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

Pest Management Science SCI

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

Journal:	*Pest Management Science*
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-througput, massive screening

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

Running title: Illumina-based pesticide resistance diagnosis

Christophe Délye, [a]* Séverine Michel, [a] Fanny Pernin, [a] Véronique Gautier, [b] Marie Gislard, [c] Charles Poncet [b] and Valérie Le Corre [a]

[a] INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France.

[b] INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales, Clermont-Ferrand, France.

[c] INRA, US 1426 GeT-PlaGe, Genotoul, Castanet-Tolosan, France

* Correspondence to:

Dr Christophe Délye

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

Phone, +33 380 693 185

Fax, +33 380 693 262

E-mail, christophe.delye@inra.fr

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.[1,2] 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.[3] 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.[4,5] 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.,[5]) or of mutations involved in the resistance of pathogens to drugs used in medical treatments (e.g.,[6,7]). Although this type of application of NGS has also been proposed for pesticide resistance diagnosis in agricultural pests,[3] 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.[8] 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.

The c.a. 64,000 sequence reads obtained were analysed through a very basic bioinformatic pipeline. The sequencing depth obtained was moderate (1.7× to 6.2× per individual plant and per codon of interest), but the mutation frequencies identified by NGS matched those obtained by Sanger sequencing of individual plants, thereby indicating that NGS can be used to detect and quantify mutations endowing pesticide resistance in pools of individuals. The second study applied the Illumina technology to the detection of mutations at one codon endowing fungicide resistance in *Zymoseptoria tritici*.[9] About 723,000 quality sequence reads were obtained and used to analyse 40 fungal isolates from various geographical origins. Despite the authors indicating that sequence reads could have been attributed to individual isolates by using appropriate tagging, the 40 isolates were analysed as one single bulk. This work illustrated the potential of the Illumina technology for pesticide resistance diagnosis. The third study sought mutations at five codons in one gene involved in insecticide resistance in *Rhyzopertha dominica*.[10] A total of 1,435 individual insects were analysed using Illumina sequencing. Using tagging, the unspecified number of quality sequence reads obtained could be attributed to each individual insect, thus allowing individual genotype determination at every codon screened. This study further confirmed the feasibility of multiple individual analysis in one single NGS run.

The three preceding pioneer works clearly demonstrated the interest and feasibility of pesticide resistance diagnosis by genotyping-by-sequencing using the NGS technologies combined with sequence read tagging that allows to sequence hosts of individuals or populations as one bulk and subsequently trace the individual or population at the source of each sequence read. The pending question is, how can this approach be most efficiently used for pesticide resistance diagnosis? Assuming that mutation(s) endowing pesticide resistance have been characterised beforehand, an ideal NGS-based resistance assay would allow to screen as many individuals of the pest considered as possible in one single NGS run. This can be

achieved by screening populations rather than individuals, and by pooling and simultaneously sequencing populations in one single NGS run. An ideal NGS-based assay would also be reliable and yield a fair estimate of the frequency of resistance-endowing mutations in pest populations. In particular, the assay must not generate false positives or false negatives. Here, the error rate of the NGS technique will determine the lowest mutation frequency that can be reliably detected. Last, an ideal NGS-based assay should not require extensive skills in bioinformatics nor depend on commercial softwares for the handling and analysis of the sequence runs, because while NGS can easily be subcontracted, downstream analysis of the huge amount of sequence data generated often remains a major stumbling block for people dealing with pesticide resistance.

The purpose of this work was to develop massive resistance diagnosis assays associating the Illumina NGS technology and a simple, robust and versatile pipe-line for the analysis of the sequence reads. We considered acetolactate-synthase (ALS) inhibitors, the second most important herbicide mode of action globally[11] and the one that selected for resistance in the highest number of weed species (160 so far).[12] We developed resistance diagnosis assays for two contrasted weed species where resistance due to mutations in the gene encoding ALS inhibitors had evolved: common groundsel (*Senecio vulgaris*), an allotetraploid species with limited genetic variation,[13] and common ragweed (*Ambrosia artemisiifolia*), a diploid invasive species originating from North America that shows a much higher genetic polymorphism.[14]

2 MATERIALS AND METHODS

The whole NGS-based resistance diagnosis procedure developed in this work is summarised in figure 1.

2.1 Plant material

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

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

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

collection, leaves were wrapped in paper towels and mailed to our lab where they were stored at -20°C.

2.2 DNA extraction and PCR amplification of ALS fragments

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

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

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

0.2 µM final concentration each. As a classical, non-proofreading *Taq* polymerase was used for PCR, three independent PCR reactions were performed and subsequently pooled for each amplicon and each population.

For groundsel analyses, overhang adapter sequences were included at the 5' end of the three F primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and of the three R primers (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-). Tailed primers were used in a first PCR as described above. For each population, all three amplicons were pooled in an equimolar mix and tagged with a population-specific index using Nextera XT Index kit (v2) (Illumina). Tagged amplicons were pooled in an equimolar mix, purified with an Agencourt AMPure XP kit (Beckman Coulter, Beverly, MA, USA), quantified using the Qubit HS kit (Invitrogen) and loaded onto one Illumina MiSeq V2 cartridge according to the manufacturer instructions. The quality of the run was checked internally using PhiX. The resulting 250 nucleotide pair-end sequences passing Illumina standard quality controls were assigned to their population of origin on the basis of the population-specific indexes using bcl2fastq v2.20.0.422 (Illumina).

For ragweed analyses, overhang adapter sequences were also included at the 5' end of the three F primers (5'-CTTTCCCTACACGACGCTCTTCCGATC-) and of the three R primers (5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-). Tailed primers were used in a first PCR as described above. For each population, all three amplicons were pooled in an equimolar mix and tagged with a population-specific, home-made six-bp index in a second, 12-cycle PCR using primers 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC and 5'-CAAGCAGAAGACGGCATACGAGAT-XXXXXX-GTGACTGGAGTTCAGACGTGT where XXXXXX is the population-specific index sequence. Tagged amplicons were purified and loaded onto one Illumina MiSeq V3 cartridge according to the manufacturer instructions.

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

2.3 Analysis of sequence reads and identification of mutations

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

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

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

homeolog copy in 50 plants carrying two copies of said homeolog) and the current error rate of the Illumina sequencing technology (0.24% per nucleotide).[21]

2.4 Checking NGS-based mutation detection using dCAPS genotyping or Sanger sequencing of individual plants

Our Illumina-based resistance diagnosis assay was designed to assess frequencies of mutant ALS alleles in each groundsel or ragweed population analysed as a bulk of 50 plants. Therefore, sequence data was not recovered for individual plants. Including the control populations, 34 groundsel populations and 25 ragweed populations containing frequencies of mutations as contrasted as possible were used to check the reliability of assessing frequencies of mutant ALS alleles using Illumina sequencing (Tables 1, 2). DNA was extracted individually from each of the 50 plants in each of the populations used to check NGS results, except the control populations that had been characterised beforehand.

In groundsel, mutations conferring resistance to ALS inhibitors were previously exclusively found at codon 197 in ALS1 and/or ALS2.[15] Individual plant genotypes at ALS1 and ALS2 codon 197 were determined using previously described homeolog-specific dCAPS assays allowing detection of mutations at codon 197.[15] This allowed to compute the frequencies of mutations at each homeolog within each population. These frequencies were compared to those obtained by NGS analysis of the corresponding bulks of 50 plants.

In ragweed, 16 populations where mutant ALS alleles had or had not been detected were selected for ALS Sanger sequencing in addition to the nine control populations that had been characterised beforehand. Each of the three amplicons analysed by NGS were generated from each individual plant in three independent PCRs using primers in Table 3 and subsequently sequenced on both strands using Sanger sequencing.

Individual plant genotypes at ALS obtained by genotyping or sequencing were used to compute the exact frequencies of mutant ALS alleles within each population. These frequencies were compared to those estimated by NGS analysis of the corresponding bulks of 50 plants.

3 RESULTS

3.1 NGS-based detection and quantification of the frequency of mutations in groundsel ALS

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

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

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

investigated. In total, eight substitutions occurring exclusively at the first two nucleotides in codon 197 in ALS1 (three substitutions) and/or ALS2 (five substitutions) were identified, with up to four different substitutions detected in one same population (Table 4). The frequencies of the substitutions assessed by the pipe-line ranged from 1.0% to 99.8% (ALS1) and from 0.9% to 99.7% (ALS2). In six populations, C-to-A transversions at the first and at the second nucleotide in ALS2 codon 197 were always detected in matching frequencies, suggesting the occurrence of one double-substitution at codon 197 in ALS2 that would cause a proline-to-asparagine amino-acid substitution. This was confirmed by Sanger sequencing of the individual plants in these populations (Tables 1, 4). Thus, a total of eight mutations causing amino-acid substitutions at ALS codon 197 were identified in varying frequencies in the groundsel populations investigated. A Pro-197-Leu substitution at ALS1 detected in 35 populations was by far the most frequent mutation in the populations analysed.

A T-to-G transversion at the third nucleotide in codon 375 causing an aspartate-to-glutamate amino-acid substitution was identified in both ALS1 and ALS2 in 89 of the 96 populations analysed, in frequencies systematically comprised between 0.9 and 1.5%. Sanger sequencing in six populations where this substitution was detected by NGS did not allow to confirm its existence. The putative Asp-375-Glu substitution was therefore considered a PCR or NGS artifact, and discarded from the analyses.

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

The ragweed MiSeq run generated 20,700,536 250-nucleotide quality reads that could be assigned to one of the 96 populations investigated. The number of quality reads assigned to one population ranged from 147,116 to 274,998, with an average value of 215,631. Each population was analysed as one pool of 50 individual plants, each carrying two ALS gene copies. Thus,

the 50 plants in one pool yielded 300 amplicons. The average expected sequencing depth was therefore of 719× for each of the 300 amplicons per pool.

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

With the 0.9% threshold for mutation detection, non-synonymous nucleotide substitutions in codons crucial for herbicide sensitivity were identified as expected in the control populations. The only exception was control population C, where a G-to-T transversion at the second nucleotide in codon 574 causing a tryptophan-to-leucine amino-acid substitution that was present with a frequency of 1% was detected with a frequency of 0.63%, i.e., below the 0.9% threshold (Table 2). No mutations were detected in 79 of the 87 field populations investigated. In the remaining eight populations (all from Occitanie), a G-to-A transition at the first nucleotide in codon 205 causing an alanine-to-threonine amino-acid substitution was detected in frequencies ranging from 0.9 to 11.5% (Table 2). Mapping the eight populations from Occitanie investigated herein revealed that six populations were clustered within a radius of roughly one km within one river valley. The other two populations could be “escapees” disseminated by, e.g., agricultural machinery, to different river valleys 2.5 km south and 7.5 north from the main resistance focus, respectively (Figure 2).

3.3 Checking NGS results using individual plant analysis

The frequencies of the mutations detected by NGS-based analysis of populations from both species as pools of 50 plants were checked by selecting pools with contrasted frequencies of mutations and analysing each plant in each of these pool individually using a double-blind procedure. In groundsel, the frequencies of the mutations detected at each ALS homeolog in 33

field populations were checked by genotyping each plant in each population individually using ALS1- and ALS2-specific dCAPS assays targeting codon 197. dCAPS assays did not discriminate among different mutations occurring at one same codon.[15] Thus, the frequencies estimated by NGS of all mutations identified at codon 197 in each homeolog were cumulated for comparison to the dCAPS results. Similarly, in ragweed, the frequencies of mutations detected in 16 field populations by NGS analysis were checked by Sanger sequencing of the full ALS sequence of each plant in each pool. The frequencies of mutations identified by Illumina sequencing of plant pools and by individual plant analysis in field populations as well as in the control populations were subsequently compared (Tables 1 and 2, Figure 3). For the 34 groundsel populations (33 field populations and one control population) and the for the 25 ragweed populations (16 field populations and nine control populations) considered, the frequency of each mutation detected by NGS analysis of one bulk of 50 plants was very strongly correlated to the frequency determined by individual plant analysis (Figure 3; Pearson's correlation coefficients values of 0.996 [P-value < 10^{-6}] and 0.979 [P-value < 10^{-6}] for groundsel and ragweed data, respectively). No false positives were identified, i.e., detection of mutation(s) in one population by NGS analysis of the corresponding plant pool was always confirmed by individual plant sequencing or genotyping. However, NGS-based analysis did not detect mutations in two groundsel populations and one ragweed population where mutations were actually present with a frequency of 1% (i.e., false negatives; Tables 1 and 2). In all three cases, the mutations in question had been detected by NGS analysis, but their estimated frequencies ranged from 0.2 to 0.7% and were therefore below the threshold set for reliable mutation detection (0.9%). On the other hand, mutations present with a frequency of 1% were successfully detected by NGS analysis in two groundsel populations and seven ragweed populations (Tables 1 and 2).

4 DISCUSSION

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

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

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

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

the grass *Apera spica-venti*,[22] a double mutation at ALS codon 122 causing a Ala-122-Tyr substitution in *Raphanus raphanistrum*[23] and two combined mutations at codons 102 and 106 at the 5-enolpyruvylshikimate-3-phosphate synthase gene that encodes another herbicide target enzyme in the grass *Eleusine indica*.[24] Double-mutations in one pesticide gene have often been reported to confer an increased resistance level and/or a broader resistance spectrum, especially in fungicides (e.g.,[25]) but also in herbicides.[24] Herbicide sensitivity bioassays are needed to establish the resistance pattern associated to the double mutation identified in groundsel.

NGS analysis correctly identified a total of four amino-acid substitutions in ragweed. Three of them were detected in North American ragweed plants used to generate control populations (Table 1). A Trp-574-Leu substitution had previously been identified in this species.[26] An Asp-376-Glu substitution had been identified in 12 species so far,[17] while a Gly-654-Asp substitution had only been reported in the grass weed *Setaria viridis* to date.[27] The last substitution, Ala-205-Thr, was carried by French plants from the field and had never been reported before. Herbicide sensitivity bioassays are underway to establish its associated resistance pattern.

4.2 Reliability of NGS-based quantification of mutations endowing pesticide resistance

The frequencies of the ALS mutations assessed by NGS analysis of 50-plants pools matched very closely those determined by analysing the corresponding individual plants in both species investigated (Figure 3), even if some differences could be observed (Tables 1 and 2). The causes at the root of these discrepancies may be differences in the amount of plant material collected among individual plants, especially in groundsel where leaf thickness is rather variable, and/or a heterogeneous efficacy of DNA extraction and/or PCR. An unequal contribution of all plants in one pool to the mixture of amplicons generated for NGS is most likely the reason why NGS analysis of plant pools failed to detect mutations present with a frequency of 1% in two

groundsel and one ragweed population (false negatives, Tables 1, 2). Indeed, these mutations were actually detected by NGS, but with an estimated frequency that was below the 0.9% threshold set for reliable mutation detection. This flaw could be overcome by using a proofreading *Taq* polymerase in the PCRs preceding NGS, which would allow lowering the mutation detection threshold, and/or by slightly decreasing the number of plants per pool (e.g., 40 plants per pool, which would yield a minimum expected mutation frequency of 1.25% on a given gene or homeolog in the pool). Considering the tremendous sequencing depth achieved in this study for every ALS amplicon sequenced, analysis of two pools of 40 plants per population investigated rather than a single pool of 50 plants could be performed for 96 populations in one single Illumina run with a codon coverage adequate for mutation detection. This way, reducing the number of plants per pool could largely be compensated for by analysing more pools per population.

4.3 NGS *vs.* genotyping or Sanger sequencing

NGS-based analysis of plant pools detected and quantified resistance-endowing mutations in 9,600 individual plants clustered into 192 pools. This approach did not allow to determine the genotypes of the individual plants constituting the pool. This could be considered a drawback in the case of recessive resistances, where differentiating between homozygous and heterozygous mutants may be of some relevance. However, for resistance management in the field, the vital information is rather the detection of mutation(s) involved in pesticide resistance in one pest population. Such mutations are an unambiguous indicator that resistance is arising, and that steps are to be taken to delay its onset.

Herein, we targeted 15 codons in one ALS gene in 96 batches of 50 plants (ragweed) and in two ALS homeologs in 96 other batches of 50 plants (groundsel), which surpasses all preceding studies having applied NGS to pesticide resistance detection.[8,9,10] This merely

necessitated 192 DNA extractions followed by the production of three amplicons, each in triplicate PCR reactions per extraction, i.e., 1,728 PCRs, and two Illumina MiSeq runs. (It should be mentioned that, given the sequencing depth achieved, one single such run would have been sufficient to analyse all 192 pools). The same study conducted by the conventional single-codon PCR-based genotyping assays used for resistance detection (e.g.,[15]) would have required 9,600 DNA extractions. It would have necessitated, first, to set up and validate 15 codon-specific genotyping assays for ragweed and 30 homeolog-and-codon-specific genotyping assays for groundsel, then, to carry out and visualise 15 genotyping assays per ragweed plant and 30 per groundsel plant, i.e., a total of 216,000 PCRs (followed by the same number of restriction enzyme digestion reactions in the case of dCAPS). If Sanger sequencing had been used instead of genotyping, this study would have necessitated the production of three amplicons in triplicate PCR reactions in ragweed and six in groundsel (three per homeolog), i.e., 121,500 PCRs, followed by 40,500 Sanger sequencing runs. This clearly illustrates the power of NGS for diagnosing resistance-endowing mutations in numerous samples.

In contrast to preceding NGS studies addressing pesticide resistance diagnosis,[9,10] our work was carried out using a rapid DNA extraction procedure based on a non-commercial, lab-made extraction buffer.[16] The rough DNA extracts obtained proved perfectly suitable for resistance detection in pools of 50 individual plants. After Illumina sequencing, assignation of the sequence reads to their originating sample was performed using the software attached to the sequencer. This can be done in the case of an in-house facility, or requested from any sequencing supplier. The downstream analyses (read mapping on reference sequences, mutation detection and quantification) required no commercial software, our pipe-line being exclusively based on freely available programs and applications.

4.4 Developments of the NGS-based method

As it is described here, our NGS-based method only targets mutations at specific codons in the ALS gene. However, the scope of the method can be extended. First, any ALS nucleotide present in the amplicons sequenced can very easily be added to the list of positions to be screened by our pipe-line. This could reveal novel mutations of interest. Second, the pipe-line developed herein can be applied to other weed or pest species. As the level of nucleotide polymorphism varies with the species, this will require to adapt the alignment stringency to the gene(s) and to the weed or pest targeted. This can be done by running the pipe-line on a set of previously characterised populations by modifying options for the Bowtie short read aligner (see MutSeeker operating instructions). Third, the sequencing depth achieved by the Illumina technology clearly allows to sequence several genes in a single run. This has been shown herein by our analysing the two ALS homeologs in groundsel. Accordingly, several distinct pesticide targets can be screened in one single NGS run. Resistance can also be due to non-target-site based mechanisms.[28] When the mutations involved in these mechanisms are known, they can be subjected to NGS-based detection. Amplifying and sequencing by NGS as many amplicons as necessary to cover the genome regions including the nucleotide positions of interest is easily feasible, keeping in mind that the length of each amplicon must be lower than that of the NGS reads (i.e., lower than 500 nucleotides for 250 nucleotide pair-end sequences). If many genome regions of interest are targeted, then a compromise is to be found between the number of amplicons to be sequenced per population and the number of population to be sequenced per NGS run so that a satisfactory sequencing depth (e.g., 10× for each amplicon obtained from each individual plant in each pool) is achieved.

5 CONCLUSIONS

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

assays can be exploited for massive resistance surveys (i.e., analysing numerous pest or weed populations) as well as for in-depth resistance detection (i.e., analysing numerous pest or weed individuals in one population to detect to onset of resistance at very early stages, when resistance management practices can still nip resistance evolution in the bud). With the continuously-increasing amount of sequence reads generated by NGS runs and the technical possibility to multiplex hundreds of pools of individuals in one single NGS run, the era of massive resistance diagnosis is clearly just dawning. As shown in our work, the tremendous amounts of sequence reads generated can be analysed without the need for extensive training in NGS data analysis, which should extend the benefit of using NGS diagnosis to a large number of scientists dealing with pest control. Actually, the throughput of assays such as the one described herein is so high that the future main issue with NGS-based resistance detection might well be the possibility to obtain massive numbers of samples from the field. This issue could be overcome by pooling populations of different pests and weed species in shared NGS runs.

ACKNOWLEDGEMENTS

This study was partly funded by Bayer CropScience France (groundsel sampling), by ANSES *via* the system for monitoring the adverse effects of plant protection products ('phytopharmacovigilance', established by the French Act on the future of agriculture of 13 October 2014) (groundsel sequencing) and by BASF SAS France (ragweed sequencing). We thank the Gentyane facility (Clermont-Ferrand, France [http://gentyane.clermont.inra.fr/]) and the GeT core facility (Toulouse, France [http://get.genotoul.fr]) for providing access to their sequencing equipment.

REFERENCES

1 Oerke E-C, Crop losses to pests. *J Agric Sci* **144**:31-43 (2006).

2 Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N and Nelson A, The global burden of pathogens and pests on major food crops. *Nature Ecol Evol* **3**:430-439 (2019).

3 R4P Network, Trends and challenges in pesticide resistance detection. *Trends Plant Sci* **21**:834-853 (2016).

4 Metzker ML, Sequencing technologies - the next generation. *Nature Rev Genet* **11**:31-46 (2010).

5 Kamps R, Brandão RD, van den Bosch BJ, Paulussen ADC, Xanthoulea S, Blok MJ and Romano A, Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. *Int J Mol Sci* **18**:Article308 (2017).

6 Sahoo MK, Lefterova MI, Yamamoto F, Waggoner JJ, Chou S, Holmes SP, Anderson MW and Pinsky BA, Detection of cytomegalovirus drug resistance mutations by next-generation sequencing. *J Clin Microbiol* **51**:3700-3710 (2013).

7 Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, Crudu V, Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC, Catanzaro A, Keim P and Engelthaler DM, Detection of low-level mixed-population drug resistance in *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS ONE* **10**:e0126626 (2015).

8 Délye C, Causse R, Gautier V, Poncet C and Michel S, Using next-generation sequencing to detect mutations endowing resistance to pesticides: application to acetolactate-synthase (ALS)-based resistance in barnyardgrass, a polyploid grass weed. *Pest Manag Sci* **71**:675-685 (2015).

9 Pieczul K and Waşowska A, The application of next-generation sequencing (NGS) for monitoring of *Zymoseptoria tritici* QoI resistance. *Crop Prot* **92**:143-147 (2017).

10 Schlipalius DI, Tuck AG, Pavic H, Daglish GJ, Nayaka MK and Ebert PR, A high-throughput system used to determine frequency and distribution of phosphine resistance across large geographical regions. *Pest Manag Sci* **75**:1091-1098 (2019).

11 Kraehmer H, Laber B, Rosinger C and Schulz A, Herbicides as weed control agents: state of the art: I. weed control research and safener technology: the path to modern agriculture. *Plant Physiol* **166**:1119-1131 (2014).

12 Heap I, International survey of herbicide-resistant weeds. Available online at: http://www.weedscience.com. Accessed 11 April 2019.

13 Zhu B-R, Barrett SCH, Zhang D-Y and Liao W-J, Invasion genetics of *Senecio vulgaris*: loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple introductions. *Biol Invasions* **19**:255-267 (2017).

14 Meyer L, Causse R, Pernin F, Scalone R, Bailly G, Chauvel B, Délye C and Le Corre V, New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species. *PLoS ONE* **12**:e0176197 (2017).

15 Délye C, Causse R, Gautier V, Poncet C and Michel S, Genetic basis, evolutionary origin and spread of resistance to herbicides inhibiting acetolactate synthase in common groundsel (*Senecio vulgaris*). *Pest Manag Sci* **72**:89-102 (2016).

16 Délye C, Matéjicek A and Gasquez J, PCR-based detection of resistance to Acetyl-CoA carboxylase-inhibiting herbicides in black-grass (*Alopecurus myosuroides* Huds) and ryegrass (*Lolium rigidum* Gaud). *Pest Manag Sci* **58**:474-478 (2002).

17 Tranel PJ, Wright TR and Heap IM, Mutations in herbicide-resistant weeds to ALS inhibitors. Available online at: http://www.weedscience.com. Accessed 11 April 2019.

18 Duggleby RG, McCourt JA and Guddat LW, Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiol Biochem* **46**:309-324 (2008).

19 Gaidatzis DA, Lerch F, Hahne F and Stadler MB, QuasR: quantify and annotate short reads in R. *Bioinformatics* **31**:1130-1132 (2015).

20 Morgan M, Pagès H, Obenchain V and Hayden N, Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 1.34.1, Available online at: http://bioconductor.org/packages/release/bioc/html/Rsamtools.html. Accessed 11 April 2019.

21 Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL and Mayer G, Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci Rep* **8**:10950 (2018).

22 Massa D, Krenz B and Gerhards R, Target-site resistance to ALS-inhibiting herbicides in *Apera spica-venti* populations is conferred by documented and previously unknown mutations. *Weed Res* **51**:294-303 (2011).

23 Han H, Yu Q, Purba E, Li M, Walsh M, Friesen S and Powles SB, A novel amino acid substitution Ala-122-Tyr in ALS confers high-level and broad resistance across ALS-inhibiting herbicides. *Pest Manag Sci* **68**:1164-1170 (2012).

24 Han H, Vila-Aiub MM, Jalaludin A, Yu Q and Powles SB, A double EPSPS gene mutation endowing glyphosate resistance shows a remarkably high resistance cost. *Plant Cell Environ* **40**:3031-3042 (2017).

25 Leroux P and Walker A-S, Multiple mechanisms account for resistance to sterol 14α-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. *Pest Manag Sci* **67**:44-59 (2011).

26 Patzoldt WL, Tranel PJ, Alexander AL and Schmitzer PR, A common ragweed population resistant to cloransulam-methyl. *Weed Sci* **49**:485-490 (2001).

27 Laplante J, Rajcan I, and Tardif FJ, Multiple allelic forms of acetohydroxyacid synthase are responsible for herbicide resistance in *Setaria viridis*. *Theor Appl Genet* **119**:577-585 (2009).

28 Délye C, Duhoux A, Pernin F, Riggins CW and Tranel PJ, Molecular mechanisms of herbicide resistance. *Weed Sci* **63**:91-115 (2015).

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 [c]			% mutant detected		Nb. Kreads per codon [c]		
Code [a]	Type [b]	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 [d]	F	0.0	0.0	18.1	6.2	7.7	78.0	67.0	17.1	4.8	9.8
RA-05 [d]	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 [d]	F	7.0	1.5	9.6	10.4	24.5	92.0	93.6	8.3	9.5	22.2
RA-12 [d]	F	0.0	0.0	18.8	18.2	21.3	7.0	11.6	20.2	20.3	20.7
RA-13 [d]	F	0.0	0.0	9.5	18.4	17.8	50.0	43.2	12.2	19.1	18.3
RA-15 [d]	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

RA-20	F	0.0	0.0	17.9	22.9	20.9	0.0	0.0	14.8	20.6	17.9
RA-22	F	0.0	0.0	22.7	7.1	13.3	0.0	0.0	17.7	7.5	11.2
RA-24	F	0.0	0.0	27.8	17.2	14.3	0.0	0.0	27.3	17.0	14.9
BZ-I	F	16.0	12.3	17.9	13.5	11.2	0.0	0.0	14.4	12.0	11.9
BZ-IV	F	0.0	0.0	16.0	11.7	19.8	0.0	0.0	18.4	11.5	19.9
BZ-VI	F	96.0	96.6	19.1	12.9	17.6	0.0	0.0	22.2	10.3	16.8
BZ-VII	F	58.0	57.9	12.9	15.5	23.9	0.0	0.0	9.7	15.1	19.9
BZ-VIII	F	100.0	99.8	29.0	35.7	22.1	1.0	0.9	26.2	38.8	23.4
BZ-04	F	2.0	2.4	4.7	21.1	12.3	0.0	0.0	4.5	19.2	12.7
BZ-05	F	1.0	1.0	10.7	10.7	14.0	0.0	0.0	13.2	8.4	13.5
BZ-09	F	51.0	57.0	16.2	7.4	7.0	0.0	0.0	14.5	6.7	6.6
BZ-13	F	5.0	1.0	8.9	2.4	4.5	0.0	0.0	10.0	2.6	4.7
BZ-24	F	0.0	0.0	38.1	6.3	18.5	2.0	0.9	32.7	4.8	15.1
BZ-38	F	2.0	1.0	12.1	22.6	21.5	1.0	0.0	11.0	20.2	21.7
BZ-57	F	0.0	0.0	7.6	6.6	5.3	4.0	3.0	8.7	7.1	6.0
BZ-59	F	91.0	88.4	11.1	9.9	7.2	7.0	2.8	14.5	10.0	8.7

BZ-61	F	49.0	46.4	9.9	11.9	8.8	22.0	18.9	9.3	11.6	9.0
BZ-66	F	5.0	5.7	18.0	32.1	10.1	1.0	0.0	15.4	31.2	11.3
BZ-67	F	7.0	2.0	19.8	13.0	0.7	0.0	0.0	19.7	13.2	0.8
BZ-69	F	4.0	2.3	17.3	12.2	7.7	0.0	0.0	21.0	12.4	7.5
BZ-71	F	4.0	2.5	17.5	18.6	7.9	0.0	0.0	16.4	17.6	7.5
BZ-73	F	20.0	31.3	5.5	6.8	12.4	5.0	9.1	11.0	6.3	10.6

[a] RA, Rhône-Alpes; BZ, Brittany.

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

[c] 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 reads were not joined.

[d] Populations with plants containing a double mutation causing a Pro-197-Asn substitution at ALS2. These populations were also subjected to Sanger sequencing of individual plants.

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

Populations		ALS				
		% mutant detected [c]		Nb. Kreads per codon [d]		
Code [a]	Type [b]	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

OC-27	F	A205T (1.0)	A205T (0.9)	29.5	60.1	25.4
OC-30	F	A205T (2.0)	A205T (1.7)	34.8	68.3	31.3
OC-43	F	A205T (7.0)	A205T (5.7)	29.5	58.7	25.9
OC-55	F	0.0	0.0	30.4	56.7	32.8
OC-60	F	A205T (3.0)	A205T (1.1)	29.9	57.2	26.8
OC-65	F	0.0	0.0	30.0	59.7	28.7
NA-25	F	0.0	0.0	28.9	71.6	27.7
NA-28	F	0.0	0.0	29.3	64.3	32.2
NA-29	F	0.0	0.0	27.4	56.2	27.2
NA-30	F	0.0	0.0	31.8	67.6	30.9
NA-39	F	0.0	0.0	30.2	66.4	27.7

[a] NA, Nouvelle Aquitaine; OC, Occitanie.

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

[c] The mutations identified are indicated with their frequency between parentheses.

[d] 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 reads were not joined.

Table 3. PCR primers.

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

[a] Nucleotides in GenBank/EMBL accession KR024410 or KR024411 (groundsel ALS1 or ALS2 partial sequences, respectively) or in accession KX870184 (ragweed full ALS sequence).

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

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

Mutation(s) [a]	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 [b] + Pro-197-Leu-2	2
Pro-197-Asn-2 [b] + Pro-197-Leu-2 + Pro-197-Ser-2	1
Pro-197-Leu-1 + Pro-197-Asn-2 [b] + Pro-197-Leu-2 + Pro-197-Ser-2	1

[a] The number following the mutation refers to the ALS homeolog (1, ALS1; 2, ALS2).

[b] Substitution caused by a double mutation.

Legends to Figures

Figure 1. Flow-chart of the NGS-based resistance diagnosis procedure.

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

Figure 3. Frequencies of mutations at the ALS gene detected using individual plant analysis (dCAPS or Sanger sequencing) plotted against the frequencies assessed using Illumina sequencing of pools of 50 plants. In groundsel, frequencies are displayed separately for each ALS homeolog. Because dCAPS genotyping does not allow to discriminate among different mutations at codon 197, the frequencies displayed are the cumulated frequencies of all mutations detected at codon 197. In ragweed, the frequencies are displayed separately for each mutation when several mutations are present in one same population.

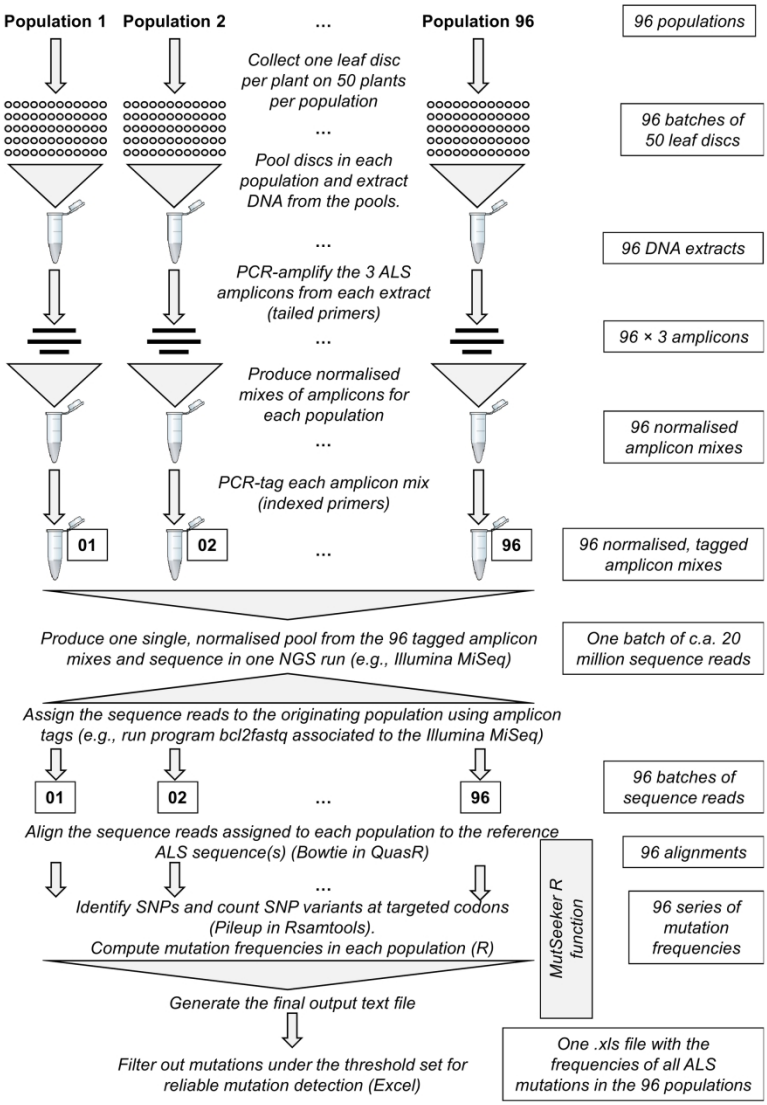

Figure 1

209x297mm (300 x 300 DPI)

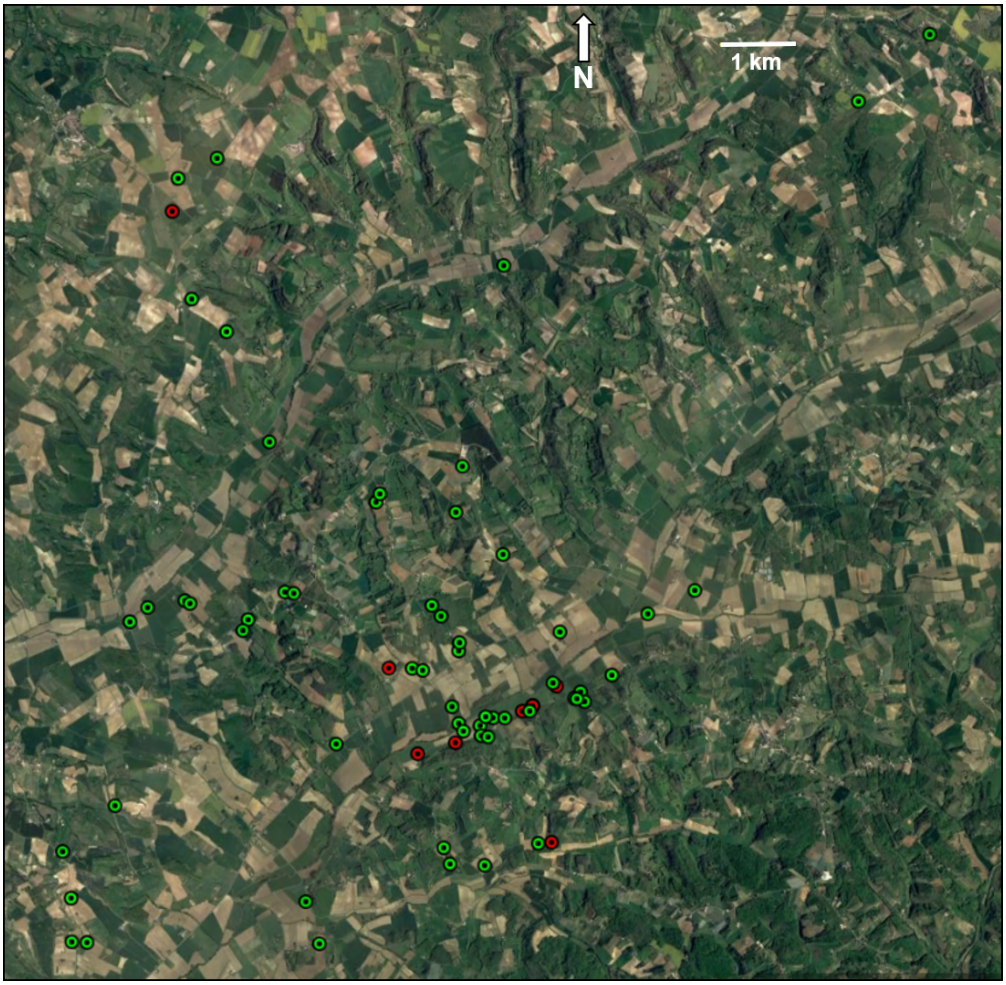

Figure 2

185x181mm (150 x 150 DPI)

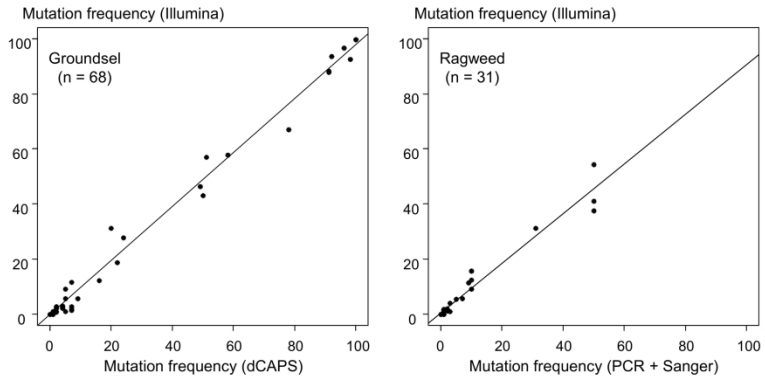

Figure 3

209x297mm (300 x 300 DPI)

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

Running title: Illumina-based pesticide resistance diagnosis

Christophe Délye, [a]* Séverine Michel, [a] Fanny Pernin, [a] Véronique Gautier, [b] Marie Gislard, [c] Charles Poncet [b] and Valérie Le Corre [a]

[a] INRA, UMR Agroécologie, 17 rue de Sully, F-21000 Dijon, France.

[b] INRA, UMR1095 Génétique, Diversité et Écophysiologie des Céréales, Clermont-Ferrand, France.

[c] INRA, US 1426 GeT-PlaGe, Genotoul, Castanet-Tolosan, France

* Correspondence to:

Dr Christophe Délye

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

Phone, +33 380 693 185

Fax, +33 380 693 262

E-mail, christophe.delye@inra.fr

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.[1,2] 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.[3] 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.[4,5] 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.,[5]) or of mutations involved in the resistance of pathogens to drugs used in medical treatments (e.g.,[6,7]). Although this type of application of NGS has also been proposed for pesticide resistance diagnosis in agricultural pests,[3] 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.[8] 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.

The c.a. 64,000 sequence reads obtained were analysed through a very basic bioinformatic pipeline. The sequencing depth obtained was moderate (1.7× to 6.2× per individual plant and per codon of interest), but the mutation frequencies identified by NGS matched those obtained by Sanger sequencing of individual plants, thereby indicating that NGS can be used to detect and quantify mutations endowing pesticide resistance in pools of individuals. The second study applied the Illumina technology to the detection of mutations at one codon endowing fungicide resistance in *Zymoseptoria tritici*.[9] About 723,000 quality sequence reads were obtained and used to analyse 40 fungal isolates from various geographical origins. Despite the authors indicating that sequence reads could have been attributed to individual isolates by using appropriate tagging, the 40 isolates were analysed as one single bulk. This work illustrated the potential of the Illumina technology for pesticide resistance diagnosis. The third study sought mutations at five codons in one gene involved in insecticide resistance in *Rhyzopertha dominica*.[10] A total of 1,435 individual insects were analysed using Illumina sequencing. Using tagging, the unspecified number of quality sequence reads obtained could be attributed to each individual insect, thus allowing individual genotype determination at every codon screened. This study further confirmed the feasibility of multiple individual analysis in one single NGS run.

The three preceding pioneer works clearly demonstrated the interest and feasibility of pesticide resistance diagnosis by genotyping-by-sequencing using the NGS technologies combined with sequence read tagging that allows to sequence hosts of individuals or populations as one bulk and subsequently trace the individual or population at the source of each sequence read. The pending question is, how can this approach be most efficiently used for pesticide resistance diagnosis? Assuming that mutation(s) endowing pesticide resistance have been characterised beforehand, an ideal NGS-based resistance assay would allow to screen as many individuals of the pest considered as possible in one single NGS run. This can be

achieved by screening populations rather than individuals, and by pooling and simultaneously sequencing populations in one single NGS run. An ideal NGS-based assay would also be reliable and yield a fair estimate of the frequency of resistance-endowing mutations in pest populations. In particular, the assay must not generate false positives or false negatives. Here, the error rate of the NGS technique will determine the lowest mutation frequency that can be reliably detected. Last, an ideal NGS-based assay should not require extensive skills in bioinformatics nor depend on commercial softwares for the handling and analysis of the sequence runs, because while NGS can easily be subcontracted, downstream analysis of the huge amount of sequence data generated often remains a major stumbling block for people dealing with pesticide resistance.

The purpose of this work was to develop massive resistance diagnosis assays associating the Illumina NGS technology and a simple, robust and versatile pipe-line for the analysis of the sequence reads. We considered acetolactate-synthase (ALS) inhibitors, the second most important herbicide mode of action globally[11] and the one that selected for resistance in the highest number of weed species (160 so far).[12] We developed resistance diagnosis assays for two contrasted weed species where resistance due to mutations in the gene encoding ALS inhibitors had evolved: common groundsel (*Senecio vulgaris*), an allotetraploid species with limited genetic variation,[13] and common ragweed (*Ambrosia artemisiifolia*), a diploid invasive species originating from North America that shows a much higher genetic polymorphism.[14]

2 MATERIALS AND METHODS

The whole NGS-based resistance diagnosis procedure developed in this work is summarised in figure 1.

2.1 Plant material

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

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

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

collection, leaves were wrapped in paper towels and mailed to our lab where they were stored at -20°C.

2.2 DNA extraction and PCR amplification of ALS fragments

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

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

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

0.2 µM final concentration each. As a classical, non-proofreading *Taq* polymerase was used for PCR, three independent PCR reactions were performed and subsequently pooled for each amplicon and each population.

For groundsel analyses, overhang adapter sequences were included at the 5' end of the three F primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and of the three R primers (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-). Tailed primers were used in a first PCR as described above. For each population, all three amplicons were pooled in an equimolar mix and tagged with a population-specific index using Nextera XT Index kit (v2) (Illumina). Tagged amplicons were pooled in an equimolar mix, purified with an Agencourt AMPure XP kit (Beckman Coulter, Beverly, MA, USA), quantified using the Qubit HS kit (Invitrogen) and loaded onto one Illumina MiSeq V2 cartridge according to the manufacturer instructions. The quality of the run was checked internally using PhiX. The resulting 250 nucleotide pair-end sequences passing Illumina standard quality controls were assigned to their population of origin on the basis of the population-specific indexes using bcl2fastq v2.20.0.422 (Illumina).

For ragweed analyses, overhang adapter sequences were also included at the 5' end of the three F primers (5'-CTTTCCCTACACGACGCTCTTCCGATC-) and of the three R primers (5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-). Tailed primers were used in a first PCR as described above. For each population, all three amplicons were pooled in an equimolar mix and tagged with a population-specific, home-made six-bp index in a second, 12-cycle PCR using primers 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC and 5'-CAAGCAGAAGACGGCATACGAGAT-XXXXXX-GTGACTGGAGTTCAGACGTGT where XXXXXX is the population-specific index sequence. Tagged amplicons were purified and loaded onto one Illumina MiSeq V3 cartridge according to the manufacturer instructions.

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

2.3 Analysis of sequence reads and identification of mutations

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

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

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

homeolog copy in 50 plants carrying two copies of said homeolog) and the current error rate of the Illumina sequencing technology (0.24% per nucleotide).[21]

2.4 Checking NGS-based mutation detection using dCAPS genotyping or Sanger sequencing of individual plants

Our Illumina-based resistance diagnosis assay was designed to assess frequencies of mutant ALS alleles in each groundsel or ragweed population analysed as a bulk of 50 plants. Therefore, sequence data was not recovered for individual plants. Including the control populations, 34 groundsel populations and 25 ragweed populations containing frequencies of mutations as contrasted as possible were used to check the reliability of assessing frequencies of mutant ALS alleles using Illumina sequencing (Tables 1, 2). DNA was extracted individually from each of the 50 plants in each of the populations used to check NGS results, except the control populations that had been characterised beforehand.

In groundsel, mutations conferring resistance to ALS inhibitors were previously exclusively found at codon 197 in ALS1 and/or ALS2.[15] Individual plant genotypes at ALS1 and ALS2 codon 197 were determined using previously described homeolog-specific dCAPS assays allowing detection of mutations at codon 197.[15] This allowed to compute the frequencies of mutations at each homeolog within each population. These frequencies were compared to those obtained by NGS analysis of the corresponding bulks of 50 plants.

In ragweed, 16 populations where mutant ALS alleles had or had not been detected were selected for ALS Sanger sequencing in addition to the nine control populations that had been characterised beforehand. Each of the three amplicons analysed by NGS were generated from each individual plant in three independent PCRs using primers in Table 3 and subsequently sequenced on both strands using Sanger sequencing.

Individual plant genotypes at ALS obtained by genotyping or sequencing were used to compute the exact frequencies of mutant ALS alleles within each population. These frequencies were compared to those estimated by NGS analysis of the corresponding bulks of 50 plants.

3 RESULTS

3.1 NGS-based detection and quantification of the frequency of mutations in groundsel ALS

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

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

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

investigated. In total, eight substitutions occurring exclusively at the first two nucleotides in codon 197 in ALS1 (three substitutions) and/or ALS2 (five substitutions) were identified, with up to four different substitutions detected in one same population (Table 4). The frequencies of the substitutions assessed by the pipe-line ranged from 1.0% to 99.8% (ALS1) and from 0.9% to 99.7% (ALS2). In six populations, C-to-A transversions at the first and at the second nucleotide in ALS2 codon 197 were always detected in matching frequencies, suggesting the occurrence of one double-substitution at codon 197 in ALS2 that would cause a proline-to-asparagine amino-acid substitution. This was confirmed by Sanger sequencing of the individual plants in these populations (Tables 1, 4). Thus, a total of eight mutations causing amino-acid substitutions at ALS codon 197 were identified in varying frequencies in the groundsel populations investigated. A Pro-197-Leu substitution at ALS1 detected in 35 populations was by far the most frequent mutation in the populations analysed.

A T-to-G transversion at the third nucleotide in codon 375 causing an aspartate-to-glutamate amino-acid substitution was identified in both ALS1 and ALS2 in 89 of the 96 populations analysed, in frequencies systematically comprised between 0.9 and 1.5%. Sanger sequencing in six populations where this substitution was detected by NGS did not allow to confirm its existence. The putative Asp-375-Glu substitution was therefore considered a PCR or NGS artifact, and discarded from the analyses.

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

The ragweed MiSeq run generated 20,700,536 250-nucleotide quality reads that could be assigned to one of the 96 populations investigated. The number of quality reads assigned to one population ranged from 147,116 to 274,998, with an average value of 215,631. Each population was analysed as one pool of 50 individual plants, each carrying two ALS gene copies. Thus,

the 50 plants in one pool yielded 300 amplicons. The average expected sequencing depth was therefore of 719× for each of the 300 amplicons per pool.

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

With the 0.9% threshold for mutation detection, non-synonymous nucleotide substitutions in codons crucial for herbicide sensitivity were identified as expected in the control populations. The only exception was control population C, where a G-to-T transversion at the second nucleotide in codon 574 causing a tryptophan-to-leucine amino-acid substitution that was present with a frequency of 1% was detected with a frequency of 0.63%, i.e., below the 0.9% threshold (Table 2). No mutations were detected in 79 of the 87 field populations investigated. In the remaining eight populations (all from Occitanie), a G-to-A transition at the first nucleotide in codon 205 causing an alanine-to-threonine amino-acid substitution was detected in frequencies ranging from 0.9 to 11.5% (Table 2). Mapping the eight populations from Occitanie investigated herein revealed that six populations were clustered within a radius of roughly one km within one river valley. The other two populations could be "escapees" disseminated by, e.g., agricultural machinery, to different river valleys 2.5 km south and 7.5 north from the main resistance focus, respectively (Figure ~~1~~2).

3.3 Checking NGS results using individual plant analysis

The frequencies of the mutations detected by NGS-based analysis of populations from both species as pools of 50 plants were checked by selecting pools with contrasted frequencies of mutations and analysing each plant in each of these pool individually using a double-blind procedure. In groundsel, the frequencies of the mutations detected at each ALS homeolog in 33

field populations were checked by genotyping each plant in each population individually using ALS1- and ALS2-specific dCAPS assays targeting codon 197. dCAPS assays did not discriminate among different mutations occurring at one same codon.[15] Thus, the frequencies estimated by NGS of all mutations identified at codon 197 in each homeolog were cumulated for comparison to the dCAPS results. Similarly, in ragweed, the frequencies of mutations detected in 16 field populations by NGS analysis were checked by Sanger sequencing of the full ALS sequence of each plant in each pool. The frequencies of mutations identified by Illumina sequencing of plant pools and by individual plant analysis in field populations as well as in the control populations were subsequently compared (Tables 1 and 2, Figure ~~2~~3). For the 34 groundsel populations (33 field populations and one control population) and the for the 25 ragweed populations (16 field populations and nine control populations) considered, the frequency of each mutation detected by NGS analysis of one bulk of 50 plants was very strongly correlated to the frequency determined by individual plant analysis (Figure ~~2~~3; Pearson's correlation coefficients values of 0.996 [P-value < 10^{-6}] and 0.979 [P-value < 10^{-6}] for groundsel and ragweed data, respectively). No false positives were identified, i.e., detection of mutation(s) in one population by NGS analysis of the corresponding plant pool was always confirmed by individual plant sequencing or genotyping. However, NGS-based analysis did not detect mutations in two groundsel populations and one ragweed population where mutations were actually present with a frequency of 1% (i.e., false negatives; Tables 1 and 2). In all three cases, the mutations in question had been detected by NGS analysis, but their estimated frequencies ranged from 0.2 to 0.7% and were therefore below the threshold set for reliable mutation detection (0.9%). On the other hand, mutations present with a frequency of 1% were successfully detected by NGS analysis in two groundsel populations and seven ragweed populations (Tables 1 and 2).

4 DISCUSSION

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

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

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

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

the grass *Apera spica-venti*,[22] a double mutation at ALS codon 122 causing a Ala-122-Tyr substitution in *Raphanus raphanistrum*[23] and two combined mutations at codons 102 and 106 at the 5-enolpyruvylshikimate-3-phosphate synthase gene that encodes another herbicide target enzyme in the grass *Eleusine indica*.[24] Double-mutations in one pesticide gene have often been reported to confer an increased resistance level and/or a broader resistance spectrum, especially in fungicides (e.g.,[25]) but also in herbicides.[24] Herbicide sensitivity bioassays are needed to establish the resistance pattern associated to the double mutation identified in groundsel.

NGS analysis correctly identified a total of four amino-acid substitutions in ragweed. Three of them were detected in North American ragweed plants used to generate control populations (Table 1). A Trp-574-Leu substitution had previously been identified in this species.[26] An Asp-376-Glu substitution had been identified in 12 species so far,[17] while a Gly-654-Asp substitution had only been reported in the grass weed *Setaria viridis* to date.[27] The last substitution, Ala-205-Thr, was carried by French plants from the field and had never been reported before. Herbicide sensitivity bioassays are underway to establish its associated resistance pattern.

4.2 Reliability of NGS-based quantification of mutations endowing pesticide resistance

The frequencies of the ALS mutations assessed by NGS analysis of 50-plants pools matched very closely those determined by analysing the corresponding individual plants in both species investigated (Figure 3~~2~~), even if some differences could be observed (Tables 1 and 2). The causes at the root of these discrepancies may be differences in the amount of plant material collected among individual plants, especially in groundsel where leaf thickness is rather variable, and/or a heterogeneous efficacy of DNA extraction and/or PCR. An unequal contribution of all plants in one pool to the mixture of amplicons generated for NGS is most likely the reason why NGS analysis of plant pools failed to detect mutations present with a

frequency of 1% in two groundsel and one ragweed population (false negatives, Tables 1, 2). Indeed, these mutations were actually detected by NGS, but with an estimated frequency that was below the 0.9% threshold set for reliable mutation detection. This flaw could be overcome by using a proofreading *Taq* polymerase in the PCRs preceding NGS, which would allow lowering the mutation detection threshold, and/or by slightly decreasing the number of plants per pool (e.g., 40 plants per pool, which would yield a minimum expected mutation frequency of 1.25% on a given gene or homeolog in the pool). Considering the tremendous sequencing depth achieved in this study for every ALS amplicon sequenced, analysis of two pools of 40 plants per population investigated rather than a single pool of 50 plants could be performed for 96 populations in one single Illumina run with a codon coverage adequate for mutation detection. This way, reducing the number of plants per pool could largely be compensated for by analysing more pools per population.

4.3 NGS *vs.* genotyping or Sanger sequencing

NGS-based analysis of plant pools detected and quantified resistance-endowing mutations in 9,600 individual plants clustered into 192 pools. This approach did not allow to determine the genotypes of the individual plants constituting the pool. This could be considered a drawback in the case of recessive resistances, where differentiating between homozygous and heterozygous mutants may be of some relevance. However, for resistance management in the field, the vital information is rather the detection of mutation(s) involved in pesticide resistance in one pest population. Such mutations are an unambiguous indicator that resistance is arising, and that steps are to be taken to delay its onset.

Herein, we targeted 15 codons in one ALS gene in 96 batches of 50 plants (ragweed) and in two ALS homeologs in 96 other batches of 50 plants (groundsel), which surpasses all preceding studies having applied NGS to pesticide resistance detection.[8,9,10] This merely

necessitated 192 DNA extractions followed by the production of three amplicons, each in triplicate PCR reactions per extraction, i.e., 1,728 PCRs, and two Illumina MiSeq runs. (It should be mentioned that, given the sequencing depth achieved, one single such run would have been sufficient to analyse all 192 pools). The same study conducted by the conventional single-codon PCR-based genotyping assays used for resistance detection (e.g.,[15]) would have required 9,600 DNA extractions. It would have necessitated, first, to set up and validate 15 codon-specific genotyping assays for ragweed and 30 homeolog-and-codon-specific genotyping assays for groundsel, then, to carry out and visualise 15 genotyping assays per ragweed plant and 30 per groundsel plant, i.e., a total of 216,000 PCRs (followed by the same number of restriction enzyme digestion reactions in the case of dCAPS). If Sanger sequencing had been used instead of genotyping, this study would have necessitated the production of three amplicons in triplicate PCR reactions in ragweed and six in groundsel (three per homeolog), i.e., 121,500 PCRs, followed by 40,500 Sanger sequencing runs. This clearly illustrates the power of NGS for diagnosing resistance-endowing mutations in numerous samples.

In contrast to preceding NGS studies addressing pesticide resistance diagnosis,[9,10] our work was carried out using a rapid DNA extraction procedure based on a non-commercial, lab-made extraction buffer.[16] The rough DNA extracts obtained proved perfectly suitable for resistance detection in pools of 50 individual plants. After Illumina sequencing, assignation of the sequence reads to their originating sample was performed using the software attached to the sequencer. This can be done in the case of an in-house facility, or requested from any sequencing supplier. The downstream analyses (read mapping on reference sequences, mutation detection and quantification) required no commercial software, our pipe-line being exclusively based on freely available programs and applications.

4.4 Developments of the NGS-based method

As it is described here, our NGS-based method only targets mutations at specific codons in the ALS gene. However, the scope of the method can be extended. First, any ALS nucleotide present in the amplicons sequenced can very easily be added to the list of positions to be screened by our pipe-line. This could reveal novel mutations of interest. Second, the pipe-line developed herein can be applied to other weed or pest species. As the level of nucleotide polymorphism varies with the species, this will require to adapt the alignment stringency to the gene(s) and to the weed or pest targeted. This can be done by running the pipe-line on a set of previously characterised populations by modifying options for the Bowtie short read aligner (see MutSeeker operating instructions). Third, the sequencing depth achieved by the Illumina technology clearly allows to sequence several genes in a single run. This has been shown herein by our analysing the two ALS homeologs in groundsel. Accordingly, several distinct pesticide targets can be screened in one single NGS run. Resistance can also be due to non-target-site based mechanisms.[28] When the mutations involved in these mechanisms are known, they can be subjected to NGS-based detection. Amplifying and sequencing by NGS as many amplicons as necessary to cover the genome regions including the nucleotide positions of interest is easily feasible, keeping in mind that the length of each amplicon must be lower than that of the NGS reads (i.e., lower than 500 nucleotides for 250 nucleotide pair-end sequences). If many genome regions of interest are targeted, then a compromise is to be found between the number of amplicons to be sequenced per population and the number of population to be sequenced per NGS run so that a satisfactory sequencing depth (e.g., 10× for each amplicon obtained from each individual plant in each pool) is achieved.

5 CONCLUSIONS

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

assays can be exploited for massive resistance surveys (i.e., analysing numerous pest or weed populations) as well as for in-depth resistance detection (i.e., analysing numerous pest or weed individuals in one population to detect to onset of resistance at very early stages, when resistance management practices can still nip resistance evolution in the bud). With the continuously-increasing amount of sequence reads generated by NGS runs and the technical possibility to multiplex hundreds of pools of individuals in one single NGS run, the era of massive resistance diagnosis is clearly just dawning. As shown in our work, the tremendous amounts of sequence reads generated can be analysed without the need for extensive training in NGS data analysis, which should extend the benefit of using NGS diagnosis to a large number of scientists dealing with pest control. Actually, the throughput of assays such as the one described herein is so high that the future main issue with NGS-based resistance detection might well be the possibility to obtain massive numbers of samples from the field. This issue could be overcome by pooling populations of different pests and weed species in shared NGS runs.

ACKNOWLEDGEMENTS

This study was partly funded by Bayer CropScience France (groundsel sampling), by ANSES *via* the system for monitoring the adverse effects of plant protection products ('phytopharmacovigilance', established by the French Act on the future of agriculture of 13 October 2014) and (groundsel sequencing) and by BASF SAS France (ragweed sequencing). We thank the Gentyane facility (Clermont-Ferrand, France [http://gentyane.clermont.inra.fr/]) and the GeT core facility (Toulouse, France [http://get.genotoul.fr]) for providing access to their sequencing equipment.

REFERENCES

1 Oerke E-C, Crop losses to pests. *J Agric Sci* **144**:31-43 (2006).

2 Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N and Nelson A, The global burden of pathogens and pests on major food crops. *Nature Ecol Evol* **3**:430-439 (2019).

3 R4P Network, Trends and challenges in pesticide resistance detection. *Trends Plant Sci* **21**:834-853 (2016).

4 Metzker ML, Sequencing technologies - the next generation. *Nature Rev Genet* **11**:31-46 (2010).

5 Kamps R, Brandão RD, van den Bosch BJ, Paulussen ADC, Xanthoulea S, Blok MJ and Romano A, Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. *Int J Mol Sci* **18**:Article308 (2017).

6 Sahoo MK, Lefterova MI, Yamamoto F, Waggoner JJ, Chou S, Holmes SP, Anderson MW and Pinsky BA, Detection of cytomegalovirus drug resistance mutations by next-generation sequencing. *J Clin Microbiol* **51**:3700-3710 (2013).

7 Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, Crudu V, Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC, Catanzaro A, Keim P and Engelthaler DM, Detection of low-level mixed-population drug resistance in *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS ONE* **10**:e0126626 (2015).

8 Délye C, Causse R, Gautier V, Poncet C and Michel S, Using next-generation sequencing to detect mutations endowing resistance to pesticides: application to acetolactate-synthase (ALS)-based resistance in barnyardgrass, a polyploid grass weed. *Pest Manag Sci* **71**:675-685 (2015).

9 Pieczul K and Waşowska A, The application of next-generation sequencing (NGS) for monitoring of *Zymoseptoria tritici* QoI resistance. *Crop Prot* **92**:143-147 (2017).

10 Schlipalius DI, Tuck AG, Pavic H, Daglish GJ, Nayaka MK and Ebert PR, A high-throughput system used to determine frequency and distribution of phosphine resistance across large geographical regions. *Pest Manag Sci* **75**:1091-1098 (2019).

11 Kraehmer H, Laber B, Rosinger C and Schulz A, Herbicides as weed control agents: state of the art: I. weed control research and safener technology: the path to modern agriculture. *Plant Physiol* **166**:1119-1131 (2014).

12 Heap I, International survey of herbicide-resistant weeds. Available online at: http://www.weedscience.com. Accessed 11 April 2019.

13 Zhu B-R, Barrett SCH, Zhang D-Y and Liao W-J, Invasion genetics of *Senecio vulgaris*: loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple introductions. *Biol Invasions* **19**:255-267 (2017).

14 Meyer L, Causse R, Pernin F, Scalone R, Bailly G, Chauvel B, Délye C and Le Corre V, New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species. *PLoS ONE* **12**:e0176197 (2017).

15 Délye C, Causse R, Gautier V, Poncet C and Michel S, Genetic basis, evolutionary origin and spread of resistance to herbicides inhibiting acetolactate synthase in common groundsel (*Senecio vulgaris*). *Pest Manag Sci* **72**:89-102 (2016).

16 Délye C, Matéjicek A and Gasquez J, PCR-based detection of resistance to Acetyl-CoA carboxylase-inhibiting herbicides in black-grass (*Alopecurus myosuroides* Huds) and ryegrass (*Lolium rigidum* Gaud). *Pest Manag Sci* **58**:474-478 (2002).

17 Tranel PJ, Wright TR and Heap IM, Mutations in herbicide-resistant weeds to ALS inhibitors. Available online at: http://www.weedscience.com. Accessed 11 April 2019.

18 Duggleby RG, McCourt JA and Guddat LW, Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiol Biochem* **46**:309-324 (2008).

19 Gaidatzis DA, Lerch F, Hahne F and Stadler MB, QuasR: quantify and annotate short reads in R. *Bioinformatics* **31**:1130-1132 (2015).

20 Morgan M, Pagès H, Obenchain V and Hayden N, Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 1.34.1, Available online at: http://bioconductor.org/packages/release/bioc/html/Rsamtools.html. Accessed 11 April 2019.

21 Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL and Mayer G, Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci Rep* **8**:10950 (2018).

22 Massa D, Krenz B and Gerhards R, Target-site resistance to ALS-inhibiting herbicides in *Apera spica-venti* populations is conferred by documented and previously unknown mutations. *Weed Res* **51**:294-303 (2011).

23 Han H, Yu Q, Purba E, Li M, Walsh M, Friesen S and Powles SB, A novel amino acid substitution Ala-122-Tyr in ALS confers high-level and broad resistance across ALS-inhibiting herbicides. *Pest Manag Sci* **68**:1164-1170 (2012).

24 Han H, Vila-Aiub MM, Jalaludin A, Yu Q and Powles SB, A double EPSPS gene mutation endowing glyphosate resistance shows a remarkably high resistance cost. *Plant Cell Environ* **40**:3031-3042 (2017).

25 Leroux P and Walker A-S, Multiple mechanisms account for resistance to sterol 14α-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. *Pest Manag Sci* **67**:44-59 (2011).

26 Patzoldt WL, Tranel PJ, Alexander AL and Schmitzer PR, A common ragweed population resistant to cloransulam-methyl. *Weed Sci* **49**:485-490 (2001).

27 Laplante J, Rajcan I, and Tardif FJ, Multiple allelic forms of acetohydroxyacid synthase are responsible for herbicide resistance in *Setaria viridis*. *Theor Appl Genet* **119**:577-585 (2009).

28 Délye C, Duhoux A, Pernin F, Riggins CW and Tranel PJ, Molecular mechanisms of herbicide resistance. *Weed Sci* **63**:91-115 (2015).

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 [3c]			% mutant detected		Nb. Kreads per codon [3c]		
Code [1a]	Type [2b]	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 [4d]	F	0.0	0.0	18.1	6.2	7.7	78.0	67.0	17.1	4.8	9.8
RA-05 [4d]	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 [4d]	F	7.0	1.5	9.6	10.4	24.5	92.0	93.6	8.3	9.5	22.2
RA-12 [4d]	F	0.0	0.0	18.8	18.2	21.3	7.0	11.6	20.2	20.3	20.7
RA-13 [4d]	F	0.0	0.0	9.5	18.4	17.8	50.0	43.2	12.2	19.1	18.3
RA-15 [4d]	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
RA-20	F	0.0	0.0	17.9	22.9	20.9	0.0	0.0	14.8	20.6	17.9
RA-22	F	0.0	0.0	22.7	7.1	13.3	0.0	0.0	17.7	7.5	11.2
RA-24	F	0.0	0.0	27.8	17.2	14.3	0.0	0.0	27.3	17.0	14.9
BZ-I	F	16.0	12.3	17.9	13.5	11.2	0.0	0.0	14.4	12.0	11.9
BZ-IV	F	0.0	0.0	16.0	11.7	19.8	0.0	0.0	18.4	11.5	19.9
BZ-VI	F	96.0	96.6	19.1	12.9	17.6	0.0	0.0	22.2	10.3	16.8
BZ-VII	F	58.0	57.9	12.9	15.5	23.9	0.0	0.0	9.7	15.1	19.9
BZ-VIII	F	100.0	99.8	29.0	35.7	22.1	1.0	0.9	26.2	38.8	23.4
BZ-04	F	2.0	2.4	4.7	21.1	12.3	0.0	0.0	4.5	19.2	12.7
BZ-05	F	1.0	1.0	10.7	10.7	14.0	0.0	0.0	13.2	8.4	13.5
BZ-09	F	51.0	57.0	16.2	7.4	7.0	0.0	0.0	14.5	6.7	6.6
BZ-13	F	5.0	1.0	8.9	2.4	4.5	0.0	0.0	10.0	2.6	4.7
BZ-24	F	0.0	0.0	38.1	6.3	18.5	2.0	0.9	32.7	4.8	15.1
BZ-38	F	2.0	1.0	12.1	22.6	21.5	1.0	0.0	11.0	20.2	21.7
BZ-57	F	0.0	0.0	7.6	6.6	5.3	4.0	3.0	8.7	7.1	6.0

BZ-59	F	91.0	88.4	11.1	9.9	7.2	7.0	2.8	14.5	10.0	8.7
BZ-61	F	49.0	46.4	9.9	11.9	8.8	22.0	18.9	9.3	11.6	9.0
BZ-66	F	5.0	5.7	18.0	32.1	10.1	1.0	0.0	15.4	31.2	11.3
BZ-67	F	7.0	2.0	19.8	13.0	0.7	0.0	0.0	19.7	13.2	0.8
BZ-69	F	4.0	2.3	17.3	12.2	7.7	0.0	0.0	21.0	12.4	7.5
BZ-71	F	4.0	2.5	17.5	18.6	7.9	0.0	0.0	16.4	17.6	7.5
BZ-73	F	20.0	31.3	5.5	6.8	12.4	5.0	9.1	11.0	6.3	10.6

[1][a] RA, Rhône-Alpes; BZ, Brittany.

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

[3][c] 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 reads were not joined.

[4][d] Populations with plants containing a double mutation causing a Pro-197-Asn substitution at ALS2. These populations were also subjected to Sanger sequencing of individual plants.

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

Populations		ALS				
		% mutant detected [3c]		Nb. Kreads per codon [4d]		
Code [1a]	Type [2b]	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
OC-27	F	A205T (1.0)	A205T (0.9)	29.5	60.1	25.4
OC-30	F	A205T (2.0)	A205T (1.7)	34.8	68.3	31.3
OC-43	F	A205T (7.0)	A205T (5.7)	29.5	58.7	25.9
OC-55	F	0.0	0.0	30.4	56.7	32.8
OC-60	F	A205T (3.0)	A205T (1.1)	29.9	57.2	26.8
OC-65	F	0.0	0.0	30.0	59.7	28.7
NA-25	F	0.0	0.0	28.9	71.6	27.7
NA-28	F	0.0	0.0	29.3	64.3	32.2
NA-29	F	0.0	0.0	27.4	56.2	27.2
NA-30	F	0.0	0.0	31.8	67.6	30.9
NA-39	F	0.0	0.0	30.2	66.4	27.7

[1][a] NA, Nouvelle Aquitaine; OC, Occitanie.

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

[3][c] The mutations identified are indicated with their frequency between parentheses.

[4][d] 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 reads were not joined.

Table 3. PCR primers.

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

[a] Nucleotides in GenBank/EMBL accession KR024410 or KR024411 (groundsel ALS1 or ALS2 partial sequences, respectively) or in accession KX870184 (ragweed full ALS sequence).

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

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

Mutation(s) [1a]	Nb. populations where found [2]
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 [2b] + Pro-197-Leu-2	2
Pro-197-Asn-2 [3b] + Pro-197-Leu-2 + Pro-197-Ser-2	1
Pro-197-Leu-1 + Pro-197-Asn-2 [2b] + Pro-197-Leu-2 + Pro-197-Ser-2	1

[1a] The number following the mutation refers to the ALS homeolog (1, ALS1; 2, ALS2).

[2b] Substitution caused by a double mutation.

Legends to Figures

Figure 1. Flow-chart of the NGS-based resistance diagnosis procedure.

Figure ~~1~~2. The 67 ragweed populations from Occitanie used for NGS-based detection of mutations endowing resistance to ALS-inhibiting herbicides. Each population is figured as a coloured dot. Green dots, populations where no mutation at the ALS gene was detected among the pool of 50 plants analysed. Red dots, populations where an Ala-205-Thr substitution was detected.

Figure ~~2~~3. Frequencies of mutations at the ALS gene detected using individual plant analysis (dCAPS or Sanger sequencing) plotted against the frequencies assessed using Illumina sequencing of pools of 50 plants. In groundsel, frequencies are displayed separately for each ALS homeolog. Because dCAPS genotyping does not allow to discriminate among different mutations at codon 197, the frequencies displayed are the cumulated frequencies of all mutations detected at codon 197. In ragweed, the frequencies are displayed separately for each mutation when several mutations are present in one same population.