MPSeqM, a tool combining multiplex PCR and high-throughput sequencing to study the polymorphism of eight *Leptosphaeria maculans* avirulence genes and its application to field surveys in France

Angélique Gautier, Valérie Laval, Marie-Hélène Balesdent

Université Paris-Saclay, INRAE, UR Bioger, Place de l'Agronomie, CS 80022, 91120 Palaiseau Cedex France

Corresponding Author email valerie.laval@inrae.fr

Keywords. Survey, virulence, *Brassica napus*, polymorphism, detection, *Leptosphaeria maculans*

Abstract

Leptosphaeria maculans is a fungal pathogen causing stem canker of oilseed rape (*Brassica napus*). The disease is mainly controlled by the deployment of varieties with major resistance genes (*Rlm*). *Rlm* genes can rapidly become ineffective following the selection of virulent isolates of the fungus, i.e. with deletions or mutations in the corresponding avirulence genes (*AvrLm*). Reasoned and durable management of *Rlm* genes relies on the detection and monitoring of virulent isolates in field populations. Based on previous knowledge of *AvrLm* gene polymorphism, we developed a tool combining multiplex PCR and Illumina sequencing to characterise allelic variants for eight *AvrLm* genes in field *L. maculans* populations.

We tested the method on DNA pools of 71 characterised *L. maculans* isolates and of leaf spots from 32 *L. maculans* isolates. After multiplex-PCR and sequencing with MiSeq technology, reads were mapped on an in-house *AvrLm* sequence database. Data were filtered using thresholds defined from control samples included in each run. Proportions of each allelic variant per gene, including deletions, perfectly correlated with expected ones. The method was then applied to around 1300 symptoms (42 pools of mainly 32 leaf spots) from nine *B. napus* fields. The proportions of virulent isolates estimated by sequencing leaf spot pools perfectly correlated with those estimated by pathotyping. In addition, the proportions of allelic variants determined at the national scale also correlated with those previously determined following individual sequencing of *AvrLm* genes in a representative collection of isolates. Finally, the method also allowed us to detect still undescribed and rare allelic variants. Despite the diversity of mechanisms generating virulent isolates and the gene-dependant diversity of *AvrLm* gene polymorphism, the method proved suitable for large-scale and regular

33 monitoring of *L. maculans* populations, which will make it possible to choose effective *Rlm*

34 genes and to detect resistance breakdowns at early stages.

35

36

INTRODUCTION

The fight against crop pests is a major issue for food safety in a changing environmental and socio-economic context. Whether chemical or genetic, control measures are more effective as they are adapted to local pest populations. These populations also are constantly evolving in response to the deployment of chemical or genetic protection solutions. For instance, pesticide-resistant individuals regularly appear in fungal or insect populations following the large-scale use of pesticides (McCulloch & Waters, 2023). The selection of such individuals in populations renders the pesticides ineffective. Similarly, plant resistance genes introduced in crop varieties and effective against fungal isolates possessing matching avirulence genes exert strong selection pressure on pathogenic populations. The selection of virulent isolates in populations rapidly decreases the efficacy of the genetic solution. Monitoring pest populations and their characterisation regarding their potential resistance to control measures is thus a main issue, not only in choosing an effective management strategy but also in detecting outbreak events and preventing their spread.

Monitoring virulent isolates is particularly important for the fungal ascomycete *Leptosphaeria maculans*, responsible for the stem canker disease of oilseed rape (*Brassica napus*). The main control measure deployed to limit the impact of the disease is genetic control, including the use of specific resistance (*Rlm*) genes. *L. maculans* populations have the potential to evolve rapidly toward virulence, which has been reported in many instances (Rouxel et al., 2003; Sprague et al., 2006, Zhang et al., 2016). Lastly, potentially effective resistance sources are scarce in oilseed rape and the few currently available *Rlm* genes must be carefully managed to avoid new breakdowns (Borhan et al., 2022; Rouxel and Balesdent, 2017). Until recently, the monitoring of *L. maculans* populations consisted of large-scale isolate samplings from

60 diseased *B. napus* plants, followed by inoculation of each isolate on a plant differential set
61 with defined *Rlm* genes (for instance, Alnajar et al., 2022; Balesdent et al., 2006, 2023;
62 Stachowiak et al., 2006; Zhang et al., 2016). These biological tests are however inadequate for
63 high-throughput analyses and for the detection of rare mutation events in populations. They
64 require adapted facilities and materials (seeds of the plant differential set, control isolates,
65 growth chambers with tight temperature and humidity regulation), and they are time- and
66 labor-intensive and thus, costly. Thus, only a limited number of isolates from a limited number
67 of sites can be analysed each year, and the survey results are obtained six to eight months
68 later. In France, large-scale surveys were done on eight to 12 sites at about 10-year intervals
69 (Balesdent et al., 2023). One full working year is needed to fully characterize populations of
70 around 80 isolates per site. The size of such samplings limits the precision of the estimates of
71 virulent isolates in a population. For instance, in a sampling of 80 isolates per site, the
72 detection of five virulent isolates (6.2%) only allows us to say that the actual proportion of
73 virulent isolates ranges between 2% and 14% (Exact CI (inrae.fr)), corresponding to very
74 distinct breakdown situations.

75 Fortunately, *L. maculans* is a fungus for which we have a detailed knowledge of the genetic
76 determinants involved in the recognition of the fungus by the plant, i.e., its avirulence (*AvrLm*)
77 genes. To date, 11 *AvrLm* genes have been cloned (Borhan et al., 2022), and the molecular
78 mechanisms by which these genes are mutated or inactivated to avoid being recognized by
79 the corresponding *Rlm* genes were analysed in detail for a few avirulence genes, such as
80 *AvrLm3* and *AvrLm4-7* (Balesdent et al., 2022). The study of the *AvrLm* gene polymorphism in
81 a French population revealed contrasting situations, depending on the gene, as regards the
82 level of polymorphism in virulent or avirulent isolates and the type of molecular events
83 responsible for the gain of virulence (Gautier et al., 2023). Some *AvrLm* genes were nearly

84 monomorphic while others were highly polymorphic, and the mutational events in virulent
85 isolates included single nucleotide mutation (SNP), inactivating mutations introducing stop
86 codons, particularly following Reaped-Induced Point mutations (RIP; Fudal et al. 2009;
87 Gladyshev, 2017), deletion of the whole gene including the surrounding region, or the masking
88 of the avirulence function by the presence of another avirulence gene. For instance, the
89 presence of at least some alleles of *AvrLm4-7* is sufficient to mask the *AvrLm3-Rlm3* or the
90 *AvrLm9-Rlm9* interactions (Plissonneau et al., 2017, Ghanbarnia et al., 2018). With this
91 knowledge, one can now envisage developing molecular tools for the survey of virulent
92 isolates at one or more avirulence loci.

93 Molecular methods were already developed for this purpose, but they are often based on
94 unsatisfactory compromises. For instance, some tools only rely on PCR amplification of the
95 avirulence gene, and the percentage of virulent isolates is deduced from the percentage of
96 isolates that fail to amplify the gene. This approach works well for *AvrLm* genes having gene
97 deletion as the main event in virulent isolates, such as *AvrLm1* (Gout et al., 2007). Even though
98 a few virulent isolates display inactivated versions of *AvrLm1*, still amplified in these isolates,
99 these events are rare, with only one such isolate among 89 in the French population (Gautier
100 et al., 2023). The lack of PCR amplification could thus be a good proxy for the frequency of
101 virulent *avrLm1* isolates. A quantitative PCR assay was developed to quantify the proportion
102 of deleted isolates for *AvrLm1*, relative to the ITS (Van de Wouw et al., 2010). In most cases
103 however, the deletion is not the only mutational event in virulent isolates, and deletions may
104 be rare or even never observed for some *AvrLm* genes (Gautier et al., 2023, Van de Wouw et
105 al., 2023).

106 Thus, *AvrLm* sequencing in individual isolates is the easiest way to access the genotype and to
107 infer its phenotype, provided the correspondence between genotype and phenotype is
108 available. In the situation of epistatic interactions between avirulence genes, such as the
109 masking effect of *AvrLm4-7* on the *AvrLm3*/*Rlm3* and *AvrLm9*/*Rlm9* interactions, the
110 sequencing of the two genes involved in the epistatic interaction should also give access to
111 the phenotype thanks to the detailed work already published on these complex interactions.

112 Different approaches have been proposed to avoid PCR amplification and sequencing of
113 *AvrLm* genes. Van de Wouw and Howlett (2014) developed a pyrosequencing method to
114 quantify the proportions of virulent isolates toward *Rlm4*. This quantitative method works on
115 pools of individuals and tags one polymorphic site in *AvrLm4-7* known to be responsible for
116 avirulence towards *Rlm4* (Parlange et al., 2009). However, this method gives sequence
117 information on a small part of the gene only (90 bp) and thus cannot be used to infer virulence
118 on *Rlm7*, due to the diversity of mutations responsible for this phenotype (Daverdin et al,
119 2012; Balesdent et al., 2022). An HRM (High-Resolution Melting analysis) method was also
120 developed to detect sequence polymorphism in *AvrLm4-7* and determine virulence toward
121 *Rlm4* and *Rlm7*. The method worked well in the situation of limited polymorphism of the gene
122 (Carpezat et al., 2014). However, with the recent large-scale use of *Rlm7,* at least in France,
123 and the diversification of sequence polymorphism in virulent isolates (Daverdin et al., 2012;
124 Balesdent et al., 2022), it became more and more complex to convert new HRM profiles into
125 virulent or avirulent alleles.

126 Finally, whole genome sequencing (WGS) of individual isolates gives direct access to the
127 sequences of the full set of *AvrLm* genes in an isolate. For instance, Chen et al. (2020) studied
128 the polymorphism in nine avirulence genes following the WGS of 162 Canadian isolates.

129 Similarly, sequences of avirulence genes in an international collection of 205 *L. maculans*
130 isolates were recently published (Van de Wouw et al., 2023). Although powerful and
131 exhaustive, such an approach cannot be performed on hundreds of isolates every year. In
132 addition, this method necessitates first to isolate and purify each isolate.

133 Here, we developed a tool based on MiSeq sequencing technology, with two levels of
134 multiplexing. First, eight *AvrLm* genes and one control gene (*Actin*) are amplified together in
135 a multiplex PCR. Second, the multiplex-PCR is done on pools of phoma leaf spots, the typical
136 *L. maculans* symptoms on *B. napus* leaves, to increase the size of the population assessed and
137 to avoid the *L. maculans* isolation step. PCR products of pooled samples are sequenced using
138 MiSeq technology and the proportions of each allele of each avirulence genes in the pooled
139 sample are determined. The tool is named here "MPSeqM tool » (Multiplex on Pools and
140 Sequencing with MiSeq). After the development and validation of the tool on DNA pools of
141 known isolates, we applied the method to samples composed of pools of leaf discs centred on
142 phoma leaf spots from experimental fields. The MPSeqM tool was applied to pools of 32 to 96
143 different leaf spots (possibly corresponding to as many different individuals, Gout et al., 2006)
144 and the results were compared to phenotyping data of isolate populations from the same
145 fields and the same sampling date (Balesdent et al., 2023). This approach not only allowed us
146 to define a robust protocol for large-scale monitoring of virulence in the field, but also
147 revealed a diversity of situations between sites, between *AvrLm* genes, and identified many
148 unknown and rare alleles.

149

150 **MATERIALS AND METHODS**

151 *Isolates*

152 Seventy-one *L. maculans* isolates were chosen, including the reference isolate JN3 (v23.1.3),
153 based on their polymorphism in their avirulence genes (Supplementary Tables S1 and S2;
154 Gautier et al., 2023). The isolate X83.51 was generated following an *in vitro*-cross aiming at
155 accumulating Avr gene deletion in a single isolate. *AvrLm1*, *AvrLm4-7*, *AvrLm11* and *AvrLmS-*
156 *Lep2* are deleted in this isolate. X83.51 was used as a negative control for these genes and in
157 serial DNA dilutions with variable proportions of deleted alleles.

158 *Field samples*

159 The oilseed rape variety Westar, devoid of resistance genes, was sown at seven sites in
160 autumn 2019 and at one site in 2020 (Table 1), with two to five plots per site. Individual leaves
161 with typical *L. maculans* leaf spots were collected in autumn (300 per site, all collected from
162 distinct plants). Part of these leaves was previously used to isolate around 80 single-pycnidium
163 isolates per site for pathotyping on a differential plant set (Balesdent et al., 2023). In parallel,
164 leaf disks (one disk per leaf) centred on a single leaf spot were cut using a cork-borer (1 cm
165 diameter). Leaf disks were stored in 24-weel microplates at -20°C until further use
166 (Supplementary Figure S1). Before DNA extraction, leaf spots from a given site were pooled
167 by 24 or 32 in one Falcon tube and freeze-dried for 24 hours. Three to six pools per site were
168 prepared (Table 1).

169 *Plant inoculation for preparation of DNA control*

170 A selection of 32 *L. maculans* isolates with known avirulence alleles was inoculated on
171 cotyledons of *B. napus* cv. Westar using standard protocols (Gautier et al., 2023;
172 Supplementary Tables S1 and S2). Fourteen days after inoculation, leaf spots that developed
173 on cotyledons were cut as described above. One spot per isolate was placed in a 50 mL Falcon

174 tube, to constitute a pool of 32 leaf spots of known *AvrLm* allele composition. The pool was
175 freeze-dried during 24 hours before DNA extraction.

176 *DNA extraction and manipulation*

177 For each of the 71 *L. maculans* isolates, fungal DNA was extracted independently from 1 mL
178 of crude conidial suspensions (between 10^8 and 10^9 spores mL^{-1}) using the DNeasy 96 Plant
179 Kit (Qiagen) following the manufacturer's instructions (Gautier et al., 2023). Freeze-dried
180 pools of leaf spots, whether from inoculations or from the fields, were crushed with a tungsten
181 carbide bead (3 mm, Qiagen) on a Retsch MM400 mixer mill for 60 seconds at 30 Hz and DNA
182 was extracted using the DNeasy Plant Mini kit (Qiagen). DNA was eluted in 100 µL of elution
183 buffer AE (Qiagen). DNA quality was assessed with NanoDrop™ One (Thermo Scientific™) and
184 DNA concentration was measured with Qubit™ 2.0 fluorimeter (Invitrogen™). After extraction,
185 DNA of pools of 32 field leaf spots from the same field were mixed equimolarly to constitute
186 "pool of pools" composed of mainly 96 leaf spots (Table 1).

187 *Control DNA samples*

188 Four different controls were included, in triplicate, in each sequencing run: (i), multiplex PCR
189 of JN3 DNA (termed JN3_DNA) used as a positive control; (ii), multiplex PCR of the pool of 32
190 leaf spots from inoculations with known isolates (termed Mix_LSpot); (iii), multiplex PCR of a
191 pool of 71 *L. maculans* DNA (termed Mix_LM); and (iv), multiplex PCR of the 11 points of a
192 dilution series for deletion quantification (termed C_dil_ser). The Mix_LM was prepared by
193 mixing 50ng of DNA of each isolate and used to test the ability to detect all allelic variants in a
194 pool (Supplementary Table S2). The dilution series was prepared by mixing DNA of isolates
195 JN3 and X83.51 in proportions varying from 0 % to 100% (with a 10% increment).

196 *Multiplex PCR and MiSeq sequencing*

197 Two technical criteria governed primer design, due to MiSeq technology and multiplexing; (i)
198 the amplified region must be smaller than 460 bp for correct paired-end subsequent analysis;
199 (ii) all amplified sequences in the multiplex PCR must be of similar length to avoid over-
200 representation of the shortest sequences (maximum of 100 bp difference). Primers fitting
201 these criteria were designed using the Primer3 website https://primer3.ut.ee/ (Table 2). The
202 PCR reaction contained 40 ng of DNA, 1X Master Mix of Type-it Microsatellite PCR Kit (Qiagen),
203 4 µL Q-Solution, and 0.25 µM of each primer (except *AvrLm2* primers at 0.2 µM and *AvrLmS-*
204 *Lep2* primers at 0.4 µM) in a total volume of 40 µL. PCR reactions were performed using an
205 Applied Biosystems 9700 thermocycler. The amplification parameters consisted of an initial
206 denaturation step at 95˚C for 5 min; then 35 cycles of denaturation at 95˚C for 60 s, annealing
207 at 60˚C for 90 s, and elongation at 72˚C for 60 s; and a final extension step at 72˚C for 10 min.
208 Three independent PCRs were performed for each sample to test the repeatability and
209 robustness of the method. The PCRs were verified on 3% agarose gels and sent to the
210 GenoToul platform (INRAE Toulouse) for Illumina MiSeq sequencing using the chemistry V3
211 (2x250pb).

212 *AvrLm allele database*

213 At the start of the project, the *L. maculans AvrLm* database included the 87 alleles and 63
214 isoforms identified following the sequencing of 11 avirulence genes in a collection of French
215 and reference isolates (Gautier et al., 2023) and available at
216 https://bioinfo.bioger.inrae.fr/portal/data-browser/public/leptosphaeria/avirulence_genes.
217 Additional *AvrLm3* and *AvrLm4-7* alleles described in Balesdent et al. (2022) were added to
218 the database.

219 *Analysis of sequence dataset: pairing, filtering, processing and assigning an allelic*
220 *classification to sequences*

221 First, for each sample, forward (R1) and reverse (R2) reads were paired using the tool Pear -
222 Paired-End read merger (Galaxy Version 0.9.6.2) of Galaxy pipeline (https://vm-galaxy-
223 prod.toulouse.inrae.fr/galaxy_main). Then, the data were processed gene by gene. For each
224 gene, only sequences with the expected primer sequences were retained. A stringent pipeline
225 (Boutigny et al., 2019) was used to filter and process these paired sequences, utilising
226 MOTHUR v.1.36.1 (Schloss et al., 2009) and parameters adapted to each gene (Supplementary
227 Text T1). Sequences with low quality (shorter or including low-quality scores) were removed
228 (Supplementary Text T1). After preprocessing the sequences, sequence files were merged,
229 and unique sequences represented by singletons or 2 reads were removed. Chimeric
230 sequences were removed with the Uchime tool (Edgar et al., 2011) available in MOTHUR, using
231 self as the reference. Uncorrected pairwise distances were determined between the
232 remaining sequences with the Needleman alignment method (Needleman et al., 1970) with
233 match 1.0, mismatch -1.0, gapopen -2.0 and gapextend 1.0 using 1000 iterations. The distance
234 matrix was then clustered with the average neighbor algorithm to assign sequences to
235 operational taxonomic units (OTUs). Using MOTHUR's classifier (Wang et al., 2007), the
236 representative sequence for each OTU was assigned to the allele of the corresponding *AvrLm*
237 gene from the custom *Lmb_avrlm_act genes* reference database using no divergence. An OTU
238 count table was obtained containing OTU number, read count per OTU per sample,
239 representative sequence name, representative sequence, and OTU representative assignation
240 (Figure 1). In a final step, the Galaxy blastn tool was used, to compare the OTU sequences
241 retained against *Lmb_avrlm_act genes* references database and distinguish the known alleles

242 with 100% identity in length and sequence to new alleles with less than 100% identity in
243 sequence.

244 *Strategy for the selection of relevant OTUs and database cleaning*

245 After the mapping step, a very high number of OTUs were identified (Table 3), most of which
246 correspond to artefacts due to PCR or the MiSeq technology. Three filters were sequentially
247 applied (Figure 1) to keep only reliable OTUs as follows: (i), keep only OTUs present in at least
248 three samples, due to the use of triplicates for each DNA sample; (ii), remove all OTUs for
249 which, whatever the sample analysed, the read numbers were below a threshold. This
250 threshold is named "Maximum read number threshold" and is gene-dependent (Table 3) (iii),
251 remove OTUs that were detected in controls but should not be present. After this OTU
252 selection step, we calculated the mean read number of the three replicates for each OTU and
253 DNA sample. Finally, the resulting count table was simplified by considering an OTU absent in
254 one sample if its proportion in this sample (i.e., read count of the OTU divided by total read
255 counts for all OTUs for this gene) was lower than a threshold ("proportion" threshold), specific
256 for each gene. This percentage threshold is determined based on the expected and observed
257 OTU proportions in controls (Table 3, Figure 1).

258 *Analysis of new AvrLm alleles*

259 When a new OTU was detected, the missing parts of the gene sequence, if any, (i.e., when
260 primers were within the genes) were completed by the corresponding 5' or 3' sequences of
261 the JN3 reference isolate to generate a new putative allele. Introns were removed and the
262 sequence was translated with the Expasy translate tool (https://web.expasy.org/translate/)
263 to generate an extrapolated, putative new isoform (Figure 1).

264 *Validation of new alleles*

265 DNA of one leaf spot pool from the Coullons site (Table 1) displaying a high proportion of new
266 OTUs was used to clone and validate putative new alleles. After *AvrLm4-7* PCR amplification
267 with external primers (Gautier et al., 2023) the PCR products were cloned into pGEM®-T Easy
268 Vector System (PROMEGA) using standard protocols. 101 colonies were recovered. The
269 *AvrLm4-7* insert was re-amplified in colonies with the plasmid standard primers SP6/T7 and
270 sequenced by Eurofins Genomics (EUROFINS, Ebersberg, Germany) using SP6 external primer.

271

272 **RESULTS**

273 **Primer design and reproducibility of multiplex PCR and sequencing**

274 We designed a set of primers fitting technical constraints and amplifying parts of *AvrLm* genes
275 containing a maximum of known polymorphic sites and all codons encoding a cysteine (Table
276 2). These primers amplified part of the genes containing between 66% and 100 % of the known
277 mutations (Table 2, Supplementary Figure S2). A few known variants were not detected, with
278 no or low consequences for the diagnosis of the allele as an avirulent or a virulent one. For
279 instance, four *AvrLm3* variants cannot be detected as mutations are outside the amplified
280 region (Table 2). However, these mutations are either in the last intron or in the peptide signal-
281 coding region, without impact on the mature protein (Supplementary Fig S2-C).

282 Using these primers, we first tested the MPSeqM tool with the concomitant amplification of
283 four, five, or nine genes in a single PCR. Sequencing results were similar whatever the number
284 of genes amplified together (data not shown) and we thus choose the nonaplex amplification
285 in all further experiments. MPSeqM of control DNAs (JN3_DNA, Mix_LSpot, and Mix_LM)

286 revealed an efficient and reproducible amplification level of all genes (Figure 2 and data not
287 shown). The less amplified gene was *AvrLm6*, followed by *AvrLmS*, still representing more than
288 1250 reads each in each control replicate.

289 **Detection and quantification of deletions**

290 To estimate the proportion of deleted alleles of a given *AvrLm* gene in a field sample, we used
291 serial dilutions of the DNA of an isolate deleted for this gene with the DNA of an isolate with
292 the gene present. The theoretical proportions of the *AvrLm* gene in the dilution series were
293 compared with the estimated proportions calculated as the total number of reads mapping to
294 this gene over the total number of reads mapping to the *L. maculans Actin* gene, which is a
295 single copy and invariant gene. The two values were proportional and allowed us to define
296 standard curves of variable shapes depending on the gene (Figure 3). We approximated the
297 curves to a polynomial equation, whose parameters were used to calculate the proportion of
298 undeleted events and, by difference, the deleted ones, in a pooled sample. For instance, the
299 expected proportion of deleted *AvrLm1* in the Mix_LSpot is 87.5% ([exact confidence interval
300 [71-96]) and the estimated proportion of the deletion calculated with the standard curve
301 equation is 84.5% (Supplementary Table S3).

302 **Threshold determination for OTUs assignation**

303 The use of triplicates for each sample and four types of controls of known allele composition
304 was essential to determine the thresholds used to keep only reliable OTUs and suppress
305 potential PCR or sequencing artefacts. Mainly, the “Maximum number of reads” threshold
306 allowed us to reduce drastically the number of OTUs (1204 OTUs in total, starting from 81024
307 OTUs kept with the criteria of presence in at least three sequencing samples, Table 3). The
308 controls with the reference JN3 isolate or with DNA mixes of known isolates also allowed us
309 to remove all OTUs still unexpectedly found in these controls, which represented 3/4 of the

310 remaining OTUs. Altogether, 300 reliable OTUs (all genes) were kept, among which 206 were
311 not in the database, 74 of which corresponded to highly mutated versions of the genes by RIP
312 (Table 3). Lastly, depending on the gene, OTUs representing less than 0.1 to 0.5% of the total
313 read counts for this gene in a field sample ("proportion" threshold) were considered absent
314 from this sample (Table 3).

315 **Validation of the method on control samples**

316 The previous steps allowed us to select reliable OTUs and to determine the proportion of
317 deleted alleles for *AvrLm1*, *AvrLm11*, and *AvrLm4-7*. Using this cleaned dataset, we compared
318 the proportion of alleles (OTUs) detected in control samples to the expected ones (Figure 4,
319 Supplementary Table S3). In the Mix_Lm control, including 71 isolates, 20 alleles out of 90
320 were not detected, but 19 of them corresponded to a rare occurrence (only one isolate among
321 71). In addition, most of them (13) were RIPped alleles. In the Mix_LSpot control, only 3 out
322 of 47 expected alleles were not detected, all corresponding to RIPped alleles for *AvrLm1* or
323 *AvrLm4-7*. All other alleles, including those present in only one isolate out of 32, were detected
324 (Supplementary Table S3). Overall, observed and expected proportions correlated very well
325 (R^2>0.98). Only one *AvrLm4-7* allele (the deleted allele) was overestimated in the Mix_LM
326 control, possibly due to the lack of amplification of some RIPped alleles. Apart from this allele,
327 all estimated values were within the exact confidence interval determined for sample sizes of
328 32 (Mix_LSpot) or 71 (Mix_LM) individuals (Figure 4).

329 **Analysis of unknown field samples: identification of new alleles**

330 The MPSeqM tool, validated on controls, was then used on 42 field samples of pooled leaf
331 spots corresponding to 1337 *L. maculans* symptoms from eight French regions (Table 1). This
332 analysis first revealed several still undescribed alleles for most of the *AvrLm* genes studied.

333 Such new alleles were considered as real alleles, not sequencing artefacts, when they
334 combined the following characteristics: (i) they were detected in some samples only, not as a
335 background signal in all samples including controls; (ii) they were consistently detected in the
336 three independent technical replicates of one given field sample; (iii) their read numbers were
337 higher than the gene-dependent threshold value determined for each gene; and (iv) they were
338 detected both in one "pool of 32 leaf spots" and in the corresponding "pool of pools", when
339 available, with consistent proportions (Figure 5). Using these criteria, a total of 206 new alleles
340 were detected.

341 **Analysis of unknown field samples: gene-by-gene analysis**

342 *AvrLm1*

343 Only one out of the five previously described *AvrLm1* alleles was detected in field samples,
344 allele *AvrLm1_0_A.* In addition, we detected two new alleles, one RIPped allele with 14
345 substitutions and one allele with a stop codon in the middle of the sequence. These two new
346 alleles were considered as virulent, as were deleted isolates whose proportions were
347 estimated using the deletion curve. Altogether, the MPSeqM analysis of field samples
348 correlated well with pathotyping; 1.8% of the population was found avirulent on *Rlm1* by
349 pathotyping (Balesdent et al., 2023), while the MPSeqM analysis of populations from the same
350 sites detected 2.1% of such isolates. With both approaches, the site with the highest
351 proportion of avirulent isolates was that of Le Rheu (Supplementary Figure S3). With the
352 MPSeqM tool, it was possible to reproducibly detect low percentages of avirulent isolates in
353 all sites, while, by phenotyping 30 isolates per site, such isolates were not detected (Balesdent
354 et al., 2023).

355 *AvrLm2*

356 No new *AvrLm2* allele was detected in field samples. The *AvrLm2_0_A* was largely prevalent,
357 with proportions ranging from 91% to 100% (average across all pools = 97.3%, Supplementary
358 Figure S4) depending on the pool. The *AvrLm2_1_A* allele was found in all sites at a low
359 frequency (0 to 8% per pool, average across all pools = 2.2%). *AvrLm2_2_A* was found in six
360 sites (10 pools out of 42; between 0 and 6.9%; average across all pools = 0.52%). The other
361 known alleles, including all avirulent ones, were never detected in the pools.

362 *AvrLm3*

363 From the 24 previously described *AvrLm3* alleles, 13 were detected in the field samples.
364 Twenty-eight new OTUs were also detected, being either rare (present in one sample only) or
365 more frequent (present in up to 23 samples), allowing us to identify in total 41 different OTUs
366 for *AvrLm3*. The new alleles could represent up to 8.5% of the OTUs in a given sample. These
367 new alleles encoded for ten putative new isoforms of the AvrLm3 protein. Based on their
368 alignment with previously described AvrLm3 isoforms, we inferred that six among them
369 potentially correspond to avirulent isoforms and four to virulent ones (Supplementary Figure
370 S5). Despite the allelic diversity detected in field samples, four alleles represented 96% of the
371 population, with allele *AvrLm3_A1* being by far the most abundant (80.2%), followed by
372 *AvrLm3_C* (11.8%) (Supplementary Figure S6), consistent with previous surveys and
373 sequencing (Gautier et al., 2023).

374 *AvrLm4-7*

375 MPSeqM results for field samples displayed 110 new *AvrLm4-7* alleles (Table 3). Most of them
376 (66) corresponded to RIPped alleles with mutations introducing stop codons in the sequence
377 and thus corresponding to virulent alleles, also unmasking *AvrLm3*. Seven sequences displayed
378 one or two bp deletions, also encoding for truncated (virulent) proteins. A few (17) were new

379 alleles of known (virulent or avirulent) isoforms, however, 25 new isoforms were detected.
380 The amino acid (AA) sequence of these new isoforms was compared to previously described
381 ones (Balesdent et al., 2023) and the corresponding phenotypes were more or less easily
382 deduced depending on the AA substitutions. One isoform putatively corresponded to an A4A7
383 (conferring avirulence on both *Rlm4* and *Rlm7*) isoform, six alleles for a3a7 isoforms
384 (conferring virulence on *Rlm4* and *Rlm7* and on *Rlm3* by masking effect), four for a3A7
385 (avirulent on *Rlm7* only) isoforms. For the remaining 14 isoforms, it remained impossible to
386 conclude.

387 *AvrLm5-9*

388 Due to primer design, alleles *AvrLm5-9_0_A* and *AvrLm5-9_0_C* (encoding for the same
389 isoform) cannot be distinguished. In field samples, seven new variants were detected. Only
390 three of them corresponded to putative new isoforms of the protein; one displayed one AA
391 change in the signal peptide compared to the AvrLm5-9_0 isoform without any modification
392 of its secretory properties (data not shown). The two others displayed only one AA change
393 each (R63L and K101Q, respectively), whose impact on the phenotype cannot be determined.
394 Among the two alleles encoding avirulent isoforms on *Rlm9*, isoform *AvrLm5-9_4* was never
395 detected in any of the 42 pools, and *AvrLm5-9_1* was found in only nine pools with very low
396 proportions (<0.27%, close to the 0.2% detection limit, 0.06% on average over all pools),
397 confirming the near-fixation of the virulence phenotype toward *Rlm9* in French *L. maculans*
398 populations (Balesdent et al., 2023 Gautier et al., 2023). Similarly, alleles with stop codons in
399 the coding sequence were never detected. Depending on the site, from two to six alleles were
400 detected (Supplementary Figure S7).

401 *AvrLm6*

402 Field samples displayed five new OTUs. They were however rare, present in one or two sites
403 only with proportions ranging from 0.6% to 4.6% (Supplementary Figure S8). Each new OTU
404 displayed only one mutation. One was in an intron; three were silent and only one of them
405 introduced a G47C substitution, and thus a new isoform. The impact of this mutation on the
406 phenotype cannot be determined yet. Allele *AvrLm6_0_A* was largely prevalent, ranging from
407 91.7% to 100% depending on the sample, followed by allele *AvrLm6_1_A* or *AvrLm6_3_A*, two
408 alleles that cannot be discriminated with the current primers (between 0% and 4.1%).

409 *AvrLm11*

410 For *AvrLm11*, two new OTUS displayed a shorter-than-expected size (one or two bp). These
411 gaps were either in the intron or after the stop, thus without impact on the isoform of the
412 protein. In addition, 27 new OTUs were detected in at least one field sample, displaying from
413 one to seven mutations along the gene. These new OTUs were often specific to one sampling
414 site (Figure 5) and, depending on the site, from one to 14 new OTUs were present. Most of
415 the mutations (72%) in these OTUs were typical of RIP mutations (C->T and G-> A). In these
416 OTUs however, the number of mutations was limited (from 1 to 7 mutations per *AvrLm11*
417 OTU) compared to what is seen in RIPped alleles for other genes like *AvrLm4-7* (from 1 to 44
418 mutations, mean=20.2, data not shown). Despite this low number of mutations per allele,
419 these RIP mutations introduced stop codons in seven OTUs at positions 29 or 74 of the protein
420 leading to truncated versions of the protein that thus corresponded to virulent alleles. For 11
421 OTUs, the mutations were silent, including one mutation in the intron or three mutations after
422 the stop codon, and corresponded to the avirulent AvrLm11_0 isoform. The remaining OTUs
423 corresponded to 13 new isoforms. One of them displayed a mutation in a codon encoding a
424 cysteine, therefore generating a de-structured protein and the probable loss of avirulence. For

425 the other 12 new isoforms, we could not determine yet the impact of these mutations on the
426 interaction phenotype. However, they represented only 0.7% of the detected alleles in
427 average, but could reach up to 7.8% in pools displaying numerous new alleles. Altogether,
428 alleles encoding the AvrLm11_0 isoform were largely prevalent in the population (94.6% in
429 average), while isoforms AvrLm11_1 or AvrLm11_2, previously described in Canada, USA or
430 New Zealand (Chen et al., 2020; Van de Wouw et al., 2023) were never detected. *AvrLm11*
431 deletion was estimated at 3.1% in average, ranging from 0% to 15.4% depending on the
432 sample.

433 *AvrLmS*

434 As already observed following the sequencing of individual isolates (Gautier et al., 2023; Van
435 de Wouw et al., 2023), a high level of polymorphism was detected for *AvrLmS*. Forty-one
436 alleles were detected here, with 27 new alleles compared to previous data (Gautier et al.,
437 2023). The proportions of each *AvrLmS* allele greatly varied between samples including within
438 one given site (Supplementary Figure S9). Two new alleles displayed one mutation introducing
439 a stop codon in the sequence, as previously detected in virulent isolates (Gorse M. and Gautier
440 A., unpublished data; Van de Wouw et al., 2023). These virulent alleles were detected in only
441 one pool from the Boisseau site at low frequency. Nineteen new alleles displayed silent
442 mutations and encoded avirulent isoforms. Three alleles displayed non-synonymous
443 mutations leading to AA substitutions in the signal peptide without modification of the mature
444 (avirulent) protein. Only three alleles displayed non-synonymous mutations (R41Q, S49L and
445 E71G) for which the impact on the phenotype could not be determined. They represented in
446 average 0.56% of the detected alleles (data not shown).

447 **Analysis of unknown field samples: comparison between 32 or 96 leaf spots in a pool**

448 We compared the proportions of each detected allele in pools of 32 leaf spots or in "pools of
449 pools" containing up to 96 leaf spots. As exemplified for *AvrLm3* for which many alleles were
450 detected, the proportion of each allele estimated from both pool sizes were highly correlated
451 (R^2=0.9996, Supplementary Figure S10). However, when focusing on alleles present at low
452 percentages in the pools (<2%) the correlation was lower (R^2=0.859), due to OTUs that were
453 not detected in pools of 96 leaf spots (detection limit) but also due to OTUs detected in pools
454 of 96 but not in pools of 32, probably corresponding to PCR or sequencing artefacts
455 (Supplementary Figure S10).

456 **Analysis of unknown field samples: comparison between genotyping and phenotyping for**
457 **virulence.**

458 Using leaf samples from the same site and the same sampling date, we compared the
459 percentages of avirulent isolates in *L. maculans* populations as estimated following inoculation
460 tests on a plant differential set (Balesdent et al., 2023) or using the MPSeqM tool (this study).
461 For most avirulence loci, the inference of the phenotype (virulence or avirulence) from the
462 allele sequences or deletion proportions was possible (see above). As regards *AvrLm3*, the
463 situation is more complex since the avirulence phenotype is masked whenever *AvrLm4-7* is
464 present, at least for some *AvrLm4-7* alleles (Balesdent et al., 2023). To determine the
465 proportion of avirulent isolates on *Rlm3*, we multiplied, for each pool, the percentage of
466 *AvrLm3* avirulent alleles by the percentage of *AvrLm4-7* alleles unable to mask the *AvrLm3-*
467 *Rlm3* interaction. Altogether, phenotypic or MPSeqM analyses gave very close estimates of
468 the proportions of avirulent isolates for the interactions investigated, including the complex
469 *AvrLm3-Rlm3* interaction (Figure 6A). For instance, site by site comparison of MPSeqM or

470 phenotypic estimation of virulent isolates on *Rlm7* perfectly correlated (Supplementary Figure
471 11).

472 **Analysis of unknown field samples: comparison between Sanger sequencing and the**
473 **MPSeqM tool**

474 We previously estimated *AvrLm* allele frequencies in French *L. maculans* population by Sanger
475 sequencing of full-length avirulence genes in a representative population of 89 isolates
476 collected in 2017 (Gautier et al., 2023). These proportions were compared with those
477 determined by the MPSeqM tool using averaged frequencies of alleles over all 42 field
478 samples. Proportions of alleles that are undistinguishable with the MiSeq primers (Like
479 *AvrLm3_A_01* and *AvrLm3_A_09*) were first merged, and similarly, proportions of RIPped and
480 deleted alleles for *AvrLm1* were merged. Altogether, 44 distinct alleles were detected for the
481 eight *AvrLm* genes in the 2017 population. The percentages of these alleles estimated with
482 the MPSeqM tool in the 2019 population or by Sanger sequencing of avirulence genes in the
483 2017 populations perfectly correlated (R^2=0.995, Figure 6B). The difference between the two
484 estimates of a given allele was low (1.8% in average). For only two *AvrLm4-7* alleles,
485 differences between the two analyses were higher than 5%, *AvrLm4-7_2_A* (an avirulent
486 allele) being lower in 2019 (16.3%) compared to 2017 (22.5%) and the deleted allele being
487 higher in 2019 (45.0%) compared to 2017 (39.3%; Figure 6B). Indeed, these differences may
488 simply reflect the evolution toward virulence of French populations between 2017 and 2019
489 due to the *Rlm7* selection pressure in France (Balesdent et al., 2022).

490 **Analysis of unknown field samples with the** MPSeqM **tool: allelic diversity per site**

491 Considering all genes, from 80 to 118 distinct OTUs were detected per site. Relative to the
492 number of leaf spots per site analysed, the site showing the highest diversity (number of

493 distinct OTUs) was that of Frelinghein (Supplementary Figure S12). Depending on the sites,
494 from 40% (Frelinghein) to 53% of the detected OTUs were new OTUs, the site with highest
495 proportion of new OTUs being that of Coullons, mainly due to the high number of new
496 *AvrLm11* OTUS found at that site (Supplementary Figure S12).

497 **Validation of the strategy including detection of new alleles**

498 One pool was used to validate new alleles found with the MPSeqM tool by cloning. In this
499 pool, 13 distinct allele of *AvrLm4-7* were detected and kept after the filtering steps with the
500 MPSeqM tool. When cloning and sequencing the PCR products after *AvrLm4-7* amplification
501 with external primers, all but two of these 13 alleles were recovered. The proportion of clones
502 with these alleles correlated the proportion of reads detected after MiSeq sequencing (data
503 not shown) and these clones represented about 50% of the sequences cloned. Among these
504 alleles, six corresponded to new alleles compared to previously published alleles, and were
505 detected in one to seven independent clones. However, the remaining 50% of the clones each
506 displayed unique alleles (i.e., each representing less than 1% of the amplified sequences) that
507 were either considered not present in the pool, because below the threshold (13 sequences),
508 or even not detected in the dataset whatever the pool (33 sequences).

509

510 **DISCUSSION**

511 In this paper, we describe the design, validation, and use of a molecular tool, termed MPSeqM,
512 for large-scale analysis of polymorphism of eight *AvrLm* genes of *L. maculans* intending to
513 determine proportions of virulent isolates toward the corresponding *B. napus Rlm* genes in
514 field samples. It combines four steps: bulk DNA extraction of *L. maculans* symptoms from the
515 field; a multiplex-PCR step amplifying eight *AvrLm* genes and *Actin* as an invariant gene;
516 illumina sequencing of the PCR products; and lastly, blasting reads on an *AvrLm* variant
517 database and estimation of proportions of each variant in the bulks. The tool was first
518 evaluated on DNA pools of known samples and was found to be versatile, reproducible,
519 quantitative, and efficient for the detection of rare and new allelic variants.

520 As regards versatility, the possibility to characterize simultaneously from four to eight *AvrLm*
521 genes with identical results opens the possibility to modify the set of genes analysed
522 depending on the characteristics of the populations to survey. For instance in France, *AvrLm2*
523 is nearly invariant with fixed virulent alleles, rendering *Rlm2* useless (Balesdent et al., 2023,
524 Gautier et al., 2023). The MPSeqM tool could thus be adapted by replacing *AvrLm2*
525 amplification with other genes of interest, like *AvrLm10* or *AvrLm14*. Alternatively, fewer
526 genes could be analysed simultaneously, to reach deeper sequencing per gene or to analyse
527 more samples for the same sequencing cost.

528 Sequencing results were very reproducible both qualitatively and quantitatively between the
529 three replicates within one sequencing run. Independent runs also gave identical estimates of
530 proportions of each allele (data not shown). The quantitative dimension of the tool was
531 evident, not only when comparing the MPSeqM results with expected proportions of each
532 alleles in control bulks of known composition, but also when comparing MPSeqM results from

533 bulks of field samples with previously established data, whatever the criteria compared
534 (proportion of virulent isolates per site, or proportions of allelic variants per gene).
535 Three main questions had to be addressed to validate the tool: (i) what could be the best
536 sample size (number of leaf spots) within each pool for field surveys, (ii) whether gene
537 deletions could be quantified, and (iii) whether unknown alleles could be detected.
538 MPSeqM results for control DNA with known isolates were used to determine the best
539 number of individual samples in one bulk. In the Mix_LM control containing DNA of 71 isolates,
540 many alleles (>20%) present in one isolate only were not detected. This value dropped to
541 around 6% in the bulk with 32 leaf spots and, more importantly, 100% of alleles encoding non-
542 RIPped sequences were detected. Lack of detection of a RIPped allele present at low
543 frequency is not a problem if the deletions (i.e. lack of amplification) can be accurately
544 detected, because both RIPped and deleted alleles correspond to virulent versions of the
545 gene. Altogether, comparison of pools of 32 or 96 leaf spots indicated a better detection of
546 alleles at low frequencies with pools of 32 leaf spots. Therefore, pools of 32 leaf spots were
547 selected to assess the MPSeqM on unknown samples.
548 The use of a dilution series with varying proportions of DNA of two isolates differing by the
549 presence or absence of some avirulence genes allowed us to quantify deletions. The data
550 (number of reads for a given avirulence gene) were first normalized with the number of reads
551 for a single-copy, conserved gene. Here the *Actin* gene was chosen as it is reliable for species-
552 specific identification of *L. maculans* within the *Leptosphaeria* species complex (Vincenot et
553 al., 2009; Voigt et al., 2005). The normalized read numbers for the deleted genes in the dilution
554 series correlated with the expected proportions. Using the dilution curves, the proportions of
555 deleted isolates in known bulks were accurately quantified for either high (*AvrLm1*), low
556 (*AvrLm11*), or intermediate (*AvrLm4-7*) proportions of deletion events.

557 Unexpected OTUs were detected at low frequencies in bulks with known *AvrLm* allele
558 composition, corresponding to artefacts generated at the PCR or sequencing steps. The
559 proportions of such reads in controls were used to establish the thresholds under which MiSeq
560 reads were suppressed. Once these thresholds were applied, the datasets from field samples
561 comprised a reduced number of allelic variants per gene, but still numerous unknown alleles.
562 We defined rules to select reliable alleles among these new sequences. Only unknown
563 sequences consistently present in the three replicates of one field sample and in a few samples
564 only were kept, allowing us to identify 206 new alleles including 74 new RIPped alleles.
565 Although not exhaustive yet, we experimentally confirmed the reality of at least some new
566 alleles by cloning PCR products from one field sample and subsequent Sanger sequencing.
567 Depending on the mutations present in these new alleles, the inference of the corresponding
568 phenotype (virulent or avirulent) was more or less easy. RIPped sequences, sequences with
569 frameshift or stop codons were considered virulent. Other variants may display silent
570 mutations or mutations without impact on the mature protein and thus, were easily converted
571 into avirulent or virulent alleles. However, a non-negligible amount (9.2%) of new alleles could
572 not be converted into a phenotype. These new alleles represented low proportions of reads
573 per sample (1.5% on average), thus without a high impact on the survey result. If an unknown
574 allele reaches high proportions in specific sites, it will be necessary to go back to isolation
575 campaigns at this site, with the aim of recovering and phenotyping isolates with that allelic
576 variant, or to assess the corresponding phenotype following the transformation of a virulent
577 isolate with a synthetic construct displaying this new allele.
578 The MPSeqM tool has some limitations. Because of the rather short sequence length, only
579 part of the full CDS length is amplified for some *AvrLm* genes, possibility preventing us from
580 detecting mutations in the 5' or 3' regions of the genes. The MPSeqM tool nevertheless

581 enabled us to estimate frequencies of virulent isolates from field samples highly correlated
582 with those determined by pathotyping. This limitation has however to be kept in mind, and
583 surveys based on isolation, pathotyping, and full-length *AvrLm* sequencing will still have to be
584 done regularly (for instance, every five years) to detect new virulent alleles, enrich the
585 sequence database, and adapt the primers to amplify new alleles. Another limitation concerns
586 *AvrLm4-7*. Some virulent isolates toward *Rlm7* do not amplify the gene with external primers
587 but could amplify internal parts of the gene with a sequence matching avirulent alleles
588 (Daverdin et al., 2012; Gautier et al., 2023). Such isolates represent 6.7% of the 2017 sampling
589 (Gautier et al., 2023) and they will be miss-identified as avirulent with the MPSeqM tool. This
590 phenomenon could explain why the MPSeqM tool detects more avirulent isolates toward
591 *Rlm7* than pathotyping (Figure 6).
592 The MPSeqM field survey revealed highly contrasted levels of polymorphism depending on
593 the *AvrLm* gene, confirming data obtained by *AvrLm* sequencing in individual isolates, either
594 from France (Gautier et al., 2023) or worldwide (Van de Wouw et al., 2023), with highly
595 polymorphic genes (*AvrLm3*, *AvrLm4-7* or *AvrLmS-Lep2*) or low polymorphic ones (*AvrLm2* or
596 *AvrLm6*). Here, by analysing a total of 1337 leaf spots in one single sequencing run, we
597 confirmed these trends but also detected a number of new, rare alleles, most of them being
598 site specific. This work thus revealed an unexpected diversity of *AvrLm* sequences for at least
599 some genes. Whether this polymorphism is pre-existing at a basal level, or whether it is locally
600 generated at each generation through meiosis (Daverdin et al., 2012), is still in question. The
601 high number of new RIPped alleles detected for *AvrLm4-7* (66 new sequences), but not for
602 other genes suggests RIP is still an active mechanism particularly frequent for this gene,
603 generating new alleles every year at meiosis.
604

605 **CONCLUSION**

606 The MPSeqM tool appears suitable for a precise characterization of *L. maculans* populations
607 as regards virulence at eight avirulence loci, without the tedious steps of *L. maculans* isolation,
608 in vitro growth and inoculation on a plant differential set. The MPSeqM tool also gives a
609 detailed and precise characterisation of the diversity of *AvrLm* genes in populations present
610 in specific situations (geographical location, specific time-point, plant genotype, plant organ).
611 We highlighted contrasting situations between sites, with some sites displaying greater
612 diversity but few new alleles, and others being less diverse but richer in new alleles. The
613 MPSeqM tool thus offers new perspectives for new research questions, such as: do some plant
614 genotypes select some allelic variants along the cropping season? How plant genotypes affect
615 allelic variants? How different are populations present in very close or distant fields? May
616 some resistance management strategies affect the allelic diversity of avirulence genes?
617 Finally, this tool, associated with scorings of stem canker severity, could help bringing
618 extensive field data to determine a "critical percentage of virulent isolates" at which the
619 resistance is in danger of being rapidly overcome, an information currently not available for
620 the *L. maculans* / *B. napus* pathosystem.

621

622 **ACKNOWLEDGMENTS**

623 This work was funded by the CASDAR project C2018-11 (Atipical). We thank all contributors
624 from the CTPS field experiments that provided infected oilseed rape leaves and Julie Noah for
625 her contribution to leaf spot recovery. The "Effectors and Pathogenesis of *L. maculans*" group
626 benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). We thank the GeT
627 Genotoul platform in Toulouse, Midi-Pyrénées for the MiSeq sequencing experiments. We are
628 grateful to the Genotoul Bioinformatics platform in Toulouse, Midi-Pyrénées for providing
629 help in computing and data storage.

630

631 **Data availability statement**

632 Data available on request from the authors.

633 **References**

634 Alnajar, D., von Tiedemann, A. & Koopmann, B. (2022) Efficacy of blackleg major resistance
635 genes in *B. napus* in Germany. *Pathogens* 11, 461.

636 Balesdent, M.H., Gautier, A., Plissonneau, C., Le Meur, L., Loiseau, A., Leflon, M. et al. (2022)
637 Twenty years of *Leptosphaeria maculans* population surveys in France suggest
638 pyramiding *Rlm3* and *Rlm7* in rapeseed is a risky resistance management strategy.
639 *Phytopathology* 112, 2126-2137.

640 Balesdent, M.H., Laval, V., Bagot, P. & Rouxel, T (2023) Large-scale population survey of
641 *Leptosphaeria maculans* in France highlights both on-going breakdowns and potentially
642 effective resistance genes in oilseed rape. *Pest Management Sciences*
643 https://doi.org/10.1002/ps.7401

644 Balesdent, M.H., Louvard, K., Pinochet, X. & Rouxel, T. (2006) A large-scale survey of races of
645 *Leptosphaeria maculans* occurring on oilseed rape in France. *European Journal of Plant*
646 *Patholo*gy 114, 53-65.

647 Borhan, M.H., Van de Wouw, A.P. & Larkan, N.J. (2022) Molecular interactions between
648 *Leptosphaeria maculans* and *Brassica* species. *Annual Review of Phytopathology* 60,
649 11.1–11.21

650 Boutigny, A.L., Gautier, A., Basler, R., Dauthieux, F., Leite, S., Valade, R. et al. (2019)
651 Metabarcoding targeting the EF1 alpha region to assess *Fusarium* diversity on cereals.
652 PLoS ONE 14(1): e0207988. https://doi.org/10.1371/journal.pone.020798.

653 Carpezat, J., Bothorel, S., Daverdin, G., Balesdent, M.H. & Leflon, M. (2014) Use of high
654 resolution melting analysis to genotype the avirulence *AvrLm4-7* gene of *Leptosphaeria*
655 *maculans*, a fungal pathogen of *Brassica napus. Annals of Applied Biology* 164, 430–444.

656 Chen, Q., Peng, G., Kutcher, R., & Yu, F. (2020) Identification of genome-wide DNA variants
657 and SNP haplotypes associated with avirulence genes in *Leptosphaeria maculans* in
658 Western Canada. *Research Square:* https://doi.org/10.21203/rs.3.rs-24766/v1.

659 Daverdin, G., Rouxel, T., Gout, L., Aubertot, J.N., Fudal, I., Meyer, M. et al. (2012) Genome
660 structure and reproductive behaviour influence the evolutionary potential of a fungal
661 phytopathogen. *PLoS Pathogens* 8, e1003020.

662 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., & Knight, R. (2011) UCHIME improves
663 sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200.
664 https://doi.org/10.1093/bioinformatics/btr381.

665 Fudal, I., Ross, S., Brun, H., Besnard, A.L., Ermel, M., Kuhn, M.L. et al. (2009) Repeat-induced
666 point mutation (RIP) as an alternative mechanism of evolution towards virulence in
667 *Leptosphaeria maculans*. *Molecular Plant-Microbe Interaction* 22, 932-941.

668 Gautier, A., Laval, V., Faure, S., Rouxel, T. & Balesdent, M.H. (2023) Polymorphism of
669 avirulence genes and adaptation to Brassica resistance genes is gene dependent in the
670 phytopathogenic fungus *Leptosphaeria maculans*. *Phytopathology*, 113, 1222-1232.

671 Ghanbarnia, K., Ma, L., Larkan, N.J., Haddadi, P., Fernando, D.W.G. & Borhan, M.H. (2018)
672 *Leptosphaeria maculans AvrLm9*: a new player in the game of hide and seek with
673 *AvrLm4-7*. *Molecular Plant Pathology* 19, 1754-1764.

674 Gladyshev, E. (2017) Repeat-Induced Point Mutation (RIP) and other genome defense
675 mechanisms in fungi. *Microbiology Spectrum* 5, 10.1128.
676 Gout, L., Eckert, M., Rouxel, T. & Balesdent, M.H. (2006) Genetic variability and distribution of
677 mating-type alleles in field populations of *Leptosphaeria maculans* from France. *Applied*
678 *and Environmental Microbiology* 72, 185-191.

679 Gout, L., Kuhn, M.L., Vincenot, L., Bernard-Samain, S., Cattolico, L., Barbetti, M. et al. (2007)
680 Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the
681 plant pathogen *Leptosphaeria maculans*. *Environmental Microbiology* 9, 2978-2992.

682 McCulloch, G. A., & Waters, J. M. (2023) Rapid adaptation in a fast-changing world: emerging
683 insights from insect genomics. *Global Change Biology* 29, 943–954.

684 Needleman, S.B., & Wunsch, C.D. (1970) A general method applicable to the search for
685 similarities in the amino acid sequence of two proteins. *Journal of Molecular Biol*ogy 48,
686 443–453.

687 Parlange, F., Daverdin, G., Fudal, I., Kuhn, M.L., Balesdent, M.H., Blaise, F. et al. (2009)
688 *Leptosphaeria maculans* avirulence gene *AvrLm4-7* confers a dual recognition specificity
689 by the *Rlm4* and *Rlm7* resistance genes of oilseed rape, and circumvents *Rlm4*-mediated
690 recognition through a single amino acid change. *Molecular Microbiology* 71, 851-863.
691 Plissonneau, C., Blaise, F., Ollivier, B., Leflon, M., Carpezat, J., Rouxel, T. et al. (2017) Unusual
692 evolutionary mechanisms to escape Effector-Triggered-Immunity in the fungal
693 phytopathogen *Leptosphaeria maculans. Molecular Ecol*ogy 26, 2183-2198.

694 Rouxel, T. & Balesdent, M.H. (2017) Life, death and rebirth of avirulence effectors in a fungal
695 pathogen of Brassica crops, *Leptosphaeria maculans. New Phytologist* 214, 526-532.

696 Rouxel, T., Penaud, A., Pinochet, X., Brun, H., Gout, L., Delourme, R. et al. (2003) A ten-year
697 survey of populations of *Leptosphaeria maculans* in France indicates a rapid adaptation
698 towards the *Rlm1* resistance gene in oilseed rape. *European Journal of Plant Pathology*
699 109, 871-881.

700 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009)
701 Introducing mothur: Open-source, platform-independent, community-supported
702 software for describing and comparing microbial communities. *Applied and*
703 *Environmental Microbiology* 75, 7537–7541.

704 Sprague, S.J., Balesdent, M.H., Brun, H., Hayden, H.L., Marcroft, S.J., Pinochet, X. et al. (2006)
705 Major gene resistance in *Brassica napus* (oilseed rape) is overcome by changes in
706 virulence of populations of *Leptosphaeria maculans* in France and Australia. *European*
707 *Journal of Plant Pathology* 114, 33–40.

708 Stachowiak, A., Olechnowicz, J., Jedryczka, M., Rouxel, T., Balesdent, M.H., Happstadius, I., et
709 al. (2006) Frequency of avirulence alleles in field populations of *Leptosphaeria maculans*
710 in Europe. *European Journal of Plant Pathology* 114, 67-75.

711 Van de Wouw, A.P., Howlett, B.J. (2012) Estimating frequencies of virulent isolates in field
712 populations of a plant pathogenic fungus, *Leptosphaeria maculans*, using high-
713 throughput pyrosequencing. *Journal of Applied Microbiology* 113, 1145–53.

714 Van de Wouw, A.P., Scanlan, J.L., Al-Mamun, H.A., Balesdent, M.H., Bousset, L., Burketová, L.
715 et al. (2023) A new set of international *Leptosphaeria maculans* isolates as a resource
716 for elucidation of the basis and evolution of blackleg disease on *Brassica napus*. *Plant*
717 *Pathology*, DOI: 10.1111/ppa.13801.

718 Van de Wouw, A.P., Stonard, J.F., Howlett, B.J., West, J.S., Fitt, B.D.L. & Atkins, S.D. (2010)
719 Determining frequencies of avirulent alleles in airborne *Leptosphaeria maculans*
720 inoculum using quantitative PCR. *Plant Pathology* 59, 809–818.

721 Vincenot, L., Balesdent, M.H., Li, H., Barbetti, M.J., Sivasithamparam, S., Gout, L. et al. (2008)
722 Occurrence of a new subclade of *Leptosphaeria biglobosa* in Western Australia.
723 *Phytopathology* 98, 321-329

724 Voigt, K., Cozijnsen, A.J., Kroymann, J., Pöggeler, S. & Howlett, B.J. (2005) Phylogenetic
725 relationships between members of the crucifer pathogenic *Leptosphaeria maculans*
726 species complex as shown by mating type (MAT1-2), actin, and beta-tubulin sequences.
727 *Molecular Phylogenetics and Evolution* 37, 541-57.

728 Wang, Q., Garrity, G.M., Tiedje, J.M., & Cole, J.R. (2007) Naive Bayesian classifier for rapid
729 assignment of rRNA sequences into the new bacterial taxonomy. *Applied and*
730 *Environmental Microbiology* 73, 5261–5267.

731 Zhang, X., Peng, P., Kutcher, R., Balesdent, M.H., Delourme, R. & Fernando, D.W.G. (2016)
732 Breakdown of *Rlm3* resistance in the *Brassica napus–Leptosphaeria maculans*
733 pathosystem in western Canada. *European Journal of Plant Patholo*gy 145, 659-674.

734

735 **Figure legends**

736 **Figure 1. Schematic representation of sequence analyses and dataset cleaning.** Grey boxes
737 correspond to output files at specific stages of the analysis. [a], this threshold ("maximum read
738 number threshold") removes background OTUs due to sequencing errors and is based on
739 unexpected values observed for control DNAs; [b], this threshold ("percentage threshold") is
740 based on the relative proportion of an OTU in an field sample and its value is based on relative
741 proportions observed in control DNAs.

742 **Figure 2. Efficiency and reproducibility of MiSeq sequencing following multiplex PCR**
743 **amplification of *Leptosphaeria maculans* DNA controls.** A, the JN3 reference, single isolate,
744 DNA; B, pooled leaf symptoms from 32 known isolates. Replicates (REP) correspond to
745 independent PCR experiments. Values are percentages of reads matching each amplified
746 gene.

747 **Figure 3. Serial dilution curves for *AvrLm* deletion quantification.** A mix of DNA of two isolates
748 differing by the presence or absence of four avirulence genes created deletion proportions
749 ranging from 0% to 100%. The observed proportion of the gene in the mix was estimated with
750 the ratio between the read numbers for this gene in the sample over the *Actin* read numbers.
751 In blue, *AvrLm1*; orange, *AvrLm4-7*; grey, *AvrLm11*; green, *AvrLmS-Lep2*. Polynomial
752 regression curves and R^2 are displayed for each gene.

753 **Figure 4. Expected and observed proportions of alleles of eight *L. maculans* avirulence genes**
754 **in DNA pools.** A, DNA of a pool of 32 leaf spots from known isolates and B, pool of DNA of 71
755 known isolates. The eight *AvrLm* genes were amplified in multiplex PCR with three
756 independent PCR replicates, and then sequenced with MiSeq. Each dot correspond to the
757 mean proportion in the pools of each allele (OTU) over all alleles of every *AvrLm* gene,
758 including deleted alleles, and the colour code on the left refers to the gene. The lower and
759 upper limits correspond to the exact confidence interval around the expected value for a mix
760 of 32 (A) or 71 (B) individuals. Detailed values are in Supplementary Table S3.

761 **Figure 5. An example of detection of new alleles.** Percentages of new OTUs over all *AvrLm11*
762 OTUs are given for the three control samples (left) and the 52 field samples from nine sites,
763 separated by the vertical blue lines. Bars in dark orange are those of pools of 32 leaf spots that
764 are also in the "pools of pools "of 96 leaf spots (pale orange).

765 **Figure 6. Validation of the MPSeqM tool on field samples**. A, Comparison between
766 percentages of avirulent isolates in French populations of *Leptosphaeria maculans,* estimated
767 by phenotyping or genotyping. Black, data from phenotyping 690 isolates from 9 sites
768 (Balesdent et al, 2023); grey, data from MultiplexPCR-Miseq analysis of 42 pools of leaf spots
769 from the same sites (including a total of 1337 leaf spots). Values are means and standard
770 errors over all sites. B, Comparison between allele frequencies (all genes) estimated by Sanger

771 sequencing of individual isolates (2017 population, Gautier et al., 2013) and by the MPSeqM
772 tool of pools of leaf spots (2019 population).

773

774 **Supplementary Table S1. List and characteristics of *Leptosphaeria maculans* isolates used**
775 **in this study**

776 **Supplementary Table S2. Genotypes of reference *Leptosphaeria maculans* isolates at eight**
777 **avirulence loci**. Isolates in bold are those present in the poolof 32 *Brassica napus* leaf spots
778 used as control.

779 **Supplementary Table S3. List and proportions of *AvrLm* alleles present in the Mix_LM and**
780 **Mix_LSpot controls and the corresponding estimated proportions with after MiSeq**
781 **sequencing.**

782 **Supplementary Figure S1. Experimental design for field samples analyses.** 1, Leaves at
783 reception in the lab, cleaned and air-dried; 2, one leaf spot per leaf is cut and placed in humid
784 conditions for pycnidia production, and 3, mono-pycnidia isolation for 4, inoculation on a plant
785 differential set to determine the phenotype. 5, standardized leaf spot cutting (one per leaf)
786 and long-term storage in 24-well plates. 7, a view of a full plate before storage until 8, pool
787 DNA extraction, multiplex PCR and MiSeq sequencing

788 **Supplementary Figure S2. Alignment of known alleles of eight *L. maculans* avirulence genes**
789 **with positions of forward and reverse MiSeq primers.** A, *AvrLm1*; B, *AvrLm2*; C, *AvrLm3*; D,
790 *AvrLm4-7*; E, *AvrLm5-9*; F*, AvrLmS-Lep2*; G*, AvrLm6*; H, *AvrLm11*. The F-primers are located
791 before the start for *AvrLm5-9* and *AvrLm6* and the R-primer is after the stop for *AvrLm11*.

792 **Supplementary Figure S3. Box plot of the frequency per sampling site of avirulent alleles for**
793 ***AvrLm1,* as estimated by MiSeq sequencing in 42 pools of leaf spots.**

794 **Supplementary Figure S4. Proportions of the three *AvrLm2* alleles detected in 42 pools of**
795 **leaf spots from 9 fields**

796 **Supplementary Figure S5 Maximum likelihood tree of all *AvrLm3* isoforms identified here.**
797 Isoforms underlined in green are known avirulent isoforms while those underlined in red are
798 known virulent isoforms**.**

799 **Supplementary Figure S6. Percentages of the 41 *AvrLm3* alleles as determined following**
800 **multiplex-PCR and MiSeq sequencing of 42 field samples**

801 **Supplementary Figure S7. Proportions of *AvrLm5-9* alleles detected in 42 pools of leaf spots**
802 **from nine fields**

803 **Supplementary Figure S8. Proportions of *AvrLm6* alleles detected in 42 pools of leaf spots**
804 **from nine fields**

805 **Supplementary Figure S9. Proportions of *AvrLmS-Lep2* alleles detected in 42 pools of leaf**
806 **spots from nine fields**

807 **Supplementary Figure S10. Relationships between percentages of each *AvrLm3* OTU per site**
808 **estimated in 32 leaf spots or from the "pools of pools" (64 to 96 leaf spots). Each dot**
809 **corresponds to the percentage of one OTU in one site.** A, all data; B, C and D, focus on high
810 (>60%), low (<20%) or very low (<2) percentages, respectively. Green circle highlights OTUs
811 that are detected at low frequencies in 24 or 32 leaf spot pools but not in 64 or 96 leaf spots.
812 The red circle highlights OTUs that are not detected in pools of 32 leaf spots but found in pools
813 of 96, thus probably corresponding to sequencing artefacts.

814 **Supplementary Figure S11. Correlation between estimates of percentage of avirulent**
815 **isolates on *Rlm7* by pathotyping and with the MPSeqM tool.** Each dot corresponds to one
816 geographical location. Samplings were done at the same place and same time but on different
817 leaves. Pathotyping was described in Balesdent et al. (2023) and was done using around 80
818 isolates per site, while the genotyping was done on 3 to 6 pools of 32 leaf spots per site. The
819 Spearman correlation coefficient is indicated.

820 **Supplementary Figure S12. Relative *AvrLm* allele diversity per site as described with the**
821 **MpSeqM tool.** A, diversity for all genes; B, diversity for *AvrLm11* only. The radar diagrams
822 represent, for each site, the relative proportion of alleles (no. of alleles at this site over total
823 no. of alleles detected in the survey) averaged by the relative proportion of leaf spots analysed
824 at this site. Dark line, overall diversity; dark grey line, diversity for "new" alleles; pale grey line,
825 diversity for site-specific new alleles. Sites as follows: Co, Coullons; Ja, Janville; Th1, Thiverval-
826 Grignon field 1; Th2, Thiverval-Grignon field 2; Boi, Boisseau; Er, Erdre-en-Anjou; Bou, Bouges-
827 le-Château; Fr, Frelinghein; LeR, Le Rheu

828 **Supplementary Text T1. Parameters used for the Mothur pipeline.**

829

830 Table 1 Characteristics of the sites sampled in autumn 2019 and 2020, and number of samples
831 analysed per site

Location	French department	No. of pools of 32 leaf spots	No. of leaf spots (no. of pools) in the "pool of pools"
Boisseau	41	6	96 (3)
Bouges le château	36	5	96 (3) and 64 (2)
Coullons	45	6	96 (3)
Erdre-en-Anjou	49	4	96 (3)
Frelinghien	59	3[a]	88 (3)
Janville	28	4	96 (3)
Le Rheu	35	6	96 (3)
Thiverval Grignon (field 1)	78	4	64 (2)
Thiverval Grignon (field 2)	78	4	64 (2)

832 [a] except for one pool with only 24 leaf spots

833

834 Table 2. Characteristics of the eight *AvrLm* and *Actin* gene regions amplified with MiSeq primers

Gene	Gene size (bp) for JN3	No. of known polymorphic sites[a]	Forward Primer for MiSeq (5'-> 3') [b]	Reverse Primer for MiSeq (5'-> 3') [c]	Amplicon size (bp)[d]	Length of the analysed part of the gene (bp)	No. (%) of known polymorphic sites in the amplicon[e]
AvrLm1	667	3	TCCGTCCAGATAAGCCCTGT	CACGAATGATGCACCAGAGGC	455	399	3 (100%)
AvrLm2	699	5	CCCAGCAACAATTTTCGGAACTCC	TTCAACTGTGTATTGGCCCTGG	433	377	5 (100%)
AvrLm3	539	25	CCAAATCTTGTTTCTATTATTTACCGCAT	GATGATCGCAGTGTACCTCGTCTTT	511	455	21 (84%)
AvrLm4-7	478	13	ACTCGCTCTCTCTATCCCTACA	CAACCACGAGTCCTTGCTTTCA	495	439	13 (100%)
AvrLm5-9	426	6	ACAACCACTCTTCTTCACAGTTT	TTAGAGATAACTAAGCTCCAATCCG	486	430	5 (83.3%)
AvrLm6	581	3	GCACGTTCCAAAGCCCTATTTA	ATCGTCATGGGCTTGGATGTTA	506	450	2 (66.7%)
AvrLm11	365	3	GCTTCATCATGCGTTTCTTGCT	CCACAGCACACTATTCACTTTAGC	504	448	3 (100%)
AvrLmS-Lep2	426	15	GGCCAACTTTATCATGCTCTACCT	GGCAACTGATGGGACCGCTC	487	431	15 (100%)
Actin partial (512F/CA14R)	374	-	ATGTGCAAGGCCGGTTTCG	GCCAGATCTTCTCCATGTCATCCCAGTT	430	374	-

835 [a] From Start to Stop, data from Gautier et al. (2023)
836 [b] all forward primers were tagged in 5' with MiSeq specific sequence CTTTCCCTACACGACGCTCTTCCGATCT
837 [c] all reverse primers were tagged in 5' with MiSeq specific sequence GGAGTTCAGACGTGTGCTCTTCCGATCT
838 [d] including the 5' and 3' tags and for JN3 isolate
839 [e] except RIP mutations; known mutations as described by Gautier et al. (2023).
840

841 Table 3. Sequencing statistics including numbers of OTUs retained at each filtering step

Gene	Mean read no. per sample (min;max)[a]	No. of OTUs in at least 3 samples	Max. read no. threshold[b]	No. of OTUs above the threshold	No. of 0TUs retained after analysis of control DNA[c]	Threshold for the detection of an OTU in a field sample[d]	Final no. of new OTUs	Final no. of new RIPed OTUs[e]	Final no. of new isoforms (full lenght)
AvrLm1	429 (0;6886)	1045	12	50	5	0.10%	2	1	0
AvrLm2	7977 (5971;14799)	18548	25	22	6	0.20%	0	0	0
AvrLm3	4251 (3045;6594)	12383	10	194	46	0.20%	28	0	10
AvrLm4-7	2183 (1242;4377)	10147	10	278	139	0.20%	110	66	25
AvrLm5-9	4398 (3283;8029)	8643	9	198	14	0.20%	7	0	3
AvrLm6	2330 (1523;4686)	4488	9	75	8	0.15%	5	0	1
AvrLm11	3815 (2692;6196)	7656	10	172	36	0.40%	27	7	13
AvrLmS-Lep2	3965 (2896;5498)	18114	8	215	46	0.50%	27	0	7
Total		81024		1204	300		206	74	59

842 [a] Data including control samples (JN3_DNA; Mix_LSpot; Mix_LM) and environmental samples (pools of leaf spots).
843 [b] All OTUs for which read counts are below these values in all samples are removed. These values are gene- and run- dependent
844 [c] based on unexpected OTUs observed in controls and with similar read counts in field samples

845 [d] This "percentage threshold" represents the percentage of read counts for an *AvrLm* OTU over all read counts for this gene in a given sample. All values
846 with OTU read counts below this threshold are replaced by zero in the final dataset.
847 [e] A sequence is considered as RIPped if it displays a few mutations from C→T or G→ A, introducing stop codons in the sequence.

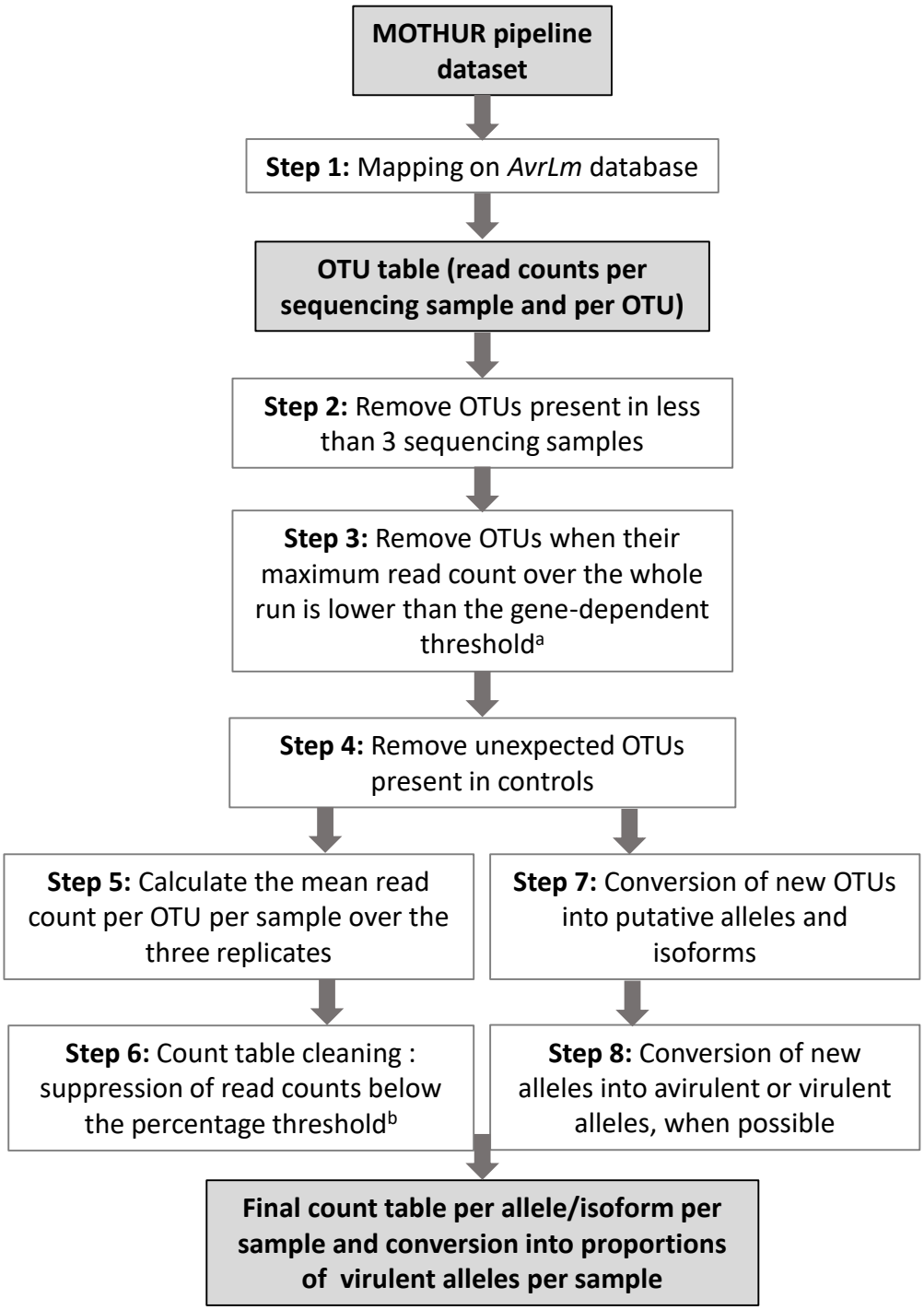

Figure 1. Schematic representation of sequence analyses and dataset cleaning. Grey boxes correspond to output files at specific stages of the analysis. [a], this threshold ("maximum read number threshold") removes background OTUs due to sequencing errors and is based on unexpected values observed for control DNAs; [b], this threshold ("percentage threshold") is based on the relative proportion of an OTU in an field sample and its value is based on relative proportions observed in control DNAs.

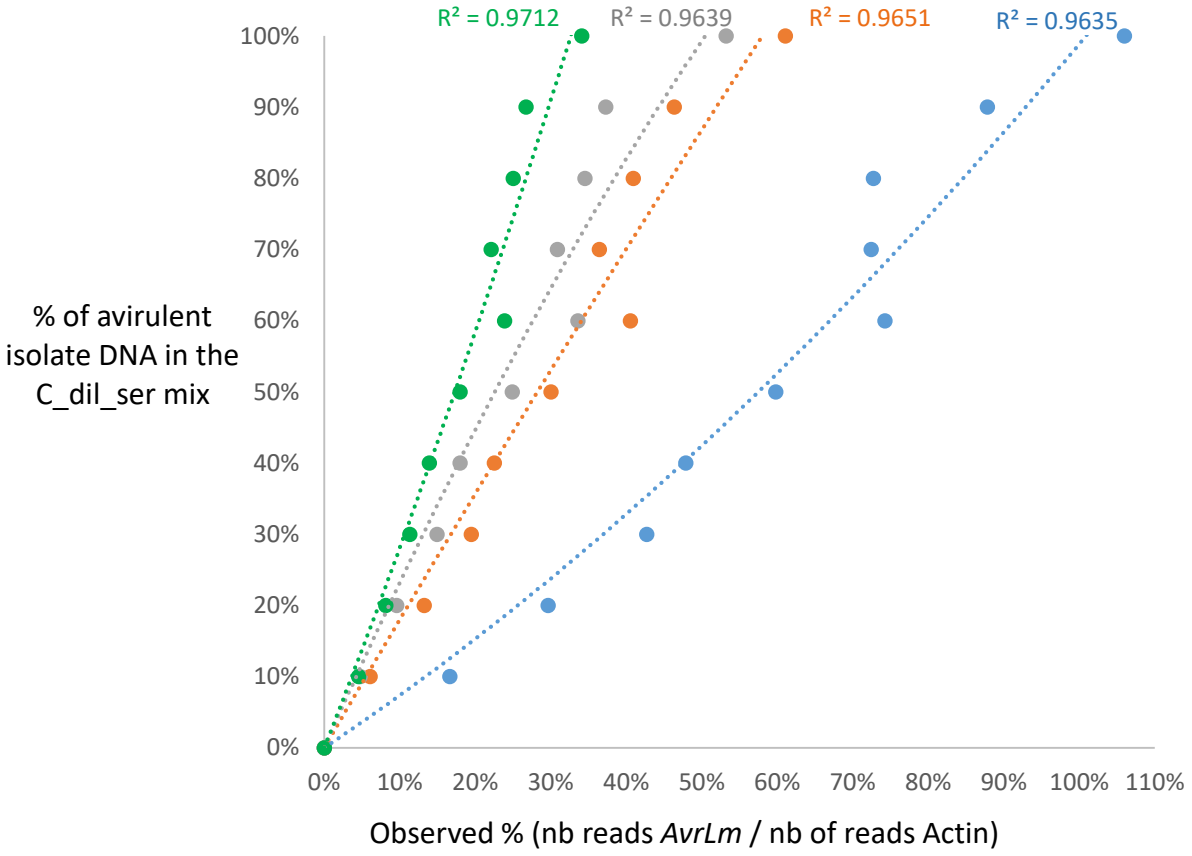

Figure 3. Serial dilution curves for *AvrLm* deletion quantification. A mix of DNA of two isolates differing by the presence or absence of four avirulence genes created deletion proportions ranging from 0% to 100%. The observed proportion of the gene in the mix was estimated with the ratio between the read numbers for this gene in the sample over the *Actin* read numbers. In blue, *AvrLm1*; orange, *AvrLm4-7*; grey, *AvrLm11*; green, *AvrLmS-Lep2*. Polynomial regression curves and R^2 are displayed for each gene.

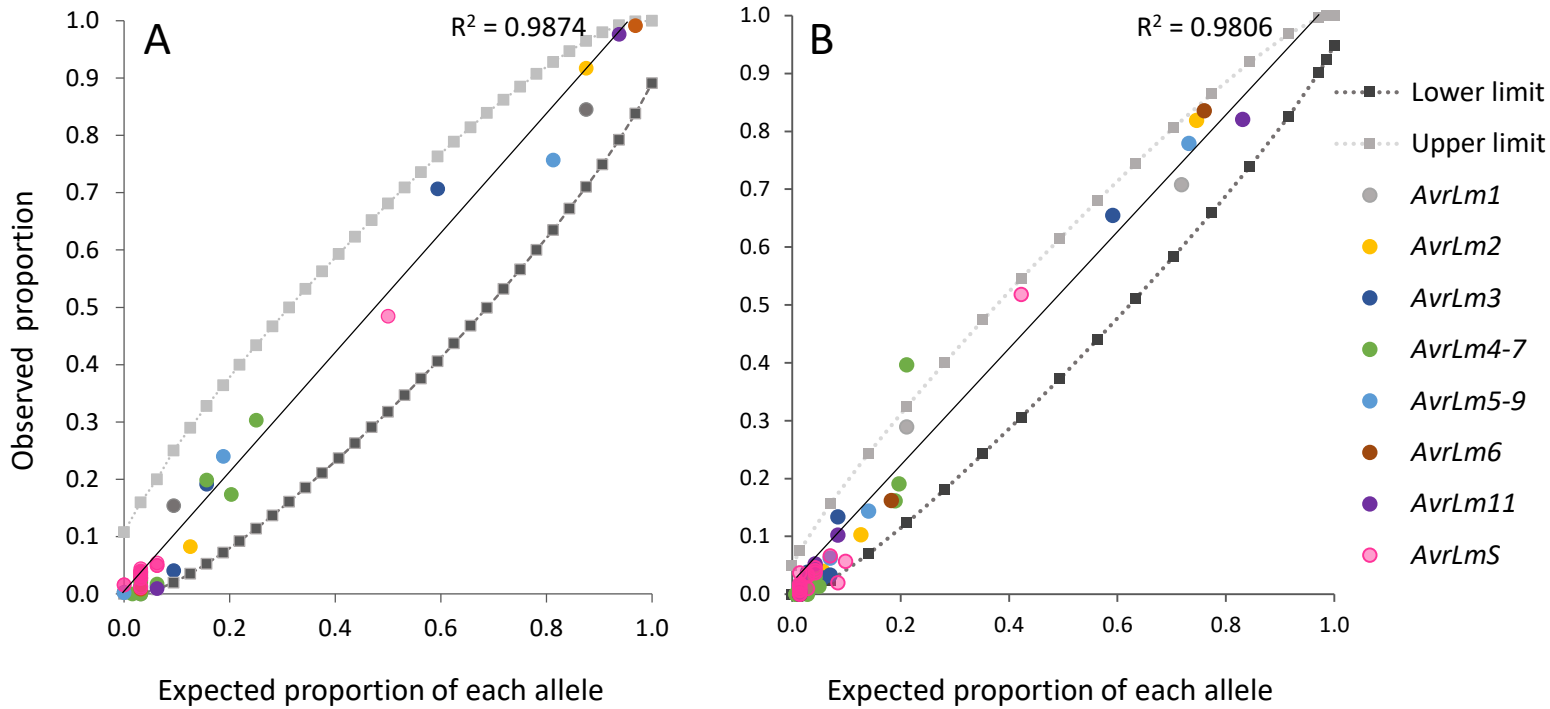

Figure 4. Expected and observed proportions of allele of eight *L. maculans* avirulence genes in DNA pools. A, DNA of a pool of 32 leaf spots from known isolates and B, pool of DNA of 71 known isolates. The eight *AvrLm* genes were amplified in multiplex PCR with three independent PCR replicates, and then sequenced with MiSeq. Each dot correspond to the mean proportion in the pools of each allele (OTU) over all alleles of every *AvrLm* gene, including deleted alleles, and the colour code on the left refers to the gene. The lower and upper limits correspond to the exact confidence interval around the expected value for a mix of 32 (A) or 71 (B) individuals. Detailed values are in Supplementary Table S3.

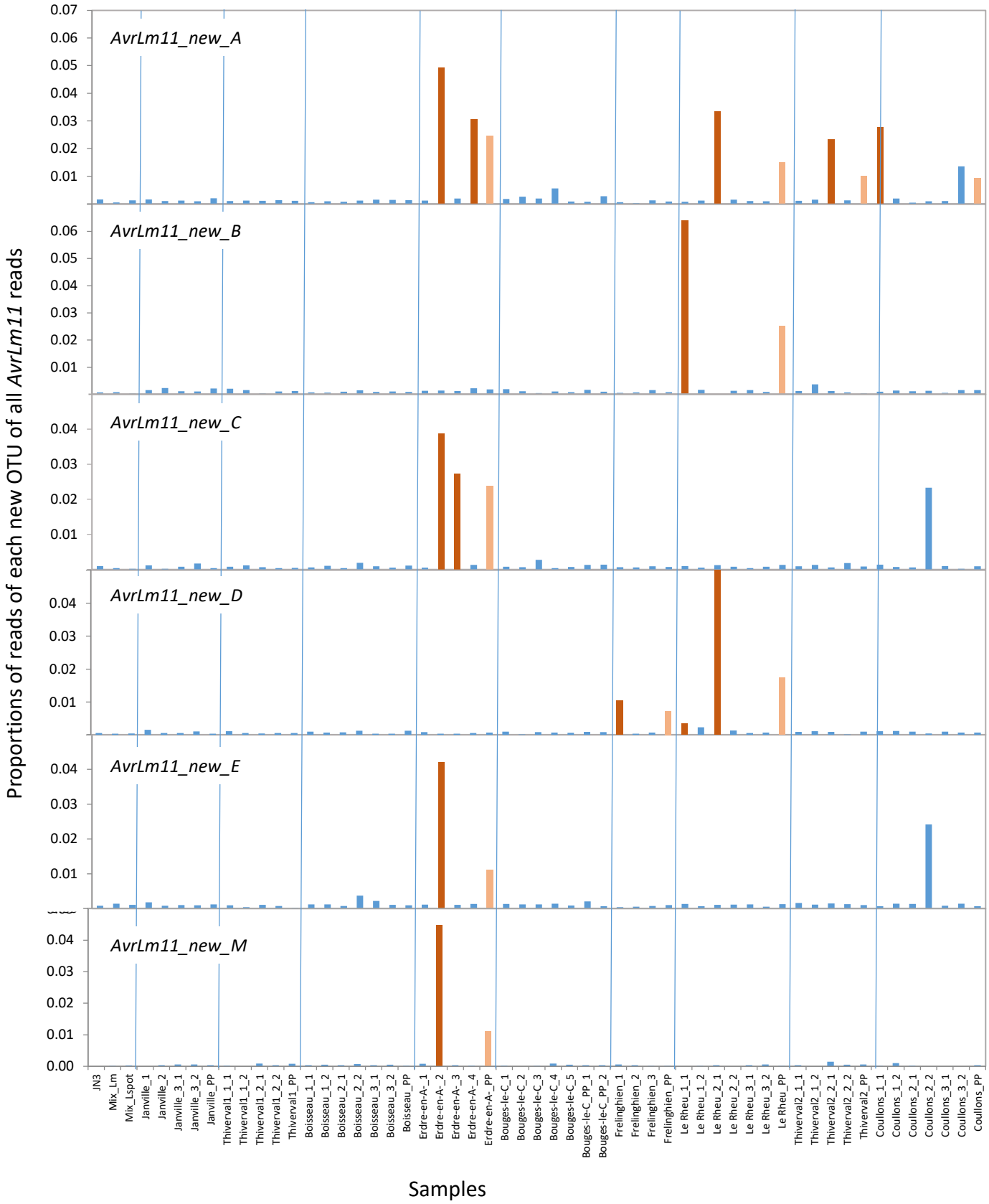

Figure 5. An example of detection of new alleles. Percentages of new OTUs over all *AvrLm11* OTUs are given for the three control samples (left) and the 52 environmental samples from nine sites, separated by the vertical blue lines. Bars in dark orange are those of pools of 32 leaf spots that are also in the "pools of pools "of 96 leaf spots (pale orange).

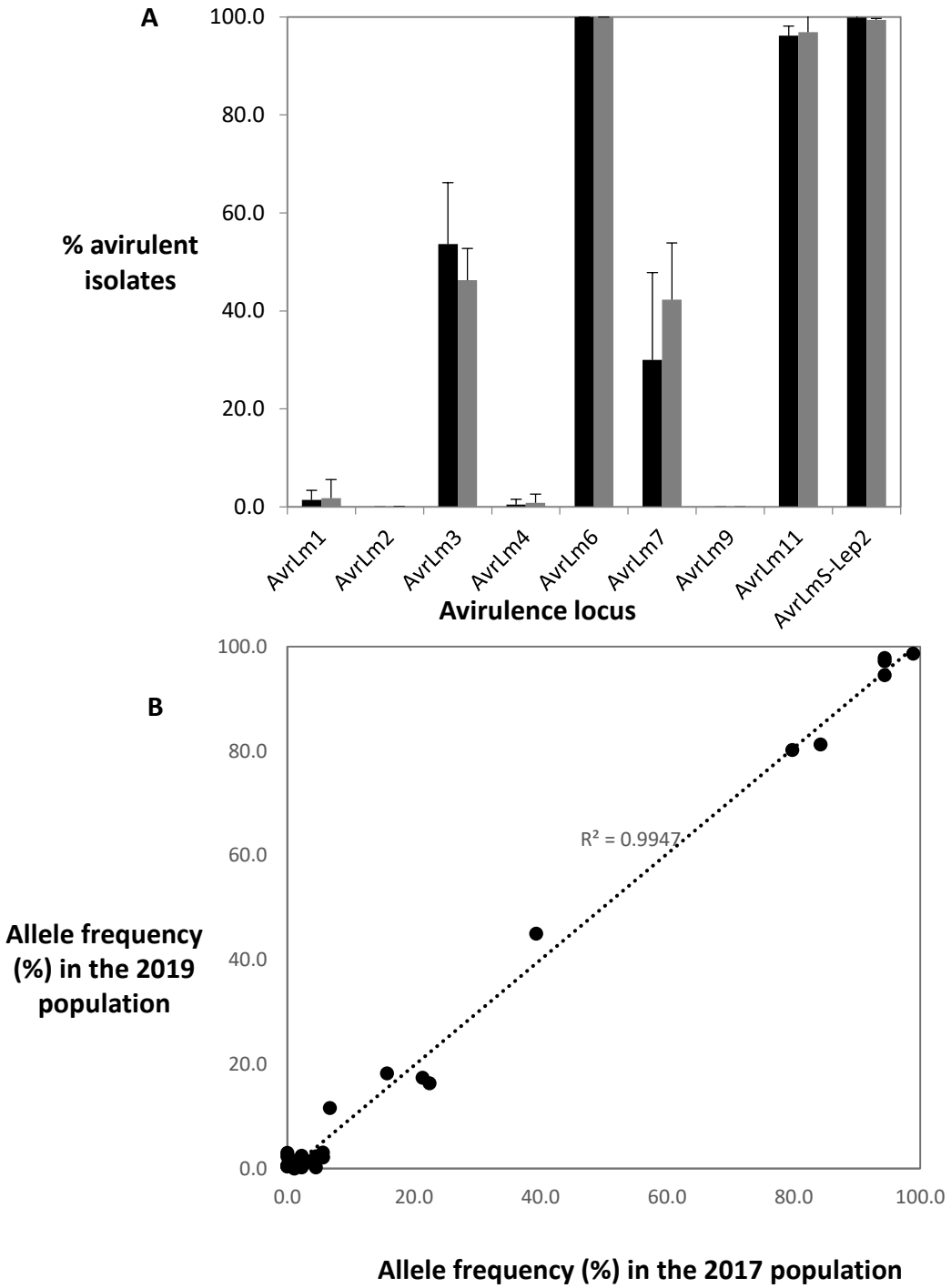

Figure 6. Validation of the MPSeqM tool on environmental samples. **A**, Comparison between percentages of avirulent isolates in French populations of *Leptosphaeria maculans,* estimated by phenotyping or genotyping. Black, data from phenotyping 690 isolates from 9 sites (Balesdent et al, 2023); grey, data from MPSeqM tool for 42 pools of leaf spots from the same sites (including a total of 1337 leaf spots). Values are means and standard errors over all sites. **B**, Comparison between allele frequencies (all genes) estimated by Sanger sequencing of individual isolates (2017 population, Gautier et al, 2013) and by the MPSeqM tool of pools of leaf spots (2019 population).