin13g18381*; *Boe-*
900 *Cin13g18382*, CNL/RLP; *BoeCin13g18388*; *BoeCin13g18389,* NL; *Boe-*
901 *Cin13g18390*). Ideally, an F1 genotype from the V3125 x 'Börner' cross with a recom-
902 bination site between the BoeCin and BoeRip haplotypes should have been identified.
903 However, such an event was not detected. It is possible that the high divergence and
904 sequence dissimilarity throughout the *Rdv1* locus causes a suppression or at least re-
905 duction of recombination events at the locus, which could explain this failure. A reduc-
906 tion of recombination frequency was, for example, also described for the *Rpv3-1* locus
907 that confers resistance to *P. viticola* in grapevine [86].

908 The fife remaining resistance genes fall, with respect to functional annotation, into two
909 types. The resistance genes *BoeCin13g18388* and *BoeCin13g18390* were not detected
910 as RGAs by RGAugury, but encode proteins similar to the *A. thaliana* resistance gene
911 *AtLRRAC1* (*At3g14460*). *AtLRRAC1* has been reported to be involved in the defense
912 against the biotrophic fungus *Golovinomyces orontii* as well as the hemibiotrophic bac-
913 teria *Pseudomonas syringae*, and might be relevant for signaling via cAMP-dependent
914 defense pathways [87]. The three other resistance genes, namely *BoeCin13g18381*, *Boe-*
915 *Cin13g18382* CC-NBS-LRR/RLP and *BoeCin13g18389* NBS-LRR (NL), encode pro-
916 teins similar to *RPPL1* (Recognition of *Peronospora Parasitica* 13-like 1). *AtRPPL1*
917 (*At3g14470*) received its annotation from *AtRPP13* (*At3g46530*), a gene that confers
918 resistance to the biotrophic oomycete *Peronospora parasitica* and which encodes an
919 NBS-LRR protein [88, 89]. In *A. thaliana*, *AtLRRAC1* and *AtRPPL1* form a cluster of two
920 neighbouring genes, a feature that is conserved in *Vitis* (Fig. 6).
921 In BoeCin and BoeRip, one of the resistance genes is classified as encoding a NBS-
922 LRR (NL) protein. Despite an identical NBS domain with only very few substitutions,
923 the LRR region varies greatly between the two alleles (*BoeCin13g18389* and
924 *BoeRip13g18554*) and is shorter in BoeRip because of a premature stop codon. Also
925 for the other four candidate resitance genes, differences between the alleles in BoeCin,
926 BoeRip and PN40024 were detected. For example, the protein encoded by *Boe-*
927 *Cin13g18382* was detected as RGA based on CC-NBS-LRR domains in addition to
928 features of an RLP. However, the genes detected at synthenic positions lack these fea-
929 tures either in part or in total.
930 Since the PN40024 sequence is defived from the most susceptible line in our compari-
931 son, the protein sequences of the fife *Rdv1* candidate resintance genes of BoeCin were
932 compared with all PN40024 protein sequences. No identical PN40024 protein sequence
933 or proteins with identical resistance domain sequences were identified. Thus, the five
934 BoeCin resistance genes are promising candidate genes for conferring resistance to
935 phylloxera.

936 Conclusions

937 The fully phase-separated genome sequence assembly BoeRC from the phylloxera re-
938 sistant rootstock ‘Börner’, and it's structural and functional gene annotation, build a
939 cornerstone for the investigation of loci from ‘Börner’ that are linked to it's various

940 valuable traits. Here, the focus was on phylloxera resistance. The sequences of the hap-
941 lotypes BoeCin conferring resistance to phylloxera and BoeRip conferring sensitivity
942 allowed to precisely map recombination sites by GBS in crucial genotypes of a mapping
943 population of V3125 x 'Börner'. Subsequently, detailed examination and comparative
944 genomics of the gene content of the *Rdv1* locus allowed to delimit the locus to a region
945 of 123 kb in BoeCin haplotype with five resistance genes as potential candidates in-
946 volved in mediating phylloxera resistance at the root. The resources generated will al-
947 low to study additional traits and have the potential to support resistance breeding for
948 the benefit of viticulture.

949 Competing interests

950 The authors declare that they have no competing interests.

951 Authors' contributions

952 Sampling and phenotyping were done at the JKI (Julius Kühn-Institute, Institute for
953 Grapevine Breeding Geilweilerhof, Siebeldingen, Germany). Sequencing and data
954 analysis were performed at Bielefeld University, Faculty of Biology & Center for Bio-
955 technology (CeBiTec) and MPIPZ (Cologne). The design of the experiments was set
956 up by LH, DH and BF. Genotype selection, sampling and DNA isolations were carried
957 out by LH and DH. PV, BH and RR accomplished library preparation and sequencing.
958 LH performed the SSR analysis. BW supervised the work at Bielefeld University. RT
959 supervised the work at JKI. RT and BW acquired project funding and wrote the project
960 proposal. All bioinformatic data analyses, creation of figures, tables and writing of the
961 manuscript were performed by BF and DH. BF and BW drafted and finalised the man-
962 uscript. All authors have read and approved the final manuscript.

963 Acknowledgements

964 The authors like to thank all members of the Genetics and Genomics of Plants team at
965 Bielefeld University as well as the members of the Julius Kühn-Institute for Grapevine
966 Breeding at Geilweilerhof.
967 The project was supported by funds from the Federal Ministry of Food and Agriculture
968 (BMEL), based on a decision by the Parliament of the Federal Republic of Germany
969 via the Federal Office for Agriculture and Food (BLE) under the innovation support

program (project acronym MureViU, 28-1-82.066-15), as well as by the EU COST Action INTEGRAPE (CA 17111). We acknowledge support for the article processing charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.
This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI).

Data availability

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

Supplementary information

Supplementary Fig. S1: Length distribution calculated over all subreads
Supplementary Fig. S2: Coverage plot of the ‘Börner’ haplotype assemblies
Supplementary Fig. S3: All-versus-all dot plot between pseudochromosomes of the *V. riparia* haplotype of 'Börner' and the *V. vinifera* reference sequence.
Supplementary Fig. S4: All-versus-all dot plot between pseudochromosomes of the *V. cinerea* haplotype of 'Börner' and the *V. vinifera* reference sequence
Supplementary Fig. S5: Display of initial assembly analysis of chr02 of BoeRip
Supplementary Fig. S6: Example for determination of allelic constitution in selected F1 genotypes

1002 Supplementary Fig. S7: Extended version of Fig. 6, "Genes of BoeCin, BoeRip,
1003 PN40024 and VitRGM at and surrounding the Rdv1 locus", transposable elements/TEs
1004 included
1005 Supplementary Fig. S8: Cumulative length of both ‘Börner’ haplotype assemblies
1006 Supplementary Fig. S9: Sequence region with BAC sequences of both haplotypes
1007 Supplementary Fig. S10: Spectra-copy number and assembly spectrum plot of the
1008 ‘Börner’ genome sequence assembly
1009 Supplementary Fig. S11: Analysis of phasing
1010 Supplementary Fig. S12: All-versus-all dot plot between pseudochromosomes of the *V.*
1011 *riparia* haplotype of 'Börner' and those of *V. riparia* Gloire de Montpellier
1012 Supplementary Fig. S13: All-versus-all dot plot between pseudochromosomes of the *V.*
1013 *cinerea* haplotype of 'Börner' and those of *V. riparia* Gloire de Montpellier
1014 Supplementary Fig. S14: Dot plot of the extended Rdv1 sequence region on pseudo-
1015 chromosome 13 between BoeRip and PN40024
1016 Supplementary Fig. S15: Dot plot of Rdv1 on pseudochromosome 13 between BoeCin
1017 and PN40024
1018
1019 Supplementary File 2 Table S1: RNA-Seq read trimming statistics
1020 Supplementary File 2 Table S2: Trimming statistics of *V. riparia* and *V. cinerea* ge-
1021 nomic reads
1022 Supplementary File 2 Table S3: Statistics of the trimmed BAC sequences
1023 Supplementary File 2 Table S4: Description of samples applied to GBS
1024 Supplementary File 2 Table S5: Primer sequences for SSR marker
1025 Supplementary File 2 Table S6: Mapping positions of the genetic SSR marker on the
1026 PN40024 genome assembly and on the ‘Börner’ haplotype assemblies BoeRip and Boe-
1027 Cin
1028 Supplementary File 2 Table S7: Distribution of the 344 genetic marker used for pseu-
1029 dochromosome validation over BoeRip and BoeCin.
1030 Supplementary File 2 Table S8: Pseudochromosome sizes of the haplotype assemblies
1031 BoeRip and BoeCin
1032 Supplementary File 2 Table S9: Assembly statistics of the BAC contigs
1033 Supplementary File 2 Table S10: Mapping statistics of the BAC contigs on the BoeRip
1034 and BoeCin haplotype assembly
1035 Supplementary File 2 Table S11: Quality statistics of Vitis genome assemblies
1036 Supplementary File 2 Table S12: Haplotype blocks of BoeRip and BoeCin
1037 Supplementary File 2 Table S13: Repeat content of the haplotypes
1038 Supplementary File 2 Table S14: GBS marker sites for the parents ‘Börner’ and V3125
1039 and selected F1 offsprings
1040 Supplementary File 2 Table S15: BoeCin, BoeRip, PN40024 and VitRGM genes in
1041 *Rdv1* region
1042
1043 Supplementary File 3 Data S1: Parameter file for PASA gene model refinement
1044

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 'Borner' (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. 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).

30. 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).

31. 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).

32. Frommer B*, et al.* Genome Sequences of Both Organelles of the Grapevine Rootstock Cultivar 'Börner'. *Microbiology Resource Announcement* **9**, e01471-01419 (2020).

33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

34. 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).

35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).

36. Camacho C*, et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

37. 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).

38. 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).

39. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* **32**, 3321-3323 (2016).

40. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-360 (2015).

41. Li H*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

42. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579 (2011).

43. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

44. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003).

45. Grabherr MG*, et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).

46. 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).

47. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Research* **12**, 656-664 (2002).

48. 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).

49. 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).

50. 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).

51. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277 (2000).

52. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

53. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

54. 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).

55. Waterhouse RM*, et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* **35**, 543-548 (2018).

56. Cannon EK, Cannon SB. Chromosome visualization tool: a whole genome viewer. *International Journal of Plant Genomics* **2011**, 373875 (2011).

57. Robinson JT*, et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26 (2011).

58. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-192 (2013).

59. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).

60. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, (2011).

61. 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).

62. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

63. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

64. Haas BJ*, et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7 (2008).

65. Haas BJ*, et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654-5666 (2003).

66. Jones P*, et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).

67. 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).

68. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164 (1991).

69. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* **8**, 298 (2007).

70. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**, 1027-1036 (2004).

71. Sillitoe I*, et al.* CATH: expanding the horizons of structure-based functional annotations for genome sequences. *Nucleic Acids Research* **47**, D280-D284 (2019).

72. Mistry J*, et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419 (2021).

73. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* **46**, D493-D496 (2018).

74. 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).

75. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).

76. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).

77. 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).

78. Ou S*, et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).

79. Xu Z, Dixon JR. Genome reconstruction and haplotype phasing using chromosome conformation capture methodologies. *Briefings in Functional Genomics* **19**, 139-150 (2020).

80. 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).

81. 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).

82. 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).

83. Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: evolution and utilization. *Annual review of phytopathology* **39**, 285-312 (2001).

84. Dupeyron M, Singh KS, Bass C, Hayward A. Evolution of Mutator transposable elements across eukaryotic diversity. *Mobile DNA* **10**, 12 (2019).

85. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics* **9**, 411-414 (2008).

86. Foria 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).

87. Bianchet C*, et al.* An Arabidopsis thaliana leucine-rich repeat protein harbors an adenylyl cyclase catalytic center and affects responses to pathogens. *Journal of Plant Physiology* **232**, 12-22 (2019).

88. 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).

89. Allen RL*, et al.* Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. *Science* **306**, 1957-1960 (2004).