

1 **The genome of *Lactuca saligna*, a wild relative of lettuce, provides insight into**
2 **non-host resistance to the downy mildew *Bremia lactucae***

3 **Short title:** *Lactuca saligna* genome

4
5 Wei Xiong¹, Lidija Berke^{1,§}, Richard Michelmore², Dirk-Jan M. van Workum³, Frank
6 F.M. Becker¹, Elio Schijlen⁴, Linda V. Bakker⁴, Sander Peters⁴, Rob van Treuren⁵,
7 Marieke Jeuken⁶, Klaas Bouwmeester¹, M. Eric Schranz^{1,*}

8
9 ¹ Biosystematics Group, Wageningen University and Research, Wageningen, The
10 Netherlands.

11 ² Genome Center and Department of Plant Sciences, University of California, Davis, United
12 States.

13 ³ Bioinformatics Group, Wageningen University, Wageningen, the Netherlands.

14 ⁴ Bioscience, Wageningen University and Research, Wageningen, The Netherlands.

15 ⁵ Centre for Genetic Resources, The Netherlands (CGN), Wageningen University and
16 Research, Wageningen, The Netherlands.

17 ⁶ Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands.

18 [§] Present address: Genetwister Technologies B.V., Wageningen, The Netherlands.

19 * Corresponding author: M. Eric Schranz (eric.schranz@wur.nl)

20
21 **ORCID IDs:** 0000-0002-4525-5969 (W.X.); 0000-0003-3842-9462 (L.B.); 0000-0002-
22 7512-592X (R.M.); 0000-0001-6247-5499 (D.W.); 0000-0002-2663-5149 (S.P.); 0000-
23 0002-0567-9548 (M.J.); 0000-0002-8141-3880 (K.B.); 0000-0001-6777-6565 (M.E.S.);
24 0000-0002-9927-2449 (R.v.T.)

25
26 **Material distribution footnote:** The author responsible for distribution of materials
27 integral to the findings presented in this article is: M. Eric Schranz
28 (eric.schranz@wur.nl)

29 **Summary**

30 *Lactuca saligna* L. is a wild relative of cultivated lettuce (*Lactuca sativa* L.), with which
31 it is partially interfertile. Hybrid progeny suffer from hybrid incompatibilities (HI),
32 resulting in reduced fertility and distorted transmission ratios. *Lactuca saligna* displays
33 broad spectrum resistance against lettuce downy mildew caused by *Bremia lactucae*
34 Regel and is considered a non-host species. This phenomenon of resistance in *L.*
35 *saligna* is called non-host resistance (NHR). One possible mechanism behind this NHR
36 is through the plant–pathogen interaction triggered by pathogen-recognition receptors,
37 including nucleotide-binding leucine-rich repeats (NLRs) and receptor-like kinases
38 (RLKs). We report a chromosome-level genome assembly of *L. saligna* (accession
39 CGN05327), leading to the identification of two large paracentric inversions (>50 Mb)
40 between *L. saligna* and *L. sativa*. Genome-wide searches delineated the major
41 resistance clusters as regions enriched in *NLRs* and *RLKs*. Three of the enriched
42 regions co-locate with previously identified NHR intervals. RNA-seq analysis of *Bremia*
43 infected lettuce identified several differentially expressed *RLKs* in NHR regions. Three
44 tandem wall-associated kinase-encoding genes (*WAKs*) in the NHR8 interval display
45 particularly high expression changes at an early stage of infection. We propose *RLKs*
46 as strong candidate(s) for determinants for the NHR phenotype of *L. saligna*.

47

48 **Keywords:** lettuce breeding, downy mildew disease, hybrid incompatibility, non-host
49 resistance, immune genes, *de novo* genome assembly, comparative genomics,
50 population genetics

51

52 INTRODUCTION

53 Lettuce (*Lactuca sativa* L.) is a leafy vegetable grown in more than 100 countries, with
54 a total yield of over 29 million tons in 2019 (FAOSTAT, 2019). One of the most
55 important goals for lettuce breeding is the introgression of durable resistance against
56 lettuce downy mildew, a destructive disease caused by the oomycete pathogen *Bremia*
57 *lactucae* Regel (Lebeda et al., 2009). Outbreak of downy mildew disease leads to
58 substantial yield and economic losses.

59 Wild relatives of lettuce are often used to introgress novel resistances (Lebeda
60 et al., 2014; Parra et al., 2016). *L. saligna*, which belongs to the secondary gene pool
61 of lettuce, is an important donor to enhance resistance to *B. lactucae* in cultivated
62 lettuce (Netzer et al., 1976; Norwood et al., 1981; Bonnier et al., 1991). *L. saligna* is a
63 diploid ($2n=2x=18$, same as lettuce) and self-pollinating species, which is partially
64 interfertile with *L. sativa* (Lebeda et al., 2007, 2019). It is broadly distributed across
65 Eurasia, from the Mediterranean region towards temperate Europe, and from the
66 Iberian Peninsula to Central Asia (Zohary, 1991; Doležalová et al., 2002; Lebeda et al.,
67 2019). *Lactuca saligna* is of particular interest to lettuce breeders as a potential
68 resistance donor due to its complete resistance to all races of *B. lactucae*. As such, it
69 is considered a non-host species to *B. lactucae* based on the definition: “All genotypes
70 of a species are resistant against all genotypes of a specific pathogen” (Bonnier et al.,
71 1991; Petrželová et al., 2011; Lebeda et al., 2009; Heath, 1981). For convenience, we
72 term this resistance phenotype of *L. saligna* as non-host resistance (NHR), which is
73 defined as strictly phenomenological and does not imply a molecular mechanism
74 (Panstruga and Moscou, 2020). To successfully introgress this NHR in lettuce cultivars
75 the gene(s) underlying the non-host resistance and the reproductive barriers observed
76 in hybrid offspring should be determined.

77 Although *L. saligna* is crossable with *L. sativa*, the F_1 plants are nearly sterile,
78 and the resulting inbred offspring (F_2 generation) show severely reduced fertility and
79 transmission ratio distortions due to hybrid incompatibility (HI) (Jeuken et al., 2001;
80 Giesbers et al., 2019). Some case of HI can be explained by the deleterious
81 combination of interspecific alleles according the Dobzhansky-Muller (DM) model
82 (Dobzhansky, 1934; Muller, 1942; Bateson, 1909). Many identified and resolved HI loci
83 are explained by a digenic deleterious epistatic interaction and often results in
84 transmission ratio distortion (TRD) (Fishman and Sweigart, 2018; Fishman and

85 McIntosh, 2019). In F₂ offspring and backcross inbred lines (BILs; i.e., single segment
86 introgression lines) of *L. saligna* x *L. sativa*, 11 HI loci were associated with TRD, six
87 of which were nullified by a paired allele from *L. saligna* (Giesbers et al., 2019). HI loci
88 may reduce the efficiency of introgression of NHR genes from *L. saligna* into *L. sativa*
89 when HI- and NHR loci are closely linked (i.e., linkage drag). In addition to HI, an
90 interspecific chromosomal rearrangement, like an inversion, will also hamper the
91 introgression of desired NHR genes via linkage drag caused by reduced recombination
92 (Hoffmann and Rieseberg, 2008; Fishman and Sweigart, 2018).

93 NHR in plants is suggested to rely on a continuum of layered defenses, including
94 both constitutive and induced resistance mechanisms (Niks and Marcel, 2009; Jones
95 and Dangl, 2006; Bettgenhaeuser et al., 2014). Previous studies propose that induced
96 NHR and host immunity rely on a similar non-self-recognition system comprising two
97 innate immunity layers: i) pattern-triggered immunity (PTI) mediated by extracellular
98 recognition of conserved non-self-molecules – called pathogen-associated molecular
99 patterns (PAMPs) – by cell surface receptors, such as diverse receptor-like kinases
100 (RLKs), and ii) host defense conferred by effector-triggered immunity (ETI) mediated
101 by *R* genes encoding intracellular nucleotide-binding leucine-rich repeat proteins
102 (NLRs) that recognize cognate pathogen-secreted effector molecules (Niks and Marcel,
103 2009; Schulze-Lefert and Panstruga, 2011; Jones and Dangl, 2006; Chisholm et al.,
104 2006). After host penetration, hyphal growth of *B. lactucae* is quickly halted in *L. saligna*
105 and consequently haustorium formation is impeded (Niks, 1987; Lebeda and Reinink,
106 1994; Zhang et al., 2009a, 2009b). To identify common loci associated with NHR to *B.*
107 *lactucae* in lettuce, Giesbers *et al.* (2018) performed mapping studies based on nine *L.*
108 *saligna* accessions from a broad range of geographic regions via multiple bidirectional
109 backcrosses: i.e., i) BC1 populations in both parental directions (F₁ x host *L. sativa*)
110 and (F₁ x non-host *L. saligna*), and ii) BC1S3 lines with three generations of inbreeding,
111 respectively. These mapping populations facilitated the identification of four epistatic
112 segments accounting for NHR in *L. saligna*: one positioned on Chromosome 4 (NHR4),
113 two on Chromosome 7 (NHR7.1 & 7.2), and another on Chromosome 8 (NHR8)
114 (Giesbers et al., 2018). It is worth noting that the NHR8 interval is closely linked to
115 HI/TRD loci, which potentially limits fine-mapping and introgression of non-host traits
116 into *L. sativa*. The genes and mechanisms underlying these four NHR loci are

117 unresolved. A high-resolution analysis of these regions is needed to unveil the genetic
118 determinants governing NHR in *L. saligna*.

119 *NLRs* in lettuce were previously identified by Christopoulou *et al.* (2015a) using
120 the *L. sativa* v6 genome. Identified *NLRs* were classified into two major groups: *TNLs*
121 with *TOLL*/interleukin-1 receptor (*TIR*) domains and *CNLs* with coiled-coil (*CC*)
122 domains, and subsequently into multiple resistance gene candidates (*RGC*) families
123 (Takken and Goverse, 2012; McHale *et al.*, 2006; Meyers *et al.*, 2003). Almost all
124 identified *NLRs* were found to reside in five major resistance clusters (*MRCs*) that co-
125 segregate with resistance to diverse pathogens (McHale *et al.*, 2009; Christopoulou *et al.*
126 *et al.*, 2015b). For example, *MRC2* on Chromosome 2 comprises multiple *RGC2* family
127 members, including the downy mildew resistance genes *Dm3*, *Dm14*, *Dm16*, and
128 *Dm18* (Shen *et al.*, 2002; Wroblewski *et al.*, 2007; Christopoulou *et al.*, 2015a). Similar
129 *MRCs* are suggested to be present in *L. saligna* based on the expected whole-genome
130 synteny, since some qualitative resistance phenotypes from *L. saligna* have been
131 mapped at single loci syntenic to *MRCs* in *L. sativa* (Giesbers *et al.*, 2017). An *L.*
132 *saligna* genome reference can facilitate synteny analysis to recognize these
133 anticipated *MRCs*.

134 Multiple *RLK* families contain members involved in a wide range of immune
135 responses in plants. Notable examples can be found in the *LRR-RLK* sub-family, such
136 as *FLS2* involved in the perception of bacterial flagellin and *IOS1* that contributes
137 towards resistance to the downy mildew *Hyaloperonospora arabidopsidis* (Zipfel *et al.*,
138 2004; Hok *et al.*, 2011). Previously, Christopoulou *et al.* (2015a) also described *LRR-*
139 *RLK* encoding genes in lettuce. Nevertheless, a specific inventory of genes encoding
140 other resistance-related *RLKs* in lettuce, such as those encoding lectin receptor
141 kinases (*LecRKs*) and wall-associated kinases (*WAKs*), is still lacking, not to mention
142 the *RLKs* of *L. saligna* (Bouwmeester *et al.*, 2011; Hu *et al.*, 2017; Zuo *et al.*, 2015;
143 Hurni *et al.*, 2015; He *et al.*, 1999).

144 Here, we report a *de novo* genome assembly of *L. saligna* (accession
145 CGN05327) using a variety of sequencing and scaffolding techniques. The assembly
146 was compiled into nine chromosomal pseudo-molecules by genetic mapping. The
147 resulting assembly enabled us to conduct diverse genomic analyses to dissect the
148 genetic determinants underlying non-host resistance in *L. saligna*. The analyses
149 provide insights of evolution into disease resistance and on host-pathogen arms race

150 in lettuce. For breeding, the gained knowledge helps to facilitate the introgression of
151 *Bremia* resistance into cultivated lettuce.

152

153 **RESULTS**

154 **Genome sequencing and assembly**

155 *L. saligna* accession CGN05327 was used to produce a reference genome for *L.*
156 *saligna* (Supplemental Note). A combination of PacBio long-read (95.4 Gb; 41X) and
157 Illumina short-read (407.4 Gb; 175X) sequencing was generated to assemble the
158 genome (Supplemental Data 1). Illumina paired-end (125 bp, PE) and mate-pair (300
159 bp, MP) reads were generated from three libraries of different insert size (200 bp, 500
160 bp and 550 bp) (Supplemental Data 1A). The *L. saligna* genome size was estimated
161 by K-mer analysis to be 2.27 Gb, which agrees with genome size estimates established
162 by flow cytometry (2.3 Gb; Doležalová et al., 2002; Zohary, 1991). K-mer analysis also
163 revealed that the genome is highly homozygous (estimated heterozygosity = 0.12%)
164 as expected for this inbreeding species (Supplemental Figure 1 and Supplemental
165 Table 1). To construct a high-quality genome of *L. saligna*, we applied a variety of
166 advanced assembly and mapping techniques (Supplemental Figure 2). An initial Canu
167 assembly (v0.5) consisted of 31,431 contigs, and the N50 number and size were 6,957
168 and 88.0 kb, respectively (Table 1; Supplemental Data 1A). Bionano fingerprinting, 10x
169 Genomics barcoding, and Dovetail Hi-C library data were sequentially applied to
170 construct the version 2 assembly, which refined the assembly to 24 super-scaffolds
171 (largest scaffold = 279.9 Mb; N50 = 146.7 Mb; Supplemental Data 2B-C).

172

173 **Linkage group anchoring and assembly assessment**

174 To generate chromosomal pseudo-molecules, we combined 417 genetic markers from
175 an F₂ population linkage map (*L. saligna* x *L. sativa*) and 19,027 syntenic markers
176 between *L. saligna* and *L. sativa* (Supplemental Table 2; Supplemental Data 3A-B).
177 This resulted in a chromosome-level assembly (v3) in which 17 out of 24 scaffolds
178 (99.8% bases, 1.75 Gb) were anchored and oriented into nine chromosomes, covering
179 ~77% of the estimated genomic sequence (1.75 of 2.27 Gb) (Supplemental Table 3;
180 Supplemental Data 2D; Supplemental Figure 3). To obtain a more complete reference
181 assembly, un-scaffolded contigs (>1000 bp) were merged to create a virtual
182 “chromosome zero.” This eventually led to a final assembly (v4) with nine

183 chromosomes plus chromosome zero, with a complete genome size of 2.17 Gb (Table
184 1; Supplemental Table 4; Supplemental Data 2E-F). This final assembly contains 91.9%
185 (1,951 out of 2,121) of the expected BUSCO (1,859 single and 92 duplicated copies)
186 eudicot gene models (Supplemental Table 5), and 92% of the 30,696 *L. saligna*
187 expressed sequence tags (ESTs) in NCBI could be aligned to the v4 assembly at 80%
188 identity and 80% coverage (Supplemental Table 6).

189

190 **Repeat and non-coding RNA annotation**

191 Our analyses estimated that 77.5% of the *L. saligna* genome consists of transposable
192 elements (TE; Table 2; Supplemental Table 7). Long terminal repeat retrotransposons
193 (LTR-RT) were the most predominant repetitive elements, comprising both Gypsy and
194 Copia retrotransposons (43.8% and 23.1% of genome, respectively) (Supplemental
195 Table 8-9). TEs were distributed across the genome, and found to be enriched in
196 regions roughly representing the pericentromeric locations (Supplemental Figure 4:
197 track E). Non-coding RNAs involved in mRNA transcription (snRNA), translation
198 (tRNAs and rRNAs), and regulation of gene expression (miRNAs) were also annotated
199 (Table 2; Supplemental Table 10).

200

201 **Gene prediction and functional annotation**

202 A combination of *de novo* search and homology support was applied for gene model
203 prediction. Most of the predicted gene models (93%) were well supported (AED > 0.5)
204 by RNA-seq data and gene homology (Supplemental Figure 5; Supplemental Data 4A).
205 In total, 42,908 gene models were retained after filtering based on coding-potential
206 (Table 2). The average coding size and exon number per gene was 1.3 kb and 5.1
207 respectively (Table 2). We further validated the potential for protein-encoding
208 sequences using domain, ortholog, and homolog databases. By combining all results,
209 40,730 genes (94.9%) had matches in at least one database (Table 2; Supplemental
210 Table 11; Supplemental Data 4B).

211

212 ***Lactuca saligna* population structure and diversity**

213 To explore the genetic diversity and population structure of *L. saligna*, we re-
214 sequenced 15 accessions representing the distribution across its native range
215 (Supplemental Table 12-13). SNPs were first called on the *L. saligna* genome

216 assembly and then filtered on missing rate (<10 %) and minor allele frequency (>0.05),
217 yielding 5,170,479 SNPs for downstream analysis (Supplemental Table 14-15). After
218 pruning the SNP dataset, we applied three complementary methods to explore the
219 structure of *L. saligna*: neighbor-joining tree building, principal component analysis
220 (PCA), and ancestry history inference. The neighbor-joining tree revealed that the *L.*
221 *saligna* population can be subdivided into three major clades that are largely congruent
222 with the geographical origins of the selected accessions (Figure 1A). This finding was
223 recapitulated by PCA (Figure 1B) and ADMIXTURE analysis (Figure 1C). These
224 analyses also uncovered the geographical origins of two accessions that were
225 previously unknown. Accession CGN05271 is implicated to be of European origin,
226 whereas CGN05282 groups with multiple accessions from the Middle East (Figure 1D).
227 It is noteworthy to mention that accession CGN05271, now found to be of European
228 origin, has been extensively utilized in many in-depth genetic studies on resistance to
229 downy mildew or reproductive barriers (Jeuken and Lindhout, 2002; den Boer, 2014;
230 Giesbers et al., 2017, 2018; Jeuken et al., 2001; Giesbers et al., 2019). Our sequenced
231 reference CGN05327 is genetically clustered with CGN05271. Finally, the leaf
232 morphology of each accession was also found in line with the *L. saligna* population
233 genetic structure (Supplemental Figure 6).

234

235 **Synteny between *L. saligna* and *L. sativa***

236 Duplication events and structural variation were identified between the *L. saligna* and
237 *L. sativa* genomes by syntenic alignments. Intra-species collinearity revealed a 3:1
238 syntenic pattern in all nine chromosomes for both species, confirming the known
239 shared whole-genome triplication event within the Asteraceae (Reyes-Chin-Wo et al.,
240 2017; Iorizzo et al., 2016) (Supplemental Figure 7). Inter-species syntenic analysis
241 revealed a high level of genome-wide collinearity between both *Lactuca* species,
242 except for two large inversions (> 50 Mb) on Chromosomes 5 and 8 (Figure 2B-C;
243 Supplemental Table 16). The observed gene density (~ 20 genes per Mb) within these
244 two inverted regions in both species suggests that they are not close to the centromere,
245 i.e., paracentric inversions (Supplemental Table 16). The ranges and positions of
246 inversions were estimated using syntenic genes at the inversion borders
247 (Supplemental Table 17). To confirm these inversions, we mapped markers derived
248 from an interspecific F₂ population to the *L. saligna* genome and compared their

249 genetic and genomic positions. This showed that the genetic position plateaued while
250 the genomic position kept increasing over the inverted region, which reflects the
251 suppressed recombination due to inversion (Supplemental Figure 8). These inversions
252 encompass of a diversity of genes, some of which encode proteins known to play key
253 roles in various biological processes, such as a methyltransferase involved in Vitamin
254 E biosynthesis and a phosphatase regulating cell wall integrity (Supplemental Table
255 18-19) (Cheng et al., 2003; Franck et al., 2018).

256

257 **Comparison of *NLR* content and distribution between *L. saligna* and *L. sativa***

258 To explore variation in the *NLR* gene family, HMMER and BLAST searches were
259 conducted against the proteomes of *L. saligna* and *L. sativa* (Supplemental Data 5A-
260 B). Retrieved amino acid sequences were first classified based on their N-terminal TIR
261 or CC domain (TNLs and CNLs, respectively) and thereafter subdivided to Resistance
262 Gene Candidate (RGC) families by phylogenetic analyses (Supplemental Figure 9;
263 Supplemental Data 5C-D). This resulted in the identification of 323 *NLRs* in *L. saligna*
264 and 364 *NLRs* in *L. sativa*. *Lactuca saligna* and *L. sativa* were found to contain a similar
265 content of both TNL- and CNL-type, i.e., 184 versus 202 (57.0%, 55.5%), and 139
266 versus 162 (43.0%, 44.5%), respectively (Table 3; Supplemental Table 20). Genomic
267 positions of MRCs previously identified in *L. sativa* were identified in the *L. saligna*
268 genome assembly using *L. sativa* orthologs (Supplemental Table 21; Supplemental
269 Data 5D). We additionally defined two *NLR*-enriched clusters (NCs) in *L. saligna* on
270 Chromosomes 4 (38.55 – 40.68 Mb) and 7 (44.01 – 44.48 Mb), hereafter named NC4
271 and NC7 (Supplemental Table 22). These two NCs were also identified in *L. sativa*,
272 but were not previously labeled as MRCs due to the absence of resistance phenotypes
273 (Christopoulou et al., 2015b). In total, 41 RGC families were identified. Seven RGC
274 families (six singletons and one multigene family) present in *L. sativa* were found
275 missing in *L. saligna*, which might be caused by the reconstructed phylogeny or they
276 may be unique to *L. sativa* (Supplemental Table 23). While *L. saligna* has a similar
277 amount of *NLRs* compared to *L. sativa* in most RGCs, we defined significant size
278 change by count and percentage difference. In this way, we observed that six and three
279 RGC families were contracted (i.e. RGC1, 4, 8, 9, 14, and 21) and expanded (i.e. RGC
280 16, 20, and 29), respectively, in this accession of *L. saligna* compared to the reference
281 genome of *L. sativa* (Supplemental Figure 9; Supplemental Table 23).

282

283 **Comparison of *RLK* genes between *L. saligna* and *L. sativa***

284 To identify genes encoding RLK proteins, we performed in-depth HMMER searches
285 against the predicted *L. saligna* and *L. sativa* proteomes. This resulted in the
286 identification of 478 and 566 *RLK* encoding genes in *L. saligna* and *L. sativa*,
287 respectively (Supplemental Table 24; Supplemental Data 6A). Sliding window analysis
288 revealed that *RLKs* are distributed on all chromosomes, with their density elevated at
289 the chromosomal ends (Figure 3: track B). *RLKs* were further classified into nine
290 subfamilies based on their extracellular domains using HMMER (Table 3;
291 Supplemental Data 6B). In both species, LRR-*RLKs* and G-type LecRKs (G-LecRKs)
292 formed the largest subfamilies. The major difference in total *RLKs* was also largely
293 accounted by these two subfamilies – with an additional 32 G-LecRKs and 48 LRR-
294 *RLKs* in *L. sativa*. The other *RLK* subfamilies were found to be of similar size in these
295 accessions of both *L. saligna* and *L. sativa*.

296

297 **Mapping HI and NHR loci on *L. saligna* genome**

298 To precisely characterize the HI and NHR regions, markers of these loci were mapped
299 to the *L. saligna* assembly (Supplemental Table 25-26). For HI, one locus was
300 positioned on Chromosome 8 (33.15–138.07 Mb) and contains the inversion identified
301 on Chromosome 8, 59–103 Mb (Figure 3: track A and D; Supplemental Figure 8)
302 (Giesbers et al., 2019). This HI region and inversion region on Chromosome 8 was
303 also adjacent to the resistance-related regions NHR8, MRC8B, and MRC8C (Figure 3:
304 track C-D; Supplemental Figure 8). For NHR, three out of four intervals either
305 overlapped with *NLR* or *RLK* hotspots. NHR7.1 was found to co-segregate with the
306 NC7 region encoding 13 *NLRs*, whereas the other three NHR intervals consist of no or
307 only one *NLR* gene (Supplemental Data 7). Moreover, mapping revealed that both
308 NHR4 and NHR8 co-locate with regions enriched in *RLK* genes (NHR4: 20 *RLKs* in
309 34.21 Mb; NHR8: 14 *RLKs* in 13.26 Mb). Especially for region NHR8, the *RLK* density
310 (1.06/Mb) was five-times higher than the genome-wide average (0.24/Mb, 422 in 1,745
311 Mb, excluding Chromosome 0). Mapping revealed a close relationship between NHR
312 regions and resistance gene hotspots, making *NLRs/RLKs* potential determinants of
313 NHR in *L. saligna*. In addition, NHR8 is also positioned near an HI segment, which may

314 prevent the introgression of the candidate resistance genes to cultivated lettuce,
315 impacting breeding for resistance.

316

317 **RNA-seq time-course analysis of *L. saligna* transcriptome in response to *Bremia***

318 To detect genes with differential expression after infection, we performed a *Bremia*
319 infection assay on leaves of *L. saligna* to generate transcriptomic data and
320 subsequently conducted a differential expression (DE) analysis. Treated and control
321 samples were collected at 8- and 24-hours post-infection (hpi). Statistical analysis of
322 quantified RNA-seq reads count identified a total of 1,268 and 1,688 differentially
323 expressed genes (DEGs) ($p_{adj} < 0.5$ and $\log_2FC > 1$) at 8 hpi and 24 hpi, respectively
324 (Supplemental Table 27; Supplemental Data 7). For both time points, the majority of
325 DEGs were up-regulated in expression, i.e. 1,222 up-regulated versus 46 down-
326 regulated genes at 8 hpi, and 1,362 up-regulated versus 326 down-regulated genes at
327 24 hpi (Supplemental Table 28). One of the most representative DEGs is
328 *Lsal_1_v1_gn_1_00001954*, showing the largest induction in expression
329 ($\log_2FC=11.72$), is a homolog of the penetration resistance gene *PEN1*, which
330 encodes a syntaxin involved in vesicle assembly for non-host resistance against
331 powdery mildew penetration in *Arabidopsis* (Collins et al., 2003).

332

333 **Enrichment analysis of identified DEGs in *L. saligna***

334 Subsequently, we applied gene ontology enrichment analysis of DEGs to explore
335 functional-related biological processes and pathways. Figure 4 shows the 20 most
336 significantly enriched terms related to DEGs at 8 hpi or 24 hpi. Sixteen out of 20
337 ontology terms were identified at both time points. Most clusters were mainly
338 associated with resistance responses, like stress perception (GO:0009620), signal
339 transduction (GO:0046777), and cell death (GO:0008219). In general, 8 hpi showed a
340 greater enrichment than 24 hpi for most top terms (Figure 4B). In contrast, three unique
341 biological clusters were found for the 24 hpi timepoint, all of which were related to
342 ribosome biogenesis (GO:0042254, GO:0042273, and ath03010) (Figure 4A-B). In
343 addition to the top 20 terms, many up-regulated genes were found to be involved in
344 plant defense, in particular in response to oomycetes, illustrating the immune response
345 of *L. saligna* upon *Bremia* infection (Supplemental Figure 10; Supplemental Table 28-
346 29). For example, these include *Lsal_1_v1_gn_9_00004094*, a homolog of the lectin

347 receptor gene *LecRK-IX.1* conferring resistance to *Phytophthora* spp. (another
348 oomycete pathogen); Lsal_1_v1_gn_8_00004656 (*SARD1*) and
349 Lsal_1_v1_gn_2_00003439 (*UGT76B1*), encoding two key regulators of salicylic acid
350 (SA) synthesis and SA mediated signaling for stress response (Wang et al., 2015; Ding
351 et al., 2016; Mohnike et al., 2021; Bauer et al., 2021). Our enrichment analysis detected
352 that DEGs at both time points post-inoculation with *Bremia* were enriched in resistance-
353 related biological processes: 8 hpi showed a stronger signal of early immune response
354 and 24 hpi showed a shift of enriched terms to extra post-transcriptional response.

355

356 **Differentially expressed genes in NHR regions at 8 hpi**

357 Based on above mapping and DE analysis results, we inspected the statistics of the
358 up-regulated genes in NHR intervals at 8 hpi to further identify candidates for
359 resistance to lettuce downy mildew. First, we calculated the DEG density per million
360 base-pair of four NHR loci and the whole genome (Supplemental Table 30). As
361 baseline, the DEG density for the whole genome was 0.70 per Mb. The NHR8 locus
362 had the highest DEG density (1.54/Mb) among all NHR intervals and was greater than
363 two-times the average density of the entire genome. Moreover, 11 DEGs located in the
364 overlapping region of NHR8 and HI may also inhibit the ability to overcome the hybrid
365 barrier. Secondly, we examined the percentage of up-regulated *RLKs* and *NLRs* (up-
366 regulated number / total number) for each NHR interval (Supplemental Table 30). The
367 percentage of differentially expressed *NLRs* was low (4.6%) across the whole genome.
368 None of the *NLRs* within the two NHR loci were differential expressed. In contrast,
369 more than 22.7% of the *RLKs* (96) were up-regulated genome-wide after *Bremia*
370 inoculation. NHR8 also displayed a high percentage of up-regulated *RLKs* (50%, seven
371 out of 14). Furthermore, we counted the number of DEGs with a large degree of change
372 ($\log_2FC > 3$) in NHR regions of interest (Supplemental Table 30). Again, NHR8 was
373 found to contain more highly expressed genes (nine) than the other three NHR regions.
374 Thus, out of four NHR loci, the statistics of DEGs strongly suggests that genes on
375 NHR8 seemed to play a critical role in the resistance to *Bremia*, especially the *RLKs*.
376 Based on these observations, we pinpointed eight DEGs located in NHR8 as
377 candidates for downy mildew resistance in *L. saligna* (Supplemental Table 31). One of
378 the candidate genes encodes a plant U-box type E3 ubiquitin ligase (PUB), of which
379 family members have been reported to play essential roles in plant defense and

380 disease resistance (González-Lamothe et al., 2006). The other candidate genes all
381 encode receptor-like kinases, i.e., one LysM-containing receptor-like kinase (LysM-
382 RK), three G-type lectin receptor kinases (G-LecRKs), and three WAKs. It is
383 noteworthy to mention that the three WAKs were tandem-arrayed, of which two were
384 highly up-regulated ($\log_2FC > 3$; Figure 4C).

385

386 **DISCUSSION**

387 ***L. saligna* reference genome and population structure**

388 In this study, we report on the *de novo* genome assembly of *L. saligna* based on long-
389 and short-read sequencing together with advanced scaffolding techniques. The
390 genome size of our *L. saligna* assembly (2.17Gb) is in line with the previously reported
391 C-value (2.3Gb) (Doležalová et al., 2002). The genomic content, such as gene space
392 and repeat content of the genome (~77%), is comparable to cultivated lettuce (Reyes-
393 Chin-Wo et al., 2017). Using SNPs called on the reference genome, population genetic
394 analysis identified three *L. saligna* sub-groups that are consistent with geography
395 (Figure 1). We also inferred the graphical origin of two genotypes derived from the
396 Jardin Botanique de Nantes, a French botanical garden, including the accession
397 CGN05271 (found to be of European origin) and accession CGN05282 (found to be of
398 Middle Eastern origin). The obtained population genetics structure is in agreement with
399 a previous clustering based on AFLP markers (Giesbers et al., 2018).

400

401 **Inversions and HI may hamper breeding with the NHR8 resistance locus**

402 Comparative genomic analysis identified two large inversions (>50 Mb) on
403 Chromosomes 5 and 8 between *L. saligna* and *L. sativa* (Figure 2). We also found that
404 the inversion on Chromosome 8 co-segregated with an HI region (Giesbers et al.,
405 2019). Genic incompatibilities associated with hybrid necrosis are often linked to
406 immune genes (Bomblies and Weigel, 2007; Fishman and Sweigart, 2018). A well-
407 described example of hybrid necrosis for lettuce is the digenic interaction between the
408 *L. saligna* allele of *Rin4*, encoding a putative negative regulator of basal plant defense,
409 and the resistance gene *Dm39* from cultivated lettuce (Jeuken et al., 2009). The HI
410 locus on Chromosome 8 was not found to be associated with the hybrid necrosis
411 phenotype (Giesbers et al., 2019). Therefore, immune gene(s) are likely not causal to
412 HI on Chromosome 8, even though we found several resistance loci (MRC and NHR)

413 close to the HI regions located on the inverted regions (Figure 3; Supplemental Figure
414 8). If the HI/TRD locus indeed resides in the inversion, then further fine mapping and
415 introgression of loci associated with HI, and potential immune genes underlying NHR
416 for that matter, will not be feasible due to the lack of recombination caused by inversion.
417 Future research could investigate whether all *L. saligna* accessions share same large
418 inversions that cause linkage drag, by combining sequencing.

419

420 ***L. sativa* contains more immune genes than the non-host *L. saligna***

421 Previous studies in *L. saligna* by genetic mapping have detected multiple loci
422 containing *NLRs* associated with its resistance phenotype, for example, the R locus
423 (*Dm39*) interacts with *Rin4*, and the R locus responds to the effector *BLR31*, which are
424 suggested not to govern the NHR phenotype (Jeuken et al., 2009; Giesbers et al.,
425 2017). The lack of knowledge on genome-wide variation in resistance genes has
426 hindered the identification of NHR determinant(s). In this paper, we comprehensively
427 inventoried *NLR* and *RLK* genes in *L. saligna* and *L. sativa* (Table 3). Our results show
428 that *L. sativa* has more *NLRs* ($364 / 323 = 1.13$) and *RLKs* ($566 / 478 = 1.18$) than *L.*
429 *saligna*. This difference could possibly be due to the genome size differences between
430 *L. sativa* and *L. saligna* ($2.5\text{Gb} / 2.3\text{Gb} = 1.09$; Doležalová et al., 2002) or the
431 incomplete sequencing and annotation. Immune genes, like *NLRs* or *RLKs*, are known
432 as the most variable genes in plants, including lettuce and its wild relatives (Karasov
433 et al., 2014; Parra et al., 2016). Due to allelic and copy number variation, the genome
434 assembly alone cannot fully capture the complete spectrum of *R* genes (Barragan and
435 Weigel, 2021). Therefore, the genetic determinant of NHR might not be identified by
436 the genome-wide searches using these reference assemblies. Target sequencing of
437 *NLRs* and *RLKs* (e.g. RenSeq and RLKSeq) can be applied to collect a more complete
438 spectrum of resistance genes (Witek et al., 2016; Lin et al., 2020).

439

440 ***RLK* and *NLR* genes associated with NHR against *Bremia***

441 To further understand the relationship between *RLK/NLRs* and NHR in *L. saligna*, we
442 mapped the four NHR loci to the *L. saligna* reference genome. Of the four NHR
443 intervals, we found that three have either elevated densities of *RLKs* (NHR4 & NHR8)
444 or *NLRs* (NHR7.1). Moreover, *RLKs* and *NLRs* do not co-occur with each other in
445 analyzed *Lactuca* species, as illustrated by NHR8 (14 *RLKs* vs zero *NLRs*) and

446 NHR7.1 (zero *RLKs* vs 13 *NLRs*) (Supplemental Table 30), which suggests that *NLRs*
447 and *RLKs* act as epistatic genes explaining NHR (Giesbers et al., 2018). Although
448 *RLKs* and *NLRs* elicit PTI and ETI respectively (Jones and Dangl, 2006), there is
449 increasing evidence that PTI and ETI are not separate phenomena and mutually
450 strengthen each other's immune response (Yuan et al., 2021; Ngou et al., 2021). This
451 could explain why the identified NHR loci in *L. saligna* involves a combination of PTI
452 and ETI.

453

454 **RNA-seq highlights a crucial role of *RLKs***

455 RNA-seq analysis of *L. saligna* leaves inoculated with *Bremia* enabled us to identify
456 DEGs related to NHR-associated plant defense responses. Multiple DEGs with high
457 levels of induced expression were found to be involved in salicylic acid (SA) synthesis
458 (SARD1 and UGT76B1) or SA-dependent penetration resistance (PEN1 and PEN3)
459 contributing to NHR in *Arabidopsis* (Supplemental Table 27-28) (Zhang et al., 2010;
460 Collins et al., 2003; Assaad et al., 2004; Mohnike et al., 2021; Bauer et al., 2021).
461 Various studies have shown that SA increases *RLK* expression in different plants
462 (Ohtake et al., 2000; Coqueiro et al., 2015). Transcriptome analysis also revealed that
463 a large portion of the DEGs at 8 hpi function in early recognition and defense signaling
464 activity, whereas DEGs at 24 hpi were found to be responsible for post-transcription
465 activity. For expression of immune genes in NHR regions, no *NLR* genes were
466 differentially expressed. This is consistent with expectations, as *NLR*-encoding genes
467 generally are lowly expressed after *Bremia* infection (Wroblewski et al., 2007). In
468 addition, a large amount of *RLKs* was differentially transcribed, which is similar as
469 described for the interaction between lettuce and the fungal pathogen *Botrytis cinerea*
470 (De Cremer et al., 2013).

471

472 **NHR8 contains *WAK* genes highly upregulated upon *Bremia* infection**

473 Among the four NHR regions, NHR8 has the highest number of differentially expressed
474 *RLKs* (Supplemental Table 30). Within it, three closely clustered wall-associated
475 kinases (*WAKs*) were of special interest because of their significant expression change
476 (Figure 4C). These three *WAK* paralogs were homologs of *Arabidopsis WAK2*, which
477 is highly expressed in leaves and can be up-regulated in expression upon pathogen
478 infection and SA application (He et al., 1999). Various studies illustrated that *WAKs*

479 provide quantitative resistance against various diseases in crops such as maize and
480 rice (Zuo et al., 2015; Hurni et al., 2015; Hu et al., 2017). For *L. saligna* infected by
481 *Bremia*, oligogalacturonides derived from damaged cell walls could be perceived by
482 WAKs to trigger PTI (Raaymakers and Van den Ackerveken, 2016; Ferrari et al., 2013;
483 Brutus et al., 2010). WAKs have also been implied in cell wall reinforcement. In rice,
484 Xa4 strengthens the cell wall by promoting cellulose synthesis and suppressing cell
485 wall loosening, thereby enhancing resistance to bacterial infection by *Xanthomonas*
486 *oryzae* (Hu et al., 2017). Hence, WAKs located on NHR8 seem to hold potential in *L.*
487 *saligna* resistance. Nevertheless, we cannot rule out the possibility that other
488 genes/factors play roles in NHR in *L. saligna*, and the expressions level of WAKs along
489 with other genes mentioned in this paper need to be further compared to their
490 homologs in susceptible lettuce cultivars or resistant introgression lines. Future fine
491 mapping and knock down/out experiments are needed to further pinpoint key factors
492 underlying the NHR in *L. saligna* using the reference genome assembly presented in
493 this paper.

494

495 **A model for NHR in *L. saligna* against lettuce downy mildew**

496 Based on our findings and previous research, we propose an NHR model for *L. saligna*
497 with the following three elements: i) The host status of *L. sativa* and *L. saligna* is partly
498 determined by the variation in orthologous RLKs involved in immunity. A specific
499 ortholog in *L. saligna* can effectively enhance resistance to colonization by *B. lactucae*.
500 A comparable role of orthologous RLKs has been observed in the interaction between
501 barley and leaf rust fungi, in which a LecRK of wild barley quantitatively enhances
502 resistance (Wang et al., 2019). ii) After non-self-recognition by RLKs, cell wall-plasma
503 membrane interactions are strengthened (Wolf, 2017), restricting intercellular hyphal
504 growth. This is in line with the reduced hyphae formation found in infected *L. saligna*
505 (Zhang et al., 2009b). In case of successful penetration, NHR to powdery mildew in
506 barley is often backed up by NLR-mediated hypersensitive response (HR) (reviewed
507 in Niks and Marcel, 2009). As for the observed NHR in *L. saligna*, this might also be
508 the case.

509

510 **MATERIALS AND METHODS**

511 **Plant materials and DNA isolation**

512 *L. saligna* accessions selected for whole-genome sequencing and resequencing were
513 obtained from the lettuce germplasm collection of the Centre for Genetic Resources,
514 The Netherlands (CGN) (Supplemental Table 12). Accession CGN05327 was
515 collected from Gerona, Spain, of which a Single Seed Descendant (SSD) was used for
516 *de novo* reference genome sequencing and assembly. Re-sequencing data of 15
517 Single Seed Decent (SSD) lines derived from *L. saligna* accessions (Supplemental
518 Table 12) were selected to represent the *L. saligna* germplasm. Seeds were stratified
519 at 4°C for three days to improve germination. Seedlings were subsequently grown in a
520 growth chamber at 17–19°C with LED light under a 16 h photoperiod and a relative
521 humidity of 75–78%. After eight weeks, plants were transplanted to larger pots
522 containing potting soil and grown under greenhouse conditions. Images of leaves (third
523 mature leaf counted from the base) of 10 accessions belonging to different subgroups
524 were taken from 15-week-old plants, which were grown in triplicate (Supplemental
525 Table 32). Tissue sampling was performed when plants were close to bolting, and DNA
526 was extracted using the protocol as in Ferguson *et al.* (2020).

527

528 **Genome sequencing**

529 A *de novo* genome assembly of *L. saligna* CGN05327 (Supplemental Figure 1) was
530 assembled using a ~21-fold coverage of long-read data generated by PacBio Sequel
531 technology (4,083,751 reads; N50 read length=16,581 bp; subread length=8,514 bp),
532 and a ~175-fold coverage of Paired-end (PE) reads obtained by Illumina mate pair
533 sequencing. The mate pair library was prepared using different insert sizes and read
534 lengths: HiSeq (200 bp insert size, 125 bp PE), HiSeq (500 bp insert size, 125 bp PE),
535 and MiSeq (550 bp insert size, 300 bp PE).

536

537 **Genome assembling and scaffolding**

538 PacBio reads were assembled using Canu and polished with Pilon (v1.20) using
539 Illumina data (Koren *et al.*, 2017; Walker *et al.*, 2014). Subsequently, multiple
540 techniques were applied to elevate the contiguity of the assembly. A follow-up
541 assembly (version 1) was scaffolded using 10x Genomics Chromium barcoding data
542 (ARC pipeline) and ~130-fold coverage BioNano optical mapping data (Yeo *et al.*,
543 2017). A Hi-C library produced by Dovetail Genomics providing ~2,553-fold coverage
544 of sequence data (429 million 2x150 bp read pairs) was used for *in vitro* proximity

545 ligation. Mis-joins in assembled contigs were corrected using the HiRise pipeline,
546 resulting into genome assembly v2 (Putnam et al., 2016).

547

548 **Assembly reconstruction by syntenic and genetic makers**

549 ALLMAPS was applied to reconstruct scaffolds of the *L. saligna* v2 assembly to
550 chromosomal linkage groups using two types of markers: 417 genetic markers (weight
551 = 2) derived from F₂ (*L. saligna* CGN05271 x *L. sativa* cv. Olof), and syntenic markers
552 (weight = 1) derived from the reciprocal best hits between *L. saligna* v2 and *L. sativa*
553 v8 (Jeuken et al., 2001; Giesbers et al., 2019; Tang et al., 2015b). Contigs (>1 kb) not
554 clustered in chromosomes were concatenated by JCVI with 100 N-content gaps to
555 generate a virtual “chromosome zero” storing left genetic content (Tang et al., 2015a).

556

557 **Genome size estimation**

558 Two paired-end Illumina libraries of *L. saligna* were used for genome size estimation
559 (~117 Gb pairs; ~932 million reads) using a k-mer count size of 23 (Supplemental
560 Table 1). Jellyfish v2.3.0 was used to count the k-mer frequency (Marçais and
561 Kingsford, 2011). Jellyfish output was used by GenomeScope (v2.0) to estimate
562 haploid genome length, percentage of repetitive DNA, and heterozygosity of the *L.*
563 *saligna* genome using the histogram file (Vurture et al., 2017).

564

565 **Genome completeness assessment**

566 Completeness of the *L. saligna* genome assembly was evaluated using multiple
567 approaches. BUSCO (v3.0.2) assessment was conducted using the
568 eudicotyledons_odb10 database (Simão et al., 2015). In addition, 226,910 ESTs of
569 diverse *Lactuca* species (retrieved on July 2019 by NCBI) were aligned to the genome
570 using GMAP (version 2019-06-10) (Wu and Watanabe, 2005). For GMAP alignment,
571 presence/absence of ESTs was determined after filtering the alignments by identity
572 and coverage at different levels of stringency using custom scripts.

573

574 **Repeat annotation**

575 Tandem Repeats Finder v4.04 was used to detect tandem repeats using the following
576 parameters: Match=2, Mismatch=7, Delta= 7, PM=80, PI=10, Minscore=50, and
577 MaxPeriod=2000 (Benson, 1999). TEs were searched using RepeatMasker v4.0.7

578 against ortholog and *de novo* databases in a serial order (Smit et al., 2019): i.e., by
579 using orthology data from Repbase and Dfam (version 20170127), and *de novo* TEs
580 library generated by RepeatModeler v2.0 and MITE-Hunter (Han and Wessler, 2010;
581 Jurka et al., 2005; Hubley et al., 2015; Price et al., 2005). Perl tool “One code to find
582 them all” was used to parse and quantify the number and position of predicted repeat
583 elements (Bailly-Bechet et al., 2014).

584

585 **Non-coding RNA annotation**

586 Non-coding RNA (ncRNA) loci were annotated according to different types. tRNAscan-
587 SE v2.0.4 was used to annotate tRNAs using eukaryote parameters (Lowe and Eddy,
588 1997). In addition, rRNA was annotated using RNAmmer v1.2 (Lagesen et al., 2007).
589 INFERNAL v1.1.2 was used to search against the Rfam database (release 14.1) to
590 detect additional miRNA, snRNA, tRNA, rRNA, and snoRNA sequences (Kalvari et al.,
591 2018; Nawrocki and Eddy, 2013). Annotations predicted by different tools were merged
592 and condensed using GenomicRanges in R v3.6 (Lawrence et al., 2013).

593

594 **Infection assays**

595 Leaves of three-week old *L. saligna* plants (accession CGN05327) were spray-
596 inoculated with a spore suspension of *B. lactucae* race Bl:21 (2.0×10^5
597 conidiospores/mL) or with sterile water. Treated plants were first kept in the dark for 4
598 h to maximize spore germination, and then incubated in a growth chamber at 15°C and
599 a 16/8 h (day/night) photoperiod. Leaf samples of inoculated and mock-treated leaves
600 were collected at 8 and 24 hpi. Leaf samples from the three biological replicates were
601 immediately frozen in liquid nitrogen and stored in -80°C until further use.

602

603 **RNA library preparation and sequencing**

604 Total RNA was isolated from 12 infection assay samples and one pooled sample
605 consisting of root and flower bud material (pooled from different floral stages) using a
606 Direct Zol RNA Miniprep Plus kit (Zymo Research) followed by DNase treatment. RNA
607 was purified by ethanol precipitation. Concentration and purity of RNA samples was
608 measured with a Nanodrop 2000c spectrophotometer and a Qubit 4.0 fluorometer
609 using a RNA Broad Range assay (Thermo Fisher Scientific). Paired-End sequencing
610 (2 x 125 bp) was performed on an Illumina HiSeq2500 platform using two flow cell
611 lanes.

612

613 **Gene prediction**

614 Gene models for protein-coding genes were annotated by combining *ab initio*
615 prediction and homology-based annotation. First, BRAKER was used to train an
616 Augustus model with RNA-seq data to predict genes *ab initio* (Hoff et al., 2016).
617 Thereafter, MAKER was applied to integrate the *ab initio* prediction with extrinsic
618 evidence: i.e., *de novo* transcripts assembled by Trinity and protein homology data
619 (Holt and Yandell, 2011; Grabherr et al., 2011). Annotation-edit-distance (AED)
620 calculated by MAKER was used to examine the quality of the genome annotation.
621 Coding-potential was calculated by CPC2 (Kang et al., 2017) to further filter out non-
622 coding transcripts (Supplemental Data 4).

623

624 **Functional annotation**

625 Potential biological function of proteins was inferred using three criteria: i) best-hit
626 matches in SwissProt, TrEMBL, and *A. thaliana* Araport11 databases using BLAST
627 v2.2.31 and DIAMOND (Buchfink et al., 2015) (E-value cut-off = 1e-5); ii) protein
628 domains/motifs identified by InterProscan against the Pfam protein database (EI-
629 Gebali et al., 2018; Zdobnov and Apweiler, 2001); and iii) gene ontology (GO) based
630 on InterPro entries. Orthology searches for pathway analysis were conducted with
631 Kofamscan (Aramaki et al., 2019) using a customized HMM database of KEGG
632 Orthologs (Kanehisa, 2000).

633

634 **Resequencing and SNPs calling**

635 Libraries of PE reads (2 x 150 bp, insert size distribution peaks at 190 bp) were
636 constructed and sequenced. Re-sequencing reads were mapped to the *de novo*
637 genome assembly using the BWA alignment tool (Li, 2013). After mapping, the
638 alignment output in SAM format was translated to the BAM format using SAMtools (Li
639 et al., 2009). Duplicated reads were marked, and read groups were assigned to the
640 remaining reads using the tools built into GATK v4.0.8.1 (Van der Auwera et al., 2013).
641 Subsequently, HaplotypeCaller and GendypeGVCFs were applied to call variants
642 (SNPs and indels, respectively) per sample and used to perform joint genotyping.
643 These results were used to generate a vcf file containing all raw SNPs and indels.
644 SelectVariants and VariantFiltration tools in GATK were used to extract biallelic SNPs,

645 which were subjected to hard-filtering for low-quality SNPs based on several scores
646 (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0,
647 and SOR > 3.0). The distribution of each of the quality scores and their cut-offs was
648 visualized in R (Supplemental Figure 11). Subsequently, SNPs were further filtered by
649 Minor Allele Frequency (MAF) and the missing rate for each SNP (MAF < 0.05, missing
650 rate > 0.1) for downstream analysis. Lastly, the filtered SNP call-set was annotated
651 with SnpEff v4.3 using default settings to predict the nucleotide change effect of every
652 SNP (Cingolani et al., 2012). Generated read data have been deposited in the
653 European Nucleotide Archive (ENA) under reference number PRJEB36060.

654

655 **Population structure analysis**

656 PLINK2 was used to prune the SNP dataset to reduce the redundancy caused by
657 linkage disequilibrium (LD) analysis for different downstream analyzes (Purcell et al.,
658 2007). Firstly, SNPhylo was used to construct a maximum likelihood (ML) phylogenetic
659 tree using 210,358 SNPs with default settings and 1,000 bootstrap replicates (window
660 size = 50 SNPs, sliding size = 10, LD < 0.1) (Lee et al., 2014). Secondly, a PCA was
661 conducted using PLINK2 on the pruned dataset of 904,930 SNPs (window size = 50
662 SNPs, sliding size = 10 SNPs, LD < 0.5). K-means clustering was performed via Eigen
663 decomposition for the PCA and visualized in R. ADMIXTURE (v1.3.0) was used to
664 deduce ancestral history and population structure using 96,804 SNPs (window size =
665 50 SNPs, sliding size = 10 SNPs, LD < 0.05) (Alexander and Lange, 2011).
666 ADMIXTURE was utilized to determine the best number of ancestral populations (K =
667 1 to 4) by cross-validation errors (Supplemental Figure 12), and then was run again
668 with the best K value with 1,000 bootstrap replicates to infer population structure.
669 Population structure results were summarized using GISCO geographical information
670 and visualized in R (Wickham, 2016; Eurostat GISCO, 2006).

671

672 **Comparative genomic analysis**

673 The longest representative transcripts were selected from *L. saligna* and *L. sativa* as
674 the basis for synteny analysis. BLAST (v2.2.31) was used to search homologous gene
675 pairs between both species. MCScanX was employed to detect syntenic blocks (E-
676 value cut-off = 1e-5, collinear block size \geq 5) between the two *Lactuca* species using
677 the top five alignment hits, which were visualized using SynVisio and JCVI (Tang et al.,

678 2015a; Wang et al., 2012; Bandi and Gutwin, 2020). A separate synteny plot was
679 created using the best-matching hit to remove noise from polyploidy and translocation
680 events. Genetic markers on Chromosomes 5 and 8 were collected, and dot plots
681 coordinated by genetic and physical positions were visualized with R v3.6.1 to validate
682 inversions detected by synteny analysis. To assess the influence of genomic
683 inversions, syntenic gene pairs located at inversion borders were searched against the
684 *A. thaliana* protein database Araport11 (<https://www.arabidopsis.org>).

685

686 **NLR identification and classification**

687 Genome-wide searches to identify *NLRs* were conducted using the genomes of *L.*
688 *saligna* (v4) and *L. sativa* (v8) (downloaded from the CoGe website; gid35223).
689 HMMER was used to search Hidden Markov Models (HMMs) for structural domains of
690 *NLRs* (E-value cut-off = 1e-10). The Pfam models used were PF00931.23 and
691 NBS_712.hmm for the NB domain, PF01582.20 and PF13676.6 for TIR, PF18052.1
692 for CC, and eight HMMs for the LRR domain (PF00560.33, PF07723.13, PF07725.13,
693 PF12799.7, PF13306.6, PF13516.6, PF13855.6, PF14580.6). NB domains identified
694 by InterProScan (see Functional annotation section) and CC motifs predicted by
695 Paircoil2 (McDonnell et al., 2006) (P scores < 0.025) were integrated with the HMMER
696 output. *NLRs* of *L. saligna* were classified into different categories (TNL/CNL and RGC
697 families) by phylogeny clustering using *NLRs* previously identified in the *L. sativa* v8
698 genome (Christopoulou et al., 2015b). RGC families with a >3 count difference and 1.5
699 ratio between two species were selected as families with major differences. Amino acid
700 sequences of NB domains were aligned with HmmerAlign (Finn et al., 2011). The
701 alignment was trimmed by trimAl using the '-gappyout' algorithm, retaining 1,367
702 residues for phylogeny construction (Capella-Gutiérrez et al., 2009). The best-hit
703 model of evolution, Blosum62+F+R10, was first selected by IQ-TREE v1.6.12 and ML
704 trees were inferred with IQ-TREE (Nguyen et al., 2015). IQTREE (-pers 0.1, -nm 500)
705 was run independently 10 times with 1,000 ultrafast bootstrap (UFBoot) replicates.
706 Finally, the 10 best ML trees inferring the tree with highest log-likelihood was selected
707 for *NLR* classification. Phylogenetic trees were visualized and annotated using iTOL
708 v6 (Letunic and Bork, 2021).

709

710 **Identification of *NLR* clusters**

711 Annotated *NLR* genes were used to determine gene intervals of MRCs on the *L.*
712 *saligna* genome. The syntenic regions of MRCs in *L. saligna* were named Isal-MRCs
713 to distinguish them from those detected in the *L. sativa* genome. An additional sliding
714 window search was performed to identify *NLR* clusters (NCs) containing more than five
715 *NLRs* (maximum 10-genes gap). Identified MRCs and NCs were visualized on the *L.*
716 *saligna* genome using Circos (Krzywinski et al., 2009).

717

718 **RLK identification and classification**

719 Sequence similarity searches against primary protein sequences were performed with
720 HMMER v3.1 using the PKinase alignment file (PF00069; E-value cut-off = 1e-10).
721 Obtained protein sequences were subsequently scanned for the presence of
722 extracellular domains using HMMER (E-value cut-off = 1e-3; Supplemental Table 23).
723 TMHMM2.0 and SCAMPI2 were used to detect transmembrane regions (Krogh et al.,
724 2001; Peters et al., 2016).

725

726 **Mapping NHR and HI regions**

727 Genetic markers previously used in assembly reconstruction were aligned to genome
728 assembly via BLAST v2.2.31 to locate the HI and NHRs regions in *L. saligna*. The
729 genomic positions of one HI and four NHR regions were subsequently plotted on the
730 *L. saligna* genome using Circos.

731

732 **RNA-seq analysis**

733 Raw RNA-seq reads were quantified on *L. saligna* transcripts using Kallisto (v.0.44.0)
734 to gain normalized transcript per million (TPM). Transcripts with a TPM value below
735 0.1 were considered not expressed. Then, DESeq2 was used to normalize the read
736 count for each gene (total read count > 3) and execute statistical analyses to determine
737 the DEGs with $p_{adj} < 0.05$ and $\log_2FC > 1$ (Love et al., 2014). Next, the read count
738 mean and SD of infected and mock samples were calculated for all DEGs. Metascape
739 was used for enrichment analysis of up-regulated genes and to render protein–protein
740 interaction networks in Cytoscape (Zhou et al., 2019; Shannon et al., 2003). To identify
741 potential candidate genes in the four NHR regions, additional counting for DE *RLK* and
742 *NLR* genes, and highly regulated genes ($|\log_2FC| > 3$) were counted separately.

743

744 **Data availability**

745 The genome assembly described in this paper, *L. saligna* v4, is available under the
746 BioProject PRJEB56287. All raw sequencing reads have been deposited in the ENA
747 database under BioProject PRJEB56288. This includes the Illumina, PacBio, 10x
748 Genomics, Bionano and Hi-C whole-genome sequences as well as RNA sequencing
749 data for genome annotation and statistical analysis of *Bremia*-infection assay. The
750 resequencing data for 15 *L. saligna* accessions are deposited under the BioProject
751 PRJEB36060, which contains data of 100 *Lactuca* accessions derived from the TKI-
752 100 project.

753

754 **Author contributions**

755 M.E.S, S.P, R.v.T, M.J conceived of the project; L.B, S.P, M.J, K.B and M.E.S designed
756 experiments; F.RM.B, E.S, and LV.B generated plant material and sequencing data;
757 L.B, and LV.B performed the genome assembling; W.X conducted the research and
758 performed the analyses; W.X wrote the manuscript with assistance from other authors;
759 all authors approved the final manuscript.

760

761

762 **ACKNOWLEDGMENTS**

763 This research was supported by a grant of the International Lettuce Genomics
764 Consortium (ILGC) funded by the Top Consortium for Knowledge and
765 Innovation Horticultural and Starting Materials (grant number 1406-039). W.X. was
766 financially supported by a fellowship by the China Scholarship Council (CSC). We
767 thank Rens Holmer, Zhaodong Hao, Deedi Sogbohossou, Nora Walden, Sandra Smit,
768 and Henri van de Geest for their support with bioinformatic analysis and genome
769 sequencing data. We would also like to thank Wenhao Li for his help with statistics and
770 R programming. We also thank Elizabeth Marie Georgian for writing support.

771

772 **Conflict of interest statement**

773 The authors declare no conflict of interest. The funders had no role in study design,
774 data collection and analysis, decision to publish, or preparation of the manuscript.

775

776 **REFERENCES**

- 777 **Alexander, D.H. and Lange, K.** (2011) Enhancements to the ADMIXTURE algorithm for individual
778 ancestry estimation. *BMC Bioinformatics* **12**: 246.
- 779 **Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H.**
780 (2019) KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold.
781 *Bioinformatics* **36**: 2251–2252.
- 782 **Assaad, F.F., Qiu, J.L., Youngs, H., Ehrhardt, D., Zimmerli, L., Kalde, M., Wanner, G., Peck, S.C.,**
783 **Edwards, H., Ramonell, K., Somerville, C.R., and Thordal-Christensen, H.** (2004) The PEN1
784 syntaxin defines a novel cellular compartment upon fungal attack and is required for the timely assembly
785 of papillae. *Mol. Biol. Cell* **15**: 5118–5129.
- 786 **Van der Auwera, G.A. et al.** (2013) From fastQ data to high-confidence variant calls: The genome
787 analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**: 11.10.1–11.10.33.
- 788 **Bailly-Bechet, M., Haudry, A., and Lerat, E.** (2014) “One code to find them all”: A perl tool to
789 conveniently parse RepeatMasker output files. *Mob. DNA* **5**: 1–15.
- 790 **Bandi, V. and Gutwin, C.** (2020) Interactive exploration of genomic conservation. In Proceedings of the
791 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI’20) (Waterloo).
- 792 **Barragan, A.C. and Weigel, D.** (2021) Plant NLR diversity: The known unknowns of pan-NLRomes.
793 *Plant Cell* **33**: 814–831.
- 794 **Bateson, W.** (1909) Heredity and variation in modern Lights. In Darwin and Modern Science, A.C.
795 Seward, ed (Cambridge University Press: Cambridge), pp. 85–101.
- 796 **Bauer, S., Mekonnen, D.W., Hartmann, M., Yildiz, I., Janowski, R., Lange, B., Geist, B., Zeier, J.,**
797 **and Schäffner, A.R.** (2021) UGT76B1, a promiscuous hub of small molecule-based immune signaling,
798 glucosylates N-hydroxypipelicolic acid, and balances plant immunity. *Plant Cell* **33**: 714–734.
- 799 **Benson, G.** (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:
800 573–580.
- 801 **Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., and Moscou, M.J.** (2014) Nonhost resistance to rust
802 pathogens – A continuation of continua. *Front. Plant Sci.* **5**: 664.
- 803 **den Boer, E., Pelgrom, K.T.B., Zhang, N.W., Visser, R.G.F., Niks, R.E., and Jeuken, M.J.W.** (2014)
804 Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. *Theor. Appl.*
805 *Genet.* **127**: 1805–1816.
- 806 **Bonnier, F.J.M., Reinink, K., and Groenwold, R.** (1991) New sources of major gene resistance in
807 *Lactuca* to *Bremia lactucae*. *Euphytica* **61**: 203–211.
- 808 **Brutus, A., Sicilia, F., Macone, A., Cervone, F., and De Lorenzo, G.** (2010) A domain swap approach
809 reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc.*
810 *Natl. Acad. Sci.* **107**: 9452–9457.
- 811 **Buchfink, B., Xie, C., and Huson, D.H.** (2015) Fast and sensitive protein alignment using DIAMOND.
812 *Nat. Methods* **12**: 59–60.
- 813 **Coqueiro, D.S.O., de Souza, A.A., Takita, M.A., Rodrigues, C.M., Kishi, L.T., and Machado, M.A.**
814 (2015). Transcriptional profile of sweet orange in response to chitosan and salicylic acid. *BMC Genomics*
815 **16**: 1–14.
- 816 **Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T.** (2009) TrimAl: A tool for automated
817 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

- 818 **Cheng, Z., Sattler, S., Maeda, H., Sakuragi, Y., Bryant, D.A., and DellaPenna, D.** (2003) Highly
819 divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in
820 cyanobacteria and photosynthetic eukaryotes. *Plant Cell* **15**: 2343–2356.
- 821 **Chisholm, S.T., Coaker, G., Day, B., and Staskawicz, B.J.** (2006) Host-microbe interactions: Shaping
822 the evolution of the plant immune response. *Cell* **124**: 803–814.
- 823 **Christopoulou, M., McHale, L.K., Kozik, A., Wo, S.R.C., Wroblewski, T., and Michelmore, R.W.**
824 (2015a) Dissection of two complex clusters of resistance genes in lettuce (*Lactuca sativa*). *Mol. Plant-*
825 *Microbe Interact.* **28**: 751–765.
- 826 **Christopoulou, M., Wo, S.R.C., Kozik, A., McHale, L.K., Truco, M.J., Wroblewski, T., and**
827 **Michelmore, R.W.** (2015b) Genome-wide architecture of disease resistance genes in lettuce. *G3 Genes,*
828 *Genomes, Genet.* **5**: 2655–2669.
- 829 **Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and**
830 **Ruden, D.M.** (2012) A program for annotating and predicting the effects of single nucleotide
831 polymorphisms, SnpEff. *Fly (Austin)*. **6**: 80–92.
- 832 **Collins, N.C., Thordal-Christensen, H., Lipka, V., Bau, S., Kombrink, E., Qiu, J.-L., Hüchelhoven,**
833 **R., Stein, M., Freialdenhoven, A., Somerville, S.C., and Schulze-Lefert, P.** (2003) SNARE-protein-
834 mediated disease resistance at the plant cell wall. *Nature* **425**: 973–977.
- 835 **De Cremer, K., Mathys, J., Vos, C., Froenicke, L., Michelmore, R.W., Cammue, B.P.A., and De**
836 **Coninck, B.** (2013) RNAseq-based transcriptome analysis of *Lactuca sativa* infected by the fungal
837 necrotroph *Botrytis cinerea*. *Plant, Cell Environ.* **36**: 1992–2007.
- 838 **Ding, P., Rekhter, D., Ding, Y., Feussner, K., Busta, L., Haroth, S., Xu, S., Li, X., Jetter, R.,**
839 **Feussner, I., and Zhang, Y.** (2016) Characterization of a pipecolic acid biosynthesis pathway required
840 for systemic acquired resistance. *Plant Cell* **28**: 2603–2615.
- 841 **Dobzhansky, T.** (1934) Studies on hybrid sterility - I. Spermatogenesis in pure and hybrid *Drosophila*
842 *pseudoobscura*. *Zeitschrift für Zellforsch. und Mikroskopische Anat.* **21**: 169–223.
- 843 **Doležalová, I., Lebeda, A., Janeček, J., Číhalíková, J., Křístková, E., and Vránová, O.** (2002)
844 Variation in chromosome numbers and nuclear DNA contents in genetic resources of *Lactuca* L. species
845 (Asteraceae). *Genet. Resour. Crop Evol.* **49**: 383–395.
- 846 **EI-Gebali, S. et al.** (2018) The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**: D427–
847 D432.
- 848 Eurostat GISCO (2006) Administrative boundaries of NUTS 2006 and COAS 2006.
- 849 FAOSTAT (2019) Food and Agriculture Organization (FAO) of the United Nations.
850 <http://www.fao.org/faostat/en/#data> (May 18, 2019)
- 851 **Ferrari, S., Savatin, D.V., Sicilia, F., Gramegna, G., Cervone, F., and De Lorenzo, G.** (2013)
852 Oligogalacturonides: Plant damage-associated molecular patterns and regulators of growth and
853 development. *Front. Plant Sci.* **4**: 49.
- 854 **Finn, R.D., Clements, J., and Eddy, S.R.** (2011) HMMER web server: Interactive sequence similarity
855 searching. *Nucleic Acids Res.* **39**: W29–W37.
- 856 **Fishman, L. and McIntosh, M.** (2019) Standard deviations: The biological bases of transmission ratio
857 distortion. *Annu. Rev. Genet.* **53**: 347–372.
- 858 **Fishman, L. and Sweigart, A.L.** (2018) When two rights make a wrong: The evolutionary genetics of
859 plant hybrid incompatibilities. *Annu. Rev. Plant Biol.* **69**: 707–731.
- 860 **Franck, C.M., Westermann, J., Bürssner, S., Lentz, R., Lituiev, D.S., and Boisson-Dernier, A.**
861 (2018) The protein phosphatases ATUNIS1 and ATUNIS2 regulate cell wall integrity in tip-growing cells.
862 *Plant Cell* **30**: 1906–1923.

- 863 **Giesbers, A.K.J., den Boer, E., Braspenning, D.N.J., Bouten, T.P.H., Specken, J.W., van**
864 **Kaauwen, M.P.W., Visser, R.G.F., Niks, R.E., and Jeuken, M.J.W.** (2018) Bidirectional backcrosses
865 between wild and cultivated lettuce identify loci involved in nonhost resistance to downy mildew. *Theor.*
866 *Appl. Genet.* **131**: 1761–1776.
- 867 **Giesbers, A.K.J., den Boer, E., Ulen, J.J.W.E.H., van Kaauwen, M.P.W., Visser, R.G.F., Niks, R.E.,**
868 **and Jeuken, M.J.W.** (2019) Patterns of transmission ratio distortion in interspecific lettuce hybrids
869 reveal a sex-independent gametophytic barrier. *Genetics* **211**: 263–276.
- 870 **Giesbers, A.K.J., Pelgrom, A.J.E., Visser, R.G.F., Niks, R.E., van den Ackerveken, G., and**
871 **Jeuken, M.J.W.** (2017) Effector-mediated discovery of a novel resistance gene against *Bremia lactucae*
872 in a nonhost lettuce species. *New Phytol.* **216**: 915–926.
- 873 **González-Lamothe, R., Tsitsigiannis, D.I., Ludwig, A.A., Panicot, M., Shirasu, K., and Jones,**
874 **J.D.G.** (2006) The U-box protein CMPG1 is required for efficient activation of defense mechanisms
875 triggered by multiple resistance genes in tobacco and tomato. *Plant Cell* **18**: 1067–1083.
- 876 **Grabherr, M.G. et al.** (2011) Full-length transcriptome assembly from RNA-Seq data without a
877 reference genome. *Nat. Biotechnol.* **29**: 644–652.
- 878 **Han, Y. and Wessler, S.R.** (2010) MITE-Hunter: a program for discovering miniature inverted-repeat
879 transposable elements from genomic sequences. *Nucleic Acids Res.* **38**: e199–e199.
- 880 **He, Z-H, Cheeseman, I., He, D and Kohorn, B.** (1999) A cluster of five cell wall-associated receptor
881 kinase genes, Wak1–5, are expressed. *Plant Mol. Biol.* **39**: 1189–1196.
- 882 **Heath, M.C.** (1981) Nonhost resistance. *Plant Dis. Control*: 201–217.
- 883 **Hok, S., Danchin, E.G.J., Allasia, V., Panabières, F., Attard, A., and Keller, H.** (2011). An
884 *Arabidopsis* (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease.
885 *Plant. Cell Environ.* **34**: 1944–1957.
- 886 **Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M.** (2016) BRAKER1:
887 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**:
888 767–769.
- 889 **Hoffmann, A.A. and Rieseberg, L.H.** (2008) Revisiting the impact of inversions in evolution: From
890 population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**:
891 21–42.
- 892 **Holt, C. and Yandell, M.** (2011) MAKER2: An annotation pipeline and genome-database management
893 tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.
- 894 **Hu, K. et al.** (2017) Improvement of multiple agronomic traits by a disease resistance gene via cell wall
895 reinforcement. *Nat. Plants* **3**: 17009.
- 896 **Huble, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler,**
897 **T.J.** (2015) The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**: D81–D89.
- 898 **Hurni, S., Scheuermann, D., Krattinger, S.G., Kessel, B., Wicker, T., Herren, G., Fitze, M.N.,**
899 **Breen, J., Presterl, T., Ouzunova, M., and Keller, B.** (2015) The maize disease resistance gene *Htn1*
900 against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proc. Natl. Acad. Sci. U.*
901 *S. A.* **112**: 8780–5.
- 902 **Iorizzo, M. et al.** (2016) A high-quality carrot genome assembly provides new insights into carotenoid
903 accumulation and asterid genome evolution. *Nat. Genet.* **48**: 657–666.
- 904 **Jeuken, M. and Lindhout, P.** (2002) *Lactuca saligna*, a non-host for lettuce downy mildew (*Bremia*
905 *lactucae*), harbors a new race-specific *Dm* gene and three QTLs for resistance. *TAG Theor. Appl. Genet.*
906 **105**: 384–391.

- 907 **Jeuken, M., van Wijk, R., Peleman, J., and Lindhout, P.** (2001) An integrated interspecific AFLP map
908 of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F2 populations. *Theor. Appl. Genet.* **103**: 638–
909 647.
- 910 **Jeuken, M.J.W. and Lindhout, P.** (2004) The development of lettuce backcross inbred lines (BILs) for
911 exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theor. Appl. Genet.* **109**: 394–401.
- 912 **Jeuken, M.J.W., Zhang, N.W., McHale, L.K., Pelgrom, K., Den Boer, E., Lindhout, P.,**
913 **Michelmore, R.W., Visser, R.G.F., and Niks, R.E.** (2009) *Rin4* causes hybrid necrosis and race-
914 specific resistance in an interspecific lettuce hybrid. *Plant Cell* **21**: 3368–3378.
- 915 **Jones, J.D.G. and Dangl, J.L.** (2006) The plant immune system. *Nature* **444**: 323–329.
- 916 **Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichewicz, J.** (2005)
917 Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**: 462–467.
- 918 **Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman,**
919 **A., Finn, R.D., and Petrov, A.I.** (2018) Rfam 13.0: Shifting to a genome-centric resource for non-
920 coding RNA families. *Nucleic Acids Res.* **46**: D335–D342.
- 921 **Kanehisa, M.** (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**: 27–30.
- 922 **Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., and Gao, G.** (2017) CPC2: A fast
923 and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**:
924 W12–W16.
- 925 **Karasov, T.L., Horton, M.W., and Bergelson, J.** (2014) Genomic variability as a driver of plant-
926 pathogen coevolution? *Curr. Opin. Plant Biol.* **18**: 24–30.
- 927 **Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M.** (2017) Canu:
928 Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome*
929 *Res.* **27**: 722–736.
- 930 **Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L.L.** (2001) Predicting transmembrane
931 protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**: 567–
932 580.
- 933 **Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and**
934 **Marra, M.A.** (2009) Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**: 1639–
935 1645.
- 936 **Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H.H., Rognes, T., and Ussery, D.W.** (2007)
937 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**: 3100–3108.
- 938 **Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T.,**
939 **and Carey, V.J.** (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**:
940 e1003118.
- 941 **Lebeda, A., Doležalová, I., Křístková, E., Kitner, M., Petrželová, I., Mieslerová, B., and Novotná,**
942 **A.** (2009) Wild *Lactuca* germplasm for lettuce breeding: Current status, gaps and challenges. *Euphytica*
943 **170**: 15–34.
- 944 **Lebeda, A., Křístková, E., Doležalová, I., Kitner, M., and Widrlechner, M.P.** (2019) Wild *Lactuca*
945 species in North America. In *North American Crop Wild Relatives*, Volume 2, pp. 131–194.
- 946 **Lebeda, A., Křístková, E., Kitner, M., Mieslerová, B., Jemelková, M., and Pink, D.A.C.** (2014) Wild
947 *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce
948 breeding. *Eur. J. Plant Pathol.* **138**: 597–640.
- 949 **Lebeda, A. and Reinink, K.** (1994) Histological characterization of resistance in *Lactuca saligna* to
950 lettuce downy mildew (*Bremia lactucae*). *Physiol. Mol. Plant Pathol.* **44**: 125–139.

- 951 **Lebeda, A., Ryder, E.J., Grube, R., Doležalová, I., and Křístková, E.** (2007) Lettuce (*Asteraceae*;
952 *Lactuca* spp). In Genetic Resources, Chromosome Engineering, and Crop Improvement: Vegetable Crops,
953 R. J. Singh, ed (CRC), pp. 377–472.
- 954 **Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H.** (2014) SNPhylo: A pipeline to construct a
955 phylogenetic tree from huge SNP data. *BMC Genomics* **15**: 1–6.
- 956 **Letunic, I. and Bork, P.** (2021) Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree
957 display and annotation. *Nucleic Acids Res.* **49**: W293–W296.
- 958 **Li, H.** (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
959 Prepr. arXiv: 1303.3997.
- 960 **Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
961 Durbin, R., and Subgroup, 1000 Genome Project Data Processing** (2009) The Sequence
962 Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- 963 **Lin, X., Armstrong, M., Baker, K., Wouters, D., Visser, R.G.F., Wolters, P.J., Hein, I., and
964 Vleeshouwers, V.G.A.A.** (2020) RLP/K enrichment sequencing; a novel method to identify receptor-like
965 protein (*RLP*) and receptor-like kinase (*RLK*) genes. *New Phytol.* **227**: 1264–1276.
- 966 **Love, M.I., Huber, W., and Anders, S.** (2014) Moderated estimation of fold change and dispersion for
967 RNA-seq data with DESeq2. *Genome Biol.* **15**: 1–21.
- 968 **Lowe, T.M. and Eddy, S.R.** (1997) tRNAscan-SE: A Program for improved detection of transfer RNA
969 genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
- 970 **Marçais, G. and Kingsford, C.** (2011) A fast, lock-free approach for efficient parallel counting of
971 occurrences of k-mers. *Bioinformatics* **27**: 764–770.
- 972 **McDonnell, A. V, Jiang, T., Keating, A.E., and Berger, B.** (2006) Paircoil2: improved prediction of
973 coiled coils from sequence. *Bioinformatics* **22**: 356–358.
- 974 **McHale, L., Tan, X., Koehl, P., and Michelmore, R.W.** (2006) Plant NBS-LRR proteins: Adaptable
975 guards. *Genome Biol.* **7**: 212.
- 976 **McHale, L.K., Truco, M.J., Kozik, A., Wroblewski, T., Ochoa, O.E., Lahre, K.A., Knapp, S.J., and
977 Michelmore, R.W.** (2009) The genomic architecture of disease resistance in lettuce. *Theor. Appl. Genet.*
978 **118**: 565–580.
- 979 **Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W.** (2003) Genome-wide
980 analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.
- 981 **Mohnike, L., Rekhter, D., Huang, W., Feussner, K., Tian, H., Herrfurth, C., Zhang, Y., and
982 Feussner, I.** (2021) The glycosyltransferase UGT76B1 modulates N-hydroxy-pipecolic acid homeostasis
983 and plant immunity. *Plant Cell* **33**: 735–749.
- 984 **Muller, H.J.** (1942) Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6**: 71–125.
- 985 **Nawrocki, E.P. and Eddy, S.R.** (2013) Infernal 1.1: 100-fold faster RNA homology searches.
986 *Bioinformatics* **29**: 2933–2935.
- 987 **Netzer, D., Globerson, D., and Sacks, J.** (1976). *Lactuca saligna* L., a new source of resistance to
988 downy mildew (*Bremia lactucae* Regel). *HortScience* **11**: 612–613.
- 989 **Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q.** (2015) IQ-TREE: A fast and effective
990 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**: 268–274.
- 991 **Niks, R.E.** (1987) Nonhost plant species as donors for resistance to pathogens with narrow host range I.
992 Determination of nonhost status. *Euphytica* **36**: 841–852.
- 993 **Niks, R.E. and Marcel, T.C.** (2009) Nonhost and basal resistance: How to explain specificity? *New*
994 *Phytol.* **182**: 817–828.

- 995 **Norwood, J.M., Johnson, A.G., and Crute, I.R.** (1981). The utilization of novel sources of resistance
996 to *Bremia lactucae* from wild *Lactuca* species. *Euphytica* **30**: 659–668.
- 997 **Ohtake, Y., Takahashi, T., and Komeda, Y.** (2000). Salicylic acid induces the expression of a number
998 of receptor-like kinase genes in *Arabidopsis thaliana*. *Plant Cell Physiol.* **41**: 1038–1044.
- 999 **Panstruga, R. and Moscou, M.J.** (2020) What is the molecular basis of nonhost resistance? *Mol. Plant-*
1000 *Microbe Interact.* **33**: 1253–1264.
- 1001 **Parra, L., Maisonneuve, B., Lebeda, A., Schut, J., Christopoulou, M., Jeuken, M., McHale, L.,**
1002 **Truco, M.J., Crute, I., and Michelmore, R.** (2016) Rationalization of genes for resistance to *Bremia*
1003 *lactucae* in lettuce. *Euphytica* **210**: 309–326.
- 1004 **Peters, C., Tsirigos, K.D., Shu, N., and Elofsson, A.** (2016) Improved topology prediction using the
1005 terminal hydrophobic helices rule. *Bioinformatics* **32**: 1158–1162.
- 1006 **Petrželová, I., Lebeda, A., and Beharav, A.** (2011) Resistance to *Bremia lactucae* in natural
1007 populations of *Lactuca saligna* from some Middle Eastern countries and France. *Ann. Appl. Biol.* **159**:
1008 442–455.
- 1009 **Price, A.L., Jones, N.C., and Pevzner, P.A.** (2005) *De novo* identification of repeat families in large
1010 genomes. *Bioinformatics* **21**: i351–i358.
- 1011 **Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J.,**
1012 **Sklar, P., De Bakker, P.I.W., Daly, M.J., and Sham, P.C.** (2007) PLINK: A tool set for whole-genome
1013 association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**: 559–575.
- 1014 **Putnam, N.H., Connell, B.O., Stites, J.C., Rice, B.J., Hartley, P.D., Sugnet, C.W., Haussler, D.,**
1015 **and Rokhsar, D.S.** (2016) Chromosome-scale shotgun assembly using an in vitro method for long-
1016 range linkage. *Genome Res.* **26**: 342–350.
- 1017 **R Core Team** (2013) R: A Language and environment for statistical computing.
- 1018 **Raaymakers, T.M. and Van den Ackerveken, G.** (2016) Extracellular recognition of oomycetes during
1019 biotrophic infection of plants. *Front. Plant Sci.* **7**: 906.
- 1020 **Reyes-Chin-Wo, S. et al.** (2017) Genome assembly with in vitro proximity ligation data and whole-
1021 genome triplication in lettuce. *Nat. Commun.* **8**: 14953.
- 1022 **Schulze-Lefert, P. and Panstruga, R.** (2011) A molecular evolutionary concept connecting nonhost
1023 resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci.* **16**: 117–125.
- 1024 **Sedlářová, M., Luhová, L., Petřivalský, M., and Lebeda, A.** (2007) Localisation and metabolism of
1025 reactive oxygen species during *Bremia lactucae* pathogenesis in *Lactuca sativa* and wild *Lactuca* spp.
1026 *Plant Physiol. Biochem.* **45**: 607–616.
- 1027 **Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,**
1028 **Schwikowski, B., and Ideker, T.** (2003) Cytoscape: A software environment for integrated models of
1029 biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.
- 1030 **Shen, K.A., Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Lavelle, D.O., Wroblewski, T., Meyers,**
1031 **B.C., and Michelmore, R.W.** (2002) *Dm3* is one member of a large constitutively expressed family of
1032 nucleotide binding site-leucine-rich repeat encoding genes. *Mol. Plant-Microbe Interact.* **15**: 251–261.
- 1033 **Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M.** (2015)
1034 BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs.
1035 *Bioinformatics* **31**: 3210–3212.
- 1036 **Smit, A., Hubley, R., and Green, P.** (2019) 2013–2015. RepeatMasker Open-4.0.
- 1037 **Takken, F.L.W. and Goverse, A.** (2012) How to build a pathogen detector: Structural basis of NB-LRR
1038 function. *Curr. Opin. Plant Biol.* **15**: 375–384.

- 1039 **Tang, H., Krishnakumar, V., and Li, J.** (2015a) jcv: JCVI utility libraries. Zenodo.
- 1040 **Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J.C., Schnable, P.S., Lyons, E., and**
1041 **Lu, J.** (2015b) ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biol.* **16**: 3.
- 1042 **Telenius, Håk., Ponder, B.A.J., Tunnacliffe, A., Pelmear, A.H., Carter, N.P., Ferguson - Smith,**
1043 **M.A., Behmel, A., Nordenskjöld, M., and Pfragner, R.** (1992) Cytogenetic analysis by chromosome
1044 painting using dop-pcr amplified flow-sorted chromosomes. *Genes, Chromosom. Cancer* **4**: 257–263.
- 1045 **Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and**
1046 **Schatz, M.C.** (2017) GenomeScope: Fast reference-free genome profiling from short reads.
1047 *Bioinformatics* **33**: 2202–2204.
- 1048 **Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng,**
1049 **Q., Wortman, J., Young, S.K., and Earl, A.M.** (2014) Pilon: An integrated tool for comprehensive
1050 microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.
- 1051 **Wang, Y., Cordewener, J.H.G., America, A.H.P., Shan, W., Bouwmeester, K., and Govers, F.**
1052 (2015). Arabidopsis lectin receptor kinases LecRK-IX.1 and LecRK-IX.2 are functional analogs in
1053 regulating *Phytophthora* resistance and plant cell death. *Mol. Plant-Microbe Interact.* **28**: 1032–1048.
- 1054 **Wang, Y., Subedi, S., de Vries, H., Doornenbal, P., Vels, A., Hensel, G., Kumlehn, J., Johnston,**
1055 **P.A., Qi, X., Blilou, I., Niks, R.E., and Krattinger, S.G.** (2019) Orthologous receptor kinases
1056 quantitatively affect the host status of barley to leaf rust fungi. *Nat. Plants* **5**: 1129–1135.
- 1057 **Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H.,**
1058 **Kissinger, J.C., and Paterson, A.H.** (2012) MCScanX: A toolkit for detection and evolutionary analysis
1059 of gene synteny and collinearity. *Nucleic Acids Res.* **40**: e49–e49.
- 1060 **Wickham, H.** (2016) ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
- 1061 **Witek, K., Jupe, F., Witek, A.I., Baker, D., Clark, M.D., and Jones, J.D.G.** (2016) Accelerated
1062 cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* **34**:
1063 656–660.
- 1064 **Wolf, S.** (2017) Plant cell wall signalling and receptor-like kinases. *Biochem. J.* **474**: 471–492.
- 1065 **Wroblewski, T., Piskurewicz, U., Tomczak, A., Ochoa, O., and Michelmore, R.W.** (2007) Silencing
1066 of the major family of NBS-LRR-encoding genes in lettuce results in the loss of multiple resistance
1067 specificities. *Plant J.* **51**: 803–818.
- 1068 **Wu, T.D. and Watanabe, C.K.** (2005) GMAP: A genomic mapping and alignment program for mRNA
1069 and EST sequences. *Bioinformatics* **21**: 1859–1875.
- 1070 **Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., Cai, B., Zhou, J.M., He, S.Y., and Xin,**
1071 **X.F.** (2021). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* **592**:
1072 105–109.
- 1073 **Yeo, S., Coombe, L., Warren, R.L., Chu, J., and Birol, I.** (2017) ARCS: Scaffolding genome drafts
1074 with linked reads. *Bioinformatics* **34**: 725–731.
- 1075 **Zdobnov, E.M. and Apweiler, R.** (2001) InterProScan--an integration platform for the signature-
1076 recognition methods in InterPro. *Bioinformatics* **17**: 847–848.
- 1077 **Zhang, N.W., Lindhout, P., Niks, R.E., and Jeuken, M.J.W.** (2009a) Genetic dissection of *Lactuca*
1078 *saligna* nonhost resistance to downy mildew at various lettuce developmental stages. *Plant Pathol.* **58**:
1079 923–932.
- 1080 **Zhang, N.W., Pelgrom, K., Niks, R.E., Visser, R.G.F.F., and Jeuken, M.J.W.** (2009b) Three
1081 combined quantitative trait loci from nonhost *Lactuca saligna* are sufficient to provide complete
1082 resistance of lettuce against *Bremia lactucae*. *Mol. Plant. Microbe. Interact.* **22**: 1160–8.

- 1083 **Zhang, Y., Xu, S., Ding, P., Wang, D., Cheng, Y.T., He, J., Gao, M., Xu, F., Li, Y., Zhu, Z., Li, X.,**
1084 **and Zhang, Y.** (2010) Control of salicylic acid synthesis and systemic acquired resistance by two
1085 members of a plant-specific family of transcription factors. *Proc. Natl. Acad. Sci.* **107**: 18220–18225.
- 1086 **Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and**
1087 **Chanda, S.K.** (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level
1088 datasets. *Nat. Commun.* **10**: 1–10.
- 1089 **Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J.D.G., Felix, G., and Boller, T.** (2004)
1090 Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**: 764–767.
- 1091 **Zohary, D.** (1991) The wild genetic resources of cultivated lettuce (*Lactuca sativa* L.). *Euphytica* **53**:
1092 31–35.
- 1093 **Zuo, W. et al.** (2015) A maize wall-associated kinase confers quantitative resistance to head smut. *Nat.*
1094 *Genet.* **47**: 151–157.
- 1095

1096 **FIGURE LEGENDS**

1097 **Figure 1** Resequencing of 15 accessions illustrates the *L. saligna* population structure. A, Neighbor-
1098 joining tree of 15 re-sequenced *L. saligna* accessions based on called SNPs. Accessions were clustered
1099 into three clades (colored in red, blue, and purple). Two accessions with unknown origins obtained from
1100 a French botanical garden are labelled by dashed lines. The black arrow indicates reference accession
1101 CGN05327 used for *de novo* sequencing. B, Principal component analysis plot of the top two-
1102 components illustrating the *L. saligna* population structure. Colors and shapes correspond to clades 1,
1103 2, and 3. C, Genetic ancestry estimation with presumed populations (K=2 and K=3) indicating the
1104 population number and evolution. Red and blue represents the two ancestral populations and the purple
1105 indicates an intermediate population between the two ancestors. D, Geographic locations of *L. saligna*
1106 accessions, colored and shaped based on population structure.

1107

1108 **Figure 2** Synteny reveals two large inversions on chromosomes 5 and 8 between *L. saligna* and *L.*
1109 *sativa*. A, Synteny of best orthologs for each chromosome between the two *Lactuca* species. Each
1110 chromosome is represented by a different color. B-C, Inverted synteny regions on chromosome 5 (purple)
1111 and 8 (pink) with 50 flanking genes shown at borders, respectively. Relative to *L. saligna*, the red lines
1112 link the first and last homologous gene pairs within inverted synteny, while the black lines indicate the
1113 first homologous pairs outside of the inversion.

1114

1115 **Figure 3** Phenotype mapping associates immune gene hotspots with NHR and HI regions. Track A,
1116 Circular ideogram of the nine pseudo-chromosomes (Mb) of the *L. saligna* assembly indicating two major
1117 inverted regions between *L. saligna* and *L. sativa*. Track B, Histogram of *RLK* density (1Mb window).
1118 Track C, *NLR* density (1Mb window) and tiles related to disease-resistance gene cluster intervals: i.e.
1119 major resistance clusters (MRCs), *NLR* clusters (NCs) with elevated density, and previously identified
1120 NHR interval. Track D, HI segment found on chromosome 8 using backcross inbred lines (*L. saligna* x
1121 *L. sativa*).

1122

1123 **Figure 4** Enrichment analysis and expression levels of DEGs in *L. saligna* upon *Bremia* infection. A,
1124 Heatmap of the top 20 ontology groups at 8 and 24 hpi. Each group comprises multiple ontology terms
1125 and is represented by the term with the best p-value. Groups are hierarchically clustered and heatmap
1126 cells are colored according transformed p-values $[-\log_{10}(p\text{-value})]$. Grey cells indicate a lack of
1127 enrichment for that term in the corresponding gene list. B, Networks of representative terms for the top
1128 20 groups. Each term is displayed by a pie chart node to illustrate the proportional number of up-
1129 regulated genes at 8 hpi (blue) and 24 hpi (red). Some groups are interconnected and form a larger
1130 network. C, Distribution of up-regulated genes across the four identified NHR regions in *L. saligna* at 8
1131 hpi. The horizontal dashed line ($y=3$) indicates the cutoff for up-regulated genes ($\log_2FC > 3$). Receptor-
1132 like kinases (*RLKs*) are indicated by red triangles, and other genes are black circles. The dashed ellipse
1133 line points out the three tandem arrayed *WAKs* on NHR8.

1134

1135 **Supplemental Figure 1** Sequencing and assembly workflow to construct the *L. saligna* reference
1136 genome. The sequencing techniques and reconstruction approaches are placed on the left. The dashed
1137 lines display contigs/scaffolds construction.

1138
1139 **Supplemental Figure 2** Genome size estimation of *L. saligna* by GenomeScope. The 21-mers were
1140 counted by Jellyfish, and output was taken by GenomeScope to estimate the genome size of *L. saligna*.
1141 The frequency (y-axis) and sequencing depth (x-axis) of 21-mer are plotted. The genome size (2.27 Gb)
1142 was estimated by the highest peak depth. len: Genome haploid length; uniq: genome uniq length; aa:
1143 homozygosity %; kcov: k-mer coverage; err: read error rate; dup: average rate of read duplication; k: k-
1144 mer length; p: ploidy.

1145
1146 **Supplemental Figure 3** ALLMAPS re-scaffolding for *L. saligna* pseudo-chromosomes using genetic
1147 and syntenic map. The *L. saligna* assembly v3 was reconstructed into 9 pseudo-chromosomes using
1148 genetic markers (weight=2) and syntenic markers (weight=1). For linkage map of each chromosome,
1149 left bar represents genes from the same syteny block, the right bar indicates the markers' position on
1150 genetic map. The two types of markers are connected to the middle ideogram of each reconstructed
1151 chromosome by orange and green lines respectively. Grey and white colors in chromosomes stand for
1152 different scaffolds from previous assembly.

1153
1154 **Supplemental Figure 4** Genomic features of the *Lactuca saligna* genome. A, Circular ideogram of the
1155 nine pseudo-chromosomes of *L. saligna* CGN05327 at Mb scale. Grey areas represent oriented regions
1156 compared to *L. sativa*. B, black regions represent inverted regions between *L. saligna* and *L. sativa*. B,
1157 Gene density (red; 1Mb window). C, GC content percentage (blue line; 1Mb window; outward). D,
1158 Density of single-nucleotide polymorphisms (SNPs) (green; 1Mb window). E, TE content percentage
1159 (purple line; 1Mb window; outward).

1160 **Supplemental Figure 5** Annotation Edit Distance (AED) cumulative fraction genome *de novo*
1161 annotation. Annotation Edit Distance (AED) indicates how well a predicted gene model is supported by
1162 biological evidence. AED values range from 0 and 1, with 0 denoting perfect agreement of the annotation
1163 to aligned evidence, and 1 denoting no evidence support for the annotation. Around 93% of the
1164 annotations having AEDs of less than 0.5, where AED smaller than 0.5 indicates a well-supported gene
1165 model.

1166
1167 **Supplemental Figure 6** Diversity of leaf shape of 10 resequenced *L. saligna* accessions. Leaf shapes
1168 of *L. saligna* accessions of European, Middle Eastern and West Asian origin reordered by clustering
1169 result (Figure 1).

1170
1171 **Supplemental Figure 7** Syntenic path dot plot of *L. sativa* versus *L. saligna* highlighting two large
1172 inversions on chromosomes 5 and 8. The y-axis represents the 9 *L. sativa* chromosomes, the x-axis
1173 represents the 9 *L. saligna* chromosomes. Overall, primary syteny is seen between all chromosomes

1174 (major diagonal line). Some off-axis synteny is also seen due to the ancient shared polyploidy between
1175 these two species. The inverted synteny blocks on chromosomes 5 and 8 are highlighted by pink boxes.
1176

1177 **Supplemental Figure 8** Genetic distance versus physical position on chromosome 5 and 8 supporting
1178 the presence of large inversions (i.e. a small genetic distance corresponding to a large physical distance).
1179 The SNP-derived markers for chromosome reconstruction were used for plotting genetic distance
1180 (based on an F2 genetic map of *L. saligna* x *L. sativa*) versus physical position in *L. saligna* on
1181 chromosome 5 (A) and 8 (B). The y-axis represents the genetic position of the genetic markers. The x-
1182 axis represents the physical position of genetic markers on the *L. saligna* genome. The inversion
1183 intervals are denoted by black lines based on the physical position of syntenic genes. The Hybrid
1184 Incompatibility (HI) regions found by TRD in BIL is located by pink line. The locus of Nonhost Resistance
1185 (NHR) fine-mapped on chromosome 8 (NHR8) is denoted by green line (Support Figure 3).
1186

1187 **Supplemental Figure 9** Circular tree of *L. saligna* nucleotide binding-leucine rich repeat receptors (NLR)
1188 generated by IQTREE. The tree was re-rooted at the midpoint between TNL and CNL clade, and ultra-
1189 fast bootstrap approximation (UFBoot) support values was calculated 1000 repetitions. Branch color
1190 represents the species: blue for *L. sativa*, red for *L. saligna*. Triangle indicates the NLRs with complete
1191 domain structure via HMMER search. Nomenclature of resistance gene candidate (RGC) family was
1192 referring to previously phylogeny of lettuce (Christopoulou et al., 2015b).
1193

1194 **Supplemental Figure 10** Top 100 enriched biological clusters of up-regulated DE genes at 8 and 24
1195 hpi. The enrichment analysis of ontology clustered the up-regulated DE genes at 8 and 24 hpi. The
1196 *Arabidopsis thaliana* gene ids were used to annotate DE genes and the plots were adjusted from the
1197 enrichment analysis results generated by Metascape platform. Heatmap of the top 100 ontology groups
1198 represented by the term with the best p-value. The terms are ordered by the hierarchically clustering
1199 result. The heatmap cells are colored by transformed p-values. Grey cells indicate the lack of enrichment
1200 for that term in the corresponding gene list. The number of each cluster or row in heatmap represents
1201 the order of clustering result.
1202

1203 **Supplemental Figure 11** Hard-filtering for biallelic SNPs variant from resequencing analysis of 15 *L.*
1204 *saligna* accessions. The distribution of six quality parameters: QualByDepth (QD), FisherStrand (FS),
1205 RMSMappingQuality (MQ), MappingQualityRankSumTest (sMQRankSum), StrandOddsRatio (SOR),
1206 ReadPosRankSumTest (ReadPosRankSum). The filtering cut-off values are denoted by black lines.
1207

1208 **Supplemental Figure 12** Cross-validation estimates the best K (population numbers) of ADMIXTURE
1209 for 15 re-sequenced *L. saligna* accessions. A good number of population (K) has a lower cross-validation
1210 (CV) error compared to other K values. K value ranges from 1 to 4 were applied to calculate the CV
1211 error. The K with lowest CV error was assumed as the best estimation for number of ancestral
1212 populations.
1213

1214 **TABLES**

1215 **Table 1** Genome assembly summary

Assembly level	Contig	Scaffold			Pseudo-molecule	
Scaffolding methods	PacBio + Illumina	Bionano + 10X Genomics	Dovetail		Genetic mapping	Merge unmapped
Version	v0.5	Unmapped	v1	v2	v3	v4
N50/number	6,957	-	307	5	4	-
N50/size	88.0 Kb	-	1.8 Mb	146.7 Mb	192.1 Mb	-
N90/number	22,020	-	928	13	8	-
N90/size	31.0 Kb	-	0.6 Mb	62.2 Mb	151.4 Mb	-
Largest contig/scaffold	794.0 Kb	1.1 Mb	8.2 Mb	279.9 Mb	279.9 Mb	-
Size of assembly (Gb)	2.03 Gb	0.42 Gb	1.75 Gb	1.75 Gb	1.75 Gb	2.17 Gb
Contig/scaffold number	31,431	6,174	1,376	24	9+7	9+1

1216

1217 **Table 2** Genome annotation summary

Genome annotation	Metrics	Stats
Gene prediction	<i>n</i> of genes	42,908
	Mean length of CDS	1,117 bp
	Mean exon number	5.1
	<i>n</i> protein-coding genes	40,730 (94.9%)
ncRNA	<i>n</i> of rRNAs	4,114
	<i>n</i> of tRNAs	1,857
	<i>n</i> of miRNAs	128
	<i>n</i> of snRNAs	329
Transposable elements	%Retrotransposons	67.8% (1.5 Gb)
	%DNA transposons	3.1% (66.9 Mb)
	%Unclassified repeats	6.6% (143.6 Mb)
	%Total	77.5% (1.7 Gb)

1218

1219

1220 **Table 3** Identification and classification of *NLRs* and *RLKs* for *L. saligna* and *L. sativa*

Family	Classification	Species	
		<i>L. saligna</i>	<i>L. sativa</i>
<i>NLR</i>	CNL ^a	139	162
	TNL	184	202
	Total	323	364
<i>RLK</i> ^b	LRR-RK	213	245
	G-LecRK	79	128
	Malectin-RK	55	50
	WAK	48	53
	CRK	35	36
	L-LecRK	29	35
	LysM-RK	12	12
	Rcc1-RK	5	5
	C-LecRK	1	1
	Other ^c	1	1
Total	478	566	

1221 ^a Including RPW8 and Rx_N type of CNLs.

1222 ^b Based on extracellular domain architecture via HMMER (Supplemental Table 24; Supplemental Dataset 5A-C).

1223 ^c According to iTAK classification the other RLKs are a RLK-Pelle_DLSV (*L. sativa*) and a RLK-Pelle_PERK-1 (*L. saligna*)
1224 (Supplemental Dataset 5A-B).

1225

1226 **SUPPLEMENTAL TABLES**

- 1227 **Supplemental Table 1** Resequencing data for genome size estimation
- 1228 **Supplemental Table 2** ALLMAPS genome reconstruction by genetic and syntenic mapping
- 1229 **Supplemental Table 3** ALLMAPS summary for the consensus map
- 1230 **Supplemental Table 4** Chromosomal length of *L. saligna* (v4)
- 1231 **Supplemental Table 5** BUSCO assessment*
- 1232 **Supplemental Table 6** Mapping ESTs of different *Lactuca* species to *L. saligna* genome
- 1233 **Supplemental Table 7** Repeat annotation summary
- 1234 **Supplemental Table 8** Categories of TEs predicted in the *L. saligna* genome
- 1235 **Supplemental Table 9** Subcategories of TEs predicted in the *L. saligna* genome
- 1236 **Supplemental Table 10** ncRNA prediction by different tools
- 1237 **Supplemental Table 11** Functional gene annotation
- 1238 **Supplemental Table 12** Passport of 15 *L. saligna* accessions
- 1239 **Supplemental Table 13** Summary of *L. saligna* re-sequencing
- 1240 **Supplemental Table 14** Summary of filtered SNPs for 15 TKI lines against *L. saligna* reference genome (v4)
- 1241 **Supplemental Table 15** Annotation of filtered SNPs by SnpEff
- 1242 **Supplemental Table 16** Chromosomal position and gene count of inversions
- 1243 **Supplemental Table 17** Genes at the border of putative breaking regions
- 1244 **Supplemental Table 18** Genes flanking interspecific inversions between *L. saligna* and *L. sativa*
- 1245 **Supplemental Table 19** ID conversion of genes flanking interspecific inversion
- 1246 **Supplemental Table 20** *NLR* gene number in *L. saligna* and *L. sativa*
- 1247 **Supplemental Table 21** Genomic position of Major Resistance Clusters (MRCs) in *Lactuca saligna*
- 1248 **Supplemental Table 22** *NLR* clusters (NCs) in the *Lactuca saligna* v4 assembly
- 1249 **Supplemental Table 23** Number of *NLRs* per RGC family in *L. sativa* and *L. saligna*
- 1250 **Supplemental Table 24** Pfam HMM motifs used for *RLK* classification
- 1251 **Supplemental Table 25** Hybrid incompatibility (HI) regions mapping in *L. saligna* v4 assembly
- 1252 **Supplemental Table 26** NHR intervals and *R* gene locus mapping in *L. saligna* v4 assembly
- 1253 **Supplemental Table 27** Number of differentially expressed genes in *L. saligna* upon *Bremia* infection
- 1254 **Supplemental Table 28** DEGs in oomycete related ontology of *Bremia*-infected *L. saligna*
- 1255 **Supplemental Table 29** Expression level of DEGs in oomycete related ontology of *Bremia*-infected *L. saligna*
- 1256 **Supplemental Table 30** Statistics of DEGs in four NHR regions of *L. saligna*
- 1257 **Supplemental Table 31** Candidate genes in NHR8 with potential *L. saligna* nonhost resistance

1258 **SUPPLEMENTAL DATASETS**

1259 **Supplemental Dataset 1**

1260 **Supplemental Datasets 2A-F**

1261 **Supplemental Datasets 3A-B**

1262 **Supplemental Datasets 4A-B**

1263 **Supplemental Datasets 5A-D**

1264 **Supplemental Datasets 6A-B**

1265 **Supplemental Dataset 7**

1266

1267

1268

1269

1270

1271

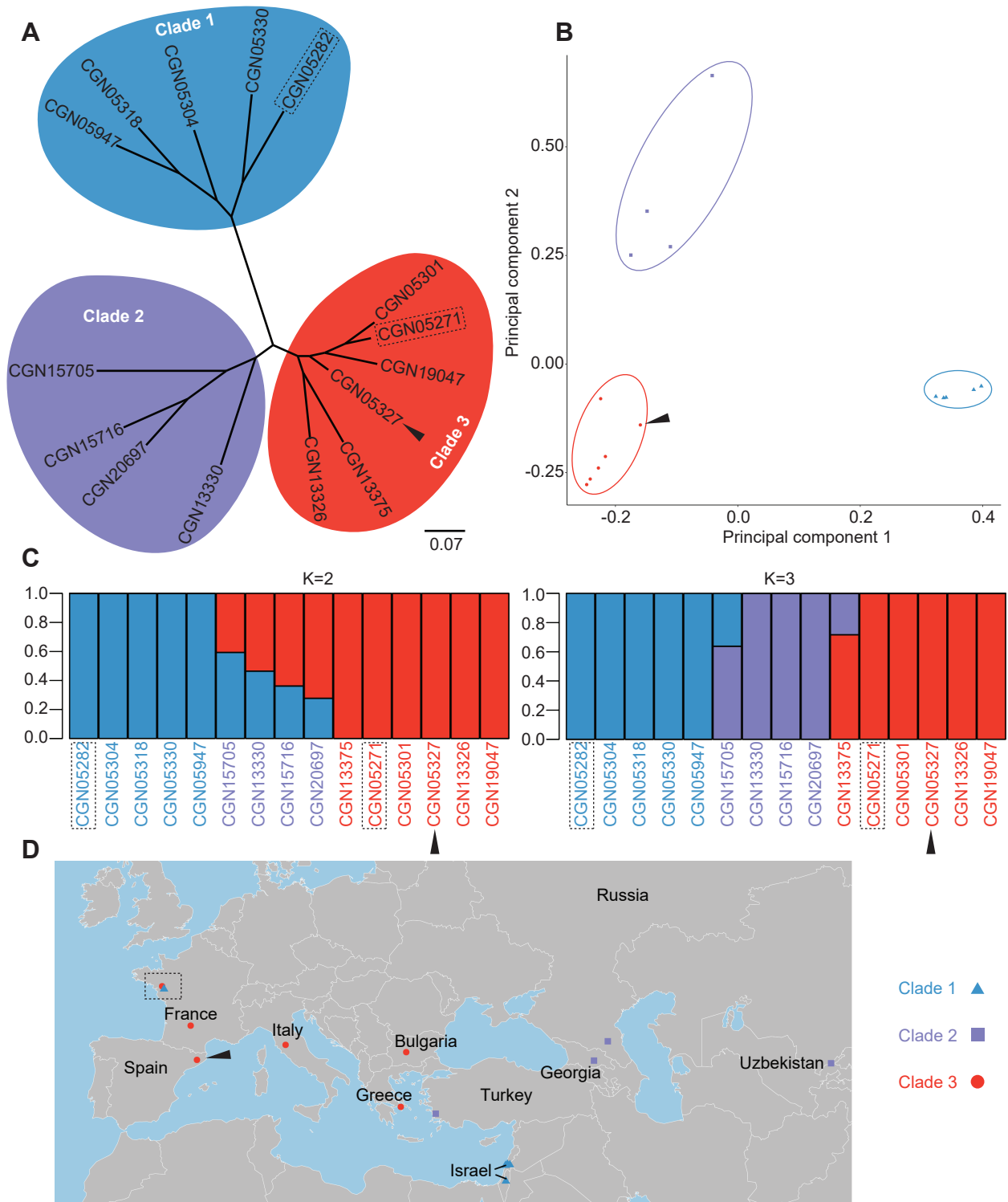
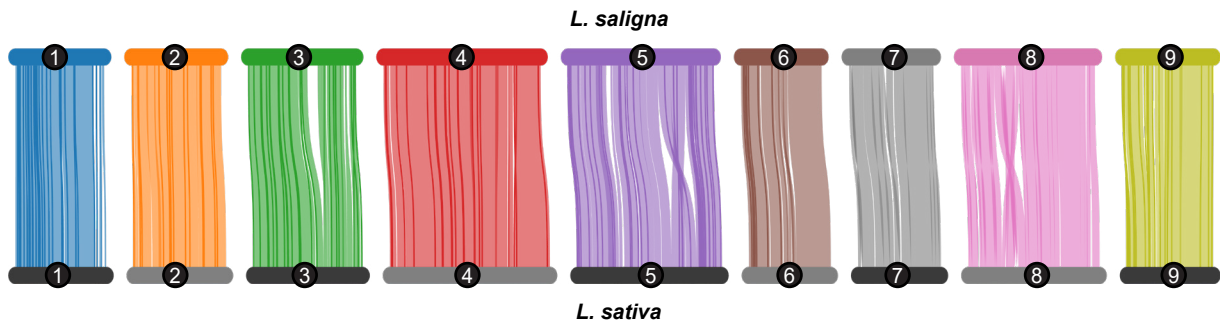
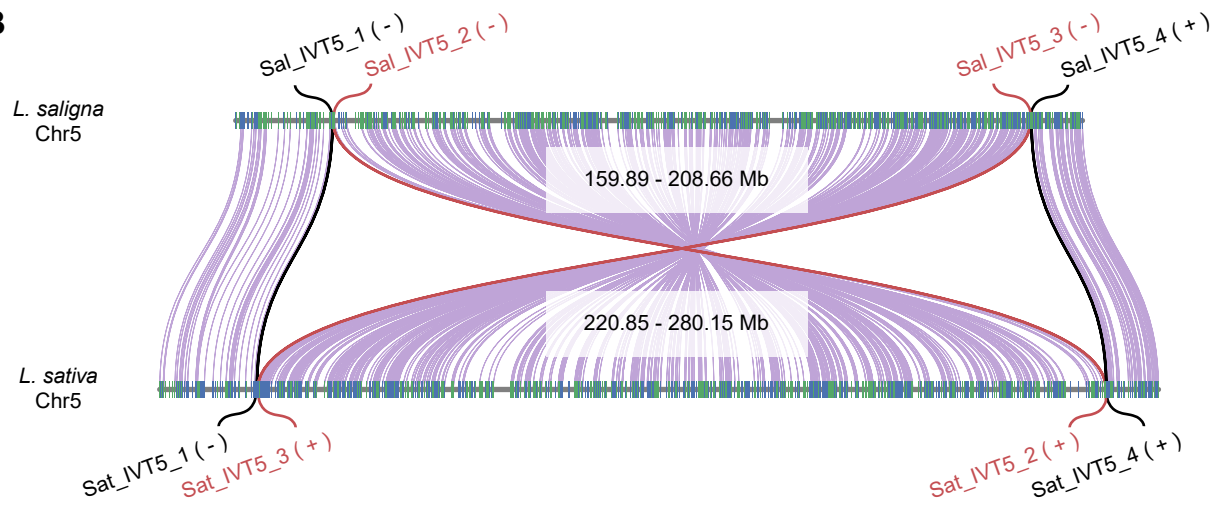


Figure 1 Resequencing of 15 accessions illustrates the *L. saligna* population structure.

A



B



C

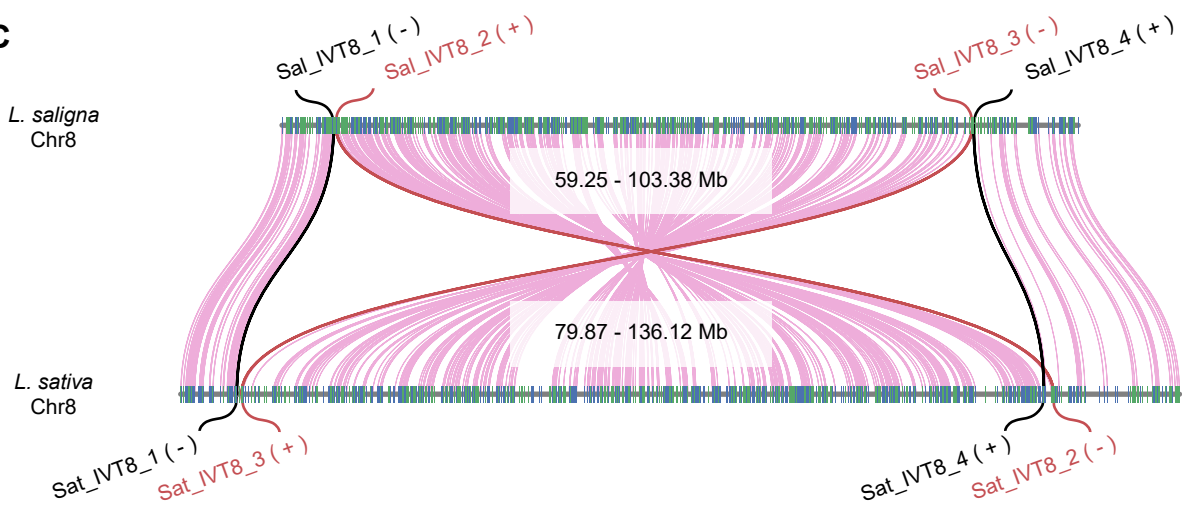


Figure 2 Synteny reveals two large inversions on chromosomes 5 and 8 between *L. saligna* and *L. sativa*.

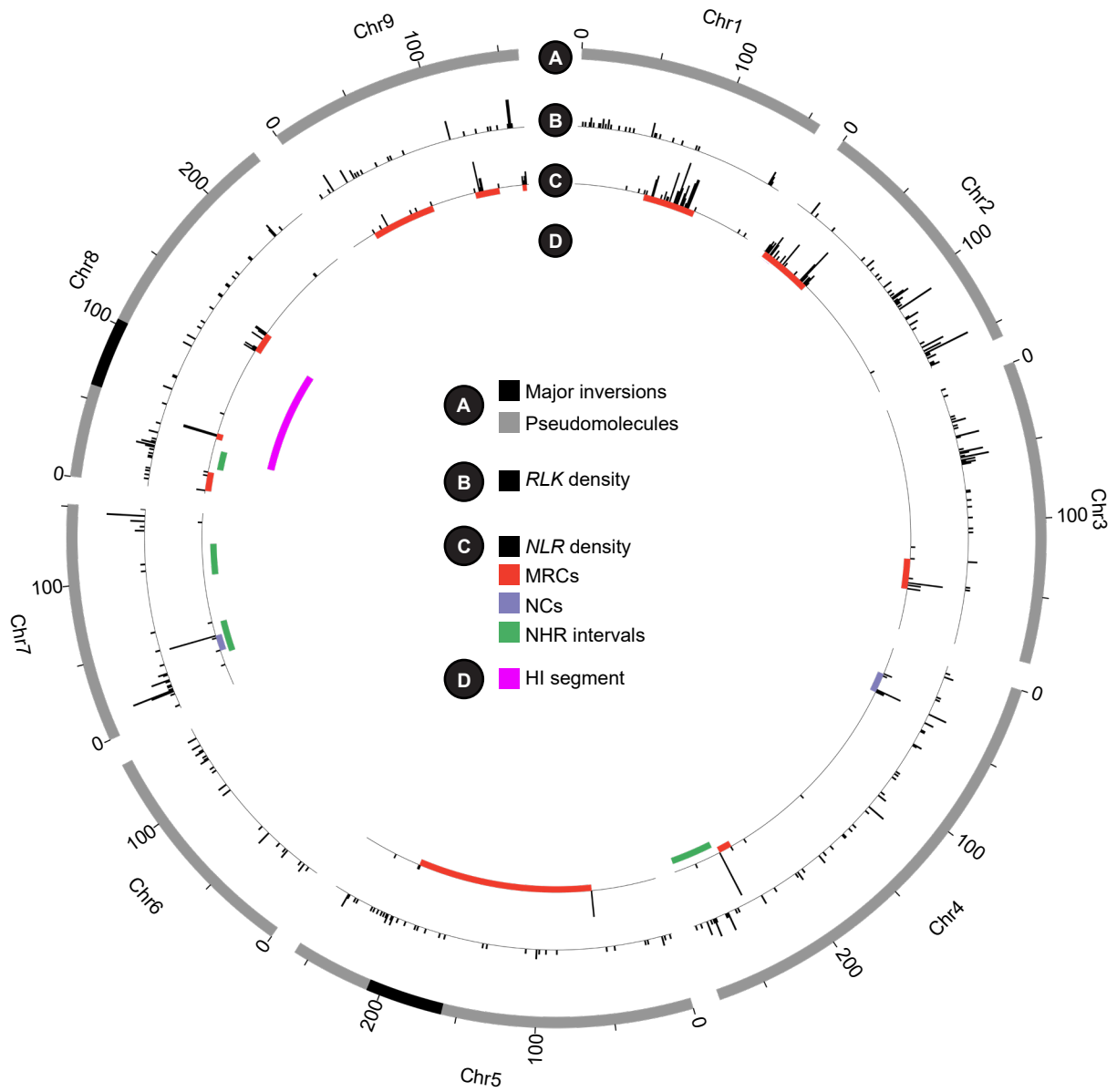


Figure 3 Phenotype mapping associates immune gene hotspots with NHR and HI regions.

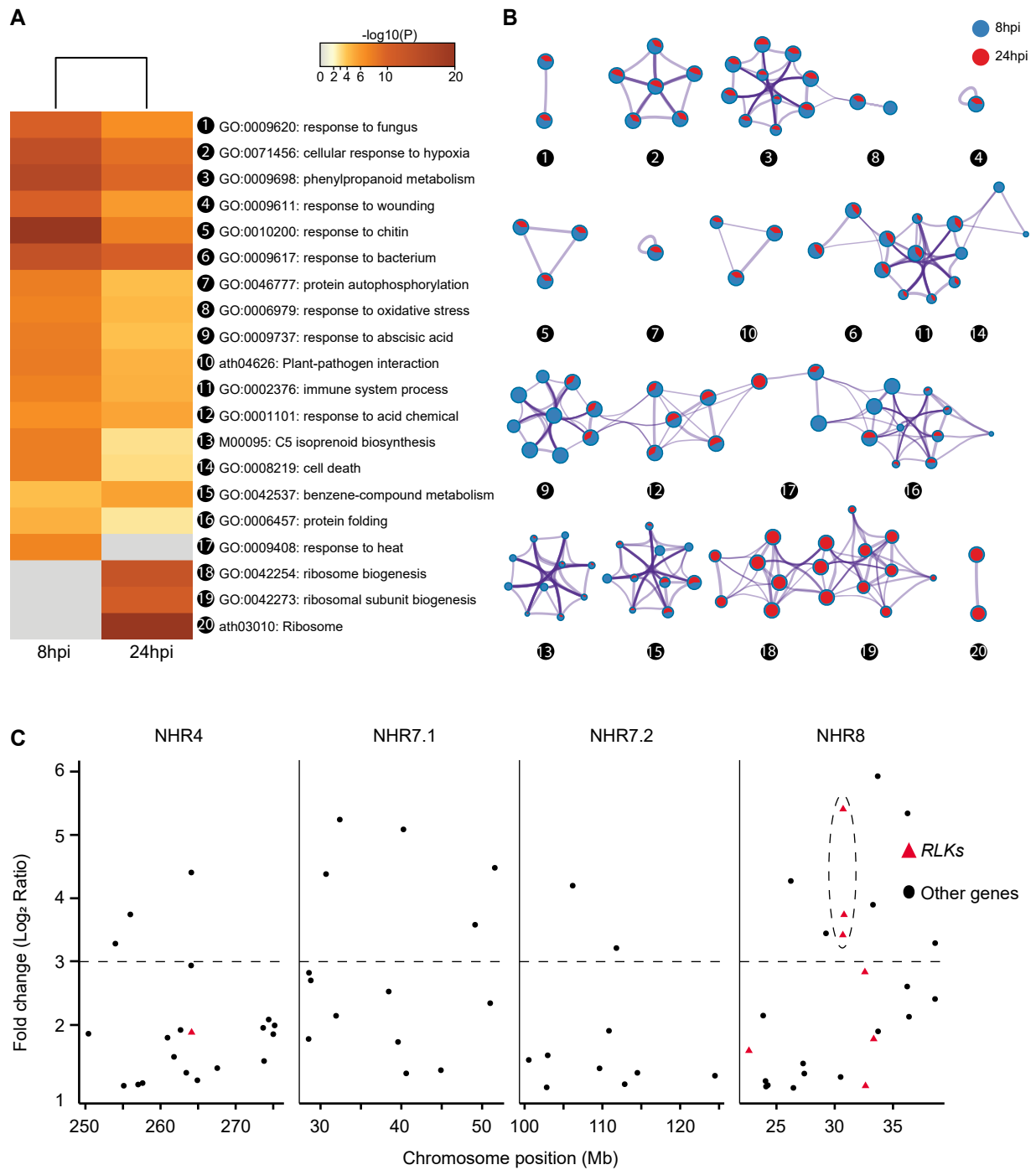
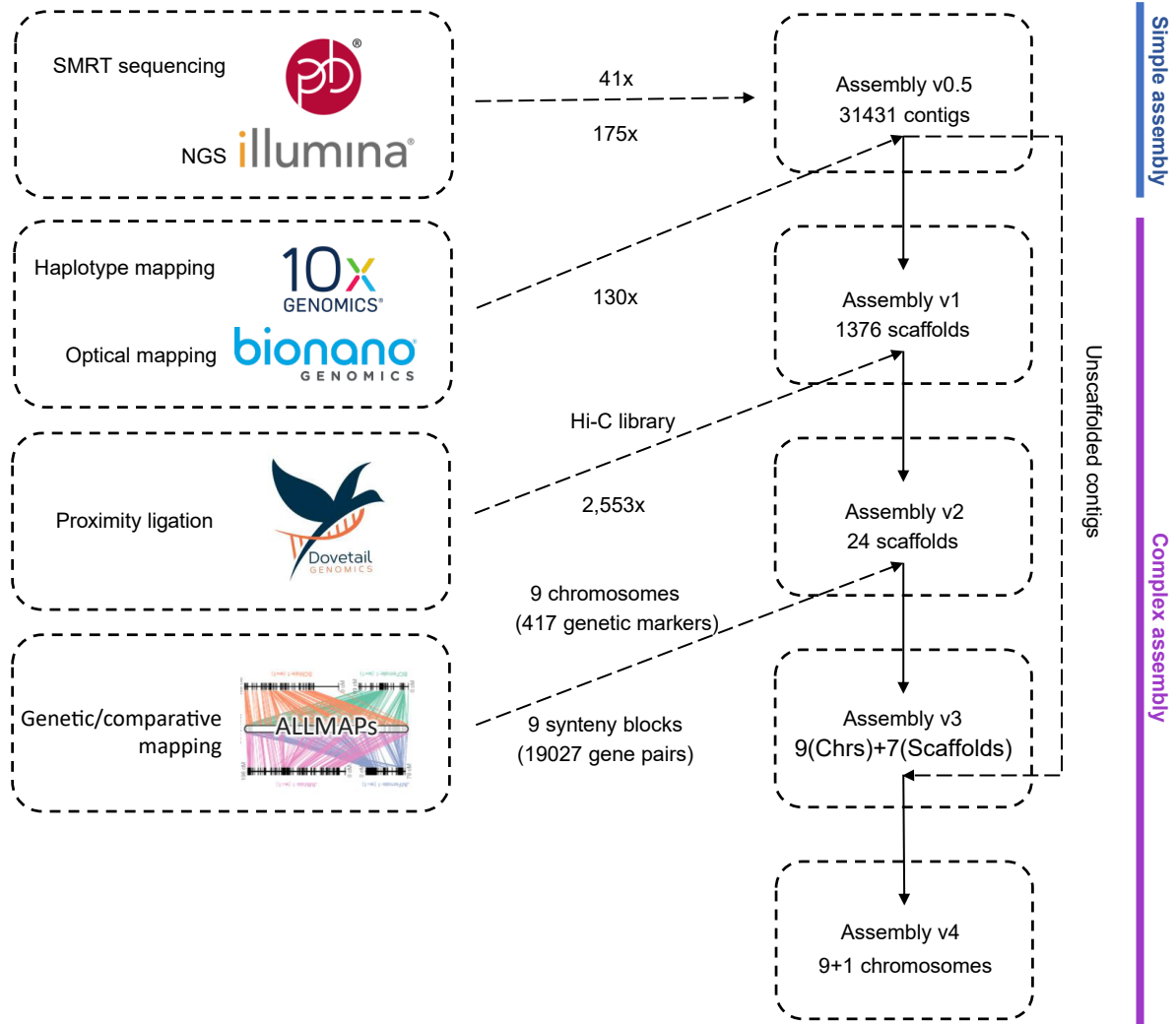


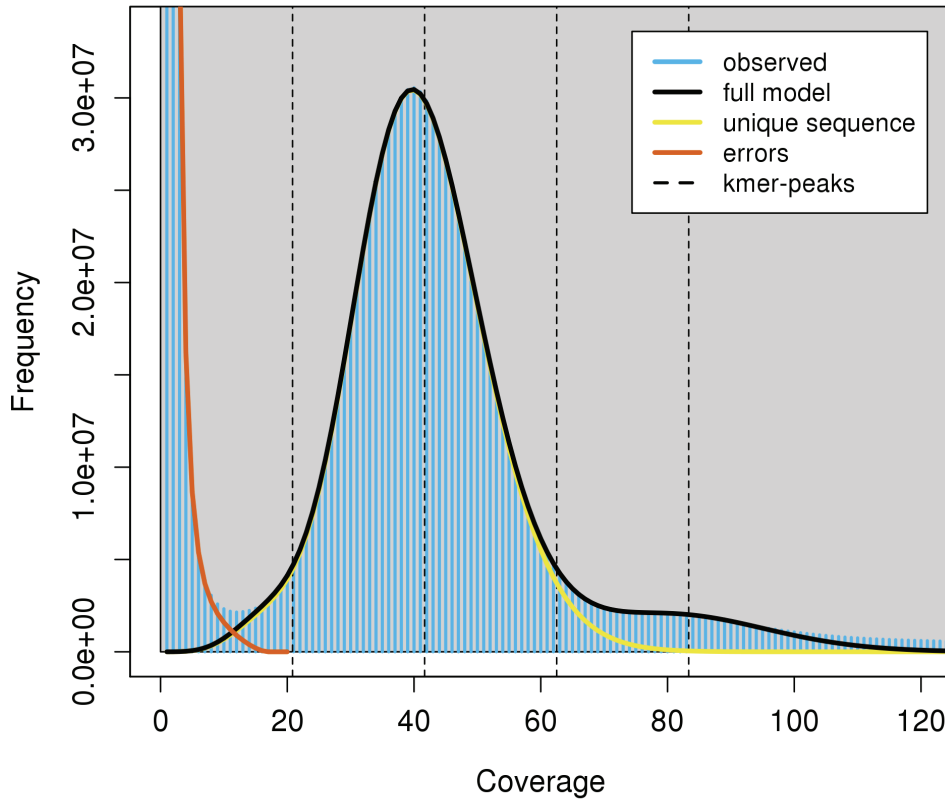
Figure 4 Enrichment analysis and expression levels of DEGs in *L. saligna* upon *Bremia* infection.



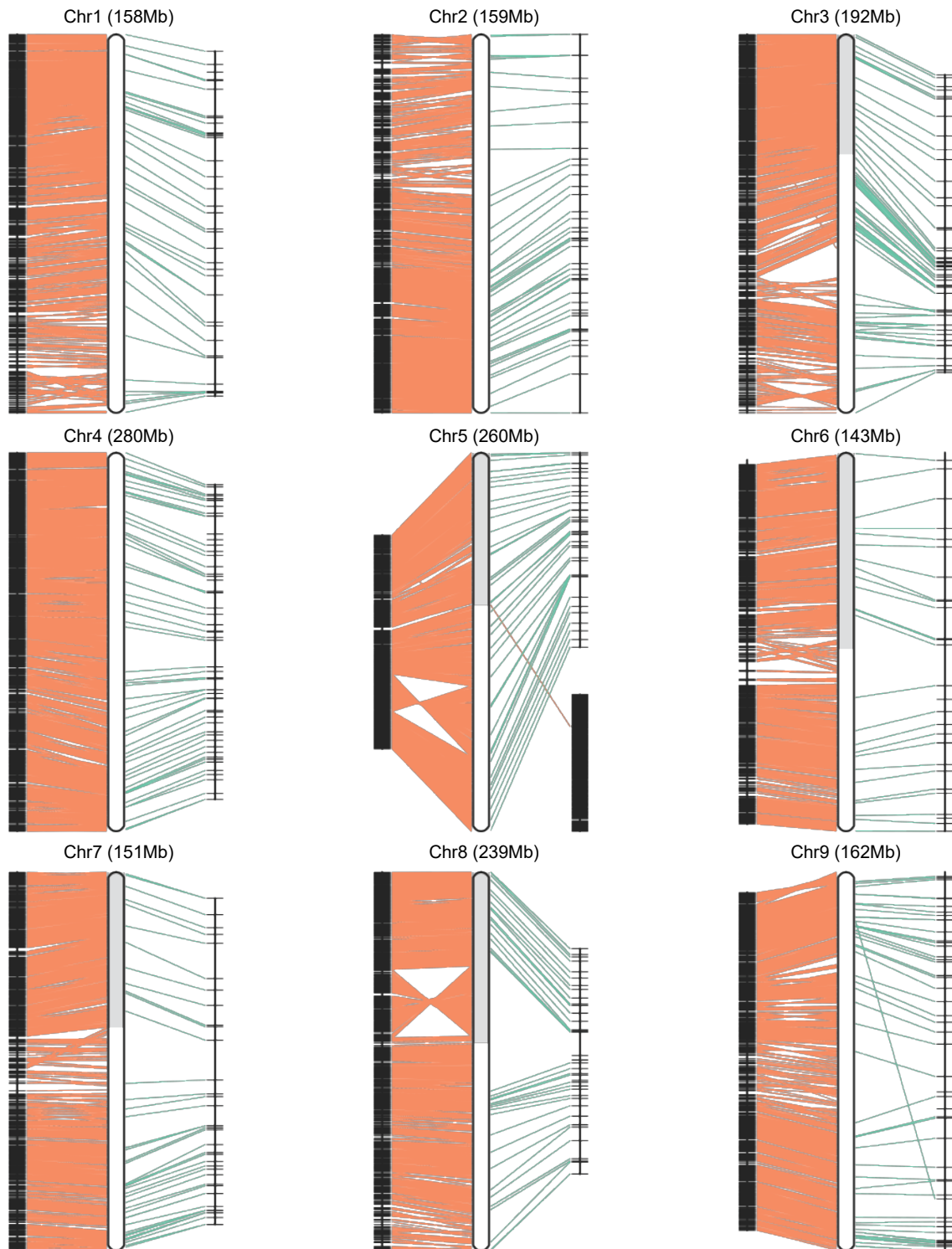
Supplemental Figure 1 Sequencing and assembly workflow to construct the *L. saligna* reference genome.

GenomeScope Profile

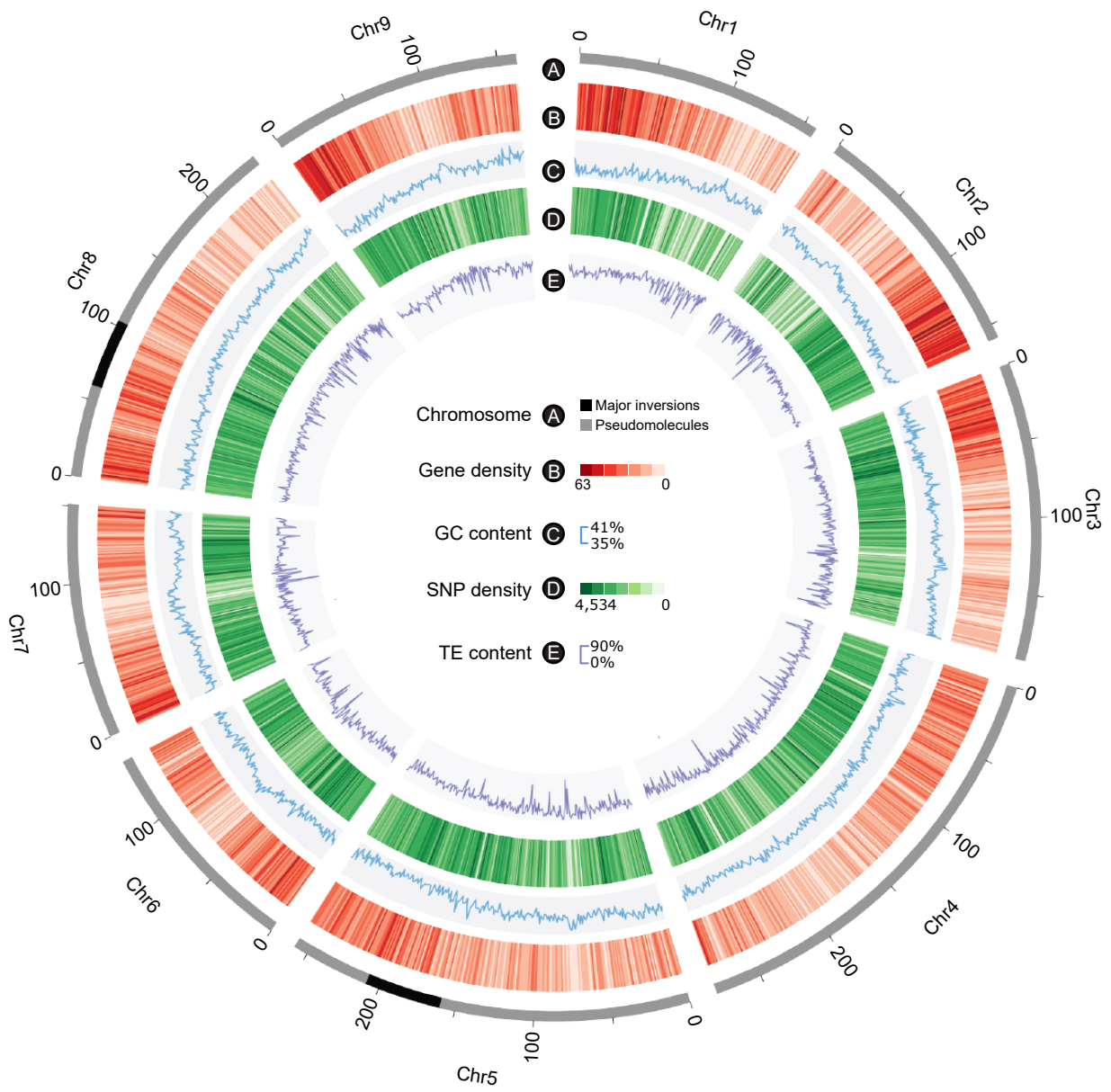
len:2,265,727,924bp uniq:33.8%
aa:99.9% ab:0.116%
kcov:20.8 err:0.168% dup:1.37 k:21 p:2



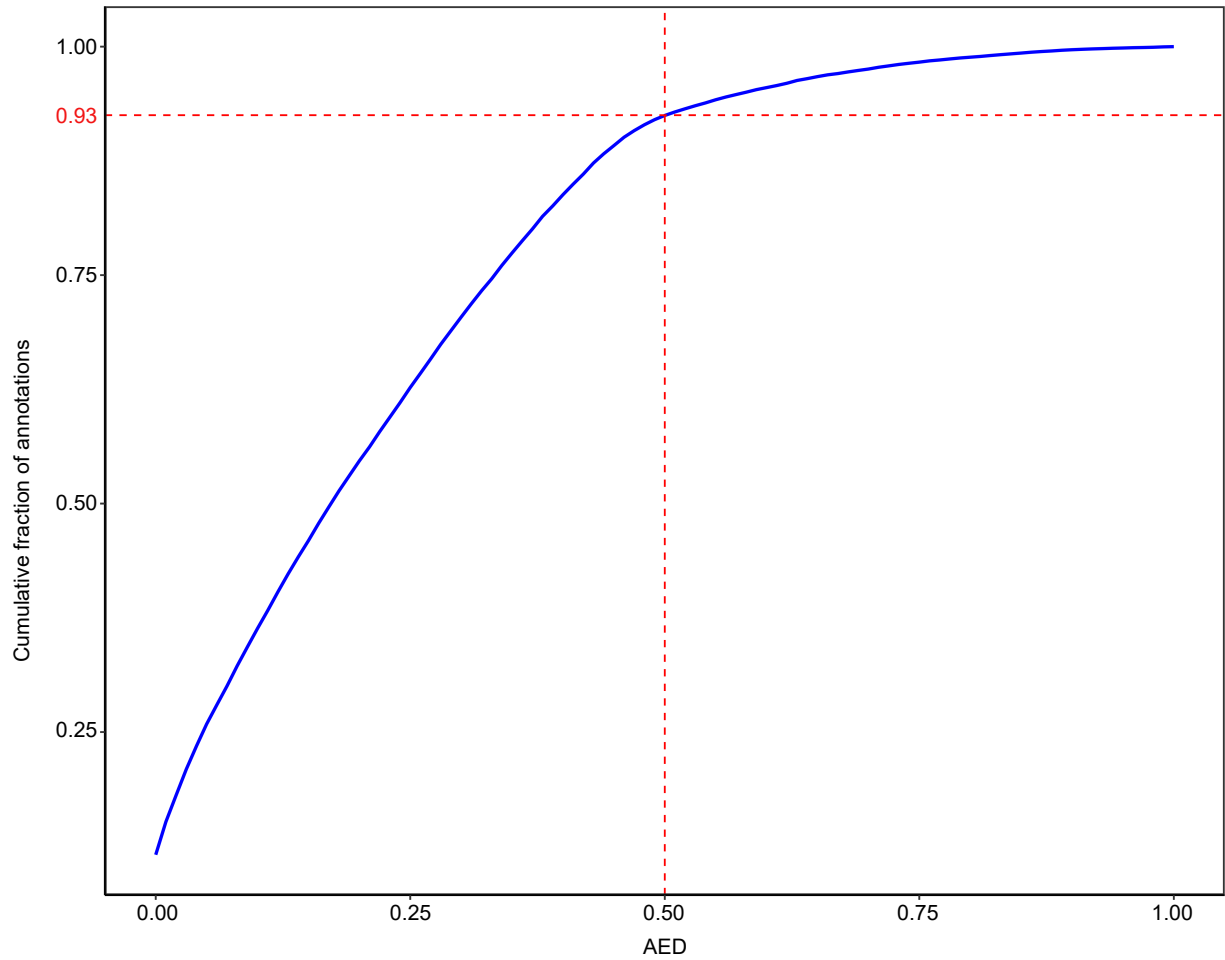
Supplemental Figure 2 Genome size estimation of *L. saligna* by GenomeScope.



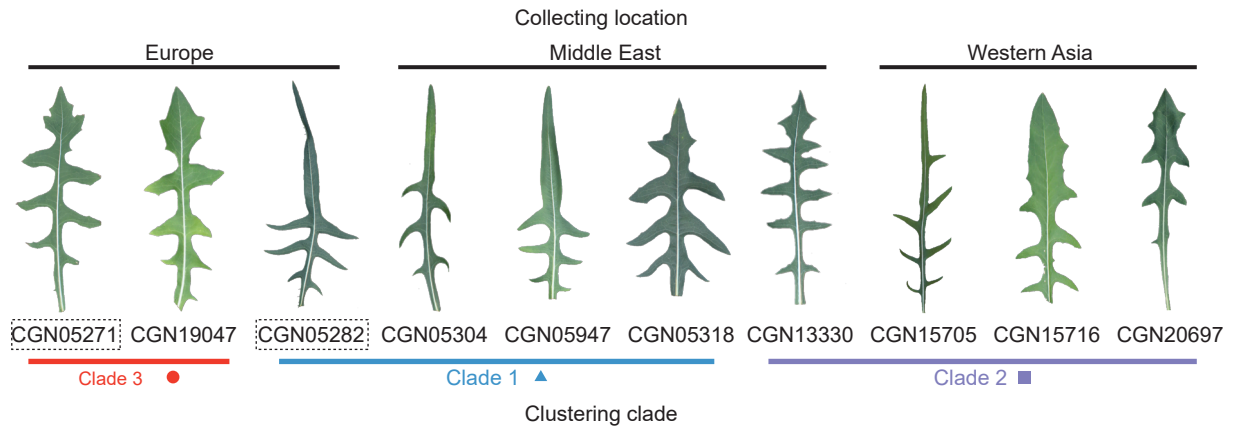
Supplemental Figure 3 ALLMAPS re-scaffolding for *L. saligna* pseudo-chromosomes using genetic and syntenic map.



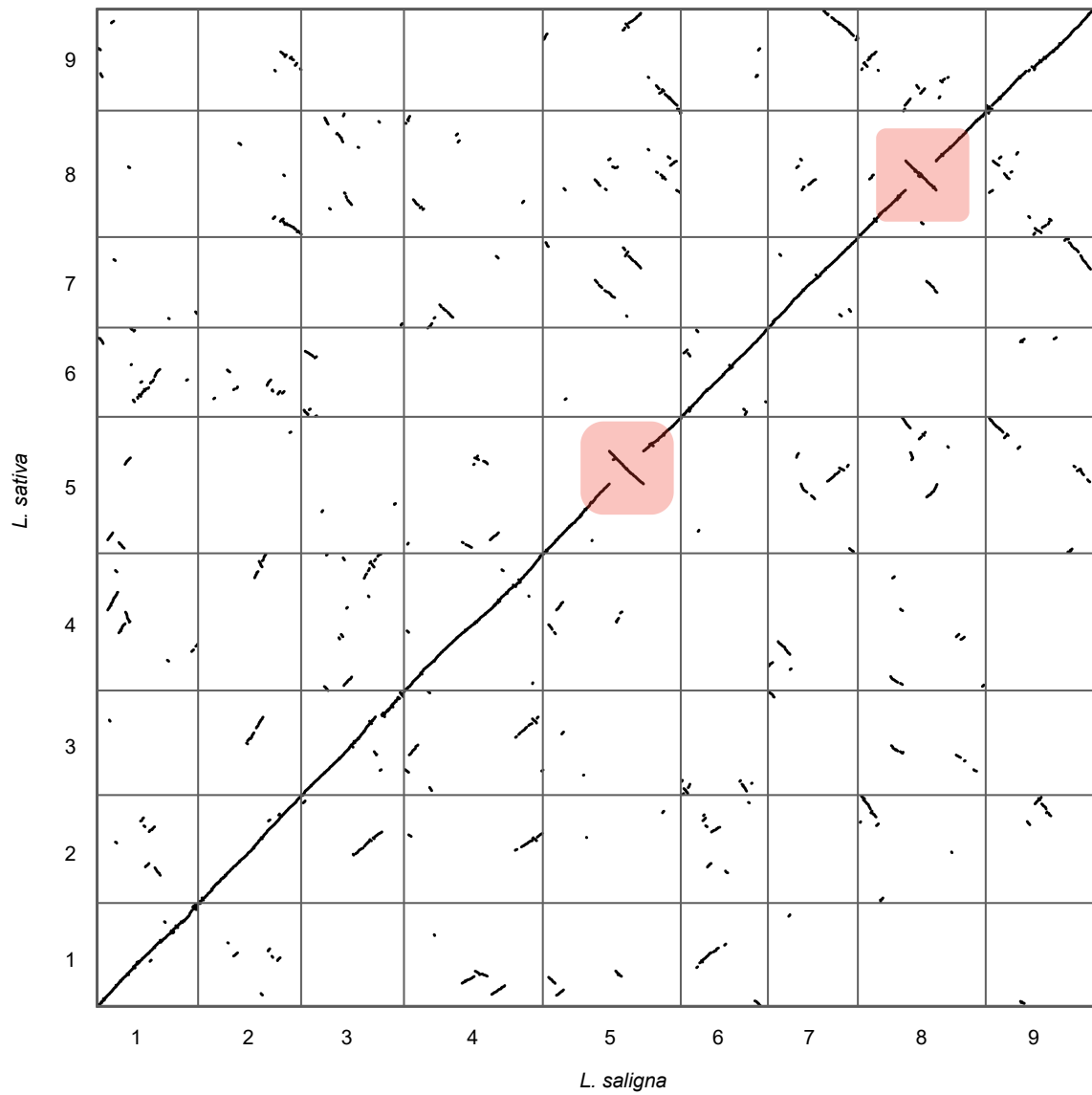
Supplemental Figure 4 Genomic features of the *Lactuca saligna* genome.



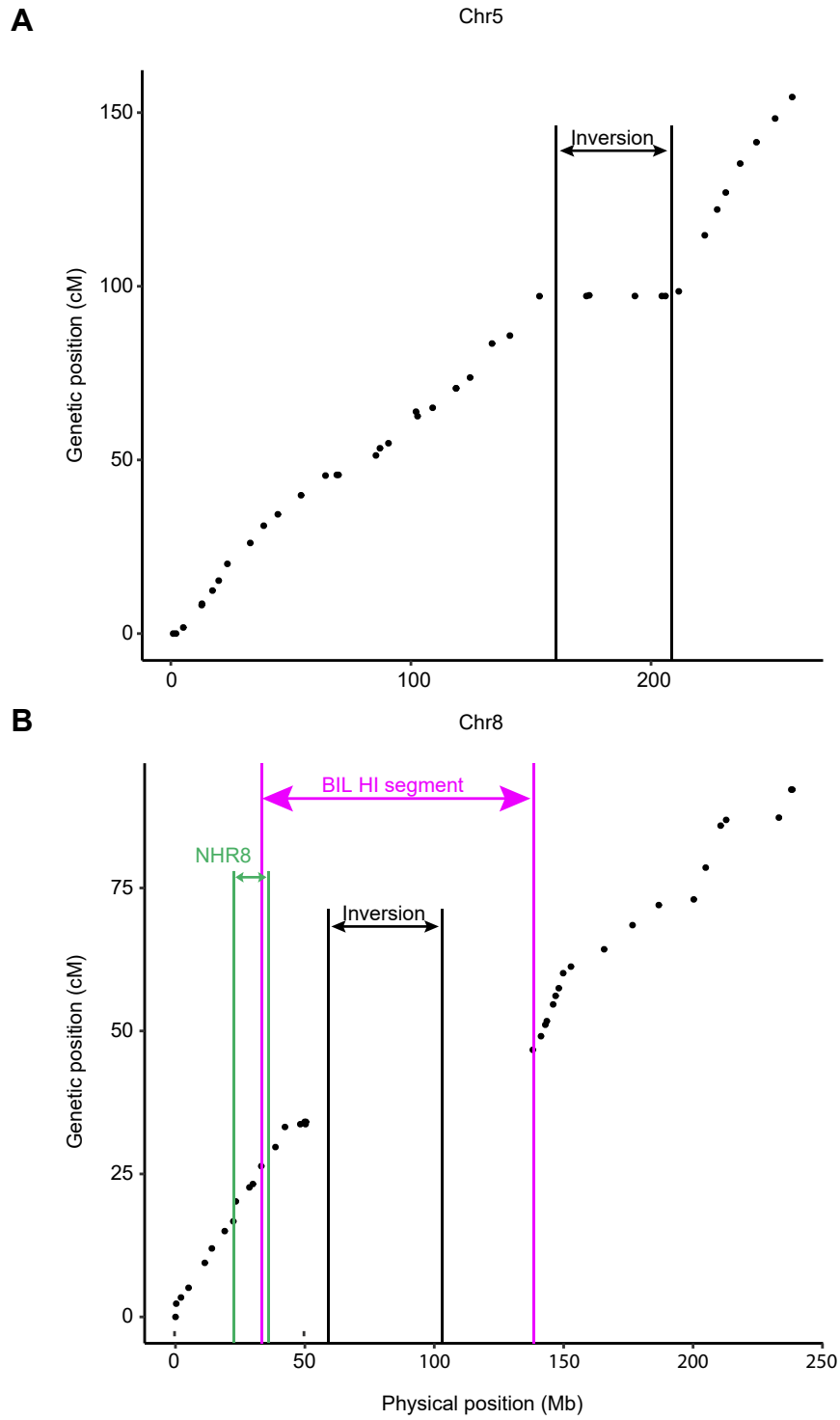
Supplemental Figure 5 Annotation Edit Distance (AED) cumulative fraction of genome *de novo* annotation.



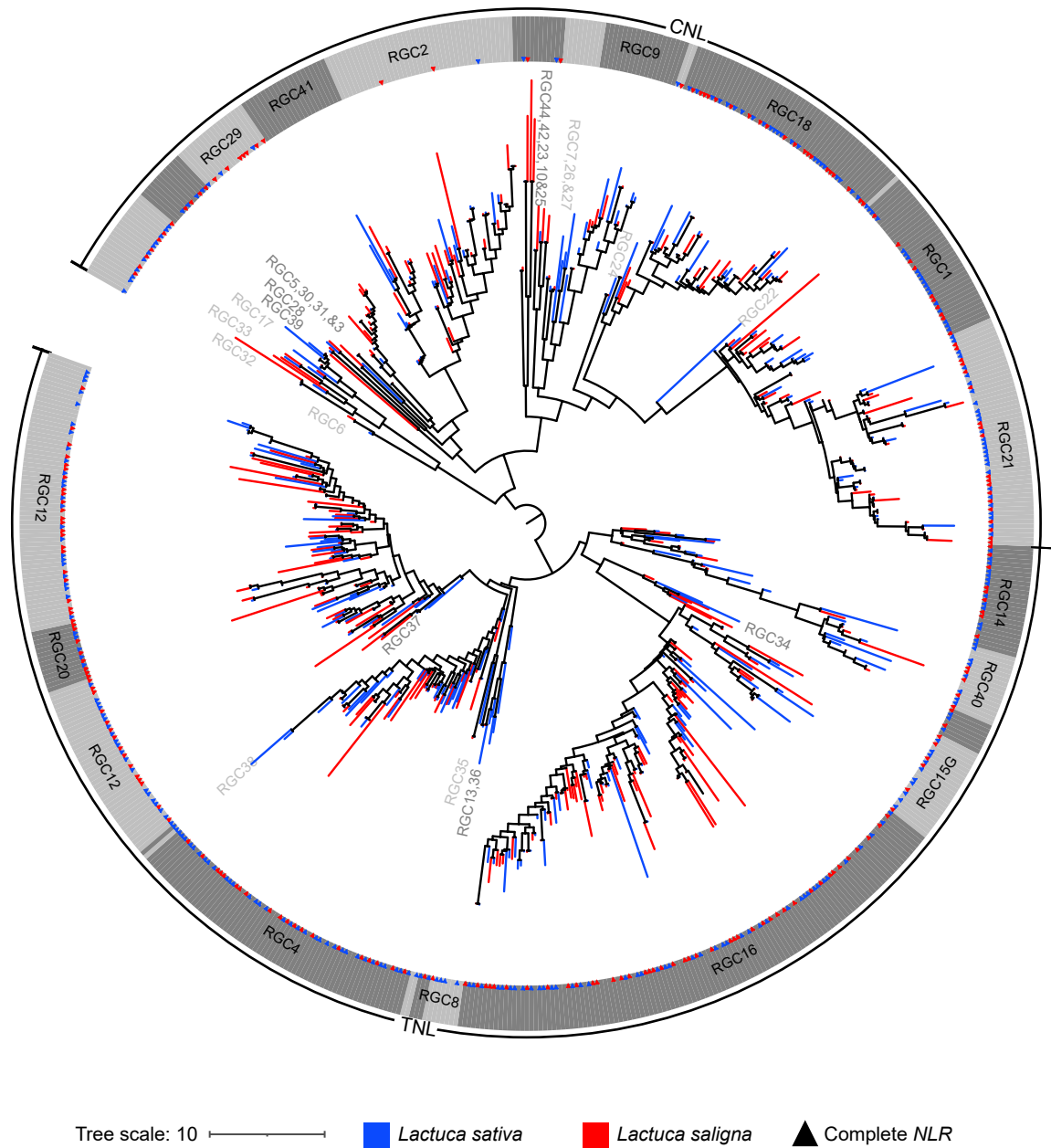
Supplemental Figure 6 Diversity of leaf shape of 10 resequenced *L. saligna* accessions.



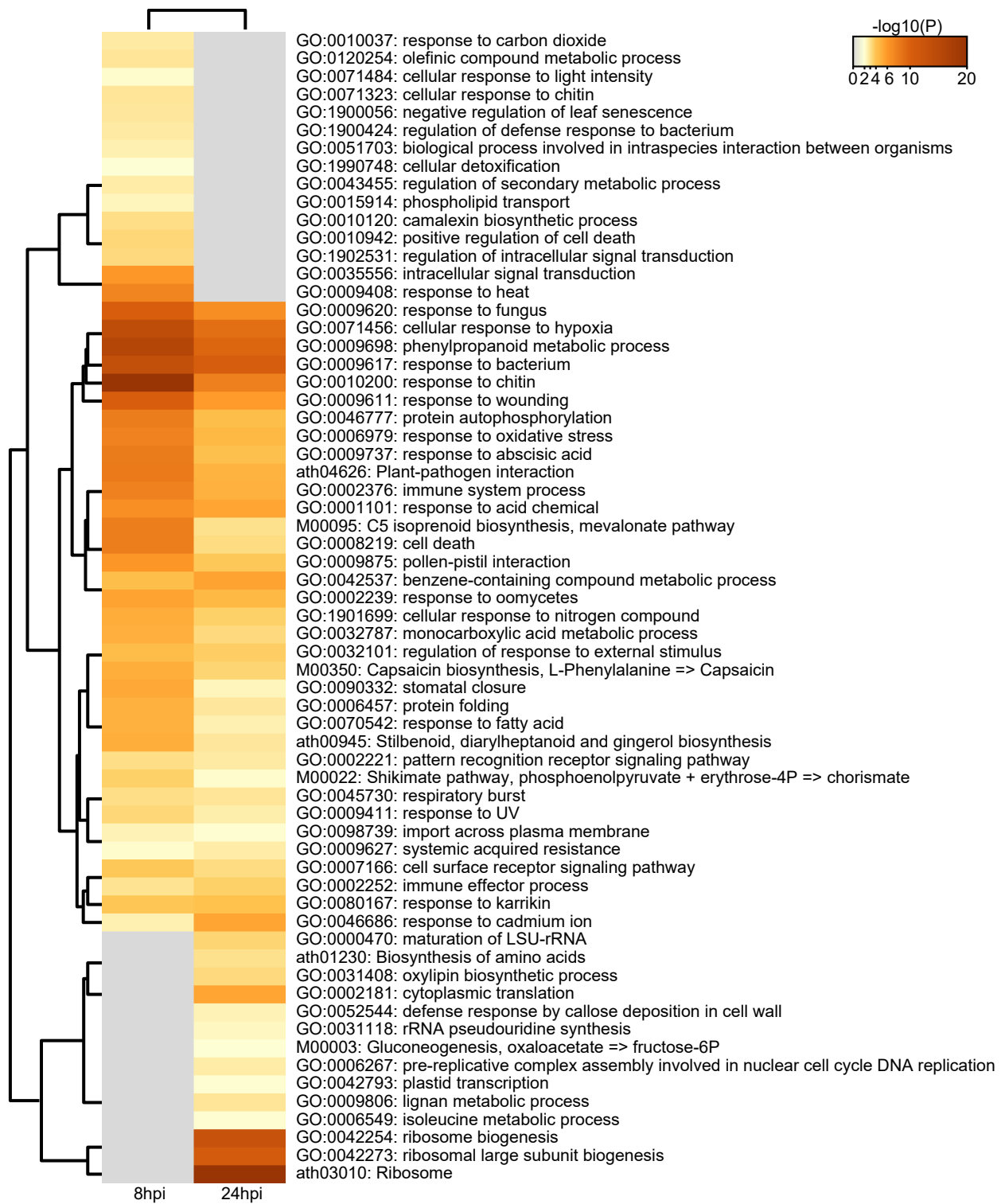
Supplemental Figure 7 Syntenic path dot plot of *L. sativa* versus *L. saligna* highlighting two large inversions on chromosomes 5 and 8.



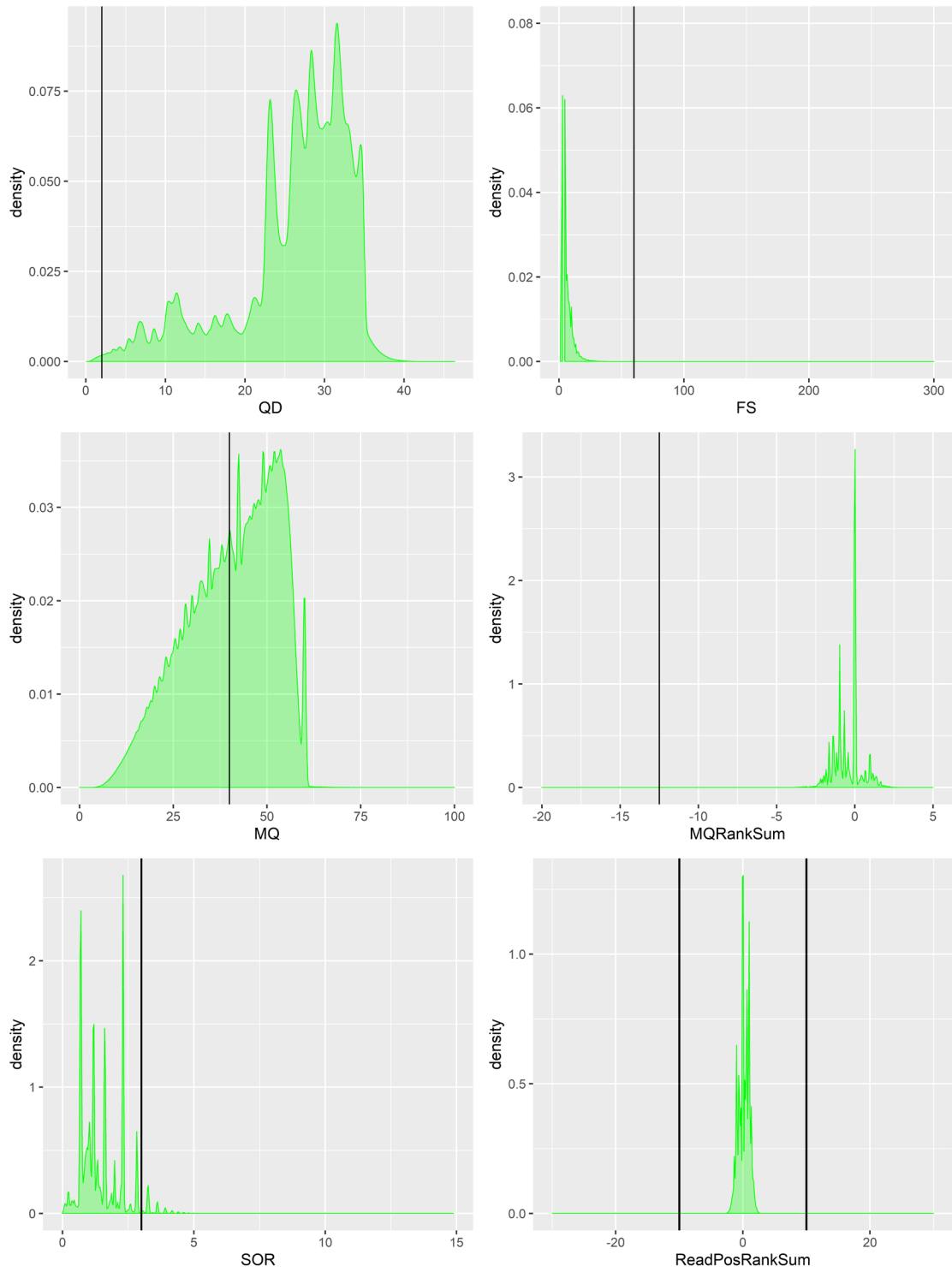
Supplemental Figure 8 Genetic distance versus physical position on chromosome 5 and 8 supporting the presence of large inversions



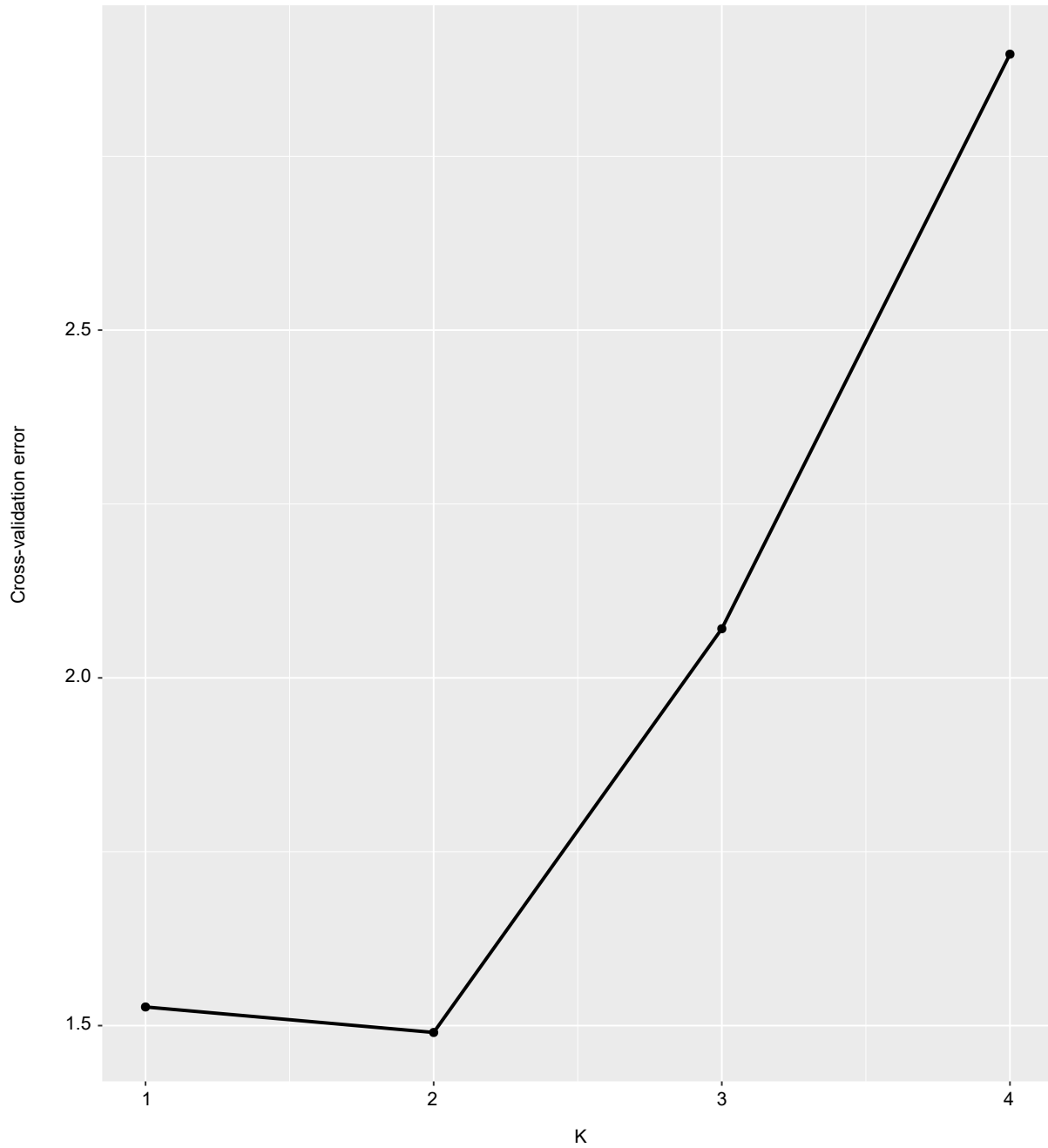
Supplemental Figure 9 Circular tree of *L. saligna* nucleotide binding-leucine rich repeat receptors (NLR) generated by IQTREE.



Supplemental Figure 10 Top 100 enriched biological clusters of up-regulated DE genes at 8 and 24 hpi.



Supplemental Figure 11 Hard-filtering for biallelic SNPs variant from resequencing analysis of 15 *L. saligna* accessions.



Supplemental Figure 12 Cross-validation estimates the best K (population numbers) of ADMIXTURE for 15 re-sequenced *L. saligna* accessions.