



**HAL**  
open science

# Harnessing the power of next-generation sequencing technologies to the purpose of high-throughput pesticide resistance diagnosis

Christophe Délye, Séverine Michel, Fanny Pernin, Véronique Gautier, Marie Gislard, Charles Poncet, Valérie Le Corre

## ► To cite this version:

Christophe Délye, Séverine Michel, Fanny Pernin, Véronique Gautier, Marie Gislard, et al.. Harnessing the power of next-generation sequencing technologies to the purpose of high-throughput pesticide resistance diagnosis. *Pest Management Science*, 2020, 76 (2), pp.543-552. 10.1002/ps.5543. hal-02553796

**HAL Id: hal-02553796**

**<https://hal.inrae.fr/hal-02553796>**

Submitted on 17 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Harnessing the power of Next-Generation Sequencing technologies to the purpose of high-throughput pesticide resistance diagnosis**

Journal:	<i>Pest Management Science</i>
Manuscript ID	PM-19-0428.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	28-Jun-2019
Complete List of Authors:	Délye, Christophe; INRA, UMR1347 Agroécologie Michel, Séverine; INRA, UMR Agroécologie Pernin, Fanny; INRA, UMR Agroécologie Gautier, Véronique; INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales Gislard, Marie; INRA, US 1426 GeT-PlaGe Poncet, Charles; INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales Le Corre, Valérie; INRA, UMR Agroécologie
Key Words:	pesticide resistance, detection, next generation sequencing, Illumina, ALS (acetolactate-synthase), AHAS (acetohydroxyacid-synthase), high-throughput, massive screening

SCHOLARONE™  
Manuscripts

1  
2  
3 1 **Harnessing the power of Next-Generation Sequencing technologies to the purpose of high-**  
4 **throughput pesticide resistance diagnosis.**  
5  
6 2  
7  
8 3  
9

10 4 Running title: Illumina-based pesticide resistance diagnosis  
11  
12 5

13  
14 6 Christophe Délye, <sup>a\*</sup> Séverine Michel, <sup>a</sup> Fanny Pernin, <sup>a</sup> Véronique Gautier, <sup>b</sup> Marie Gislard, <sup>c</sup>

15  
16 7 Charles Poncet <sup>b</sup> and Valérie Le Corre <sup>a</sup>  
17  
18  
19 8

20  
21 9

22  
23 10

24  
25 11

26  
27 12

28  
29 13

30  
31 14

32  
33 14 <sup>a</sup> INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France.  
34

35 15 <sup>b</sup> INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales, Clermont-Ferrand,  
36 France.  
37 16

38  
39 17 <sup>c</sup> INRA, US 1426 GeT-PlaGe, Genotoul, Castanet-Tolosan, France  
40  
41 18

42  
43  
44 19 \* Correspondence to:

45  
46 20 Dr Christophe Délye

47  
48 21 INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France

49  
50 22 Phone, +33 380 693 185

51  
52 23 Fax, +33 380 693 262

53  
54 24 E-mail, christophe.delye@inra.fr  
55  
56  
57  
58  
59  
60

1  
2  
3 **Abstract**  
4

5  
6 BACKGROUND: Next Generation Sequencing (NGS) technologies offer tremendous  
7  
8 possibilities for high-throughput pesticide resistance diagnosis *via* massive genotyping-by-  
9  
10 sequencing. Herein, we used Illumina sequencing combined with a simple, non-commercial  
11  
12 bioinformatics pipe-line to seek mutations involved in herbicide resistance in two weeds.  
13

14  
15 RESULTS: DNA was extracted from 96 pools of 50 plants for each species. Three amplicons  
16  
17 encompassing 15 ALS (acetolactate-synthase) codons crucial for herbicide resistance were  
18  
19 amplified from each DNA extract. Above 18 and 20 million quality 250-nucleotide sequence  
20  
21 reads were obtained for groundsel (*Senecio vulgaris*, tetraploid) and ragweed (*Ambrosia*  
22  
23 *artemisiifolia*, diploid), respectively. Herbicide resistance-endowing mutations were identified  
24  
25 in 45 groundsel and in eight ragweed field populations. The mutations detected and their  
26  
27 frequencies assessed by NGS were checked by individual plant genotyping or Sanger  
28  
29 sequencing. NGS results were fully confirmed, except in three instances out of 12 where  
30  
31 mutations present at a frequency of 1% were detected below the threshold set for reliable  
32  
33 mutation detection.  
34  
35

36  
37 CONCLUSION: Analysing 9,600 plants requested 192 DNA extractions followed by 1,728  
38  
39 PCRs and two Illumina runs. Equivalent results obtained by individual analysis would have  
40  
41 necessitated 9,600 individual DNA extractions followed by 216,000 genotyping PCRs, or by  
42  
43 121,500 PCRs and 40,500 Sanger sequence runs. This clearly demonstrates the interest and  
44  
45 power of NGS-based detection of pesticide resistance from pools of individuals for diagnosing  
46  
47 resistance in massive numbers of individuals.  
48  
49

50  
51  
52  
53 **Key-words.** Acetolactate-synthase (ALS), diagnosis, genotyping-by-sequencing, herbicide,  
54  
55 Illumina, resistance.  
56  
57  
58  
59  
60

## 1 INTRODUCTION

Chemical pesticides (herbicides, insecticides, fungicides) are currently key tools for efficient pest management and preservation of the global food security.<sup>1,2</sup> However, the reckless use of chemical pest control has promoted the evolution of pesticide resistance in numerous weeds, pests and plant pathogens, thereby jeopardising efficient pest control. Achieving accurate resistance detection, especially at early stages in the onset of resistance evolution, is crucial to adapt pest management practices and sustain pesticide efficiency. A variety of methods can be used for this purpose.<sup>3</sup> Among those, molecular assays targeting mutations at the root of resistance mechanisms show the highest potential for rapid routine detection of resistance in pest populations. This potential has been tremendously increased with the advent of the Next Generation Sequencing (NGS) technologies.<sup>4,5</sup> NGS generates huge amounts of sequence data that enable the identification of polymorphisms across whole genomes as well as the simultaneous detection of many mutations in a massive number of samples. As mutation detection is achieved by sequencing whole gene regions of interest, NGS-based search for polymorphisms can also reveal new mutations of potential relevance. NGS potentialities are already exploited in the medical field for the detection *via* genotyping-by-sequencing of genetic markers for human diseases (e.g.,<sup>5</sup>) or of mutations involved in the resistance of pathogens to drugs used in medical treatments (e.g.,<sup>6,7</sup>). Although this type of application of NGS has also been proposed for pesticide resistance diagnosis in agricultural pests,<sup>3</sup> the use of NGS techniques for this purpose is still in its infancy. To the best of our knowledge, only three studies have applied NGS to the purpose of pesticide resistance detection so far. The first one described the detection of mutations at seven codons of one gene involved in herbicide resistance in two polyploid *Echinochloa* species using the now obsolete 454 sequencing technology.<sup>8</sup> A total of 1,120 individuals plants were screened as 28 bulks, each corresponding to one field populations. 454 reads corresponding to each population were identified by specific short sequence tags.

1  
2  
3 74 The c.a. 64,000 sequence reads obtained were analysed through a very basic bioinformatic pipe-  
4  
5 75 line. The sequencing depth obtained was moderate ( $1.7\times$  to  $6.2\times$  per individual plant and per  
6  
7 76 codon of interest), but the mutation frequencies identified by NGS matched those obtained by  
8  
9 77 Sanger sequencing of individual plants, thereby indicating that NGS can be used to detect and  
10  
11 78 quantify mutations endowing pesticide resistance in pools of individuals. The second study  
12  
13 79 applied the Illumina technology to the detection of mutations at one codon endowing fungicide  
14  
15 80 resistance in *Zymoseptoria tritici*.<sup>9</sup> About 723,000 quality sequence reads were obtained and  
16  
17 81 used to analyse 40 fungal isolates from various geographical origins. Despite the authors  
18  
19 82 indicating that sequence reads could have been attributed to individual isolates by using  
20  
21 83 appropriate tagging, the 40 isolates were analysed as one single bulk. This work illustrated the  
22  
23 84 potential of the Illumina technology for pesticide resistance diagnosis. The third study sought  
24  
25 85 mutations at five codons in one gene involved in insecticide resistance in *Rhizopertha*  
26  
27 86 *dominica*.<sup>10</sup> A total of 1,435 individual insects were analysed using Illumina sequencing. Using  
28  
29 87 tagging, the unspecified number of quality sequence reads obtained could be attributed to each  
30  
31 88 individual insect, thus allowing individual genotype determination at every codon screened.  
32  
33 89 This study further confirmed the feasibility of multiple individual analysis in one single NGS  
34  
35 90 run.  
36  
37  
38  
39  
40  
41

42 91 The three preceding pioneer works clearly demonstrated the interest and feasibility of  
43  
44 92 pesticide resistance diagnosis by genotyping-by-sequencing using the NGS technologies  
45  
46 93 combined with sequence read tagging that allows to sequence hosts of individuals or  
47  
48 94 populations as one bulk and subsequently trace the individual or population at the source of  
49  
50 95 each sequence read. The pending question is, how can this approach be most efficiently used  
51  
52 96 for pesticide resistance diagnosis? Assuming that mutation(s) endowing pesticide resistance  
53  
54 97 have been characterised beforehand, an ideal NGS-based resistance assay would allow to screen  
55  
56 98 as many individuals of the pest considered as possible in one single NGS run. This can be  
57  
58  
59  
60

1  
2  
3 99 achieved by screening populations rather than individuals, and by pooling and simultaneously  
4  
5 100 sequencing populations in one single NGS run. An ideal NGS-based assay would also be  
6  
7 101 reliable and yield a fair estimate of the frequency of resistance-endowing mutations in pest  
8  
9 102 populations. In particular, the assay must not generate false positives or false negatives. Here,  
10  
11 103 the error rate of the NGS technique will determine the lowest mutation frequency that can be  
12  
13 104 reliably detected. Last, an ideal NGS-based assay should not require extensive skills in  
14  
15 105 bioinformatics nor depend on commercial softwares for the handling and analysis of the  
16  
17 106 sequence runs, because while NGS can easily be subcontracted, downstream analysis of the  
18  
19 107 huge amount of sequence data generated often remains a major stumbling block for people  
20  
21 108 dealing with pesticide resistance.  
22  
23  
24  
25

26 109 The purpose of this work was to develop massive resistance diagnosis assays associating  
27  
28 110 the Illumina NGS technology and a simple, robust and versatile pipe-line for the analysis of the  
29  
30 111 sequence reads. We considered acetolactate-synthase (ALS) inhibitors, the second most  
31  
32 112 important herbicide mode of action globally<sup>11</sup> and the one that selected for resistance in the  
33  
34 113 highest number of weed species (160 so far).<sup>12</sup> We developed resistance diagnosis assays for  
35  
36 114 two contrasted weed species where resistance due to mutations in the gene encoding ALS  
37  
38 115 inhibitors had evolved: common groundsel (*Senecio vulgaris*), an allotetraploid species with  
39  
40 116 limited genetic variation,<sup>13</sup> and common ragweed (*Ambrosia artemisiifolia*), a diploid invasive  
41  
42 117 species originating from North America that shows a much higher genetic polymorphism.<sup>14</sup>  
43  
44  
45  
46  
47  
48

## 49 119 **2 MATERIALS AND METHODS**

50  
51 120  
52  
53 121 The whole NGS-based resistance diagnosis procedure developed in this work is summarised in  
54  
55 122 figure 1.  
56  
57  
58  
59  
60

## 124 **2.1 Plant material**

125 For each of the two species studied, plant material consisted of 96 populations of 50 plants each.  
126 Common groundsel can infest all types of crops. It has evolved resistance to ALS inhibitors in  
127 France in both field crops and vineyards.<sup>15</sup> Accordingly, populations were collected in 81 wheat  
128 fields in the Brittany region in 2016 and in 14 vineyards in the Rhône-Alpes region in 2015.  
129 The 96<sup>th</sup> population was a control consisting of 50 mutant and non-mutant plants which ALS  
130 had been previously sequenced.<sup>15</sup> Plants were selected so that the control contained two mutant  
131 ALS alleles at a frequency of 2% each (Table 1).

132 In France, common ragweed particularly infests sunflower, soybean and maize. ALS  
133 inhibitors are pivotal for the chemical control of ragweed in these crops, but unsatisfactory  
134 control was recently reported in several regions. Populations were collected in two regions  
135 faced with very high levels of ragweed infestation: Occitanie in 2016 and 2017 (67 populations)  
136 and Nouvelle-Aquitaine in 2017 (20 populations). The remaining nine populations analysed  
137 were controls. Each consisted of 50 mutant and non-mutant plants which ALS had been  
138 sequenced beforehand, chosen so that the controls contained known frequencies of different  
139 mutations at the ALS gene. Plants carrying an Ala-205-Thr substitution originated from our  
140 resistance monitoring in France. Seeds from North American ragweed plants carrying a Trp-  
141 574-Leu mutation were kindly provided by Dr Jeff Stachler (North Dakota State University).  
142 After ALS sequencing, two additional mutations (Asp-376-Glu and Gly-654-Asp) were found  
143 in some of the plants obtained from these seeds (Table 2).

144 All the fields sampled had been sprayed with ALS inhibitors during the years preceding  
145 our sampling and were selected on the basis of the recurrent presence of high numbers of  
146 individuals from the species of interest. Each field was geolocated. Within each field, one leaf  
147 was collected on 50 plants of the species of interest scattered all over the infested area. After

1  
2  
3 148 collection, leaves were wrapped in paper towels and mailed to our lab where they were stored  
4  
5 149 at -20°C.  
6  
7

8 150

## 151 **2.2 DNA extraction and PCR amplification of ALS fragments**

152 One disc was punched out of the 50 leaves in each groundsel or ragweed population using a 2-  
153 mm diameter hollow punch. The 50 discs were placed together in one 2 mL microtube  
154 containing one steel bead (3 mm diameter). Tubes were closed, frozen in liquid nitrogen during  
155 at least one minute and placed in a bead mill homogeniser (TissueLyser II, Qiagen) at 1,800  
156 oscillations.min<sup>-1</sup> for 1 min 30 s. 400 µL extraction buffer<sup>16</sup> were added to each tube, and leaf  
157 powder was suspended in the buffer by 30 sec vortexing. Tubes were immediately incubated in  
158 a water bath 5 min at 95°C, cooled on ice and centrifugated 2 min at 20,000 g. The resulting  
159 rough DNA extracts were stored at -20°C.

160 Dilutions (1/10) of DNA extracts were directly used for PCR. PCR mixes were as  
161 described.<sup>16</sup> The PCR programs consisted of 3 min at 95°C, followed by 37 cycles of 5 sec at  
162 95°C, 10 sec at 60°C and 30 sec at 72°C. Currently, eight ALS codons are known to be involved  
163 in herbicide resistance selected for in the field (Ala122, Pro197, Ala205, Asp376, Arg377,  
164 Trp574, Ala653 and Gly654, standardised to *Arabidopsis thaliana* ALS sequence).<sup>17</sup> Seven  
165 additional codons have been shown to be involved in changes in herbicide sensitivity in  
166 artificial selection experiments (Gly121, Met124, Val196, Arg199, Asp375, Val571 and  
167 Phe578).<sup>18</sup> For each species investigated, three pair of PCR primers were designed to generate  
168 amplicons encompassing all these positions (Table 3).

169 An allotetraploid, groundsel contains two homeolog ALS genes (ALS1 and ALS2) only  
170 differing by 19 nucleotide substitutions.<sup>15</sup> Primers used to amplify groundsel DNA matched  
171 both ALS1 and ALS2 sequences, but the amplicons expected contained four (amplicon 1), one  
172 (amplicon 2) and five (amplicon 3) homeolog-specific nucleotides. All primers were used at

1  
2  
3 173 0.2  $\mu$ M final concentration each. As a classical, non-proofreading *Taq* polymerase was used for  
4  
5 174 PCR, three independent PCR reactions were performed and subsequently pooled for each  
6  
7  
8 175 amplicon and each population.

9  
10 176 For groundsel analyses, overhang adapter sequences were included at the 5' end of the  
11  
12 177 three F primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and of the three  
13  
14 178 R primers (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-). Tailed primers  
15  
16  
17 179 were used in a first PCR as described above. For each population, all three amplicons were  
18  
19 180 pooled in an equimolar mix and tagged with a population-specific index using Nextera XT  
20  
21 181 Index kit (v2) (Illumina). Tagged amplicons were pooled in an equimolar mix, purified with an  
22  
23  
24 182 Agencourt AMPure XP kit (Beckman Coulter, Beverly, MA, USA), quantified using the Qubit  
25  
26 183 HS kit (Invitrogen) and loaded onto one Illumina MiSeq V2 cartridge according to the  
27  
28 184 manufacturer instructions. The quality of the run was checked internally using PhiX. The  
29  
30 185 resulting 250 nucleotide pair-end sequences passing Illumina standard quality controls were  
31  
32  
33 186 assigned to their population of origin on the basis of the population-specific indexes using  
34  
35 187 bcl2fastq v2.20.0.422 (Illumina).

36  
37 188 For ragweed analyses, overhang adapter sequences were also included at the 5' end of  
38  
39 189 the three F primers (5'-CTTCCCTACACGACGCTCTTCCGATC-) and of the three R  
40  
41 190 primers (5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-). Tailed primers were used in a  
42  
43  
44 191 first PCR as described above. For each population, all three amplicons were pooled in an  
45  
46 192 equimolar mix and tagged with a population-specific, home-made six-bp index in a second, 12-  
47  
48  
49 193 cycle PCR using primers 5'-  
50  
51 194 AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGAC and 5'-  
52  
53 195 CAAGCAGAAGACGGCATAACGAGAT-XXXXXX-GTGACTGGAGTTCAGACGTGT  
54  
55  
56 196 where XXXXXX is the population-specific index sequence. Tagged amplicons were purified  
57  
58 197 and loaded onto one Illumina MiSeq V3 cartridge according to the manufacturer instructions.  
59  
60

1  
2  
3 198 The quality of the run was checked internally using PhiX. The resulting 250 nucleotide pair-  
4  
5 199 end sequences passing Illumina standard quality controls were assigned to their population of  
6  
7  
8 200 origin as described above.  
9

10 201  
11  
12 202 **2.3 Analysis of sequence reads and identification of mutations**  
13  
14 203 Illumina sequence data were analysed using “MutSeeker”, a custom function written in the R  
15  
16 204 programming language and implementing packages available in the Bioconductor project  
17  
18 205 (bioconductor.org). This function, together with the R script for operating it, detailed operating  
19  
20 206 instructions and one training data set for each of the two species studied here, are available from  
21  
22  
23 207 the [Dataverse](https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/BCZF3S) repository at:  
24  
25  
26 208 <https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/BCZF3S>  
27

28 209 Briefly, the 250-nucleotide paired reads were not joined. Reads were aligned to  
29  
30 210 reference ALS sequences (GenBank/EMBL accession KR024410 and KR024411 for groundsel  
31  
32 211 ALS1 and ALS2 partial sequences, respectively, or KX870184 for ragweed full ALS sequence)  
33  
34 212 using the short read aligner Bowtie as implemented in the package QuasR.<sup>19</sup> SNP calling was  
35  
36 213 performed using the pileup query in the Rsamtools package.<sup>20</sup> All codons targeted were covered  
37  
38 214 by either the forward or the reverse read in the 250-nucleotide read pairs matching the  
39  
40 215 corresponding amplicon, except codons Asp375, Asp376 and Arg377 that were covered by both  
41  
42 216 reads. For these three codons, SNPs calls obtained from forward and reverse reads were pooled.  
43  
44 217 In a final step, variant counts were calculated at every nucleotide position where mutations  
45  
46 218 would cause amino-acid changes in the 15 ALS codons implicated in herbicide resistance  
47  
48 219 (Table 3). Thus, analysis yielded the frequencies of single-nucleotide substitutions in every pool  
49  
50 220 of 50 plants analysed.  
51  
52  
53  
54

55 221 The threshold for the detection of mutations was set to 0.9% considering the minimum  
56  
57 222 expected frequency of mutations in a pool of 50 plants of 1% (i.e., one mutation on one ALS  
58  
59  
60

1  
2  
3 223 homeolog copy in 50 plants carrying two copies of said homeolog) and the current error rate of  
4  
5 224 the Illumina sequencing technology (0.24% per nucleotide).<sup>21</sup>  
6  
7  
8 225

9  
10 226 **2.4 Checking NGS-based mutation detection using dCAPS genotyping or Sanger**  
11  
12 227 **sequencing of individual plants**  
13

14 228 Our Illumina-based resistance diagnosis assay was designed to assess frequencies of mutant  
15  
16 229 ALS alleles in each groundsel or ragweed population analysed as a bulk of 50 plants. Therefore,  
17  
18 230 sequence data was not recovered for individual plants. Including the control populations, 34  
19  
20 231 groundsel populations and 25 ragweed populations containing frequencies of mutations as  
21  
22 232 contrasted as possible were used to check the reliability of assessing frequencies of mutant ALS  
23  
24 233 alleles using Illumina sequencing (Tables 1, 2). DNA was extracted individually from each of  
25  
26 234 the 50 plants in each of the populations used to check NGS results, except the control  
27  
28 235 populations that had been characterised beforehand.  
29  
30  
31

32  
33 236 In groundsel, mutations conferring resistance to ALS inhibitors were previously  
34  
35 237 exclusively found at codon 197 in ALS1 and/or ALS2.<sup>15</sup> Individual plant genotypes at ALS1  
36  
37 238 and ALS2 codon 197 were determined using previously described homeolog-specific dCAPS  
38  
39 239 assays allowing detection of mutations at codon 197.<sup>15</sup> This allowed to compute the frequencies  
40  
41 240 of mutations at each homeolog within each population. These frequencies were compared to  
42  
43 241 those obtained by NGS analysis of the corresponding bulks of 50 plants.  
44  
45

46  
47 242 In ragweed, 16 populations where mutant ALS alleles had or had not been detected were  
48  
49 243 selected for ALS Sanger sequencing in addition to the nine control populations that had been  
50  
51 244 characterised beforehand. Each of the three amplicons analysed by NGS were generated from  
52  
53 245 each individual plant in three independent PCRs using primers in Table 3 and subsequently  
54  
55 246 sequenced on both strands using Sanger sequencing.  
56  
57  
58  
59  
60

1  
2  
3 247 Individual plant genotypes at ALS obtained by genotyping or sequencing were used to  
4  
5 248 compute the exact frequencies of mutant ALS alleles within each population. These frequencies  
6  
7  
8 249 were compared to those estimated by NGS analysis of the corresponding bulks of 50 plants.  
9

10 250

### 11 251 **3 RESULTS**

12 252

#### 13 253 **3.1 NGS-based detection and quantification of the frequency of mutations in groundsel**

##### 14 254 **ALS**

15 255 The groundsel MiSeq run generated 18,084,459 250-nucleotide quality reads that could be  
16 256 assigned to one of the 96 populations investigated. The number of quality reads assigned to one  
17 257 population ranged from 35,366 to 498,841, with an average value of 188,380. Each population  
18 258 was analysed as one pool of 50 individual plants, each carrying two copies of two ALS  
19 259 homeologs. Three categories of amplicons were generated per ALS homeolog (amplicons 1, 2  
20 260 and 3, Table 1). As each plant carried two copies of each of two ALS homeologs, one plant  
21 261 yielded four amplicons in each category, i.e., 12 amplicons. Thus, the 50 plants in one pool  
22 262 yielded 600 amplicons. The average expected sequencing depth was therefore of 314 $\times$  for each  
23 263 of the 600 amplicons per pool.

24 264 The pipe-line assigned the reads to ALS1 or ALS2 on the basis of the homeolog-specific  
25 265 nucleotide(s) present in all amplicons. The lowest number of reads covering one codon in one  
26 266 population was 735 (ALS1 codons 571 to 654 in population BZ-67, Table 1), which allowed  
27 267 an expected sequencing depth of 7.3 $\times$  for each of these three codons in each individual plant in  
28 268 the pool. This was deemed sufficient for mutation detection.

29 269 With the 0.9% threshold for mutation detection, non-synonymous nucleotide  
30 270 substitutions were identified at two codons crucial for herbicide sensitivity. Substitutions were  
31 271 identified at codon 197 in the control population and in 45 of the 95 field populations

1  
2  
3 272 investigated. In total, eight substitutions occurring exclusively at the first two nucleotides in  
4  
5 273 codon 197 in ALS1 (three substitutions) and/or ALS2 (five substitutions) were identified, with  
6  
7  
8 274 up to four different substitutions detected in one same population (Table 4). The frequencies of  
9  
10 275 the substitutions assessed by the pipe-line ranged from 1.0% to 99.8% (ALS1) and from 0.9%  
11  
12 276 to 99.7% (ALS2). In six populations, C-to-A transversions at the first and at the second  
13  
14 277 nucleotide in ALS2 codon 197 were always detected in matching frequencies, suggesting the  
15  
16 278 occurrence of one double-substitution at codon 197 in ALS2 that would cause a proline-to-  
17  
18 279 asparagine amino-acid substitution. This was confirmed by Sanger sequencing of the individual  
19  
20 280 plants in these populations (Tables 1, 4). Thus, a total of eight mutations causing amino-acid  
21  
22 281 substitutions at ALS codon 197 were identified in varying frequencies in the groundsel  
23  
24 282 populations investigated. A Pro-197-Leu substitution at ALS1 detected in 35 populations was  
25  
26 283 by far the most frequent mutation in the populations analysed.

27  
28  
29  
30 284 A T-to-G transversion at the third nucleotide in codon 375 causing an aspartate-to-  
31  
32 285 glutamate amino-acid substitution was identified in both ALS1 and ALS2 in 89 of the 96  
33  
34 286 populations analysed, in frequencies systematically comprised between 0.9 and 1.5%. Sanger  
35  
36 287 sequencing in six populations where this substitution was detected by NGS did not allow to  
37  
38 288 confirm its existence. The putative Asp-375-Glu substitution was therefore considered a PCR  
39  
40 289 or NGS artifact, and discarded from the analyses.

41  
42  
43  
44  
45 290

### 291 **3.2 NGS-based detection and quantification of the frequency of mutations in ragweed ALS**

42  
43  
44  
45  
46  
47 292 The ragweed MiSeq run generated 20,700,536 250-nucleotide quality reads that could be  
48  
49 293 assigned to one of the 96 populations investigated. The number of quality reads assigned to one  
50  
51 294 population ranged from 147,116 to 274,998, with an average value of 215,631. Each population  
52  
53 295 was analysed as one pool of 50 individual plants, each carrying two ALS gene copies. Thus,  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 296 the 50 plants in one pool yielded 300 amplicons. The average expected sequencing depth was  
4  
5 297 therefore of 719× for each of the 300 amplicons per pool.  
6

7  
8 298 The lowest number of reads covering one codon in one population was 15,253 (codons  
9  
10 299 571 to 654, not shown). This allowed an expected sequencing depth of 152× for each of these  
11  
12 300 five codons in each individual plant in the corresponding pool, which was more than sufficient  
13  
14 301 for mutation detection.  
15

16  
17 302 With the 0.9% threshold for mutation detection, non-synonymous nucleotide  
18  
19 303 substitutions in codons crucial for herbicide sensitivity were identified as expected in the  
20  
21 304 control populations. The only exception was control population C, where a G-to-T transversion  
22  
23 305 at the second nucleotide in codon 574 causing a tryptophan-to-leucine amino-acid substitution  
24  
25 306 that was present with a frequency of 1% was detected with a frequency of 0.63%, i.e., below  
26  
27 307 the 0.9% threshold (Table 2). No mutations were detected in 79 of the 87 field populations  
28  
29 308 investigated. In the remaining eight populations (all from Occitanie), a G-to-A transition at the  
30  
31 309 first nucleotide in codon 205 causing an alanine-to-threonine amino-acid substitution was  
32  
33 310 detected in frequencies ranging from 0.9 to 11.5% (Table 2). Mapping the eight populations  
34  
35 311 from Occitanie investigated herein revealed that six populations were clustered within a radius  
36  
37 312 of roughly one km within one river valley. The other two populations could be “escapees”  
38  
39 313 disseminated by, e.g., agricultural machinery, to different river valleys 2.5 km south and 7.5  
40  
41 314 north from the main resistance focus, respectively (Figure 2).  
42  
43  
44  
45  
46  
47  
48

### 316 **3.3 Checking NGS results using individual plant analysis**

49  
50  
51 317 The frequencies of the mutations detected by NGS-based analysis of populations from both  
52  
53 318 species as pools of 50 plants were checked by selecting pools with contrasted frequencies of  
54  
55 319 mutations and analysing each plant in each of these pool individually using a double-blind  
56  
57 320 procedure. In groundsel, the frequencies of the mutations detected at each ALS homeolog in 33  
58  
59  
60

1  
2  
3 321 field populations were checked by genotyping each plant in each population individually using  
4  
5 322 ALS1- and ALS2-specific dCAPS assays targeting codon 197. dCAPS assays did not  
6  
7 323 discriminate among different mutations occurring at one same codon.<sup>15</sup> Thus, the frequencies  
8  
9 324 estimated by NGS of all mutations identified at codon 197 in each homeolog were cumulated  
10  
11 325 for comparison to the dCAPS results. Similarly, in ragweed, the frequencies of mutations  
12  
13 326 detected in 16 field populations by NGS analysis were checked by Sanger sequencing of the  
14  
15 327 full ALS sequence of each plant in each pool. The frequencies of mutations identified by  
16  
17 328 Illumina sequencing of plant pools and by individual plant analysis in field populations as well  
18  
19 329 as in the control populations were subsequently compared (Tables 1 and 2, Figure 3). For the  
20  
21 330 34 groundsel populations (33 field populations and one control population) and the for the 25  
22  
23 331 ragweed populations (16 field populations and nine control populations) considered, the  
24  
25 332 frequency of each mutation detected by NGS analysis of one bulk of 50 plants was very strongly  
26  
27 333 correlated to the frequency determined by individual plant analysis (Figure 3; Pearson's  
28  
29 334 correlation coefficients values of 0.996 [P-value < 10<sup>-6</sup>] and 0.979 [P-value < 10<sup>-6</sup>] for groundsel  
30  
31 335 and ragweed data, respectively). No false positives were identified, i.e., detection of mutation(s)  
32  
33 336 in one population by NGS analysis of the corresponding plant pool was always confirmed by  
34  
35 337 individual plant sequencing or genotyping. However, NGS-based analysis did not detect  
36  
37 338 mutations in two groundsel populations and one ragweed population where mutations were  
38  
39 339 actually present with a frequency of 1% (i.e., false negatives; Tables 1 and 2). In all three cases,  
40  
41 340 the mutations in question had been detected by NGS analysis, but their estimated frequencies  
42  
43 341 ranged from 0.2 to 0.7% and were therefore below the threshold set for reliable mutation  
44  
45 342 detection (0.9%). On the other hand, mutations present with a frequency of 1% were  
46  
47 343 successfully detected by NGS analysis in two groundsel populations and seven ragweed  
48  
49 344 populations (Tables 1 and 2).  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 346 **4 DISCUSSION**

347 We successfully applied the Illumina NGS technology to the detection of mutations endowing  
348 resistance to herbicides. PCR amplification of three regions in the ALS gene that harbour the  
349 15 codons known to date to be crucial for herbicide sensitivity were screened in a total of 9,600  
350 plants from two species, the allotetraploid groundsel and the diploid ragweed.

### 352 **4.1 Reliability of NGS-based detection of mutations endowing pesticide resistance**

353 Analysis of the Illumina sequence reads obtained for the 192 pools of 50 plants each  
354 using the pipe-line described herein enabled the reliable detection of mutations in both species  
355 analysed. Although we used a non-proofreading *Taq* polymerase in our PCRs, no false positive  
356 sample was identified when checking NGS results by individual analysis of the plants in the  
357 pools. There was one case of detection of a non-existing nucleotide change that would have  
358 caused an Asp375Glu amino-acid in groundsel and was disproved by Sanger sequencing.  
359 However, this case was suspicious from the start. When detected, the putative substitution was  
360 systematically present in both ALS homeologs, with frequencies that were always lower than  
361 1.5%. Occurrence of this artifact illustrates the necessity to carefully check the first NGS results  
362 using previously characterised populations before embarking into massive sampling analysis,  
363 and to confirm any NGS-based detection of putative new mutations by other techniques (e.g.,  
364 Sanger sequencing).

365 In the case of the tetraploid groundsel, the mutations detected were correctly assigned  
366 to ALS1 or to ALS2 on the sole basis of a handful of homeolog-specific nucleotides, which  
367 demonstrates the feasibility of analysing several pesticide targets in pest populations in one  
368 single NGS run. NGS analysis also enabled to identify a double-mutation at codon 197 that  
369 caused a Pro-197-Asn amino-acid substitution. Previous reports of double-mutations in  
370 herbicide target sites are scarce. They include a similar Pro-197-Asn amino-acid substitution in

1  
2  
3 371 the grass *Apera spica-venti*,<sup>22</sup> a double mutation at ALS codon 122 causing a Ala-122-Tyr  
4  
5 372 substitution in *Raphanus raphanistrum*<sup>23</sup> and two combined mutations at codons 102 and 106  
6  
7  
8 373 at the 5-enolpyruvylshikimate-3-phosphate synthase gene that encodes another herbicide target  
9  
10 374 enzyme in the grass *Eleusine indica*.<sup>24</sup> Double-mutations in one pesticide gene have often been  
11  
12 375 reported to confer an increased resistance level and/or a broader resistance spectrum, especially  
13  
14 376 in fungicides (e.g.,<sup>25</sup>) but also in herbicides.<sup>24</sup> Herbicide sensitivity bioassays are needed to  
15  
16 377 establish the resistance pattern associated to the double mutation identified in groundsel.

17  
18  
19 378 NGS analysis correctly identified a total of four amino-acid substitutions in ragweed.  
20  
21 379 Three of them were detected in North American ragweed plants used to generate control  
22  
23 380 populations (Table 1). A Trp-574-Leu substitution had previously been identified in this  
24  
25 381 species.<sup>26</sup> An Asp-376-Glu substitution had been identified in 12 species so far,<sup>17</sup> while a Gly-  
26  
27 382 654-Asp substitution had only been reported in the grass weed *Setaria viridis* to date.<sup>27</sup> The last  
28  
29 383 substitution, Ala-205-Thr, was carried by French plants from the field and had never been  
30  
31 384 reported before. Herbicide sensitivity bioassays are underway to establish its associated  
32  
33 385 resistance pattern.  
34  
35  
36  
37  
38  
39

#### 40 387 **4.2 Reliability of NGS-based quantification of mutations endowing pesticide resistance**

41  
42 388 The frequencies of the ALS mutations assessed by NGS analysis of 50-plants pools matched  
43  
44 389 very closely those determined by analysing the corresponding individual plants in both species  
45  
46 390 investigated (Figure 3), even if some differences could be observed (Tables 1 and 2). The causes  
47  
48 391 at the root of these discrepancies may be differences in the amount of plant material collected  
49  
50 392 among individual plants, especially in groundsel where leaf thickness is rather variable, and/or  
51  
52 393 a heterogeneous efficacy of DNA extraction and/or PCR. An unequal contribution of all plants  
53  
54 394 in one pool to the mixture of amplicons generated for NGS is most likely the reason why NGS  
55  
56 395 analysis of plant pools failed to detect mutations present with a frequency of 1% in two  
57  
58  
59  
60

1  
2  
3 396 groundsel and one ragweed population (false negatives, Tables 1, 2). Indeed, these mutations  
4  
5 397 were actually detected by NGS, but with an estimated frequency that was below the 0.9%  
6  
7 398 threshold set for reliable mutation detection. This flaw could be overcome by using a  
8  
9  
10 399 proofreading *Taq* polymerase in the PCRs preceding NGS, which would allow lowering the  
11  
12 400 mutation detection threshold, and/or by slightly decreasing the number of plants per pool (e.g.,  
13  
14 401 40 plants per pool, which would yield a minimum expected mutation frequency of 1.25% on a  
15  
16  
17 402 given gene or homeolog in the pool). Considering the tremendous sequencing depth achieved  
18  
19 403 in this study for every ALS amplicon sequenced, analysis of two pools of 40 plants per  
20  
21 404 population investigated rather than a single pool of 50 plants could be performed for 96  
22  
23 405 populations in one single Illumina run with a codon coverage adequate for mutation detection.  
24  
25  
26 406 This way, reducing the number of plants per pool could largely be compensated for by analysing  
27  
28 407 more pools per population.  
29  
30  
31 408

### 32 33 409 **4.3 NGS vs. genotyping or Sanger sequencing**

34  
35 410 NGS-based analysis of plant pools detected and quantified resistance-endowing mutations in  
36  
37 411 9,600 individual plants clustered into 192 pools. This approach did not allow to determine the  
38  
39 412 genotypes of the individual plants constituting the pool. This could be considered a drawback  
40  
41 413 in the case of recessive resistances, where differentiating between homozygous and  
42  
43 414 heterozygous mutants may be of some relevance. However, for resistance management in the  
44  
45 415 field, the vital information is rather the detection of mutation(s) involved in pesticide resistance  
46  
47 416 in one pest population. Such mutations are an unambiguous indicator that resistance is arising,  
48  
49 417 and that steps are to be taken to delay its onset.  
50  
51  
52

53  
54 418 Herein, we targeted 15 codons in one ALS gene in 96 batches of 50 plants (ragweed)  
55  
56 419 and in two ALS homeologs in 96 other batches of 50 plants (groundsel), which surpasses all  
57  
58 420 preceding studies having applied NGS to pesticide resistance detection.<sup>8,9,10</sup> This merely  
59  
60

1  
2  
3 421 necessitated 192 DNA extractions followed by the production of three amplicons, each in  
4  
5 422 triplicate PCR reactions per extraction, i.e., 1,728 PCRs, and two Illumina MiSeq runs. (It  
6  
7 423 should be mentioned that, given the sequencing depth achieved, one single such run would have  
8  
9 424 been sufficient to analyse all 192 pools). The same study conducted by the conventional single-  
10  
11 425 codon PCR-based genotyping assays used for resistance detection (e.g.,<sup>15</sup>) would have required  
12  
13 426 9,600 DNA extractions. It would have necessitated, first, to set up and validate 15 codon-  
14  
15 427 specific genotyping assays for ragweed and 30 homeolog-and-codon-specific genotyping  
16  
17 428 assays for groundsel, then, to carry out and visualise 15 genotyping assays per ragweed plant  
18  
19 429 and 30 per groundsel plant, i.e., a total of 216,000 PCRs (followed by the same number of  
20  
21 430 restriction enzyme digestion reactions in the case of dCAPS). If Sanger sequencing had been  
22  
23 431 used instead of genotyping, this study would have necessitated the production of three  
24  
25 432 amplicons in triplicate PCR reactions in ragweed and six in groundsel (three per homeolog),  
26  
27 433 i.e., 121,500 PCRs, followed by 40,500 Sanger sequencing runs. This clearly illustrates the  
28  
29 434 power of NGS for diagnosing resistance-endowing mutations in numerous samples.

30  
31  
32  
33  
34  
35 435 In contrast to preceding NGS studies addressing pesticide resistance diagnosis,<sup>9,10</sup> our  
36  
37 436 work was carried out using a rapid DNA extraction procedure based on a non-commercial, lab-  
38  
39 437 made extraction buffer.<sup>16</sup> The rough DNA extracts obtained proved perfectly suitable for  
40  
41 438 resistance detection in pools of 50 individual plants. After Illumina sequencing, assignment of  
42  
43 439 the sequence reads to their originating sample was performed using the software attached to the  
44  
45 440 sequencer. This can be done in the case of an in-house facility, or requested from any  
46  
47 441 sequencing supplier. The downstream analyses (read mapping on reference sequences,  
48  
49 442 mutation detection and quantification) required no commercial software, our pipe-line being  
50  
51 443 exclusively based on freely available programs and applications.  
52  
53  
54  
55  
56  
57

#### 58 445 **4.4 Developments of the NGS-based method**

59  
60

1  
2  
3 446 As it is described here, our NGS-based method only targets mutations at specific codons in the  
4  
5 447 ALS gene. However, the scope of the method can be extended. First, any ALS nucleotide  
6  
7 448 present in the amplicons sequenced can very easily be added to the list of positions to be  
8  
9 449 screened by our pipe-line. This could reveal novel mutations of interest. Second, the pipe-line  
10  
11 450 developed herein can be applied to other weed or pest species. As the level of nucleotide  
12  
13 451 polymorphism varies with the species, this will require to adapt the alignment stringency to the  
14  
15 452 gene(s) and to the weed or pest targeted. This can be done by running the pipe-line on a set of  
16  
17 453 previously characterised populations by modifying options for the Bowtie short read aligner  
18  
19 454 (see MutSeeker operating instructions). Third, the sequencing depth achieved by the Illumina  
20  
21 455 technology clearly allows to sequence several genes in a single run. This has been shown herein  
22  
23 456 by our analysing the two ALS homeologs in groundsel. Accordingly, several distinct pesticide  
24  
25 457 targets can be screened in one single NGS run. Resistance can also be due to non-target-site  
26  
27 458 based mechanisms.<sup>28</sup> When the mutations involved in these mechanisms are known, they can  
28  
29 459 be subjected to NGS-based detection. Amplifying and sequencing by NGS as many amplicons  
30  
31 460 as necessary to cover the genome regions including the nucleotide positions of interest is easily  
32  
33 461 feasible, keeping in mind that the length of each amplicon must be lower than that of the NGS  
34  
35 462 reads (i.e., lower than 500 nucleotides for 250 nucleotide pair-end sequences). If many genome  
36  
37 463 regions of interest are targeted, then a compromise is to be found between the number of  
38  
39 464 amplicons to be sequenced per population and the number of population to be sequenced per  
40  
41 465 NGS run so that a satisfactory sequencing depth (e.g., 10× for each amplicon obtained from  
42  
43 466 each individual plant in each pool) is achieved.  
44  
45  
46  
47  
48  
49  
50

467

## 468 **5 CONCLUSIONS**

469 This work illustrates the power and throughput of the NGS technologies for the detection of  
470 mutations endowing pesticide resistance. The capacity of NGS-based resistance diagnosis

1  
2  
3 471 assays can be exploited for massive resistance surveys (i.e., analysing numerous pest or weed  
4  
5 472 populations) as well as for in-depth resistance detection (i.e., analysing numerous pest or weed  
6  
7 473 individuals in one population to detect to onset of resistance at very early stages, when  
8  
9 474 resistance management practices can still nip resistance evolution in the bud). With the  
10  
11 475 continuously-increasing amount of sequence reads generated by NGS runs and the technical  
12  
13 476 possibility to multiplex hundreds of pools of individuals in one single NGS run, the era of  
14  
15 477 massive resistance diagnosis is clearly just dawning. As shown in our work, the tremendous  
16  
17 478 amounts of sequence reads generated can be analysed without the need for extensive training  
18  
19 479 in NGS data analysis, which should extend the benefit of using NGS diagnosis to a large number  
20  
21 480 of scientists dealing with pest control. Actually, the throughput of assays such as the one  
22  
23 481 described herein is so high that the future main issue with NGS-based resistance detection might  
24  
25 482 well be the possibility to obtain massive numbers of samples from the field. This issue could  
26  
27 483 be overcome by pooling populations of different pests and weed species in shared NGS runs.  
28  
29  
30  
31  
32  
33  
34

### 35 485 **ACKNOWLEDGEMENTS**

36  
37 486 This study was partly funded by Bayer CropScience France (groundsel sampling), by ANSES  
38  
39 487 *via* the system for monitoring the adverse effects of plant protection products  
40  
41 488 ('phytopharmacovigilance', established by the French Act on the future of agriculture of 13  
42  
43 489 October 2014) (groundsel sequencing) and by BASF SAS France (ragweed sequencing). We  
44  
45 490 thank the Gentyane facility (Clermont-Ferrand, France [<http://gentyane.clermont.inra.fr/>]) and  
46  
47 491 the GeT core facility (Toulouse, France [<http://get.genotoul.fr/>]) for providing access to their  
48  
49 492 sequencing equipment.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

493 **REFERENCES**

- 494 1 Oerke E-C, Crop losses to pests. *J Agric Sci* **144**:31-43 (2006).
- 495 2 Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N and Nelson A, The  
496 global burden of pathogens and pests on major food crops. *Nature Ecol Evol* **3**:430-439  
497 (2019).
- 498 3 R4P Network, Trends and challenges in pesticide resistance detection. *Trends Plant Sci*  
499 **21**:834-853 (2016).
- 500 4 Metzker ML, Sequencing technologies - the next generation. *Nature Rev Genet* **11**:31-  
501 46 (2010).
- 502 5 Kamps R, Brandão RD, van den Bosch BJ, Paulussen ADC, Xanthoulea S, Blok MJ  
503 and Romano A, Next-generation sequencing in oncology: genetic diagnosis, risk  
504 prediction and cancer classification. *Int J Mol Sci* **18**:Article308 (2017).
- 505 6 Sahoo MK, Lefterova MI, Yamamoto F, Waggoner JJ, Chou S, Holmes SP, Anderson  
506 MW and Pinsky BA, Detection of cytomegalovirus drug resistance mutations by next-  
507 generation sequencing. *J Clin Microbiol* **51**:3700-3710 (2013).
- 508 7 Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, Crudu V,  
509 Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC, Catanzaro A, Keim  
510 P and Engelthaler DM, Detection of low-level mixed-population drug resistance in  
511 *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS ONE*  
512 **10**:e0126626 (2015).
- 513 8 Délye C, Causse R, Gautier V, Poncet C and Michel S, Using next-generation  
514 sequencing to detect mutations endowing resistance to pesticides: application to  
515 acetolactate-synthase (ALS)-based resistance in barnyardgrass, a polyploid grass weed.  
516 *Pest Manag Sci* **71**:675-685 (2015).

- 1  
2  
3 517 9 Pieczul K and Wařowska A, The application of next-generation sequencing (NGS) for  
4  
5 518 monitoring of *Zymoseptoria tritici* QoI resistance. *Crop Prot* **92**:143-147 (2017).  
6  
7  
8 519 10 Schlipalius DI, Tuck AG, Pavic H, Daghish GJ, Nayaka MK and Ebert PR, A high-  
9  
10 520 throughput system used to determine frequency and distribution of phosphine resistance  
11  
12 521 across large geographical regions. *Pest Manag Sci* **75**:1091-1098 (2019).  
13  
14  
15 522 11 Kraehmer H, Laber B, Rosinger C and Schulz A, Herbicides as weed control agents:  
16  
17 523 state of the art: I. weed control research and safener technology: the path to modern  
18  
19 524 agriculture. *Plant Physiol* **166**:1119-1131 (2014).  
20  
21  
22 525 12 Heap I, International survey of herbicide-resistant weeds. Available online at:  
23  
24 526 <http://www.weedscience.com>. Accessed 11 April 2019.  
25  
26 527 13 Zhu B-R, Barrett SCH, Zhang D-Y and Liao W-J, Invasion genetics of *Senecio vulgaris*:  
27  
28 528 loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple  
29  
30 529 introductions. *Biol Invasions* **19**:255-267 (2017).  
31  
32  
33 530 14 Meyer L, Causse R, Pernin F, Scalone R, Bailly G, Chauvel B, Délye C and Le Corre  
34  
35 531 V, New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant  
36  
37 532 *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species.  
38  
39 533 *PLoS ONE* **12**:e0176197 (2017).  
40  
41  
42 534 15 Délye C, Causse R, Gautier V, Poncet C and Michel S, Genetic basis, evolutionary  
43  
44 535 origin and spread of resistance to herbicides inhibiting acetolactate synthase in common  
45  
46 536 groundsel (*Senecio vulgaris*). *Pest Manag Sci* **72**:89-102 (2016).  
47  
48  
49 537 16 Délye C, Matějicek A and Gasquez J, PCR-based detection of resistance to Acetyl-CoA  
50  
51 538 carboxylase-inhibiting herbicides in black-grass (*Alopecurus myosuroides* Huds) and  
52  
53 539 ryegrass (*Lolium rigidum* Gaud). *Pest Manag Sci* **58**:474-478 (2002).  
54  
55  
56 540 17 Tranel PJ, Wright TR and Heap IM, Mutations in herbicide-resistant weeds to ALS  
57  
58 541 inhibitors. Available online at: <http://www.weedscience.com>. Accessed 11 April 2019.  
59  
60

- 1  
2  
3 542 18 Duggleby RG, McCourt JA and Guddat LW, Structure and mechanism of inhibition of  
4  
5 543 plant acetohydroxyacid synthase. *Plant Physiol Biochem* **46**:309-324 (2008).  
6  
7  
8 544 19 Gaidatzis DA, Lerch F, Hahne F and Stadler MB, QuasR: quantify and annotate short  
9  
10 545 reads in R. *Bioinformatics* **31**:1130-1132 (2015).  
11  
12 546 20 Morgan M, Pagès H, Obenchain V and Hayden N, Rsamtools: binary alignment (BAM),  
13  
14 547 FASTA, variant call (BCF), and tabix file import. R package version 1.34.1, Available  
15  
16 548 online at: <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>.  
17  
18 549 Accessed 11 April 2019.  
19  
20  
21 550 21 Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL and Mayer G,  
22  
23 551 Systematic evaluation of error rates and causes in short samples in next-generation  
24  
25 552 sequencing. *Sci Rep* **8**:10950 (2018).  
26  
27  
28 553 22 Massa D, Krenz B and Gerhards R, Target-site resistance to ALS-inhibiting herbicides  
29  
30 554 in *Apera spica-venti* populations is conferred by documented and previously unknown  
31  
32 555 mutations. *Weed Res* **51**:294-303 (2011).  
33  
34  
35 556 23 Han H, Yu Q, Purba E, Li M, Walsh M, Friesen S and Powles SB, A novel amino acid  
36  
37 557 substitution Ala-122-Tyr in ALS confers high-level and broad resistance across ALS-  
38  
39 558 inhibiting herbicides. *Pest Manag Sci* **68**:1164-1170 (2012).  
40  
41  
42 559 24 Han H, Vila-Aiub MM, Jalaludin A, Yu Q and Powles SB, A double EPSPS gene  
43  
44 560 mutation endowing glyphosate resistance shows a remarkably high resistance cost.  
45  
46 561 *Plant Cell Environ* **40**:3031-3042 (2017).  
47  
48  
49 562 25 Leroux P and Walker A-S, Multiple mechanisms account for resistance to sterol 14 $\alpha$ -  
50  
51 563 demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. *Pest Manag*  
52  
53 564 *Sci* **67**:44-59 (2011).  
54  
55  
56 565 26 Patzoldt WL, Tranel PJ, Alexander AL and Schmitzer PR, A common ragweed  
57  
58 566 population resistant to cloransulam-methyl. *Weed Sci* **49**:485-490 (2001).  
59  
60

- 1  
2  
3 567 27 Laplante J, Rajcan I, and Tardif FJ, Multiple allelic forms of acetohydroxyacid synthase  
4  
5 568 are responsible for herbicide resistance in *Setaria viridis*. *Theor Appl Genet* **119**:577-  
6  
7 569 585 (2009).  
8  
9  
10 570 28 Délye C, Duhoux A, Pernin F, Riggins CW and Tranel PJ, Molecular mechanisms of  
11  
12 571 herbicide resistance. *Weed Sci* **63**:91-115 (2015).  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

572 **Table 1.** Groundsel populations used for validation of Illumina-based detection of resistance using dCAPS genotyping.

Populations		ALS1					ALS2				
		% mutant detected		Nb. Kreads per codon <sup>c</sup>			% mutant detected		Nb. Kreads per codon <sup>c</sup>		
Code <sup>a</sup>	Type <sup>b</sup>	dCAPS	Illumina	121-205	375-377	571-654	dCAPS	Illumina	121-205	375-377	571-654
Control	C	2.0	2.0	14.0	17.7	21.0	2.0	2.2	12.9	16.2	24.1
RA-01	F	0.0	0.0	25.1	9.8	14.9	0.0	0.0	26.2	9.2	12.5
RA-02	F	0.0	0.0	22.3	5.1	23.1	24.0	27.9	23.8	5.3	21.1
RA-03 <sup>d</sup>	F	0.0	0.0	18.1	6.2	7.7	78.0	67.0	17.1	4.8	9.8
RA-05 <sup>d</sup>	F	2.0	2.9	7.5	17.1	31.1	98.0	92.6	8.3	18.2	32.4
RA-08	F	0.0	0.0	19.2	5.7	20.7	100.0	99.7	19.8	5.2	22.6
RA-10 <sup>d</sup>	F	7.0	1.5	9.6	10.4	24.5	92.0	93.6	8.3	9.5	22.2
RA-12 <sup>d</sup>	F	0.0	0.0	18.8	18.2	21.3	7.0	11.6	20.2	20.3	20.7
RA-13 <sup>d</sup>	F	0.0	0.0	9.5	18.4	17.8	50.0	43.2	12.2	19.1	18.3
RA-15 <sup>d</sup>	F	9.0	5.7	32.1	28.4	15.9	91.0	87.9	32.8	34.1	15.7
RA-18	F	0.0	0.0	24.6	36.1	12.3	0.0	0.0	25.9	36.3	18.0
RA-19	F	0.0	0.0	12.7	27.7	9.8	0.0	0.0	13.3	27.7	9.5

1												
2												
3	RA-20	F	0.0	0.0	17.9	22.9	20.9	0.0	0.0	14.8	20.6	17.9
4												
5	RA-22	F	0.0	0.0	22.7	7.1	13.3	0.0	0.0	17.7	7.5	11.2
6												
7	RA-24	F	0.0	0.0	27.8	17.2	14.3	0.0	0.0	27.3	17.0	14.9
8												
9	BZ-I	F	16.0	12.3	17.9	13.5	11.2	0.0	0.0	14.4	12.0	11.9
10												
11	BZ-IV	F	0.0	0.0	16.0	11.7	19.8	0.0	0.0	18.4	11.5	19.9
12												
13	BZ-VI	F	96.0	96.6	19.1	12.9	17.6	0.0	0.0	22.2	10.3	16.8
14												
15	BZ-VII	F	58.0	57.9	12.9	15.5	23.9	0.0	0.0	9.7	15.1	19.9
16												
17	BZ-VIII	F	100.0	99.8	29.0	35.7	22.1	1.0	0.9	26.2	38.8	23.4
18												
19	BZ-04	F	2.0	2.4	4.7	21.1	12.3	0.0	0.0	4.5	19.2	12.7
20												
21	BZ-05	F	1.0	1.0	10.7	10.7	14.0	0.0	0.0	13.2	8.4	13.5
22												
23	BZ-09	F	51.0	57.0	16.2	7.4	7.0	0.0	0.0	14.5	6.7	6.6
24												
25	BZ-13	F	5.0	1.0	8.9	2.4	4.5	0.0	0.0	10.0	2.6	4.7
26												
27	BZ-24	F	0.0	0.0	38.1	6.3	18.5	2.0	0.9	32.7	4.8	15.1
28												
29	BZ-38	F	2.0	1.0	12.1	22.6	21.5	1.0	0.0	11.0	20.2	21.7
30												
31	BZ-57	F	0.0	0.0	7.6	6.6	5.3	4.0	3.0	8.7	7.1	6.0
32												
33	BZ-59	F	91.0	88.4	11.1	9.9	7.2	7.0	2.8	14.5	10.0	8.7
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												

1												
2												
3	BZ-61	F	49.0	46.4	9.9	11.9	8.8	22.0	18.9	9.3	11.6	9.0
4												
5	BZ-66	F	5.0	5.7	18.0	32.1	10.1	1.0	0.0	15.4	31.2	11.3
6												
7	BZ-67	F	7.0	2.0	19.8	13.0	0.7	0.0	0.0	19.7	13.2	0.8
8												
9	BZ-69	F	4.0	2.3	17.3	12.2	7.7	0.0	0.0	21.0	12.4	7.5
10												
11	BZ-71	F	4.0	2.5	17.5	18.6	7.9	0.0	0.0	16.4	17.6	7.5
12												
13	BZ-73	F	20.0	31.3	5.5	6.8	12.4	5.0	9.1	11.0	6.3	10.6
14												
15												
16												

573 <sup>a</sup> RA, Rhône-Alpes; BZ, Brittany.

574 <sup>b</sup> C, control population (artificial population consisting of 50 mutant and non-mutant plants which ALS had been sequenced beforehand, chosen so  
575 that the controls contained known frequencies of ALS mutations); F, population collected in the field.

576 <sup>c</sup> Number of kilo-reads (i.e., thousands of reads) covering the codons in the interval mentioned (e.g., 121-205: ALS codons 121 to 205). Paired  
577 reads were not joined.

578 <sup>d</sup> Populations with plants containing a double mutation causing a Pro-197-Asn substitution at ALS2. These populations were also subjected to  
579 Sanger sequencing of individual plants.

580 **Table 2.** Ragweed populations used for validation of Illumina-based detection of resistance using Sanger sequencing.

Populations		ALS				
Code <sup>a</sup>	Type <sup>b</sup>	% mutant detected <sup>c</sup>		Nb. Kreads per codon <sup>d</sup>		
		Sanger	Illumina	121-205	375-377	571-654
ControlA	C	D376E (1.0) + W574L (1.0)	D376E (1.3) + W574L (1.4)	21.2	52.6	23.7
ControlB	C	D376E (1.0) + W574L (50.0)	D376E (0.9) + W574L (54.3)	32.7	15.7	21.9
ControlC	C	D376E (10.0) + W574L (1.0)	D376E (9.1)	21.4	42.2	20.3
ControlD	C	A205T (1.0)	A205T (1.3)	26.3	59.1	22.6
ControlE	C	A205T (50.0)	A205T (41.0)	29.4	16.3	22.8
ControlF	C	A205T (10.0) + D376E (1.0)	A205T (15.8) + D376E (1.1)	26.7	62.3	19.2
ControlG	C	W574L (1.0) + G654D (31.0)	W574L (1.1) + G654D (31.2)	25.3	57.7	22.0
ControlH	C	W574L (50.0) + G654D (3.0)	W574L (37.6) + G654D (4.0)	30.1	16.4	25.8
ControlI	C	W574L (10.0)	W574L (12.4)	33.0	65.7	25.9
OC-01	F	A205T (9.0)	A205T (11.5)	29.6	63.1	25.6
OC-05	F	A205T (5.0)	A205T (5.6)	26.4	52.8	21.6
OC-11	F	0.0	0.0	26.9	48.0	27.1
OC-12	F	A205T (2.0)	A205T (2.0)	24.5	62.3	21.5
OC-15	F	A205T (2.0)	A205T (1.2)	28.1	60.7	27.7

1	OC-27	F	A205T (1.0)	A205T (0.9)	29.5	60.1	25.4
2							
3	OC-30	F	A205T (2.0)	A205T (1.7)	34.8	68.3	31.3
4							
5	OC-43	F	A205T (7.0)	A205T (5.7)	29.5	58.7	25.9
6							
7	OC-55	F	0.0	0.0	30.4	56.7	32.8
8							
9	OC-60	F	A205T (3.0)	A205T (1.1)	29.9	57.2	26.8
10							
11	OC-65	F	0.0	0.0	30.0	59.7	28.7
12							
13	NA-25	F	0.0	0.0	28.9	71.6	27.7
14							
15	NA-28	F	0.0	0.0	29.3	64.3	32.2
16							
17	NA-29	F	0.0	0.0	27.4	56.2	27.2
18							
19	NA-30	F	0.0	0.0	31.8	67.6	30.9
20							
21	NA-39	F	0.0	0.0	30.2	66.4	27.7
22							
23							
24							
25							

26 581 <sup>a</sup> NA, Nouvelle Aquitaine; OC, Occitanie.

27  
28 582 <sup>b</sup> C, control population (artificial population consisting of 50 mutant and non-mutant plants which ALS had been sequenced beforehand, chosen so  
29  
30  
31 583 that the controls contained known frequencies of ALS mutations); F, population collected in the field.

32  
33 584 <sup>c</sup> The mutations identified are indicated with their frequency between parentheses.

34  
35 585 <sup>d</sup> Number of kilo-reads (i.e., thousands of reads) covering the codons in the interval mentioned (e.g., 121-205: ALS codons 121 to 205). Paired  
36  
37 586 reads were not joined.

587 **Table 3.** PCR primers.

Species	Code	Sequence (3'-5')	Target <sup>a</sup>	Expected amplicon size	ALS codons of interest encompassed <sup>b</sup>
Groundsel	ALSEN8F	CTTTGGAACGTGAAGGTGTCACC	42-64	350 bp (Amplicon 1)	Gly121, <b>Ala122</b> , Met124, Val196,
	ALSEN9R	ATTATGTTTAGTAATCGAGCGCG	391-369		<b>Pro197</b> , Arg199, <b>Ala205</b>
Groundsel	ALSEN4F	CTTCAAATGCTTGGAATGCATGG	761-783	306 bp (Amplicon 2)	Asp375, <b>Asp376</b> , <b>Arg377</b>
	ALSEN8R	TGCTCATCTAATTCCTTCCTCCAAG	1065-1041		
Groundsel	ALSEN7F	TGAATAATCAGCATTGTTGGGTATGGTG	1413-1438	310 bp (Amplicon 3)	Val571, <b>Trp574</b> , Phe578, <b>Ala653</b> ,
	ALSEN7R	CCATCACCTTCAGTAATCACGTC	1722-1700		<b>Gly654</b>
Ragweed	ALAMB9F	CTTTGGAACGTGAAGGCGTAACCG	375-398	346 bp Amplicon 1)	Gly121, <b>Ala122</b> , Met124, Val196,
	ALAMB10R	TGTTTAGTAATGGAACGTGTTACCTC	720-695		<b>Pro197</b> , Arg199, <b>Ala205</b>
Ragweed	ALAMB10F	TTGCATATGCTTGGGATGCATGG	1094-1116	243 bp (Amplicon 2)	Asp375, <b>Asp376</b> , <b>Arg377</b>
	ALAMB4R	CAAAATCTCGTTAAGCCCCTGTAAC	1336-1312		
Ragweed	ALAMB11F	TAACAATCAGCATTGTTGGGTATGGTGG	1747-1772	383 bp (Amplicon 3)	Val571, <b>Trp574</b> , Phe578, <b>Ala653</b> ,
	ALAMB2R	CATTCAAACCGACAAACTGCTTAC	>2121-2104		<b>Gly654</b>

588 <sup>a</sup> Nucleotides in GenBank/EMBL accession KR024410 or KR024411 (groundsel ALS1 or ALS2 partial sequences, respectively) or in accession  
589 KX870184 (ragweed full ALS sequence).

590 <sup>b</sup> Numbered after *Arabidopsis thaliana* ALS sequence (Genbank/EMBL accession X51514). Codons in **bold** have been implicated in herbicide  
591 resistance in the field.<sup>17</sup> Other codons have been implicated in herbicide resistance in artificial selection experiments.<sup>18</sup>

**Table 4.** Combinations of mutations detected in the 96 groundsel field populations analysed.

Mutation(s) <sup>a</sup>	Nb. populations where found
Pro-197-Leu-1	22
Pro-197-Leu-2	2
Pro-197-Ser-1	2
Pro-197-Ser-2	2
Pro-197-Thr-1	1
Pro-197-Thr-2	1
Pro-197-Leu-1 + Pro-197-Arg-2	3
Pro-197-Leu-1 + Pro-197-Leu-2	5
Pro-197-Asn-2 + Pro-197-Leu-2	2
Pro-197-Leu-1 + Pro-197-Arg-2 + Pro-197-Leu-2	2
Pro-197-Leu-1 + Pro-197-Asn-2 <sup>b</sup> + Pro-197-Leu-2	2
Pro-197-Asn-2 <sup>b</sup> + Pro-197-Leu-2 + Pro-197-Ser-2	1
Pro-197-Leu-1 + Pro-197-Asn-2 <sup>b</sup> + Pro-197-Leu-2 + Pro-197-Ser-2	1

<sup>a</sup> The number following the mutation refers to the ALS homeolog (1, ALS1; 2, ALS2).

<sup>b</sup> Substitution caused by a double mutation.

1 595 **Legends to Figures**

2  
3 596

4  
5 597 **Figure 1.** Flow-chart of the NGS-based resistance diagnosis procedure.

6  
7 598

8  
9 599 **Figure 2.** The 67 ragweed populations from Occitanie used for NGS-based detection of  
10 600 mutations endowing resistance to ALS-inhibiting herbicides. Each population is figured as a  
11 601 coloured dot. Green dots, populations where no mutation at the ALS gene was detected among  
12 602 the pool of 50 plants analysed. Red dots, populations where an Ala-205-Thr substitution was  
13 603 detected.

14  
15 604

16  
17 605 **Figure 3.** Frequencies of mutations at the ALS gene detected using individual plant analysis  
18 606 (dCAPS or Sanger sequencing) plotted against the frequencies assessed using Illumina  
19 607 sequencing of pools of 50 plants. In groundsel, frequencies are displayed separately for each  
20 608 ALS homeolog. Because dCAPS genotyping does not allow to discriminate among different  
21 609 mutations at codon 197, the frequencies displayed are the cumulated frequencies of all  
22 610 mutations detected at codon 197. In ragweed, the frequencies are displayed separately for each  
23 611 mutation when several mutations are present in one same population.

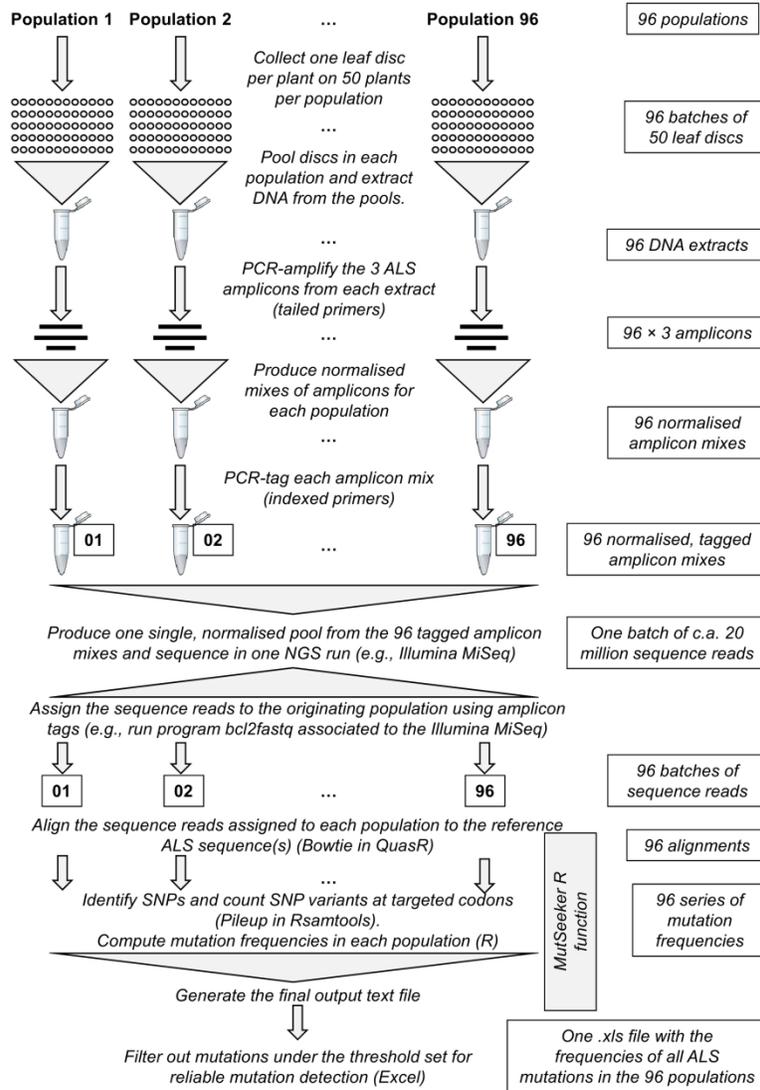


Figure 1

209x297mm (300 x 300 DPI)

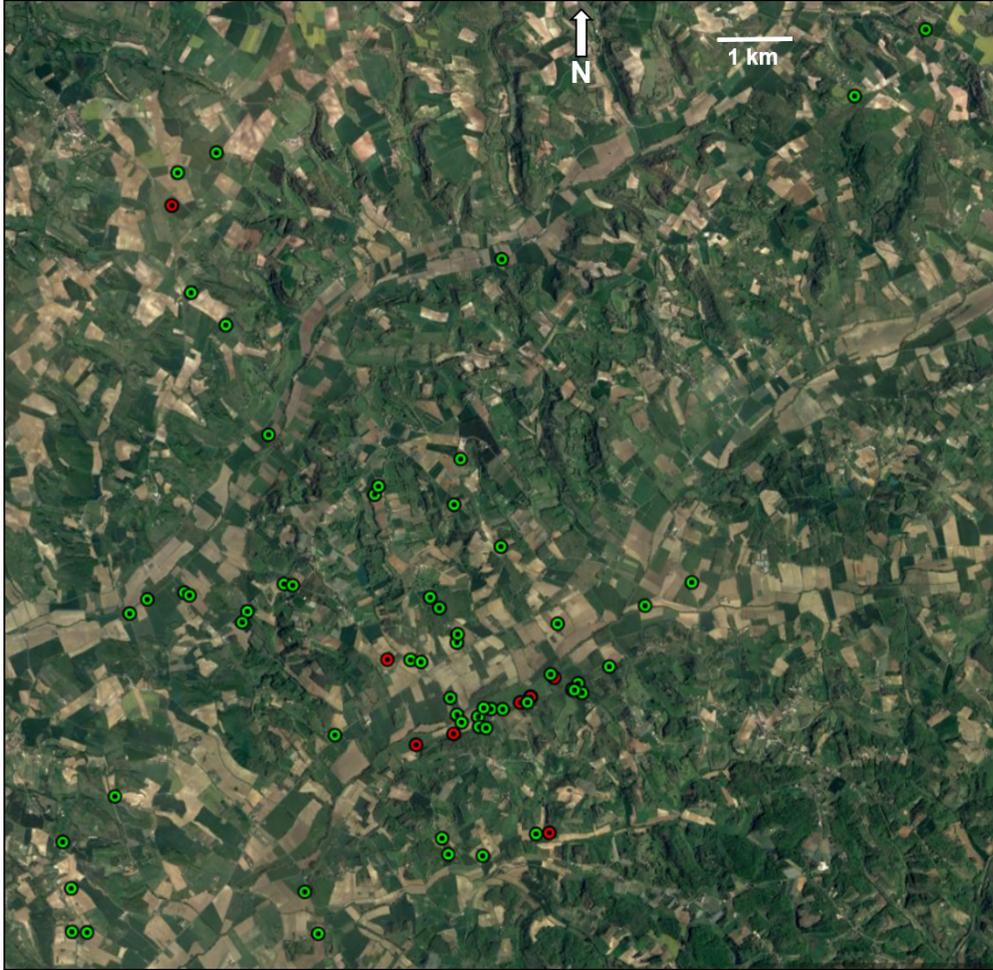


Figure 2

185x181mm (150 x 150 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

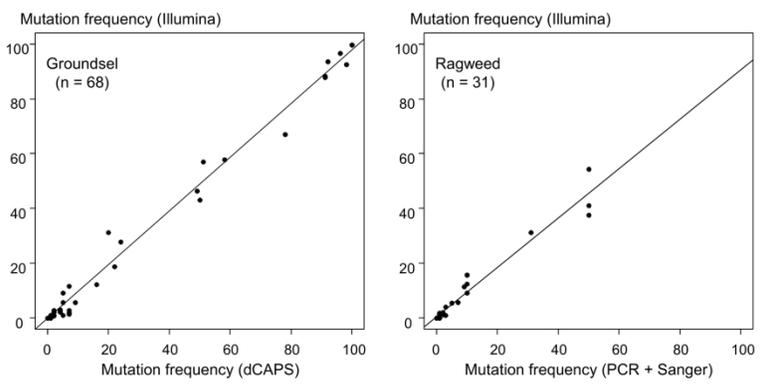


Figure 3

209x297mm (300 x 300 DPI)

1  
2  
3 1 **Harnessing the power of Next-Generation Sequencing technologies to the purpose of high-**  
4  
5 2 **throughput pesticide resistance diagnosis.**  
6  
7  
8 3  
9

10 4 Running title: Illumina-based pesticide resistance diagnosis  
11  
12 5

13  
14 6 Christophe Délye, <sup>a\*</sup> Séverine Michel, <sup>a</sup> Fanny Pernin, <sup>a</sup> Véronique Gautier, <sup>b</sup> Marie Gislard, <sup>c</sup>

15  
16 7 Charles Poncet <sup>b</sup> and Valérie Le Corre <sup>a</sup>  
17  
18  
19 8

20  
21 9

22  
23 10

24  
25 11

26  
27 12

28  
29 13

30  
31 14

32  
33 14 <sup>a</sup> INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France.  
34

35 15 <sup>b</sup> INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales, Clermont-Ferrand,  
36  
37 16 France.  
38

39  
40 17 <sup>c</sup> INRA, US 1426 GeT-PlaGe, Genotoul, Castanet-Tolosan, France  
41

42  
43 18

44 19 \* Correspondence to:

45  
46 20 Dr Christophe Délye

47  
48 21 INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France  
49

50  
51 22 Phone, +33 380 693 185  
52

53  
54 23 Fax, +33 380 693 262  
55

56 24 E-mail, christophe.delye@inra.fr  
57  
58  
59  
60

**Abstract**

**BACKGROUND:** Next Generation Sequencing (NGS) technologies offer tremendous possibilities for high-throughput pesticide resistance diagnosis *via* massive genotyping-by-sequencing. Herein, we used Illumina sequencing combined with a simple, non-commercial bioinformatics pipe-line to seek mutations involved in herbicide resistance in two weeds.

**RESULTS:** DNA was extracted from 96 pools of 50 plants for each species. Three amplicons encompassing 15 ALS (acetolactate-synthase) codons crucial for herbicide resistance were amplified from each DNA extract. Above 18 and 20 million quality 250-nucleotide sequence reads were obtained for groundsel (*Senecio vulgaris*, tetraploid) and ragweed (*Ambrosia artemisiifolia*, diploid), respectively. Herbicide resistance-endowing mutations were identified in 45 groundsel and in eight ragweed field populations. The mutations detected and their frequencies assessed by NGS were checked by individual plant genotyping or Sanger sequencing. NGS results were fully confirmed, except in three instances out of 12 where mutations present at a frequency of 1% were detected below the threshold set for reliable mutation detection.

**CONCLUSION:** Analysing 9,600 plants requested 192 DNA extractions followed by 1,728 PCRs and two Illumina runs. Equivalent results obtained by individual analysis would have necessitated 9,600 individual DNA extractions followed by 216,000 genotyping PCRs, or by 121,500 PCRs and 40,500 Sanger sequence runs. This clearly demonstrates the interest and power of NGS-based detection of pesticide resistance from pools of individuals for diagnosing resistance in massive numbers of individuals.

**Key-words.** Acetolactate-synthase (ALS), diagnosis, genotyping-by-sequencing, herbicide, Illumina, resistance.

## 1 INTRODUCTION

Chemical pesticides (herbicides, insecticides, fungicides) are currently key tools for efficient pest management and preservation of the global food security.<sup>1,2</sup> However, the reckless use of chemical pest control has promoted the evolution of pesticide resistance in numerous weeds, pests and plant pathogens, thereby jeopardising efficient pest control. Achieving accurate resistance detection, especially at early stages in the onset of resistance evolution, is crucial to adapt pest management practices and sustain pesticide efficiency. A variety of methods can be used for this purpose.<sup>3</sup> Among those, molecular assays targeting mutations at the root of resistance mechanisms show the highest potential for rapid routine detection of resistance in pest populations. This potential has been tremendously increased with the advent of the Next Generation Sequencing (NGS) technologies.<sup>4,5</sup> NGS generates huge amounts of sequence data that enable the identification of polymorphisms across whole genomes as well as the simultaneous detection of many mutations in a massive number of samples. As mutation detection is achieved by sequencing whole gene regions of interest, NGS-based search for polymorphisms can also reveal new mutations of potential relevance. NGS potentialities are already exploited in the medical field for the detection *via* genotyping-by-sequencing of genetic markers for human diseases (e.g.,<sup>5</sup>) or of mutations involved in the resistance of pathogens to drugs used in medical treatments (e.g.,<sup>6,7</sup>). Although this type of application of NGS has also been proposed for pesticide resistance diagnosis in agricultural pests,<sup>3</sup> the use of NGS techniques for this purpose is still in its infancy. To the best of our knowledge, only three studies have applied NGS to the purpose of pesticide resistance detection so far. The first one described the detection of mutations at seven codons of one gene involved in herbicide resistance in two polyploid *Echinochloa* species using the now obsolete 454 sequencing technology.<sup>8</sup> A total of 1,120 individuals plants were screened as 28 bulks, each corresponding to one field populations. 454 reads corresponding to each population were identified by specific short sequence tags.

1  
2  
3 74 The c.a. 64,000 sequence reads obtained were analysed through a very basic bioinformatic pipe-  
4  
5 75 line. The sequencing depth obtained was moderate ( $1.7\times$  to  $6.2\times$  per individual plant and per  
6  
7 76 codon of interest), but the mutation frequencies identified by NGS matched those obtained by  
8  
9 77 Sanger sequencing of individual plants, thereby indicating that NGS can be used to detect and  
10  
11 78 quantify mutations endowing pesticide resistance in pools of individuals. The second study  
12  
13 79 applied the Illumina technology to the detection of mutations at one codon endowing fungicide  
14  
15 80 resistance in *Zymoseptoria tritici*.<sup>9</sup> About 723,000 quality sequence reads were obtained and  
16  
17 81 used to analyse 40 fungal isolates from various geographical origins. Despite the authors  
18  
19 82 indicating that sequence reads could have been attributed to individual isolates by using  
20  
21 83 appropriate tagging, the 40 isolates were analysed as one single bulk. This work illustrated the  
22  
23 84 potential of the Illumina technology for pesticide resistance diagnosis. The third study sought  
24  
25 85 mutations at five codons in one gene involved in insecticide resistance in *Rhizopertha*  
26  
27 86 *dominica*.<sup>10</sup> A total of 1,435 individual insects were analysed using Illumina sequencing. Using  
28  
29 87 tagging, the unspecified number of quality sequence reads obtained could be attributed to each  
30  
31 88 individual insect, thus allowing individual genotype determination at every codon screened.  
32  
33 89 This study further confirmed the feasibility of multiple individual analysis in one single NGS  
34  
35 90 run.  
36  
37  
38  
39  
40

41  
42 91 The three preceding pioneer works clearly demonstrated the interest and feasibility of  
43  
44 92 pesticide resistance diagnosis by genotyping-by-sequencing using the NGS technologies  
45  
46 93 combined with sequence read tagging that allows to sequence hosts of individuals or  
47  
48 94 populations as one bulk and subsequently trace the individual or population at the source of  
49  
50 95 each sequence read. The pending question is, how can this approach be most efficiently used  
51  
52 96 for pesticide resistance diagnosis? Assuming that mutation(s) endowing pesticide resistance  
53  
54 97 have been characterised beforehand, an ideal NGS-based resistance assay would allow to screen  
55  
56 98 as many individuals of the pest considered as possible in one single NGS run. This can be  
57  
58  
59  
60

1  
2  
3 99 achieved by screening populations rather than individuals, and by pooling and simultaneously  
4  
5 100 sequencing populations in one single NGS run. An ideal NGS-based assay would also be  
6  
7 101 reliable and yield a fair estimate of the frequency of resistance-endowing mutations in pest  
8  
9 102 populations. In particular, the assay must not generate false positives or false negatives. Here,  
10  
11 103 the error rate of the NGS technique will determine the lowest mutation frequency that can be  
12  
13 104 reliably detected. Last, an ideal NGS-based assay should not require extensive skills in  
14  
15 105 bioinformatics nor depend on commercial softwares for the handling and analysis of the  
16  
17 106 sequence runs, because while NGS can easily be subcontracted, downstream analysis of the  
18  
19 107 huge amount of sequence data generated often remains a major stumbling block for people  
20  
21 108 dealing with pesticide resistance.  
22  
23  
24  
25

26 109 The purpose of this work was to develop massive resistance diagnosis assays associating  
27  
28 110 the Illumina NGS technology and a simple, robust and versatile pipe-line for the analysis of the  
29  
30 111 sequence reads. We considered acetolactate-synthase (ALS) inhibitors, the second most  
31  
32 112 important herbicide mode of action globally<sup>11</sup> and the one that selected for resistance in the  
33  
34 113 highest number of weed species (160 so far).<sup>12</sup> We developed resistance diagnosis assays for  
35  
36 114 two contrasted weed species where resistance due to mutations in the gene encoding ALS  
37  
38 115 inhibitors had evolved: common groundsel (*Senecio vulgaris*), an allotetraploid species with  
39  
40 116 limited genetic variation,<sup>13</sup> and common ragweed (*Ambrosia artemisiifolia*), a diploid invasive  
41  
42 117 species originating from North America that shows a much higher genetic polymorphism.<sup>14</sup>  
43  
44  
45  
46  
47  
48

## 49 119 2 MATERIALS AND METHODS

50  
51 120  
52  
53 121 The whole NGS-based resistance diagnosis procedure developed in this work is summarised in  
54  
55 122 figure 1.  
56  
57  
58  
59  
60

## 124 **2.1 Plant material**

125 For each of the two species studied, plant material consisted of 96 populations of 50 plants each.  
126 Common groundsel can infest all types of crops. It has evolved resistance to ALS inhibitors in  
127 France in both field crops and vineyards.<sup>15</sup> Accordingly, populations were collected in 81 wheat  
128 fields in the Brittany region in 2016 and in 14 vineyards in the Rhône-Alpes region in 2015.  
129 The 96<sup>th</sup> population was a control consisting of 50 mutant and non-mutant plants which ALS  
130 had been previously sequenced.<sup>15</sup> Plants were selected so that the control contained two mutant  
131 ALS alleles at a frequency of 2% each (Table 1).

132 In France, common ragweed particularly infests sunflower, soybean and maize. ALS  
133 inhibitors are pivotal for the chemical control of ragweed in these crops, but unsatisfactory  
134 control was recently reported in several regions. Populations were collected in two regions  
135 faced with very high levels of ragweed infestation: Occitanie in 2016 and 2017 (67 populations)  
136 and Nouvelle-Aquitaine in 2017 (20 populations). The remaining nine populations analysed  
137 were controls. Each consisted of 50 mutant and non-mutant plants which ALS had been  
138 sequenced beforehand, chosen so that the controls contained known frequencies of different  
139 mutations at the ALS gene. Plants carrying an Ala-205-Thr substitution originated from our  
140 resistance monitoring in France. Seeds from North American ragweed plants carrying a Trp-  
141 574-Leu mutation were kindly provided by Dr Jeff Stachler (North Dakota State University).  
142 After ALS sequencing, two additional mutations (Asp-376-Glu and Gly-654-Asp) were found  
143 in some of the plants obtained from these seeds (Table 2).

144 All the fields sampled had been sprayed with ALS inhibitors during the years preceding  
145 our sampling and were selected on the basis of the recurrent presence of high numbers of  
146 individuals from the species of interest. Each field was geolocated. Within each field, one leaf  
147 was collected on 50 plants of the species of interest scattered all over the infested area. After

1  
2  
3 148 collection, leaves were wrapped in paper towels and mailed to our lab where they were stored  
4  
5 149 at -20°C.  
6  
7

8 150

## 151 **2.2 DNA extraction and PCR amplification of ALS fragments**

152 One disc was punched out of the 50 leaves in each groundsel or ragweed population using a 2-  
153 mm diameter hollow punch. The 50 discs were placed together in one 2 mL microtube  
154 containing one steel bead (3 mm diameter). Tubes were closed, frozen in liquid nitrogen during  
155 at least one minute and placed in a bead mill homogeniser (TissueLyser II, Qiagen) at 1,800  
156 oscillations.min<sup>-1</sup> for 1 min 30 s. 400 µL extraction buffer<sup>16</sup> were added to each tube, and leaf  
157 powder was suspended in the buffer by 30 sec vortexing. Tubes were immediately incubated in  
158 a water bath 5 min at 95°C, cooled on ice and centrifugated 2 min at 20,000 g. The resulting  
159 rough DNA extracts were stored at -20°C.

160 Dilutions (1/10) of DNA extracts were directly used for PCR. PCR mixes were as  
161 described.<sup>16</sup> The PCR programs consisted of 3 min at 95°C, followed by 37 cycles of 5 sec at  
162 95°C, 10 sec at 60°C and 30 sec at 72°C. Currently, eight ALS codons are known to be involved  
163 in herbicide resistance selected for in the field (Ala122, Pro197, Ala205, Asp376, Arg377,  
164 Trp574, Ala653 and Gly654, standardised to *Arabidopsis thaliana* ALS sequence).<sup>17</sup> Seven  
165 additional codons have been shown to be involved in changes in herbicide sensitivity in  
166 artificial selection experiments (Gly121, Met124, Val196, Arg199, Asp375, Val571 and  
167 Phe578).<sup>18</sup> For each species investigated, three pair of PCR primers were designed to generate  
168 amplicons encompassing all these positions (Table 3).

169 An allotetraploid, groundsel contains two homeolog ALS genes (ALS1 and ALS2) only  
170 differing by 19 nucleotide substitutions.<sup>15</sup> Primers used to amplify groundsel DNA matched  
171 both ALS1 and ALS2 sequences, but the amplicons expected contained four (amplicon 1), one  
172 (amplicon 2) and five (amplicon 3) homeolog-specific nucleotides. All primers were used at

1  
2  
3 173 0.2  $\mu$ M final concentration each. As a classical, non-proofreading *Taq* polymerase was used for  
4  
5 174 PCR, three independent PCR reactions were performed and subsequently pooled for each  
6  
7  
8 175 amplicon and each population.

9  
10 176 For groundsel analyses, overhang adapter sequences were included at the 5' end of the  
11  
12 177 three F primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and of the three  
13  
14 178 R primers (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-). Tailed primers  
15  
16  
17 179 were used in a first PCR as described above. For each population, all three amplicons were  
18  
19 180 pooled in an equimolar mix and tagged with a population-specific index using Nextera XT  
20  
21 181 Index kit (v2) (Illumina). Tagged amplicons were pooled in an equimolar mix, purified with an  
22  
23  
24 182 Agencourt AMPure XP kit (Beckman Coulter, Beverly, MA, USA), quantified using the Qubit  
25  
26 183 HS kit (Invitrogen) and loaded onto one Illumina MiSeq V2 cartridge according to the  
27  
28 184 manufacturer instructions. The quality of the run was checked internally using PhiX. The  
29  
30 185 resulting 250 nucleotide pair-end sequences passing Illumina standard quality controls were  
31  
32  
33 186 assigned to their population of origin on the basis of the population-specific indexes using  
34  
35 187 bcl2fastq v2.20.0.422 (Illumina).

36  
37 188 For ragweed analyses, overhang adapter sequences were also included at the 5' end of  
38  
39 189 the three F primers (5'-CTTCCCTACACGACGCTCTTCCGATC-) and of the three R  
40  
41 190 primers (5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-). Tailed primers were used in a  
42  
43  
44 191 first PCR as described above. For each population, all three amplicons were pooled in an  
45  
46 192 equimolar mix and tagged with a population-specific, home-made six-bp index in a second, 12-  
47  
48  
49 193 cycle PCR using primers 5'-  
50  
51 194 AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGAC and 5'-  
52  
53 195 CAAGCAGAAGACGGCATAACGAGAT-XXXXXX-GTGACTGGAGTTCAGACGTGT  
54  
55  
56 196 where XXXXXX is the population-specific index sequence. Tagged amplicons were purified  
57  
58 197 and loaded onto one Illumina MiSeq V3 cartridge according to the manufacturer instructions.  
59  
60

198 The quality of the run was checked internally using PhiX. The resulting 250 nucleotide pair-  
199 end sequences passing Illumina standard quality controls were assigned to their population of  
200 origin as described above.

201

### 202 2.3 Analysis of sequence reads and identification of mutations

203 Illumina sequence data were analysed using ~~the~~ “MutSeeker”, a custom function written in the  
204 R programming language ~~and~~ implementing packages available in the Bioconductor project  
205 (bioconductor.org). This function, together with the R script for operating it, detailed operating  
206 instructions and one training data set for each of the two species studied here, are available from  
207 the [Dataverse](#) repository at:  
208 <https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/BCZF3S>

209 Briefly, ~~T~~he 250-nucleotide paired reads were not joined. ~~All~~ Reads were aligned to  
210 reference ALS sequences (GenBank/EMBL accession KR024410 and KR024411 for groundsel  
211 ALS1 and ALS2 partial sequences, respectively, or KX870184 for ragweed full ALS sequence)  
212 using the short read aligner Bowtie as implemented in the package qQuasR.<sup>19</sup> SNP calling was  
213 performed using the pileup query in the Rsamtools package.<sup>20</sup> All codons targeted were covered  
214 by either the forward or the reverse read in the 250-nucleotide read pairs matching the  
215 corresponding amplicon, except codons Asp375, Asp376 and Arg377 that were covered by both  
216 reads. For these three codons, SNPs calls obtained from forward and reverse reads were pooled.  
217 In a final step, variant counts were calculated at every nucleotide position where mutations  
218 would cause amino-acid changes in the 15 ALS codons implicated in herbicide resistance  
219 (Table 3). Thus, analysis yielded the frequencies of single-nucleotide substitutions in every pool  
220 of 50 plants analysed.

221 The threshold for the detection of mutations was set to 0.9% considering the minimum  
222 expected frequency of mutations in a pool of 50 plants of 1% (i.e., one mutation on one ALS

1  
2  
3 223 homeolog copy in 50 plants carrying two copies of said homeolog) and the current error rate of  
4  
5 224 the Illumina sequencing technology (0.24% per nucleotide).<sup>21</sup>  
6  
7  
8 225

9  
10 226 **2.4 Checking NGS-based mutation detection using dCAPS genotyping or Sanger**  
11  
12 227 **sequencing of individual plants**  
13

14 228 Our Illumina-based resistance diagnosis assay was designed to assess frequencies of mutant  
15 229 ALS alleles in each groundsel or ragweed population analysed as a bulk of 50 plants. Therefore,  
16 230 sequence data was not recovered for individual plants. Including the control populations, 34  
17 231 groundsel populations and 25 ragweed populations containing frequencies of mutations as  
18  
19 232 contrasted as possible were used to check the reliability of assessing frequencies of mutant ALS  
20 233 alleles using Illumina sequencing (Tables 1, 2). DNA was extracted individually from each of  
21 234 the 50 plants in each of the populations used to check NGS results, except the control  
22 235 populations that had been characterised beforehand.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 236 In groundsel, mutations conferring resistance to ALS inhibitors were previously  
34 237 exclusively found at codon 197 in ALS1 and/or ALS2.<sup>15</sup> Individual plant genotypes at ALS1  
35 238 and ALS2 codon 197 were determined using previously described homeolog-specific dCAPS  
36 239 assays allowing detection of mutations at codon 197.<sup>15</sup> This allowed to compute the frequencies  
37 240 of mutations at each homeolog within each population. These frequencies were compared to  
38 241 those obtained by NGS analysis of the corresponding bulks of 50 plants.  
39  
40  
41  
42  
43  
44  
45

46 242 In ragweed, 16 populations where mutant ALS alleles had or had not been detected were  
47 243 selected for ALS Sanger sequencing in addition to the nine control populations that had been  
48 244 characterised beforehand. Each of the three amplicons analysed by NGS were generated from  
49 245 each individual plant in three independent PCRs using primers in Table 3 and subsequently  
50 246 sequenced on both strands using Sanger sequencing.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 247 Individual plant genotypes at ALS obtained by genotyping or sequencing were used to  
4  
5 248 compute the exact frequencies of mutant ALS alleles within each population. These frequencies  
6  
7  
8 249 were compared to those estimated by NGS analysis of the corresponding bulks of 50 plants.  
9

10 250

### 12 251 **3 RESULTS**

14 252

#### 17 253 **3.1 NGS-based detection and quantification of the frequency of mutations in groundsel**

##### 19 254 **ALS**

21 255 The groundsel MiSeq run generated 18,084,459 250-nucleotide quality reads that could be  
22  
23  
24 256 assigned to one of the 96 populations investigated. The number of quality reads assigned to one  
25  
26 257 population ranged from 35,366 to 498,841, with an average value of 188,380. Each population  
27  
28 258 was analysed as one pool of 50 individual plants, each carrying two copies of two ALS  
29  
30 259 homeologs. Three categories of amplicons were generated per ALS homeolog (amplicons 1, 2  
31  
32 and 3, Table 1). As each plant carried two copies of each of two ALS homeologs, one plant  
33  
34 260 yielded four amplicons in each category, i.e., 12 amplicons. Thus, the 50 plants in one pool  
35  
36 261 yielded 600 amplicons. The average expected sequencing depth was therefore of 314× for each  
37  
38 262 of the 600 amplicons per pool.  
39  
40 263

42 264 The pipe-line assigned the reads to ALS1 or ALS2 on the basis of the homeolog-specific  
43  
44 265 nucleotide(s) present in all amplicons. The lowest number of reads covering one codon in one  
45  
46 266 population was 735 (ALS1 codons 571 to 654 in population BZ-67, Table 1), which allowed  
47  
48 267 an expected sequencing depth of 7.3× for each of these three codons in each individual plant in  
49  
50 268 the pool. This was deemed sufficient for mutation detection.  
51  
52

53 269 With the 0.9% threshold for mutation detection, non-synonymous nucleotide  
54  
55 270 substitutions were identified at two codons crucial for herbicide sensitivity. Substitutions were  
56  
57 271 identified at codon 197 in the control population and in 45 of the 95 field populations  
58  
59  
60

1  
2  
3 272 investigated. In total, eight substitutions occurring exclusively at the first two nucleotides in  
4  
5 273 codon 197 in ALS1 (three substitutions) and/or ALS2 (five substitutions) were identified, with  
6  
7  
8 274 up to four different substitutions detected in one same population (Table 4). The frequencies of  
9  
10 275 the substitutions assessed by the pipe-line ranged from 1.0% to 99.8% (ALS1) and from 0.9%  
11  
12 276 to 99.7% (ALS2). In six populations, C-to-A transversions at the first and at the second  
13  
14 277 nucleotide in ALS2 codon 197 were always detected in matching frequencies, suggesting the  
15  
16 278 occurrence of one double-substitution at codon 197 in ALS2 that would cause a proline-to-  
17  
18 279 asparagine amino-acid substitution. This was confirmed by Sanger sequencing of the individual  
19  
20 280 plants in these populations (Tables 1, 4). Thus, a total of eight mutations causing amino-acid  
21  
22 281 substitutions at ALS codon 197 were identified in varying frequencies in the groundsel  
23  
24 282 populations investigated. A Pro-197-Leu substitution at ALS1 detected in 35 populations was  
25  
26 283 by far the most frequent mutation in the populations analysed.

27  
28  
29  
30 284 A T-to-G transversion at the third nucleotide in codon 375 causing an aspartate-to-  
31  
32 285 glutamate amino-acid substitution was identified in both ALS1 and ALS2 in 89 of the 96  
33  
34 286 populations analysed, in frequencies systematically comprised between 0.9 and 1.5%. Sanger  
35  
36 287 sequencing in six populations where this substitution was detected by NGS did not allow to  
37  
38 288 confirm its existence. The putative Asp-375-Glu substitution was therefore considered a PCR  
39  
40 289 or NGS artifact, and discarded from the analyses.

41  
42  
43  
44  
45 290

### 291 **3.2 NGS-based detection and quantification of the frequency of mutations in ragweed ALS**

42  
43  
44  
45  
46  
47 292 The ragweed MiSeq run generated 20,700,536 250-nucleotide quality reads that could be  
48  
49 293 assigned to one of the 96 populations investigated. The number of quality reads assigned to one  
50  
51 294 population ranged from 147,116 to 274,998, with an average value of 215,631. Each population  
52  
53 295 was analysed as one pool of 50 individual plants, each carrying two ALS gene copies. Thus,  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 296 the 50 plants in one pool yielded 300 amplicons. The average expected sequencing depth was  
4  
5 297 therefore of 719× for each of the 300 amplicons per pool.  
6  
7

8 298 The lowest number of reads covering one codon in one population was 15,253 (codons  
9  
10 299 571 to 654, not shown). This allowed an expected sequencing depth of 152× for each of these  
11  
12 300 five codons in each individual plant in the corresponding pool, which was more than sufficient  
13  
14 301 for mutation detection.  
15

16  
17 302 With the 0.9% threshold for mutation detection, non-synonymous nucleotide  
18  
19 303 substitutions in codons crucial for herbicide sensitivity were identified as expected in the  
20  
21 304 control populations. The only exception was control population C, where a G-to-T transversion  
22  
23 305 at the second nucleotide in codon 574 causing a tryptophan-to-leucine amino-acid substitution  
24  
25 306 that was present with a frequency of 1% was detected with a frequency of 0.63%, i.e., below  
26  
27 307 the 0.9% threshold (Table 2). No mutations were detected in 79 of the 87 field populations  
28  
29 308 investigated. In the remaining eight populations (all from Occitanie), a G-to-A transition at the  
30  
31 309 first nucleotide in codon 205 causing an alanine-to-threonine amino-acid substitution was  
32  
33 310 detected in frequencies ranging from 0.9 to 11.5% (Table 2). Mapping the eight populations  
34  
35 311 from Occitanie investigated herein revealed that six populations were clustered within a radius  
36  
37 312 of roughly one km within one river valley. The other two populations could be “escapees”  
38  
39 313 disseminated by, e.g., agricultural machinery, to different river valleys 2.5 km south and 7.5  
40  
41 314 north from the main resistance focus, respectively (Figure 12).  
42  
43  
44  
45  
46  
47  
48

### 49 316 **3.3 Checking NGS results using individual plant analysis**

50  
51 317 The frequencies of the mutations detected by NGS-based analysis of populations from both  
52  
53 318 species as pools of 50 plants were checked by selecting pools with contrasted frequencies of  
54  
55 319 mutations and analysing each plant in each of these pool individually using a double-blind  
56  
57 320 procedure. In groundsel, the frequencies of the mutations detected at each ALS homeolog in 33  
58  
59  
60

1  
2  
3 321 field populations were checked by genotyping each plant in each population individually using  
4  
5 322 ALS1- and ALS2-specific dCAPS assays targeting codon 197. dCAPS assays did not  
6  
7 323 discriminate among different mutations occurring at one same codon.<sup>15</sup> Thus, the frequencies  
8  
9  
10 324 estimated by NGS of all mutations identified at codon 197 in each homeolog were cumulated  
11  
12 325 for comparison to the dCAPS results. Similarly, in ragweed, the frequencies of mutations  
13  
14 326 detected in 16 field populations by NGS analysis were checked by Sanger sequencing of the  
15  
16 327 full ALS sequence of each plant in each pool. The frequencies of mutations identified by  
17  
18 328 Illumina sequencing of plant pools and by individual plant analysis in field populations as well  
19  
20  
21 329 as in the control populations were subsequently compared (Tables 1 and 2, Figure 23). For the  
22  
23 330 34 groundsel populations (33 field populations and one control population) and the for the 25  
24  
25 331 ragweed populations (16 field populations and nine control populations) considered, the  
26  
27 332 frequency of each mutation detected by NGS analysis of one bulk of 50 plants was very strongly  
28  
29 333 correlated to the frequency determined by individual plant analysis (Figure 23; Pearson's  
30  
31 334 correlation coefficients values of 0.996 [P-value < 10<sup>-6</sup>] and 0.979 [P-value < 10<sup>-6</sup>] for groundsel  
32  
33 335 and ragweed data, respectively). No false positives were identified, i.e., detection of mutation(s)  
34  
35 336 in one population by NGS analysis of the corresponding plant pool was always confirmed by  
36  
37 337 individual plant sequencing or genotyping. However, NGS-based analysis did not detect  
38  
39 338 mutations in two groundsel populations and one ragweed population where mutations were  
40  
41 339 actually present with a frequency of 1% (i.e., false negatives; Tables 1 and 2). In all three cases,  
42  
43 340 the mutations in question had been detected by NGS analysis, but their estimated frequencies  
44  
45 341 ranged from 0.2 to 0.7% and were therefore below the threshold set for reliable mutation  
46  
47 342 detection (0.9%). On the other hand, mutations present with a frequency of 1% were  
48  
49 343 successfully detected by NGS analysis in two groundsel populations and seven ragweed  
50  
51 344 populations (Tables 1 and 2).  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 346 4 DISCUSSION

347 We successfully applied the Illumina NGS technology to the detection of mutations endowing  
348 resistance to herbicides. PCR amplification of three regions in the ALS gene that harbour the  
349 15 codons known to date to be crucial for herbicide sensitivity were screened in a total of 9,600  
350 plants from two species, the allotetraploid groundsel and the diploid ragweed.

### 352 4.1 Reliability of NGS-based detection of mutations endowing pesticide resistance

353 Analysis of the Illumina sequence reads obtained for the 192 pools of 50 plants each  
354 using the pipe-line described herein enabled the reliable detection of mutations in both species  
355 analysed. Although we used a non-proofreading *Taq* polymerase in our PCRs, no false positive  
356 sample was identified when checking NGS results by individual analysis of the plants in the  
357 pools. There was one case of detection of a non-existing nucleotide change that would have  
358 caused an Asp375Glu amino-acid in groundsel and was disproved by Sanger sequencing.  
359 However, this case was suspicious from the start. When detected, the putative substitution was  
360 systematically present in both ALS homeologs, with frequencies that were always lower than  
361 1.5%. Occurrence of this artifact illustrates the necessity to carefully check the first NGS results  
362 using previously characterised populations before embarking into massive sampling analysis,  
363 and to confirm any NGS-based detection of putative new mutations by other techniques (e.g.,  
364 Sanger sequencing).

365 In the case of the tetraploid groundsel, the mutations detected were correctly assigned  
366 to ALS1 or to ALS2 on the sole basis of a handful of homeolog-specific nucleotides, which  
367 demonstrates the feasibility of analysing several pesticide targets in pest populations in one  
368 single NGS run. NGS analysis also enabled to identify a double-mutation at codon 197 that  
369 caused a Pro-197-Asn amino-acid substitution. Previous reports of double-mutations in  
370 herbicide target sites are scarce. They include a similar Pro-197-Asn amino-acid substitution in

1  
2  
3 371 the grass *Apera spica-venti*,<sup>22</sup> a double mutation at ALS codon 122 causing a Ala-122-Tyr  
4  
5 372 substitution in *Raphanus raphanistrum*<sup>23</sup> and two combined mutations at codons 102 and 106  
6  
7 373 at the 5-enolpyruvylshikimate-3-phosphate synthase gene that encodes another herbicide target  
8  
9 374 enzyme in the grass *Eleusine indica*.<sup>24</sup> Double-mutations in one pesticide gene have often been  
10  
11 375 reported to confer an increased resistance level and/or a broader resistance spectrum, especially  
12  
13 376 in fungicides (e.g.,<sup>25</sup>) but also in herbicides.<sup>24</sup> Herbicide sensitivity bioassays are needed to  
14  
15 377 establish the resistance pattern associated to the double mutation identified in groundsel.

16  
17  
18  
19 378 NGS analysis correctly identified a total of four amino-acid substitutions in ragweed.  
20  
21 379 Three of them were detected in North American ragweed plants used to generate control  
22  
23 380 populations (Table 1). A Trp-574-Leu substitution had previously been identified in this  
24  
25 381 species.<sup>26</sup> An Asp-376-Glu substitution had been identified in 12 species so far,<sup>17</sup> while a Gly-  
26  
27 382 654-Asp substitution had only been reported in the grass weed *Setaria viridis* to date.<sup>27</sup> The last  
28  
29 383 substitution, Ala-205-Thr, was carried by French plants from the field and had never been  
30  
31 384 reported before. Herbicide sensitivity bioassays are underway to establish its associated  
32  
33 385 resistance pattern.  
34  
35  
36  
37  
38  
39

#### 40 387 **4.2 Reliability of NGS-based quantification of mutations endowing pesticide resistance**

41  
42 388 The frequencies of the ALS mutations assessed by NGS analysis of 50-plants pools matched  
43  
44 389 very closely those determined by analysing the corresponding individual plants in both species  
45  
46 390 investigated (Figure 32), even if some differences could be observed (Tables 1 and 2). The  
47  
48 391 causes at the root of these discrepancies may be differences in the amount of plant material  
49  
50 392 collected among individual plants, especially in groundsel where leaf thickness is rather  
51  
52 393 variable, and/or a heterogeneous efficacy of DNA extraction and/or PCR. An unequal  
53  
54 394 contribution of all plants in one pool to the mixture of amplicons generated for NGS is most  
55  
56 395 likely the reason why NGS analysis of plant pools failed to detect mutations present with a  
57  
58  
59  
60

1  
2  
3 396 frequency of 1% in two groundsel and one ragweed population (false negatives, Tables 1, 2).  
4  
5 397 Indeed, these mutations were actually detected by NGS, but with an estimated frequency that  
6  
7 398 was below the 0.9% threshold set for reliable mutation detection. This flaw could be overcome  
8  
9  
10 399 by using a proofreading *Taq* polymerase in the PCRs preceding NGS, which would allow  
11  
12 400 lowering the mutation detection threshold, and/or by slightly decreasing the number of plants  
13  
14 401 per pool (e.g., 40 plants per pool, which would yield a minimum expected mutation frequency  
15  
16 402 of 1.25% on a given gene or homeolog in the pool). Considering the tremendous sequencing  
17  
18 403 depth achieved in this study for every ALS amplicon sequenced, analysis of two pools of 40  
19  
20 404 plants per population investigated rather than a single pool of 50 plants could be performed for  
21  
22 405 96 populations in one single Illumina run with a codon coverage adequate for mutation  
23  
24 406 detection. This way, reducing the number of plants per pool could largely be compensated for  
25  
26 407 by analysing more pools per population.  
27  
28  
29  
30  
31  
32

### 33 409 **4.3 NGS vs. genotyping or Sanger sequencing**

34  
35 410 NGS-based analysis of plant pools detected and quantified resistance-endowing mutations in  
36  
37 411 9,600 individual plants clustered into 192 pools. This approach did not allow to determine the  
38  
39 412 genotypes of the individual plants constituting the pool. This could be considered a drawback  
40  
41 413 in the case of recessive resistances, where differentiating between homozygous and  
42  
43 414 heterozygous mutants may be of some relevance. However, for resistance management in the  
44  
45 415 field, the vital information is rather the detection of mutation(s) involved in pesticide resistance  
46  
47 416 in one pest population. Such mutations are an unambiguous indicator that resistance is arising,  
48  
49 417 and that steps are to be taken to delay its onset.  
50  
51  
52

53  
54 418 Herein, we targeted 15 codons in one ALS gene in 96 batches of 50 plants (ragweed)  
55  
56 419 and in two ALS homeologs in 96 other batches of 50 plants (groundsel), which surpasses all  
57  
58 420 preceding studies having applied NGS to pesticide resistance detection.<sup>8,9,10</sup> This merely  
59  
60

1  
2  
3 421 necessitated 192 DNA extractions followed by the production of three amplicons, each in  
4  
5 422 triplicate PCR reactions per extraction, i.e., 1,728 PCRs, and two Illumina MiSeq runs. (It  
6  
7 423 should be mentioned that, given the sequencing depth achieved, one single such run would have  
8  
9  
10 424 been sufficient to analyse all 192 pools). The same study conducted by the conventional single-  
11  
12 425 codon PCR-based genotyping assays used for resistance detection (e.g.,<sup>15</sup>) would have required  
13  
14 426 9,600 DNA extractions. It would have necessitated, first, to set up and validate 15 codon-  
15  
16 427 specific genotyping assays for ragweed and 30 homeolog-and-codon-specific genotyping  
17  
18 428 assays for groundsel, then, to carry out and visualise 15 genotyping assays per ragweed plant  
19  
20 429 and 30 per groundsel plant, i.e., a total of 216,000 PCRs (followed by the same number of  
21  
22 430 restriction enzyme digestion reactions in the case of dCAPS). If Sanger sequencing had been  
23  
24 431 used instead of genotyping, this study would have necessitated the production of three  
25  
26 432 amplicons in triplicate PCR reactions in ragweed and six in groundsel (three per homeolog),  
27  
28 433 i.e., 121,500 PCRs, followed by 40,500 Sanger sequencing runs. This clearly illustrates the  
29  
30 434 power of NGS for diagnosing resistance-endowing mutations in numerous samples.  
31  
32  
33  
34

35 435 In contrast to preceding NGS studies addressing pesticide resistance diagnosis,<sup>9,10</sup> our  
36  
37 436 work was carried out using a rapid DNA extraction procedure based on a non-commercial, lab-  
38  
39 437 made extraction buffer.<sup>16</sup> The rough DNA extracts obtained proved perfectly suitable for  
40  
41 438 resistance detection in pools of 50 individual plants. After Illumina sequencing, assignment of  
42  
43 439 the sequence reads to their originating sample was performed using the software attached to the  
44  
45 440 sequencer. This can be done in the case of an in-house facility, or requested from any  
46  
47 441 sequencing supplier. The downstream analyses (read mapping on reference sequences,  
48  
49 442 mutation detection and quantification) required no commercial software, our pipe-line being  
50  
51 443 exclusively based on freely available programs and applications.  
52  
53  
54  
55

444

#### 445 **4.4 Developments of the NGS-based method**

1  
2  
3 446 As it is described here, our NGS-based method only targets mutations at specific codons in the  
4  
5 447 ALS gene. However, the scope of the method can be extended. First, any ALS nucleotide  
6  
7 448 present in the amplicons sequenced can very easily be added to the list of positions to be  
8  
9 449 screened by our pipe-line. This could reveal novel mutations of interest. Second, the pipe-line  
10  
11 450 developed herein can be applied to other weed or pest species. As the level of nucleotide  
12  
13 451 polymorphism varies with the species, this will require to adapt the alignment stringency to the  
14  
15 452 gene(s) and to the weed or pest targeted. This can be done by running the pipe-line on a set of  
16  
17 453 previously characterised populations by modifying options for the Bowtie short read aligner  
18  
19 454 (see MutSeeker operating instructions). Third, the sequencing depth achieved by the Illumina  
20  
21 455 technology clearly allows to sequence several genes in a single run. This has been shown herein  
22  
23 456 by our analysing the two ALS homeologs in groundsel. Accordingly, several distinct pesticide  
24  
25 457 targets can be screened in one single NGS run. Resistance can also be due to non-target-site  
26  
27 458 based mechanisms.<sup>28</sup> When the mutations involved in these mechanisms are known, they can  
28  
29 459 be subjected to NGS-based detection. Amplifying and sequencing by NGS as many amplicons  
30  
31 460 as necessary to cover the genome regions including the nucleotide positions of interest is easily  
32  
33 461 feasible, keeping in mind that the length of each amplicon must be lower than that of the NGS  
34  
35 462 reads (i.e., lower than 500 nucleotides for 250 nucleotide pair-end sequences). If many genome  
36  
37 463 regions of interest are targeted, then a compromise is to be found between the number of  
38  
39 464 amplicons to be sequenced per population and the number of population to be sequenced per  
40  
41 465 NGS run so that a satisfactory sequencing depth (e.g., 10× for each amplicon obtained from  
42  
43 466 each individual plant in each pool) is achieved.  
44  
45  
46  
47  
48  
49  
50

467

## 468 **5 CONCLUSIONS**

469 This work illustrates the power and throughput of the NGS technologies for the detection of  
470 mutations endowing pesticide resistance. The capacity of NGS-based resistance diagnosis

1  
2  
3 471 assays can be exploited for massive resistance surveys (i.e., analysing numerous pest or weed  
4  
5 472 populations) as well as for in-depth resistance detection (i.e., analysing numerous pest or weed  
6  
7 473 individuals in one population to detect to onset of resistance at very early stages, when  
8  
9 474 resistance management practices can still nip resistance evolution in the bud). With the  
10  
11 475 continuously-increasing amount of sequence reads generated by NGS runs and the technical  
12  
13 476 possibility to multiplex hundreds of pools of individuals in one single NGS run, the era of  
14  
15 477 massive resistance diagnosis is clearly just dawning. As shown in our work, the tremendous  
16  
17 478 amounts of sequence reads generated can be analysed without the need for extensive training  
18  
19 479 in NGS data analysis, which should extend the benefit of using NGS diagnosis to a large number  
20  
21 480 of scientists dealing with pest control. Actually, the throughput of assays such as the one  
22  
23 481 described herein is so high that the future main issue with NGS-based resistance detection might  
24  
25 482 well be the possibility to obtain massive numbers of samples from the field. This issue could  
26  
27 483 be overcome by pooling populations of different pests and weed species in shared NGS runs.  
28  
29  
30  
31  
32  
33  
34

## 35 485 **ACKNOWLEDGEMENTS**

36  
37 486 This study was partly funded by Bayer CropScience France (groundsel sampling), by ANSES  
38  
39 487 via the system for monitoring the adverse effects of plant protection products  
40  
41 488 (‘phytopharmacovigilance’, established by the French Act on the future of agriculture of 13  
42  
43 489 October 2014) and (groundsel sequencing) and by BASF SAS France (ragweed sequencing).  
44  
45  
46 490 We thank the Gentyane facility (Clermont-Ferrand, France [<http://gentyane.clermont.inra.fr/>])  
47  
48 491 and the GeT core facility (Toulouse, France [<http://get.genotoul.fr/>]) for providing access to their  
49  
50 492 sequencing equipment.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

493 **REFERENCES**

- 494 1 Oerke E-C, Crop losses to pests. *J Agric Sci* **144**:31-43 (2006).
- 495 2 Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N and Nelson A, The  
496 global burden of pathogens and pests on major food crops. *Nature Ecol Evol* **3**:430-439  
497 (2019).
- 498 3 R4P Network, Trends and challenges in pesticide resistance detection. *Trends Plant Sci*  
499 **21**:834-853 (2016).
- 500 4 Metzker ML, Sequencing technologies - the next generation. *Nature Rev Genet* **11**:31-  
501 46 (2010).
- 502 5 Kamps R, Brandão RD, van den Bosch BJ, Paulussen ADC, Xanthoulea S, Blok MJ  
503 and Romano A, Next-generation sequencing in oncology: genetic diagnosis, risk  
504 prediction and cancer classification. *Int J Mol Sci* **18**:Article308 (2017).
- 505 6 Sahoo MK, Lefterova MI, Yamamoto F, Waggoner JJ, Chou S, Holmes SP, Anderson  
506 MW and Pinsky BA, Detection of cytomegalovirus drug resistance mutations by next-  
507 generation sequencing. *J Clin Microbiol* **51**:3700-3710 (2013).
- 508 7 Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, Crudu V,  
509 Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC, Catanzaro A, Keim  
510 P and Engelthaler DM, Detection of low-level mixed-population drug resistance in  
511 *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS ONE*  
512 **10**:e0126626 (2015).
- 513 8 Délye C, Causse R, Gautier V, Poncet C and Michel S, Using next-generation  
514 sequencing to detect mutations endowing resistance to pesticides: application to  
515 acetolactate-synthase (ALS)-based resistance in barnyardgrass, a polyploid grass weed.  
516 *Pest Manag Sci* **71**:675-685 (2015).

- 1  
2  
3 517 9 Pieczul K and Wařowska A, The application of next-generation sequencing (NGS) for  
4  
5 518 monitoring of *Zymoseptoria tritici* QoI resistance. *Crop Prot* **92**:143-147 (2017).  
6  
7  
8 519 10 Schlipalius DI, Tuck AG, Pavic H, Daghish GJ, Nayaka MK and Ebert PR, A high-  
9  
10 520 throughput system used to determine frequency and distribution of phosphine resistance  
11  
12 521 across large geographical regions. *Pest Manag Sci* **75**:1091-1098 (2019).  
13  
14 522 11 Kraehmer H, Laber B, Rosinger C and Schulz A, Herbicides as weed control agents:  
15  
16 523 state of the art: I. weed control research and safener technology: the path to modern  
17  
18 524 agriculture. *Plant Physiol* **166**:1119-1131 (2014).  
19  
20  
21 525 12 Heap I, International survey of herbicide-resistant weeds. Available online at:  
22  
23 526 <http://www.weedscience.com>. Accessed 11 April 2019.  
24  
25  
26 527 13 Zhu B-R, Barrett SCH, Zhang D-Y and Liao W-J, Invasion genetics of *Senecio vulgaris*:  
27  
28 528 loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple  
29  
30 529 introductions. *Biol Invasions* **19**:255-267 (2017).  
31  
32  
33 530 14 Meyer L, Causse R, Pernin F, Scalone R, Bailly G, Chauvel B, Délye C and Le Corre  
34  
35 531 V, New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant  
36  
37 532 *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species.  
38  
39 533 *PLoS ONE* **12**:e0176197 (2017).  
40  
41  
42 534 15 Délye C, Causse R, Gautier V, Poncet C and Michel S, Genetic basis, evolutionary  
43  
44 535 origin and spread of resistance to herbicides inhibiting acetolactate synthase in common  
45  
46 536 groundsel (*Senecio vulgaris*). *Pest Manag Sci* **72**:89-102 (2016).  
47  
48  
49 537 16 Délye C, Matějček A and Gasquez J, PCR-based detection of resistance to Acetyl-CoA  
50  
51 538 carboxylase-inhibiting herbicides in black-grass (*Alopecurus myosuroides* Huds) and  
52  
53 539 ryegrass (*Lolium rigidum* Gaud). *Pest Manag Sci* **58**:474-478 (2002).  
54  
55  
56 540 17 Tranel PJ, Wright TR and Heap IM, Mutations in herbicide-resistant weeds to ALS  
57  
58 541 inhibitors. Available online at: <http://www.weedscience.com>. Accessed 11 April 2019.  
59  
60

- 1  
2  
3 542 18 Duggleby RG, McCourt JA and Guddat LW, Structure and mechanism of inhibition of  
4  
5 543 plant acetohydroxyacid synthase. *Plant Physiol Biochem* **46**:309-324 (2008).  
6  
7  
8 544 19 Gaidatzis DA, Lerch F, Hahne F and Stadler MB, QuasR: quantify and annotate short  
9  
10 545 reads in R. *Bioinformatics* **31**:1130-1132 (2015).  
11  
12 546 20 Morgan M, Pagès H, Obenchain V and Hayden N, Rsamtools: binary alignment (BAM),  
13  
14 547 FASTA, variant call (BCF), and tabix file import. R package version 1.34.1, Available  
15  
16 548 online at: <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>.  
17  
18 549 Accessed 11 April 2019.  
19  
20  
21 550 21 Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL and Mayer G,  
22  
23 551 Systematic evaluation of error rates and causes in short samples in next-generation  
24  
25 552 sequencing. *Sci Rep* **8**:10950 (2018).  
26  
27  
28 553 22 Massa D, Krenz B and Gerhards R, Target-site resistance to ALS-inhibiting herbicides  
29  
30 554 in *Apera spica-venti* populations is conferred by documented and previously unknown  
31  
32 555 mutations. *Weed Res* **51**:294-303 (2011).  
33  
34  
35 556 23 Han H, Yu Q, Purba E, Li M, Walsh M, Friesen S and Powles SB, A novel amino acid  
36  
37 557 substitution Ala-122-Tyr in ALS confers high-level and broad resistance across ALS-  
38  
39 558 inhibiting herbicides. *Pest Manag Sci* **68**:1164-1170 (2012).  
40  
41  
42 559 24 Han H, Vila-Aiub MM, Jalaludin A, Yu Q and Powles SB, A double EPSPS gene  
43  
44 560 mutation endowing glyphosate resistance shows a remarkably high resistance cost.  
45  
46 561 *Plant Cell Environ* **40**:3031-3042 (2017).  
47  
48  
49 562 25 Leroux P and Walker A-S, Multiple mechanisms account for resistance to sterol 14 $\alpha$ -  
50  
51 563 demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. *Pest Manag*  
52  
53 564 *Sci* **67**:44-59 (2011).  
54  
55  
56 565 26 Patzoldt WL, Tranel PJ, Alexander AL and Schmitzer PR, A common ragweed  
57  
58 566 population resistant to cloransulam-methyl. *Weed Sci* **49**:485-490 (2001).  
59  
60

- 1  
2  
3 567 27 Laplante J, Rajcan I, and Tardif FJ, Multiple allelic forms of acetohydroxyacid synthase  
4  
5 568 are responsible for herbicide resistance in *Setaria viridis*. *Theor Appl Genet* **119**:577-  
6  
7 569 585 (2009).  
8  
9  
10 570 28 Délye C, Duhoux A, Pernin F, Riggins CW and Tranel PJ, Molecular mechanisms of  
11  
12 herbicide resistance. *Weed Sci* **63**:91-115 (2015).  
13 571  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

572 **Table 1.** Groundsel populations used for validation of Illumina-based detection of resistance using dCAPS genotyping.

Populations		ALS1					ALS2				
Code <sup>4a</sup>	Type	% mutant detected		Nb. Kreads per codon <sup>3c</sup>			% mutant detected		Nb. Kreads per codon <sup>3c</sup>		
		dCAPS	Illumina	121-205	375-377	571-654	dCAPS	Illumina	121-205	375-377	571-654
Control	C	2.0	2.0	14.0	17.7	21.0	2.0	2.2	12.9	16.2	24.1
RA-01	F	0.0	0.0	25.1	9.8	14.9	0.0	0.0	26.2	9.2	12.5
RA-02	F	0.0	0.0	22.3	5.1	23.1	24.0	27.9	23.8	5.3	21.1
RA-03 <sup>4d</sup>	F	0.0	0.0	18.1	6.2	7.7	78.0	67.0	17.1	4.8	9.8
RA-05 <sup>4d</sup>	F	2.0	2.9	7.5	17.1	31.1	98.0	92.6	8.3	18.2	32.4
RA-08	F	0.0	0.0	19.2	5.7	20.7	100.0	99.7	19.8	5.2	22.6
RA-10 <sup>4d</sup>	F	7.0	1.5	9.6	10.4	24.5	92.0	93.6	8.3	9.5	22.2
RA-12 <sup>4d</sup>	F	0.0	0.0	18.8	18.2	21.3	7.0	11.6	20.2	20.3	20.7
RA-13 <sup>4d</sup>	F	0.0	0.0	9.5	18.4	17.8	50.0	43.2	12.2	19.1	18.3
RA-15 <sup>4d</sup>	F	9.0	5.7	32.1	28.4	15.9	91.0	87.9	32.8	34.1	15.7
RA-18	F	0.0	0.0	24.6	36.1	12.3	0.0	0.0	25.9	36.3	18.0

1												
2												
3	RA-19	F	0.0	0.0	12.7	27.7	9.8	0.0	0.0	13.3	27.7	9.5
4												
5	RA-20	F	0.0	0.0	17.9	22.9	20.9	0.0	0.0	14.8	20.6	17.9
6												
7	RA-22	F	0.0	0.0	22.7	7.1	13.3	0.0	0.0	17.7	7.5	11.2
8												
9	RA-24	F	0.0	0.0	27.8	17.2	14.3	0.0	0.0	27.3	17.0	14.9
10												
11	BZ-I	F	16.0	12.3	17.9	13.5	11.2	0.0	0.0	14.4	12.0	11.9
12												
13	BZ-IV	F	0.0	0.0	16.0	11.7	19.8	0.0	0.0	18.4	11.5	19.9
14												
15	BZ-VI	F	96.0	96.6	19.1	12.9	17.6	0.0	0.0	22.2	10.3	16.8
16												
17	BZ-VII	F	58.0	57.9	12.9	15.5	23.9	0.0	0.0	9.7	15.1	19.9
18												
19	BZ-VIII	F	100.0	99.8	29.0	35.7	22.1	1.0	0.9	26.2	38.8	23.4
20												
21	BZ-04	F	2.0	2.4	4.7	21.1	12.3	0.0	0.0	4.5	19.2	12.7
22												
23	BZ-05	F	1.0	1.0	10.7	10.7	14.0	0.0	0.0	13.2	8.4	13.5
24												
25	BZ-09	F	51.0	57.0	16.2	7.4	7.0	0.0	0.0	14.5	6.7	6.6
26												
27	BZ-13	F	5.0	1.0	8.9	2.4	4.5	0.0	0.0	10.0	2.6	4.7
28												
29	BZ-24	F	0.0	0.0	38.1	6.3	18.5	2.0	0.9	32.7	4.8	15.1
30												
31	BZ-38	F	2.0	1.0	12.1	22.6	21.5	1.0	0.0	11.0	20.2	21.7
32												
33	BZ-57	F	0.0	0.0	7.6	6.6	5.3	4.0	3.0	8.7	7.1	6.0
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												

1												
2												
3	BZ-59	F	91.0	88.4	11.1	9.9	7.2	7.0	2.8	14.5	10.0	8.7
4												
5	BZ-61	F	49.0	46.4	9.9	11.9	8.8	22.0	18.9	9.3	11.6	9.0
6												
7	BZ-66	F	5.0	5.7	18.0	32.1	10.1	1.0	0.0	15.4	31.2	11.3
8												
9	BZ-67	F	7.0	2.0	19.8	13.0	0.7	0.0	0.0	19.7	13.2	0.8
10												
11	BZ-69	F	4.0	2.3	17.3	12.2	7.7	0.0	0.0	21.0	12.4	7.5
12												
13	BZ-71	F	4.0	2.5	17.5	18.6	7.9	0.0	0.0	16.4	17.6	7.5
14												
15	BZ-73	F	20.0	31.3	5.5	6.8	12.4	5.0	9.1	11.0	6.3	10.6
16												
17												
18												
19	573	<sup>1</sup> <u>a</u> RA, Rhône-Alpes; BZ, Brittany.										
20												
21	574	<sup>2</sup> <u>b</u> C, control population (artificial population consisting of 50 mutant and non-mutant plants which ALS had been sequenced beforehand, chosen										
22												
23	575	so that the controls contained known frequencies of ALS mutations); F, population collected in the field.										
24												
25												
26	576	<sup>3</sup> <u>c</u> Number of kilo-reads (i.e., thousands of reads) covering the codons in the interval mentioned (e.g., 121-205: ALS codons 121 to 205). Paired										
27												
28	577	reads were not joined.										
29												
30												
31	578	<sup>4</sup> <u>d</u> Populations with plants containing a double mutation causing a Pro-197-Asn substitution at ALS2. These populations were also subjected to										
32												
33	579	Sanger sequencing of individual plants.										
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												

580 **Table 2.** Ragweed populations used for validation of Illumina-based detection of resistance using Sanger sequencing.

Populations		ALS				
Code <sup>1a</sup>	Type	% mutant detected <sup>3c</sup>		Nb. Kreads per codon <sup>4d</sup>		
		Sanger	Illumina	121-205	375-377	571-654
<u>Control</u> A	C	D376E (1.0) + W574L (1.0)	D376E (1.3) + W574L (1.4)	21.2	52.6	23.7
<u>Control</u> B	C	D376E (1.0) + W574L (50.0)	D376E (0.9) + W574L (54.3)	32.7	15.7	21.9
<u>Control</u> C	C	D376E (10.0) + W574L (1.0)	D376E (9.1)	21.4	42.2	20.3
<u>Control</u> D	C	A205T (1.0)	A205T (1.3)	26.3	59.1	22.6
<u>Control</u> E	C	A205T (50.0)	A205T (41.0)	29.4	16.3	22.8
<u>Control</u> F	C	A205T (10.0) + D376E (1.0)	A205T (15.8) + D376E (1.1)	26.7	62.3	19.2
<u>Control</u> G	C	W574L (1.0) + G654D (31.0)	W574L (1.1) + G654D (31.2)	25.3	57.7	22.0
<u>Control</u> H	C	W574L (50.0) + G654D (3.0)	W574L (37.6) + G654D (4.0)	30.1	16.4	25.8
<u>Control</u> I	C	W574L (10.0)	W574L (12.4)	33.0	65.7	25.9
OC-01	F	A205T (9.0)	A205T (11.5)	29.6	63.1	25.6
OC-05	F	A205T (5.0)	A205T (5.6)	26.4	52.8	21.6
OC-11	F	0.0	0.0	26.9	48.0	27.1
OC-12	F	A205T (2.0)	A205T (2.0)	24.5	62.3	21.5

1	OC-15	F	A205T (2.0)	A205T (1.2)	28.1	60.7	27.7
2							
3	OC-27	F	A205T (1.0)	A205T (0.9)	29.5	60.1	25.4
4							
5	OC-30	F	A205T (2.0)	A205T (1.7)	34.8	68.3	31.3
6							
7	OC-43	F	A205T (7.0)	A205T (5.7)	29.5	58.7	25.9
8							
9							
10	OC-55	F	0.0	0.0	30.4	56.7	32.8
11							
12	OC-60	F	A205T (3.0)	A205T (1.1)	29.9	57.2	26.8
13							
14	OC-65	F	0.0	0.0	30.0	59.7	28.7
15							
16							
17	NA-25	F	0.0	0.0	28.9	71.6	27.7
18							
19	NA-28	F	0.0	0.0	29.3	64.3	32.2
20							
21	NA-29	F	0.0	0.0	27.4	56.2	27.2
22							
23							
24	NA-30	F	0.0	0.0	31.8	67.6	30.9
25							
26	NA-39	F	0.0	0.0	30.2	66.4	27.7
27							

581 <sup>1-a</sup> NA, Nouvelle Aquitaine; OC, Occitanie.

582 <sup>2-b</sup> C, control population (artificial population consisting of 50 mutant and non-mutant plants which ALS had been sequenced beforehand, chosen  
583 so that the controls contained known frequencies of ALS mutations); F, population collected in the field.

584 <sup>3-c</sup> The mutations identified are indicated with their frequency between parentheses.

585 <sup>4-d</sup> Number of kilo-reads (i.e., thousands of reads) covering the codons in the interval mentioned (e.g., 121-205: ALS codons 121 to 205). Paired  
586 reads were not joined.

587 **Table 3.** PCR primers.

Species	Code	Sequence (3'-5')	Target <sup>a</sup>	Expected amplicon size	ALS codons of interest encompassed <sup>b</sup>
Groundsel	ALSEN8F	CTTTGGAACGTGAAGGTGTCACC	42-64	350 bp (Amplicon 1)	Gly121, <b>Ala122</b> , Met124, Val196,
	ALSEN9R	ATTATGTTTAGTAATCGAGCGCG	391-369		<b>Pro197</b> , Arg199, <b>Ala205</b>
Groundsel	ALSEN4F	CTTCAAATGCTTGGAATGCATGG	761-783	306 bp (Amplicon 2)	Asp375, <b>Asp376</b> , <b>Arg377</b>
	ALSEN8R	TGCTCATCTAATTCCTTCCTCCAAG	1065-1041		
Groundsel	ALSEN7F	TGAATAATCAGCATTGTTGGGTATGGTG	1413-1438	310 bp (Amplicon 3)	Val571, <b>Trp574</b> , Phe578, <b>Ala653</b> ,
	ALSEN7R	CCATCACCTTCAGTAATCACGTC	1722-1700		<b>Gly654</b>
Ragweed	ALAMB9F	CTTTGGAACGTGAAGGCGTAACCG	375-398	346 bp Amplicon 1)	Gly121, <b>Ala122</b> , Met124, Val196,
	ALAMB10R	TGTTTAGTAATGGAACGTGTTACCTC	720-695		<b>Pro197</b> , Arg199, <b>Ala205</b>
Ragweed	ALAMB10F	TTGCATATGCTTGGGATGCATGG	1094-1116	243 bp (Amplicon 2)	Asp375, <b>Asp376</b> , <b>Arg377</b>
	ALAMB4R	CAAAATCTCGTTAAGCCCCTGTAAC	1336-1312		
Ragweed	ALAMB11F	TAACAATCAGCATTGTTGGGTATGGTGG	1747-1772	383 bp (Amplicon 3)	Val571, <b>Trp574</b> , Phe578, <b>Ala653</b> ,
	ALAMB2R	CATTCAAACCGACAAACTGCTTAC	>2121-2104		<b>Gly654</b>

588 <sup>a</sup> Nucleotides in GenBank/EMBL accession KR024410 or KR024411 (groundsel ALS1 or ALS2 partial sequences, respectively) or in accession  
589 KX870184 (ragweed full ALS sequence).

590 <sup>b</sup> Numbered after *Arabidopsis thaliana* ALS sequence (Genbank/EMBL accession X51514). Codons in **bold** have been implicated in herbicide  
591 resistance in the field.<sup>17</sup> Other codons have been implicated in herbicide resistance in artificial selection experiments.<sup>18</sup>

**Table 4.** Combinations of mutations detected in the 96 groundsel field populations analysed.

Mutation(s) <sup>1a</sup>	Nb. populations where found <sup>2</sup>
Pro-197-Leu-1	22
Pro-197-Leu-2	2
Pro-197-Ser-1	2
Pro-197-Ser-2	2
Pro-197-Thr-1	1
Pro-197-Thr-2	1
Pro-197-Leu-1 + Pro-197-Arg-2	3
Pro-197-Leu-1 + Pro-197-Leu-2	5
Pro-197-Asn-2 + Pro-197-Leu-2	2
Pro-197-Leu-1 + Pro-197-Arg-2 + Pro-197-Leu-2	2
Pro-197-Leu-1 + Pro-197-Asn-2 <sup>2_b</sup> + Pro-197-Leu-2	2
Pro-197-Asn-2 <sup>3_b</sup> + Pro-197-Leu-2 + Pro-197-Ser-2	1
Pro-197-Leu-1 + Pro-197-Asn-2 <sup>2_b</sup> + Pro-197-Leu-2 + Pro-197-Ser-	1
2	

<sup>1\_a</sup> The number following the mutation refers to the ALS homeolog (1, ALS1; 2, ALS2).

<sup>2\_b</sup> Substitution caused by a double mutation.

1 595 **Legends to Figures**

2  
3 596

4  
5 597 **Figure 1. Flow-chart of the NGS-based resistance diagnosis procedure.**

6  
7  
8 598

9  
10 599 **Figure 12.** The 67 ragweed populations from Occitanie used for NGS-based detection of  
11  
12 600 mutations endowing resistance to ALS-inhibiting herbicides. Each population is figured as a  
13  
14 601 coloured dot. Green dots, populations where no mutation at the ALS gene was detected among  
15  
16 602 the pool of 50 plants analysed. Red dots, populations where an Ala-205-Thr substitution was  
17  
18  
19 603 detected.

20  
21  
22 604

23  
24 605 **Figure 23.** Frequencies of mutations at the ALS gene detected using individual plant analysis  
25  
26 606 (dCAPS or Sanger sequencing) plotted against the frequencies assessed using Illumina  
27  
28 607 sequencing of pools of 50 plants. In groundsel, frequencies are displayed separately for each  
29  
30 608 ALS homeolog. Because dCAPS genotyping does not allow to discriminate among different  
31  
32 609 mutations at codon 197, the frequencies displayed are the cumulated frequencies of all  
33  
34 610 mutations detected at codon 197. In ragweed, the frequencies are displayed separately for each  
35  
36  
37 611 mutation when several mutations are present in one same population.