Searching for introgression of herbicide resistance in weedy sunflowers by hybridization with cultivated sunflowers.

Nawfel Benseddik[1] | Constance Bonhomme[2] | Théo Fagegaltier[3] | Nicolas Geffroy[4]

[1]nawfel.benseddik@supagro.fr student from the Institut agro Montpellier

[2]constance.bonhomme@supagro.fr student from the Institut agro Montpellier

[3] theo.fagegaltier@supagro.fr student from the Institut agro Montpellier

[4] nicolas.geffroy@supagro.fr student from the Institut agro Montpellier

ABSTRACT

Globalization leads to flows of animal and plant species between countries, which can result in biological invasions difficult to control. Europe has recently experienced the installation of weedy sunflower (*Helianthus annuus L.*) in the agricultural ecosystems. Our study focuses on the French Lauragais region where they were formally identified in 2004. These weeds are responsible for yield reductions of up to 50%. To tackle this problem, systems such as Clearfield® or Clearfield Plus® have been commercialized. They combine an herbicide against weedy sunflowers and seeds of cultivated sunflowers containing an allele of resistance against this herbicide. However, weedy sunflowers are usually able to hybridize with cultivated populations. There is therefore a risk that the resistance allele could be transferred from the crop compartment to the weedy compartment. Thus, the herbicides could become ineffective. In this study, we looked for the presence of these alleles in weedy sunflowers populations collected in the Lauragais. By sequencing with a quantitative approach (NGS Illumina), we determined allelic frequencies of these mutations. Our research has shown that the Clearfield Plus® resistance allele (A122T) is present in high proportions in northern populations. It's the same for the Clearfield® resistance allele (A205V). The study of the differentiation in allele frequencies (FST) between populations revealed that these two mutations would have been introgressed into two fields in the North before spreading.

KEYWORDS

ECOLOGICAL GENETICS, HERBICIDE RESISTANCE, HYBRIDIZATION, SUNFLOWER, WEED

1 | INTRODUCTION

Sunflowers (*Helianthus annuus L.*) were domesticated in North America and spread to Europe for agricultural purposes during the XIX century. Wild *H. annuus* are absent in Europe, but weedy forms started to appear in agricultural fields in the South West of France, especially in the Lauragais, one of the main producing regions in France (Muller, Roumet and Lecomte, 2012; Marie-Hélène, Marie and Vincent, no date). Globalization is at the origin of this spread. Indeed, many improvements in trading ways like an increasing ease of trade and other pathway for trade (on land, air and water) has led to frequent biological invasion (Hulme, 2009).

Cultivated sunflowers have a single large flower head. They are tall (between 1.0 m and 1.30 m on average) and unbranched. Their seeds are rich in oil (oleic or linoleic fatty acid, depending on the variety) and remain attached to the flower head once they have reached maturity. Their morphological characteristics are the result of centuries of domestication and selection (Figure 1A) (Roumet, 2011). Weedy sunflowers are branched and generally reach a greater height. They have several smaller flower heads (2 to 3 times smaller on average). When they reach maturity, a majority of seeds break off from the flower head. Once on the ground, these seeds do not germinate at the same time. Being affected by the dormancy phenomenon, this allows them to present the same generation of individuals over several reproduction periods. Unlike the cultivated ones, these seeds do not have the properties to extract good quality oil (Figure 1B) (Roumet, 2011). Sunflowers are concerned by the phenomenon of sporophytic self-incompatibility (Gandhi *et al.*, 2005). It is a barrier to autogamy and helps to value the creation of diversity within the population. Nevertheless, the new generations of cultivated sunflowers gradually lose this phenotypic trait (Astiz *et al.*, 2011).

Weedy sunflowers are very competitive. Their sprawl in cultivated fields has led to a reduce solar interception for cultivated sunflowers because of the height of weedy sunflower (Husse *et al.*, 2016). In addition, the dormancy of their seeds (Corbineau, Bagniol and Côme, 1990) combines with their inability to hold back their seeds (Roumet, 2011), lead to a storing of seeds in the soils. Consequently, they can resist to phenomena which would have affected a full-grown plant. Finally, their ability to form bushes that prevent the passage of agricultural machinery, combined with the impossibility of using their seeds, can result in yield losses of up to 50% (Muller *et al.*, 2009).

A reduction of the production of sunflower oil could be detrimental for human consumption (Salas, Bootello and Garcés, 2015) and for the development of agrofuel (Perea-Moreno, Manzano-Agugliaro and Perea-Moreno, 2018). All of this has led to the use of Weed Killers aimed at those weedy sunflowers. The technology proposed by the company Clearfield® was developed after the discovery of wild resistant sunflowers to imidazolines herbicide in a soja field (Al-Khatib *et al.*, 1998; White *et al.*, 2002). Thus, Clearfield® (CL) herbicide corresponds to the PULSAR40® herbicide made from imazamox (Pfenning, Palfay and Guillet, 2008). To make sure it aimed right, the cultivated sunflowers were provided with a resistance gene to the Clearfield® Weed Killer (Tan *et al.*, 2005). The resistance allele corresponds to a version of the AHAS1 gene coding for a subunit of the acetohydroxyacid synthase enzyme involved in the biosynthesis of certain amino acids (as valine, leucine, isoleucine) (Umbarger, 1978; Duggleby and Pang, 2000; Pfenning, Palfay and Guillet, 2008).

Two substitutions could be at the origin of this allele, more precisely two C/T transitions. On the protein this results in two amino acid changes: a leucine instead of a proline (called P197L afterward) and a valine instead of an alanine (called A205V afterward) which render the herbicide-binding site unable to function (Tan *et al.*, 2005; Pfenning, Palfay and Guillet, 2008). This system was commercialized in 2003 and seven years after, a new system called Clearfield Plus® (CLPlus) is marketed in its turn (Tan *et al.*, 2005; Weston, McNevin and Carlson, no date). This system is based on the same principle but offers a more effective herbicide, named Passat® Plus, with associated resistant seeds. The genomic variation is again in the AHAS1 gene but in a single A/G transition. On the protein this corresponds to the replacement of an alanine by a threonine (called A122T afterward) (Kolkman *et al.*, 2004; Sala *et al.*, 2008). The CL allele confers resistance in the homozygous state but is partial in the heterozygous state. The CLPlus allele provides resistance in the homozygous and heterozygous state. Since the CL and CLPlus alleles are located on the same gene, it is not possible to obtain CL and CLPlus homozygous individuals (Weston, McNevin and Carlson, no date).

Weedy sunflowers have the potential to hybridize with cultivated ones (Massinga *et al.*, 2003; Whitney, Randell and Rieseberg, 2006). Hybridization suggests a mixing of the genomes of the two compartments (Wei, Qian and Ma, 1999). For the sunflower, the gene flow is notably dependent on the action of pollinating insects, which in turn is conditioned by other factors such as distance and air currents (Muhammad, 2021) (Bozic *et al.*, 2015) (Couvillon, Schürch and Ratnieks, 2014). The gene flow is estimated as a 30 m radius around the field (Massinga *et al.*, 2003) as others calculated a maximum range of 1 km with a maximum of gene flow when sunflowers are separated from each other by 3 meters (Arias and Rieseberg, 1994). A CL allele definitely increases the fitness (viability x fertility) of sunflowers in contact with the herbicide (Vrbnicanin *et al.*, 2017). Therefore, if one of the two CL alleles enters the weed compartment and succeeds in establishing itself in this population because of the selective advantage it confers, this could undermine the efficiency and reliability of the Clearfield® system. More than that, the Clearfield® technology could be put to an end. There was an instance of this happening to other plants like many canola (Knispel *et al.*, 2008). They quantified the heredity of this allele on herbicide resistant canola and found that the hybridization between a sensible and resistant form could produce progeny with the resistance.

Currently, no study has shown that the resistance allele has transferred into the weed compartment (Bozic *et al.*, 2019). The aim of this study is to further analyze if one of these resistant alleles has crossed from the cultivated compartment to the weedy forms. To answer this question, we used sunflowers capitulum harvested at the end of August in fields from the Lauragais (Annex 1). To have an allelic frequency representative of the effective pollen cloud, we tried to work over the maximum of samples. All the experimental process is therefore based on bulks of seeds. Next, we did an amplification of a part of the AHAS1 gene by PCR. The PCR primers were designed to amplify a locus comprising the three Clearfield® substitutions (A122T, P197L and A205V). Then, the PCR products were sequenced with NGS Illumina® technology and analyzed with bioinformatics tools.

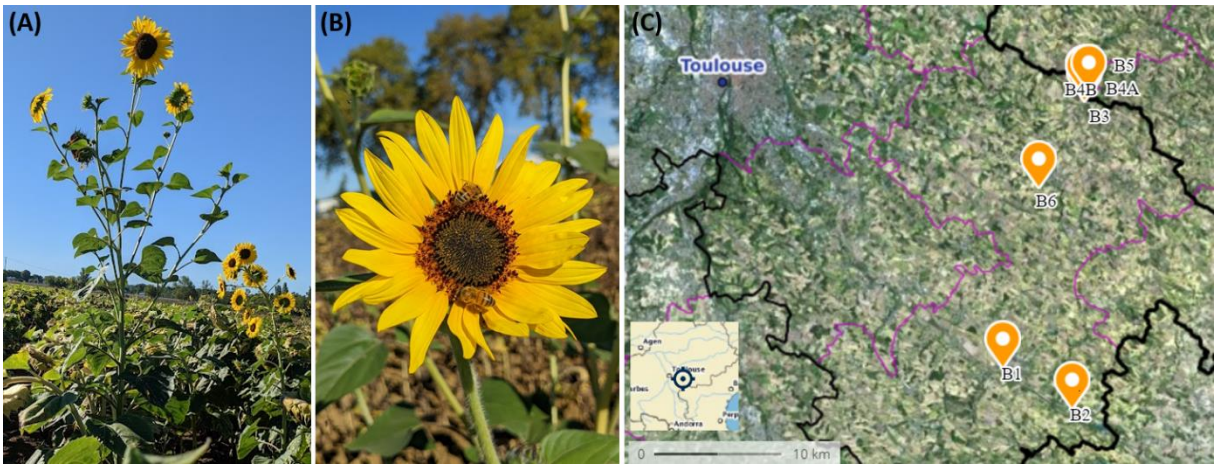


FIGURE 1: (A) Weedy sunflower on the field. (B) Weedy sunflower head. (C) Map of sampled fields.

2 | MATERIAL AND METHODS

2.1 | Material

For our research, we did a harvest among seven fields (Figure 1C). We collected one capitulum per weedy sunflower for a total of thirty plants in each field. All the seeds harvested were distributed among seven bulks. Each bulk corresponds to a specific field (Annex 1) all located in the Lauragais region.

Each bulk are composed of 10 seeds per capitulum, adding up to 300 seeds per field. We included two controls designed using seeds of a cultivated sunflower variety, provided by a company. The negative control represented by seeds from a sensitive variety to CL herbicide and a positive control represented by seeds from a CL variety. The negative control was performed with 120 sensible seeds. The positive control was performed by diluting one CL seed among 119 sensitive seeds. This dilution was intended to avoid contamination during sequencing. Each control was composed of 30 seeds.

2.2 | DNA extraction & purification

For each batch of seeds, a grinding was carried out. This grinding is done by pouring the seeds into liquid nitrogen to facilitate the crushing and to preserve the state of their cells to limit the chemical degradation of nucleic acids. Once crushed, the seeds were divided into tubes per batch of 60 with a total of 4 to 6 tubes per bulk.

For each bulk, the crushed seeds were put in contact with a solution of sodium dodecyl sulphate (SDS) and stirred at 65°C for about 30 minutes. Potassium acetate (AcK) was also added afterward as well as an antioxidant (β-mercaptoethanol). Those steps made it first possible to disaggregate the cellular compounds. AcK lead to ANDsb precipitation and the β-mercaptoethanol limits the oxidation, phenomena that is present in oily seeds such as sunflower seeds. The tubes were then centrifuged at 5°C for 10 minutes to recover a clarified lysate. The lysates were pooled to obtain 2 tubes per bulk. Isopropanol was added to the previously distributed lysates to allow selective precipitation of the nucleic acids. The tubes were then centrifuged in the same way to separate the macromolecules from the liquid phase. The supernatant was then removed and taken up with ultrapure water. Chloroform and

phenol were added to the solution to allow separation of the remaining water and lipids and nucleic acids. For each tube, the aqueous phase containing the nucleic acids were placed in a new tube with a new addition of chloroform. It allows to eliminate the phenol which can inhibit the action of the Taq polymerase (Davalieva and Efremov, 2010). The tubes were then emulsified and centrifuged. The aqueous phase was recovered and placed in a plate well. This plate then underwent a final purification step by a binding process on silica. This process is based on the principle of electrostatic attraction of nucleic acids. Then storage of the tubes was necessary at -20°C (Annex 2).

2.3 | PCR amplification: Resistance allele, choice of primers and PCR conditions

Several mutations in the AHAS gene lead to the resistance to AHAS inhibiting herbicides in plants (Kolkman *et al.*, 2004; Tan *et al.*, 2005; Sala *et al.*, 2008). Mutations are located on chromosome 9 and lead to amino acid substitutions in the primary sequence of the AHAS enzyme. CLPlus resistant AHAS has an alanine to threonine transition at position 122 (A122T). Regarding CL resistance, AHAS could have a proline to leucine transition at position 197 (P197L) or/and an alanine to valine transition at position 205 (A205V) (Figure 2A and 2B) (Kolkman *et al.*, 2004; Tan *et al.*, 2005; Sala *et al.*, 2008). Thus, choice of primers must include these three SNPs. In addition, the choice of primers must consider two other conditions: the fragment to be sequenced must be less than 500 bp due to the length of the Illumina fragments that are 250 base pairs long. As there are two reads from each side of the fragment if the sequence is longer than 500 base pairs some nucleotides will not be sequenced. Besides, the primers must be compatible with NGS Illumina technology. Following these constraints, the PCR was performed with primers chosen with the NCBI's tools pick primers (Figure 2C). Thus, these primers provide a sequence of 336 nucleotides in length. Finally, we have added Illumina tails for DNA fixation during the sequencing as well as shifters (Figure 2C).

The PCR was fully performed during our JRL-AT project of four months. For each bulk with a touchdown 66 to 56°C in 14 runs with 30 more runs at 56°C. The volume of each well was 25 µL containing 3 µL of DNA, 12.5 µL of buffer Taq3G (10X and containing dNTP), 0.6 of MgCl2 25 mM, 0.1 µL of each primer, 0.1 of Taq3G (5 unit/µL) and 8.6 µL of water. Taq3G was used because our DNA solution most likely contained many phenols that are inhibitors of a standard Taq (Taq2G).

2.4 | Sequence analysis

The sequences obtained from the Illumina sequencing, called reads, were analyzed using bioinformatics tools. First, a cleaning of the reads was necessary. This cleaning allowed to remove from the dataset the reads that were too small and therefore not exploitable as well as those whose sequencing had included fragments of shifters or Illumina tails. To achieved this, we used the program Cutadapt (MARTIN, Marcel, 2011), included in a snakemake pipeline developed by the hosting team (https://github.com/BioInfo-GE2POP-BLE/CAPTURE_SNAKEMAKE_WORKFLOWS). Afterwards, reads obtained were mapped onto the entire sunflower genome (*Helianthus annuus L.* genome available on NCBI website). This mapping step is important to check the specificity of the primers. Especially since our study focuses on a specific part of the genome (AHAS1 gene on chromosome 9). We again use the snakemake pipeline to run this step, that relies on the program bwa-mem (MARTIN, Marcel, 2011) for mapping. Finally, we focused

on reads mapped on chromosome 9, using samtools. Then, we searched the mapped reads to count the number of reads including the sequence with the CLPlus mutation (A122T), each of the two potential CL mutations (P197L and A205V) and both CL mutations together on the same read. To achieve this, we used the function « grep » to find the sequences (pattern) or their reverse complement. In addition, we counted the number of reads that covered each mutation to estimate the frequency of the mutations in each of our bulks.

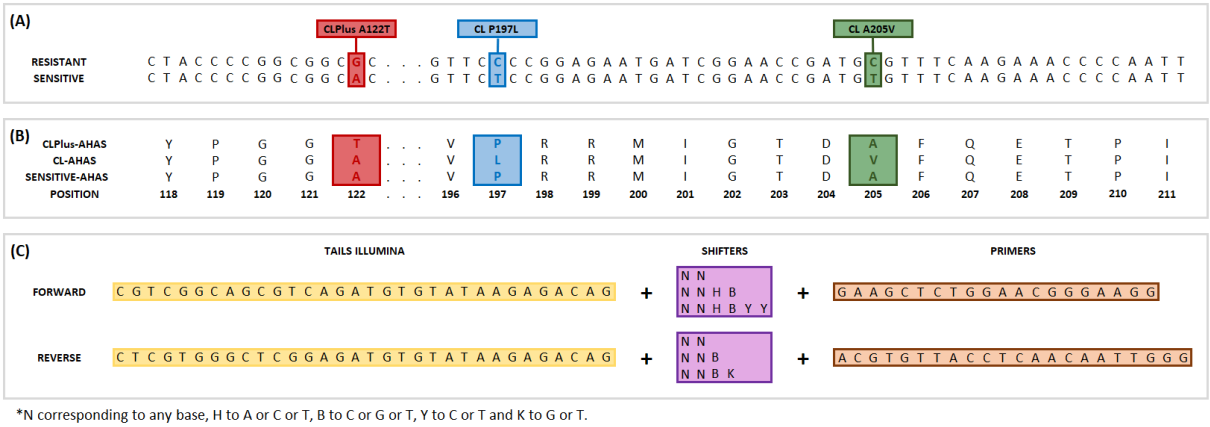


FIGURE 2: (A) Representation of the three SNPs. (B) Representation of the amino acids substitution for CLPlus, CL and sensitive AHAS. (C) Primer's design.

3 | RESULTS

3.1 | Quantity and quality DNA control

Given the difficulty of extracting DNA from seeds, we wanted to check for the presence of DNA. An electrophoresis has been done for that. The topmost band at 60,000 pairs of bases corresponds to the DNA (Figure 3A). This is due to the mechanical processing which breaks the DNA into small pieces of a minimum size of 60,000 pair of bases. The differences in fluorescence between the bulks show that the DNA was not extracted with the same efficiency for each bulk and especially for the control groups that have DNA in lower concentration (weak fluorescence). The fragments below correspond to the ribosomal RNA's. The lowermost fragments correspond to the smallest RNAs. Electrophoresis is a direct scan of the presence of DNA. To confirm these results the absorbance of the samples between two hundred and four hundred nanometers was measured using spectroscopy (Figure 3B) (Annex 3). This is an indirect scan of the presence of DNA in our samples. The peak at 260 nanometres corresponds to the maximum absorption of DNA. The purity of DNA is evaluated with the 260/280 ratio which assess the presence of contaminants (organic solvents, RNA, residual proteins) in the DNA solution. A 260/280 ratio greater than or equal to 1.7 is expected in the case of pure double-stranded DNA. Another electrophoresis was performed to ensure that during the PCR, the primers had amplified the desired sequence i.e., the AHAS gene on chromosome 9. The results show that our sequence (336 bp) has been amplified rightly in each bulk (Figure 3C).

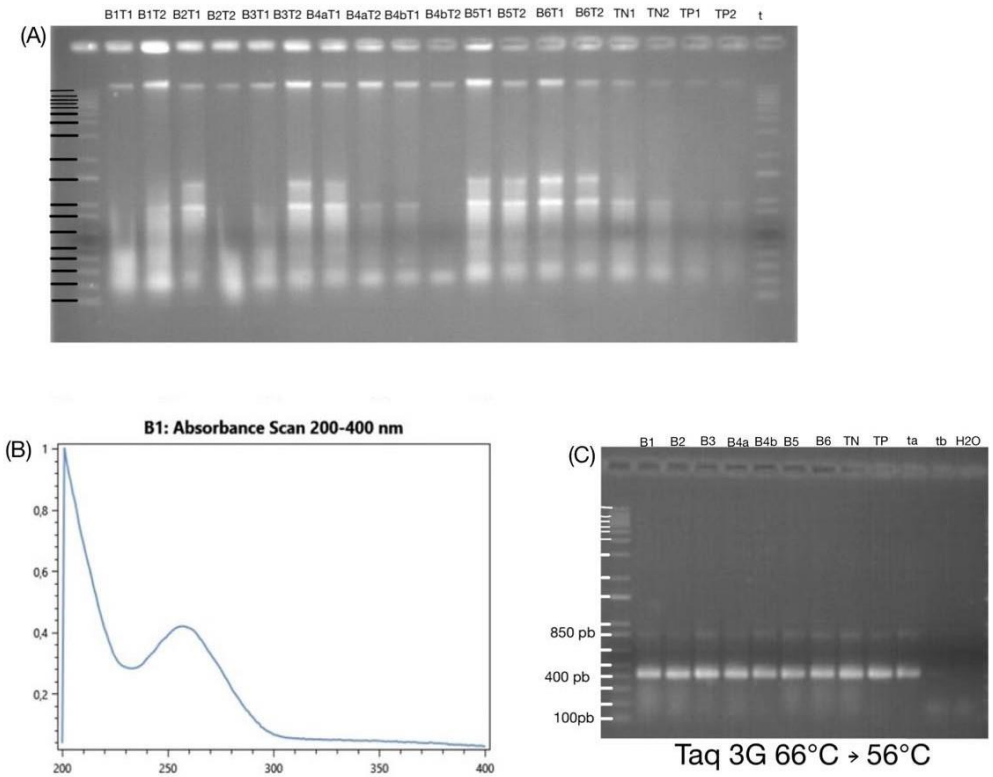


FIGURE 3: Results to confirm the quality of DNA in our samples and its amplification. (A) Electrophoresis of all 7 bulks split into two tubes. (B) Absorbance scan 200-400nm for the bulk n°2. (C) Electrophoresis of the PCR results done with Taq3G enzyme using touch down from 66°C to 56°C.

3.2 | Sequencing and population dynamics

PCR products were sequenced with Illumina MiSeq technology. Thus, approximately 1,520,276 reads of about 250 bp were obtained. By using bioinformatics tools, we did a data cleaning and 1,487,850 reads remained. Then, about 99.7% of the reads were mapped on the whole genome which reveals a good quality of the sequencing. Mapping reveals that a significant part of the reads was mapped on chromosome 9 but also on chromosomes 2 and 6. This is explained by the fact that the gene coding for the AHAS enzyme belongs to a multigenic family (AHAS1, 2 and 3 are paralogous genes) (Kolkman *et al.*, 2004) . These results indicate that the choice of primers was not optimal, because it was not sufficiently specific to the targeted fragment on the AHAS1 gene on chromosome 9.

However, extracted data from the bioinformatics processing of the sequences allows us to estimate allelic. Then, for each polymorphism (A122T, P197L and A205V), we calculated the number of reads containing the mutation and the total number of reads covering that genome position. Thus, the ratio of the number of reads including the mutation over the total number of reads at that position allows us to obtain an estimation of the allelic frequencies at each of these three loci frequencies (Figure 4A). This was performed for each locus in each bulk (B1, B2, ..., B6). The allelic frequencies and percentages obtained for negative controls allow us to evaluate the background noise of the sequencing (Allelic frequencies < 0.65%).

The extracts from each bulk were divided into two tubes (T1, T2). To ensure that the distribution of nucleic acids is homogeneous for both tubes, we made a correlation graph for

each mutation (Figure 4B). A homogeneous distribution is reflected by a directing coefficient close to 1 (with p-values<2%).

Thereafter, we represented for each bulk the allelic frequency of each locus as a pie chart. These diagrams were placed on the map of the sampling locations (Figure 4C). This representation allows us to formulate several hypotheses. First, we notice that the allelic frequencies of the CLPlus polymorphism (A122T) seems to be higher in the northern populations B3, B4A and B4B. The same is true for the CL polymorphism (A205V) in the populations B3, B4A, B4B, B5 and B6. Thus, we can assume that the introgression of the CLPlus polymorphism (A122T) occurred in the B4B population and that of the CL polymorphism (A205V) in the B4A population. Thus, to test these hypotheses, we calculated the differentiation in allele frequencies (FST) between each pair of populations (Annex 4). Then, we checked which ones were significant via a Chi-square test. Therefore, we observe that the B4B population is very differentiated for the CLPlus polymorphism (A122T) (FST of 0.34 with B1, 0.88 with B2, 0.13 with B3, 0.28 with B4A, 0.17 with B5 and 0.34 with B6). On the contrary, the other populations are very close genetically for this polymorphism (FST < 0.09). For the CL mutation (P197L), all populations are very poorly differentiated from each other (FST < 0.02). Finally, for the CL mutation (A205V), the B4A population is the most differentiated (FST of 0.16 with B1, 0.16 with B2, 0.12 with B5 and 0.1 with B6). However, it is very close to B3 (FST=0.03). All other populations are very close to each other for this polymorphism (FST < 0.06).

Then, to highlight the process of gene flow and its incidence among these fields in the same region, it is possible to plot the distances between the bulks according to their genetic proximity (FST) for each locus (Figure 5A, 5B and 5C). However, there is no significant correlation for all the bulks.

The frequency of the CLPlus allele fluctuate between bulks. The bulk with the highest percentage of CLPlus allele is B4B. The presence of this allele tends to decrease with distance from this plot. To translate these observations mathematically, we established a simple linear regression taking as a variable the distance to the plot of the B4B bulk (predictor variable) and the FST of each bulk associated with the B4B bulk (Figure 5D). The linear regression gives a p-value lower than the statistical limit of 0.05 (p-value=0.037<0.05), which allows us to consider a significant link between the two variables mentioned above. The test of the significance of the linear link between the two variables was then carried out by evaluating the homoscedasticity, normality, independence, and linearity of the residuals (respectively by the Fligner-Killeen, Shapiro-Wilk and Durbin-Watson tests). It can therefore be considered that there is a link between the distance to the B4B bulk plot and the genetic proximity for the CLPlus allele. Among the sunflower fields studied, it can be assumed that the CLPlus allele appeared in the weed populations of the bulk B4B plot and that subsequently, this allele reached the neighboring fields by pollen or seed gene flow. Since the allele confers a selective advantage, it is expected that it will increase in frequency in the surrounding populations and continue to spread. However, since gene flow depends on several environmental factors (insects, animals, air currents), we do not expect that this diffusion will be homogeneous and proportional. Using the same process of observations and statistical transcriptions, it is found that the FST for the A205V polymorphism is affected by a significant effect of distance to the B4A plot (p-value = 0.02) (Figure 5E). Considering the two findings above, it can be assumed that the plots in field 4 have had an important genetic influence on the surrounding plots. It seems that alleles of selective interest (concerning herbicide sensitivity) appeared in

the weed populations at the level of field 4 and spread afterwards.

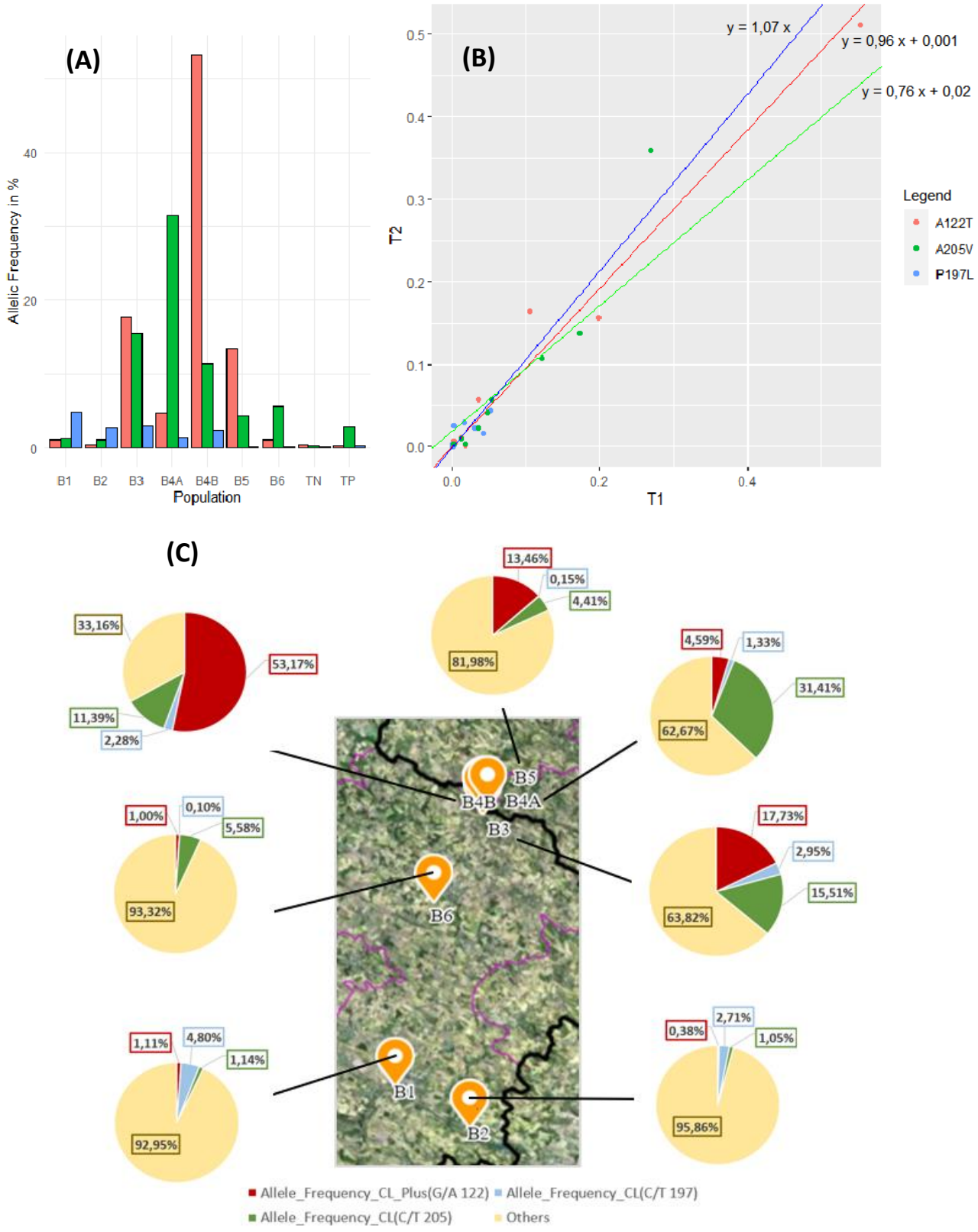


FIGURE 3: (A) Histogram of allele frequencies by mutation and by bulk (Red = A122T, Blue = P197L and Green = A205V). (B) Correlation lines between the allelic frequencies of tubes 1 and 2 by mutations. (C) Map of the different bulks and proportions of allelic frequencies in the form of a pie chart.

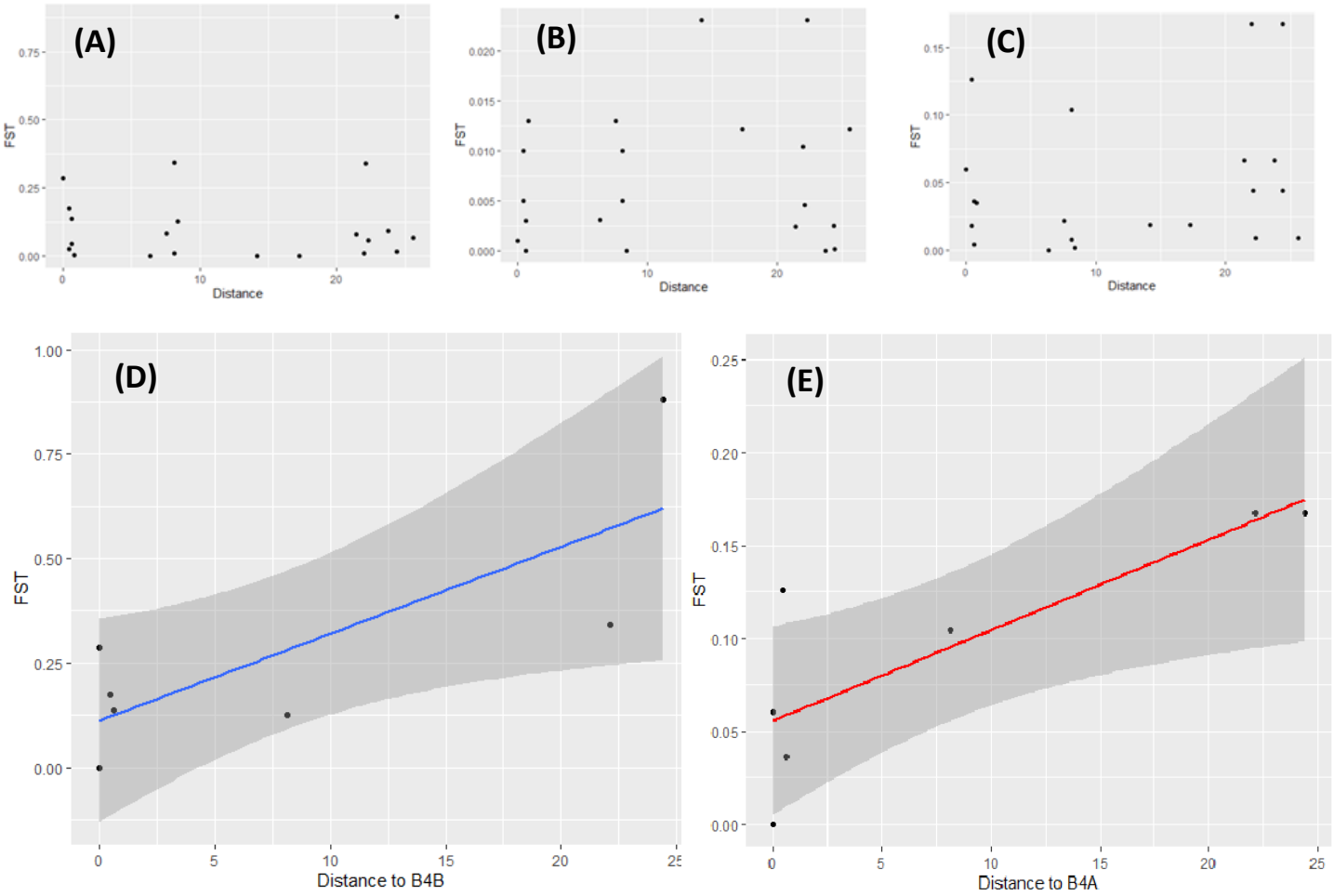


FIGURE 5 2: FST as a function of the distance between each bulk (A) CLPlus A122T (B) CL P197L (C) CL A205V, distance in km.(D) FST of CLPlus mutation as function of the distance to B4B. (E) FST of CL mutation (A205V) as function of the distance to B4B.

4 | DISCUSSION

The Clearfield® allele has been found in a few copies with both mutations (P197L and A205V) but for insufficient percentages (background noise). However, A205V mutation was found in sufficient percentages to be considered. This SNP alone can offer an acceptable tolerance to imidazolinone compounds (Tan *et al.*, 2005; Sala *et al.*, 2008). Furthermore, CLPlus allele has clearly been found in our results with sometimes an allele frequency of up to 50% (Annex 5). This had not been proved before. Nevertheless, CL were discovered in low percentages in the positive control which was composed of one supposed CL seed and one hundred and nineteen sensitive seeds. Besides, electrophoresis (Figure 3A) attests to the small amount of DNA extracted. This alerts to the reliability of this dilution. It was supposed to avoid contamination during sequencing but using only resistant seeds might have been preferable. For example, extraction from a large number of CL seeds followed by dilution of this DNA in solution (to limit the risk of contamination during the amplification and sequencing steps) would have been preferable to obtain sufficient quantities of DNA. Thus, our results must be put into perspective. The lack of

positive controls could lead to believe that we have misidentified the mutations and that it is located elsewhere on the AHAS gene.

Regarding the study of differentiation between populations (FST), it would have been judicious to determine other factors of divergence such as the technical itinerary and the variety used in addition to the distance. In addition, the use of the same machines for several farmers is also a factor to considered. This results in the flow of seeds between fields over greater distances than those allowed by pollinators carrying pollen. These transfers become part of the gene flow and contribute to the complexity of its reading through the graphs. The use of cultivated sunflowers coming from the same fields as the weedy ones would have been decisive to estimate the gene flow. We could have computed the FST between weedy and cultivated sunflowers on CL et CLPlus polymorphism and on another neutral site to determine the selective advantage of this allele.

These results are unique compared to what was found before. We performed the extractions on sunflower seeds and not on plant samples (Hübner *et al.*, 2022), we therefore did not use the herbicide (Bozic *et al.*, 2019). This process is laborious, in particular, the choice of the lethal dose to be applied to discriminate between resistant and non-resistant plants is based solely on the vendor's recommendations. In addition, the choice of herbicide (CL or CLPlus system) is an irreversible choice because one herbicide can potentially kill a sunflower plant but not the other. Nevertheless, one herbicide application allows the selection of resistant plants at the applied rate, which limits the number of samples to be considered thereafter. Furthermore, a plant that has resisted an application of herbicide does not systematically contain a resistance allele resulting from introgression from the cultivated compartment to the wild compartment. Resistance may be derived from other genomic properties and the spontaneous formation of the CLPlus allele (a single SNP) in the weed population is also a hypothesis to be considered. Taking into account our results, it is difficult to know whether the proven presence of the CLPlus resistance allele among our samples is the result of a simple natural mutation with the main selection pressure being the application of the Passat® Plus herbicide (Singh *et al.*, 2020), or an introgression of the allele by a gene flow coming from the cultivated plot. This can be possibly answered by searching for another specific allele of the cultivated sunflowers to check if there is any hybridization between weedy and cultivated plants. We focused on genes that are specific of the cultivated phenotype (Blackman *et al.*, 2011). Allele of domestication and selection such as high oleic allele could be a possible choice. However, this gene is duplicated several times in the genome which prevents us from using it for sequencing. Nevertheless, this step should be considered to strengthen the analysis.

If our DNA extraction method was unique, the results appear to be equally unique. A Serbian study observed weedy sprout resistant to herbicide applications but they did not find the Clearfield® allele in their genome (Bozic *et al.*, 2019).

Due to the lack of time to carry out the experiments, some methods could be improved. It is the case of the PCR. The Taq 3G used for the experiments is less sensitive to phenol but also less specific. It is more prone to replications errors. This reduces the reliability of the amplification since a technical mutation may have appeared during it. A solution was to use sprouts which are less concentrated in phenols. In our first version of the protocol, we thought about using the sprouts of seeds to extract DNA, but the germination never took place. We noticed that seeds were either empty (we chose non-empty seeds for the DNA extraction then) or they presented a black stain underneath. We hypothesize that a large

number of seeds may have been oxidated or immature. Besides, the problem of seeds dormancy in wild populations (need for an elongated period of cold to initiate germination) may have been the reason why our seeds never germinate. Moreover, PCR primers were not specific enough. Primers amplified other loci located on different chromosomes. These loci were all associated with the expression of the AHAS1 enzyme but are not affected by the polymorphisms in the Clearfield® systems. For a more rigorous PCR, the size of the primers (and/or the locus) should have been increased to improve their specificity.

To enhance the robustness of our analyzes and confirm the Illumina sequencing of bulks with high percentages of CLPlus allele presence, another sequencing method should be used. One possibility is individual Sanger sequencing on new samples from the plot associated with the bulk considered. In the same way, it is interesting to do a phenotypic test on these same seeds. A germination followed by an application of Passat® Plus would allow to verify the expression of the CLPlus allele present in these weeds. To lean in favour of one hypothesis or another, it would be advisable to first obtain information on the use (quantitative and qualitative aspects) of the Clearfield® system (CL or CLPlus) by farmers in the Lauragais.

Finally, regarding the FST, it would have been interesting to collect seeds from the cultivated sunflowers on each plot. The establishment of the FST between cultivated and weedy plants would have made it possible to highlight a potential genetic proximity that would support the introgression hypothesis.

5 | CONCLUSION

Allele frequencies of up to 50% were found for the Clearfield Plus locus (A122T) in our weed populations. This frequency fluctuated between fields, but the Northern ones seemed to be more concerned by the potential introgression of this allele. In fact, the field B4B could be the origin of the introgression of this gene. In the same way, we found that the Clearfield locus (A205V) could have been introgressed in the field B4A. Thus, these resistant alleles may be present in the weed samples which if it is the case should put an end to the Clearfield technology.

ACKNOWLEDGEMENT

First of all, we want to thank our tutors, Laurene Gay and Sylvain Santoni, who gave us great insight on how to approach the problem. Sylvain Santoni were also crucial in the experiment from the extraction to the DNA to the DNA sequencing. He provided all the materials, helped us design them as well as correcting them. Laurene Gay was more crucial in the bioinformatic analysis.

We were also helped by Marjolaine Roussel in the conception of our samples to make them statistically significant. Aline Rocher helped us in the germination of our seeds which sadly did not.

Jacques David provided us with the base of our statistical analysis essential for our project.

We think that the JRLAT, created and supervised by Jacque David and Jean François Martin, and all the courses that it provided to us were useful in all aspects of our paper.

AUTHORS CONTRIBUTION

NB, CB, NG and TF took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. SS designed the primers and the experiments. LG did the bioinformatics.

DATA AVAILABILITY STATEMENT

Available upon request

REFERENCES

Al-Khatib, K. *et al.* (1998) 'Imazethapyr resistance in common sunflower (Helianthus annuus)', *Weed Science*, 46(4), pp. 403–407. Available at: https://doi.org/10.1017/S0043174500090809.

Arias, D.M. and Rieseberg, L.H. (1994) 'Gene flow between cultivated and wild sunflowers', *Theoretical and Applied Genetics*, 89(6), pp. 655–660. Available at: https://doi.org/10.1007/BF00223700.

Astiz, V. *et al.* (2011) 'Self-compatibility in modern hybrids of sunflower (Helianthus annuus L.). Fruit set in open and self-pollinated (bag isolated) plants grown in two different locations', *Helia*, 34, pp. 129–138. Available at: https://doi.org/10.2298/HEL1154129A.

Blackman, B.K. *et al.* (2011) 'Sunflower domestication alleles support single domestication center in eastern North America', *Proceedings of the National Academy of Sciences*, 108(34), pp. 14360–14365. Available at: https://doi.org/10.1073/pnas.1104853108.

Bozic, D. *et al.* (2015) 'Gene Flow from Herbicide-Resistant Sunflower Hybrids to Weedy Sunflower', *Journal of Plant Diseases and Protection*, 122(4), pp. 183–188. Available at: https://doi.org/10.1007/BF03356548.

Bozic, D. *et al.* (2019) 'Studies on gene flow from herbicide resistant to weedy sunflower', *Genetika*, 51(1), pp. 287–298. Available at: https://doi.org/10.2298/GENSR1901287B.

Corbineau, F., Bagniol, S. and Côme, D. (1990) 'Sunflower (helianthus Annuus L.) Seed Dormancy and Its Regulation by Ethylene', *Israel Journal of Botany*, 39(4–6), pp. 313–325. Available at: https://doi.org/10.1080/0021213X.1990.10677156.

Couvillon, M.J., Schürch, R. and Ratnieks, F.L.W. (2014) 'Waggle Dance Distances as Integrative Indicators of Seasonal Foraging Challenges', *PLOS ONE*, 9(4), p. e93495. Available at: https://doi.org/10.1371/journal.pone.0093495.

Davalieva, K. and Efremov, G.D. (2010) 'Influence of Salts and Pcr Inhibitors on the Amplification Capacity of Three Thermostable Dna Polymerases', *Macedonian Journal of Chemistry and Chemical Engineering*, 29(1), pp. 57–62.

Duggleby, R. and Pang, S. (2000) 'Acetohydroxyacid Synthase', *Journal of biochemistry and molecular biology*, 33, pp. 1–36.

Gandhi, S.D. *et al.* (2005) 'The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower', *Theoretical and Applied Genetics*, 111(4), pp. 619–629. Available at: https://doi.org/10.1007/s00122-005-1934-7.

Hübner, S. *et al.* (2022) 'Wild sunflower goes viral: Citizen science and comparative genomics allow tracking the origin and establishment of invasive sunflower in the Levant', *Molecular Ecology*, 31(7), pp. 2061–2072. Available at: https://doi.org/10.1111/mec.16380.

Hulme, P.E. (2009) 'Trade, transport and trouble: managing invasive species pathways in an era of globalization', *Journal of Applied Ecology*, 46(1), pp. 10–18. Available at: https://doi.org/10.1111/j.1365-2664.2008.01600.x.

Husse, S. *et al.* (2016) 'Larger yields of mixtures than monocultures of cultivated grassland species match with asynchrony in shoot growth among species but not with increased light interception', *Field Crops Research*, 194, pp. 1–11. Available at: https://doi.org/10.1016/j.fcr.2016.04.021.

Knispel, A.L. *et al.* (2008) 'Gene flow and multiple herbicide resistance in escaped canola populations', *Weed Science*, 56(1), pp. 72–80. Available at: https://doi.org/10.1614/WS-07-097.1.

Kolkman, J.M. *et al.* (2004) 'Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower', *Theoretical and Applied Genetics*, 109(6), pp. 1147–1159. Available at: https://doi.org/10.1007/s00122-004-1716-7.

Marie-Hélène, M., Marie, R. and Vincent, L. (no date) 'Les tournesols adventices : un exemple d'évolution d'une mauvaise-herbe apparentée à une espèce cultivée – Synthèse des recherches menées sur les populations adventices de tournesol en France depuis 2004', p. 9.

Massinga, R.A. *et al.* (2003) 'Gene flow from imidazolinone-resistant domesticated sunflower to wild relatives', *Weed Science*, 51(6), pp. 854–862. Available at: https://doi.org/10.1614/WS-03-032R.

Muhammad, W. (2021) 'Pollinator Community of Sunflower (Helianthus annus L.) and its Role in Crop Reproductive Success', *Asian Journal of Agriculture and Biology* [Preprint]. Available at: https://doi.org/10.35495/ajab.2020.07.398.

Muller, M.-H. *et al.* (2009) 'Occurrence, distribution and distinctive morphological traits of weedy Helianthus annuus L. populations in Spain and France', *Genetic Resources and Crop Evolution*, 56(6), pp. 869–877. Available at: https://doi.org/10.1007/s10722-009-9409-3.

Muller, M.-H., Roumet, M. and Lecomte, V. (2012) 'Les tournesols adventices : un exemple d'évolution d'une mauvaise-herbe apparentée à une espèce cultivée – Synthèse des recherches menées sur les populations adventices de tournesol en France depuis 2004', *Oléagineux, Corps gras, Lipides*, 19(5), pp. 304–312. Available at: https://doi.org/10.1051/ocl.2012.0465.

Perea-Moreno, M.-A., Manzano-Agugliaro, F. and Perea-Moreno, A.-J. (2018) 'Sustainable Energy Based on Sunflower Seed Husk Boiler for Residential Buildings', *Sustainability*, 10(10), p. 3407. Available at: https://doi.org/10.3390/su10103407.

Pfenning, M., Palfay, G. and Guillet, T. (2008) 'The CLEARFIELD® technology – A new broad-spectrum post-emergence weed control system for European sunflower growers', *Journal of Plant Diseases and Protection*, 21.

Roumet, M. (2011) *Les populations adventices de tournesol (Helianthus annuus L.) : Flux génique et adaptation à l'agrosystème*. Université de Montpellier II.

Sala, C.A. *et al.* (2008) 'Molecular and biochemical characterization of an induced mutation conferring imidazolinone resistance in sunflower', *Theoretical and Applied Genetics*, 118(1), pp. 105–112. Available at: https://doi.org/10.1007/s00122-008-0880-6.

Salas, J.J., Bootello, M.A. and Garcés, R. (2015) '14 - Food Uses of Sunflower Oils', in E. Martínez-Force, N.T. Dunford, and J.J. Salas (eds) *Sunflower*. AOCS Press, pp. 441–464. Available at: https://doi.org/10.1016/B978-1-893997-94-3.50020-9.

Singh, V. *et al.* (2020) 'First case of glyphosate resistance in weedy sunflower (Helianthus annuus)', *Pest Management Science*, 76(11), pp. 3685–3692. Available at: https://doi.org/10.1002/ps.5917.

Tan, S. *et al.* (2005) 'Imidazolinone-tolerant crops: history, current status and future', *Pest Management Science*, 61(3), pp. 246–257. Available at: https://doi.org/10.1002/ps.993.

Umbarger, H.E. (1978) 'Amino Acid Biosynthesis and its Regulation', *Annual Review of Biochemistry*, 47(1), pp. 533–606. Available at: https://doi.org/10.1146/annurev.bi.47.070178.002533.

Vrbnicanin, S.P. *et al.* (2017) 'Fitness studies on invasive weedy sunflower populations from Serbia', *Romanian Biotechnological Letters*, 22(2), p. 9.

Wei, W., Qian, Y.Q. and Ma, K.P. (1999) 'Gene flow between transgenic crops and their wild related species', *Acta Botanica Sinica*, 41(4), pp. 343–348.

Weston, B., McNevin, G. and Carlson, D. (no date) 'Clearfield® Plus Technology in Sunflowers', p. 6.

White, A.D. *et al.* (2002) 'Common sunflower resistance to acetolactate synthase–inhibiting herbicides', *Weed Science*, 50(4), pp. 432–437. Available at: https://doi.org/10.1614/0043-1745(2002)050[0432:CSRTAS]2.0.CO;2.

Whitney, K.D., Randell, R.A. and Rieseberg, L.H. (2006) 'Adaptive Introgression of Herbivore Resistance Traits in the Weedy Sunflower Helianthus annuus.', *The American Naturalist*, 167(6), pp. 794–807. Available at: https://doi.org/10.1086/504606.

Annex 1: Table showing the bulks, their locations, and their infestation levels

Site_name	*GPS_lat*	*GPS_long*	*GPS_code*	*GPS_num*	*Location*	*Remarks*
Bulk 1	43°23.824N	1°42.248E	165	43.397067, 1.704133	Villefranche de lauragais	Two fields, abundant weeds between the two, + sporadic in the field
Bulk 2	43°22.096N	1°46.311E	166	43.368267, 1.771850	Avignonet-Lauragais	
Bulk 3	43°34.911N	1°47.048E	167	43.581850, 1.784133	Loubens-Lauragais	Very lightly infested
Bulk 4a	43°35.251N	1°46.979E	168	43.587517, 1.782983	Villeneuve-lès-Lavaur	Two fields on either side of a stream, medium infestation
Bulk 4b	43°35.251N	1°46.979E	168	43.587517, 1.782983	Villeneuve-lès-Lavaur	Two fields on either side of a stream, medium infestation
Bulk 5	43°35.337N	1°47.308E	169	43.588694, 1.788220	Villeneuve-lès-Lavaur	Very infested
Bulk 6	43°31.313N	1°44.389E	170	43.521883, 1.739817	Caraman	highly infested, weeds at different stages of maturity (dry or in flower)

Annex 3: Results of the spectrophotometry analysis

Report

TECAN.

Blanking type	Individual blanking
Result name	NA-Quant_20221108-1
Date	08/11/2022 15:42:35

Measurement parameters

Range: A1-H2; Sample: dsDNA; Ratio: 260/230, 260/280;

Blanking results

Well	230 [nm]	260 [nm]	280 [nm]	310 [nm]
A1	0,0726	0,0557	0,0520	0,0467
A2	0,0697	0,0557	0,0519	0,0471
B1	0,0869	0,0652	0,0607	0,0546
B2	0,0724	0,0565	0,0529	0,0479
C1	0,0728	0,0569	0,0535	0,0483
C2	0,0671	0,0544	0,0514	0,0468
D1	0,0724	0,0569	0,0535	0,0484
D2	0,0768	0,0608	0,0571	0,0516
E1	0,0765	0,0597	0,0556	0,0501
E2	0,0679	0,0542	0,0511	0,0466
F1	0,0722	0,0557	0,0523	0,0474
F2	0,0654	0,0518	0,0490	0,0447
G1	0,0707	0,0554	0,0522	0,0468
G2	0,0689	0,0540	0,0507	0,0463
H1	0,0749	0,0588	0,0556	0,0499
H2	0,0659	0,0539	0,0509	0,0463

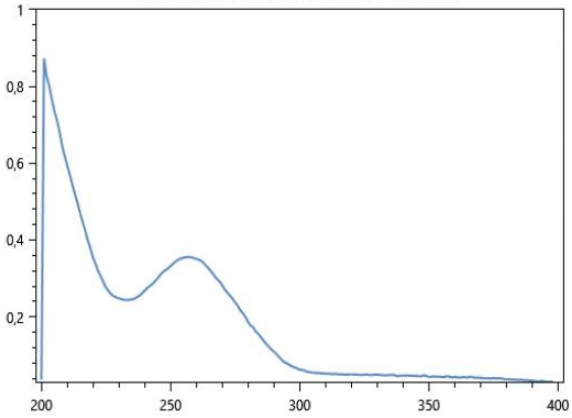
Measurement results

Well	260	230	280	ng/µl	Ratio 260/230	Ratio 260/280
A1	0,3004	0,1958	0,1383	300,35	1,53	2,17
A2	0,0953	0,0873	0,0417	95,25	1,09	2,29
B1	0,3614	0,2327	0,1669	361,42	1,55	2,17
B2	0,1948	0,1439	0,0888	194,78	1,35	2,19
C1	0,2140	0,1753	0,0998	213,97	1,22	2,14
C2	0,1289	0,1198	0,0599	128,88	1,08	2,15
D1	0,1443	0,1489	0,0686	144,30	0,97	2,10
D2	0,0586	0,0741	0,0269	58,56	0,79	2,18
E1	0,0449	0,0635	0,0199	44,93	0,71	2,26
E2	0,2054	0,1519	0,0929	205,43	1,35	2,21
F1	0,0262	0,0555	0,0113	26,19	0,47	2,32
F2	0,1639	0,1402	0,0774	163,87	1,17	2,12
G1	0,2815	0,1815	0,1300	281,52	1,55	2,17
G2	0,0006	0,0002	-0,0003	0,56	2,71	-2,03
H1	0,1710	0,1354	0,0790	170,97	1,26	2,16
H2	0,0011	0,0022	0,0000	1,13	0,52	24,13

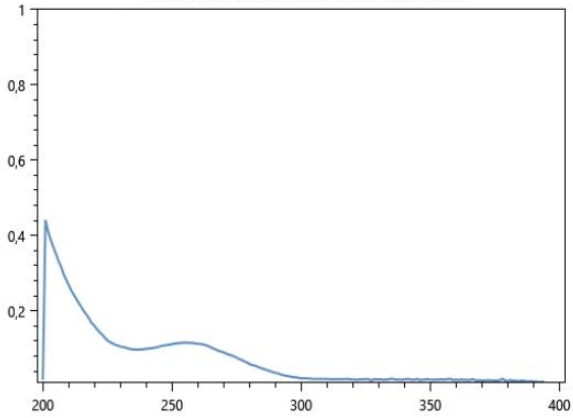
Report

TECAN.

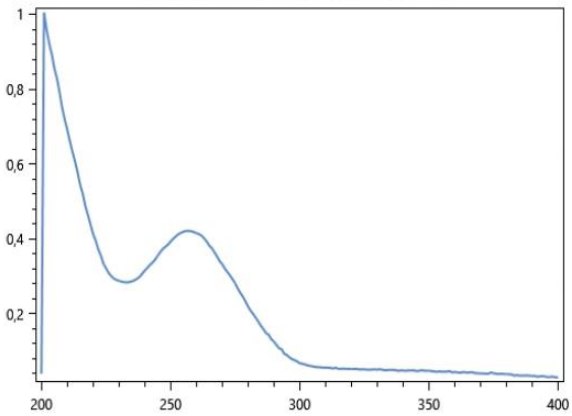
A1: Absorbance Scan 200-400 nm



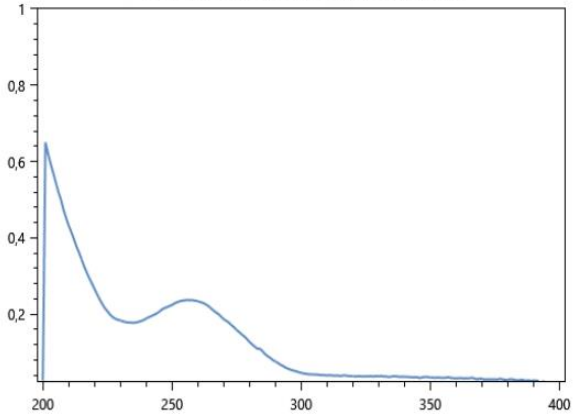
A2: Absorbance Scan 200-400 nm



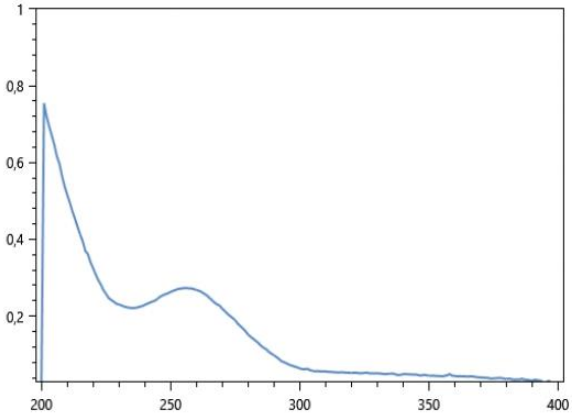
B1: Absorbance Scan 200-400 nm



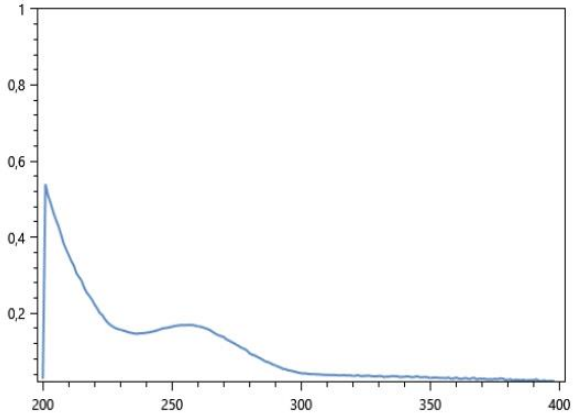
B2: Absorbance Scan 200-400 nm



C1: Absorbance Scan 200-400 nm



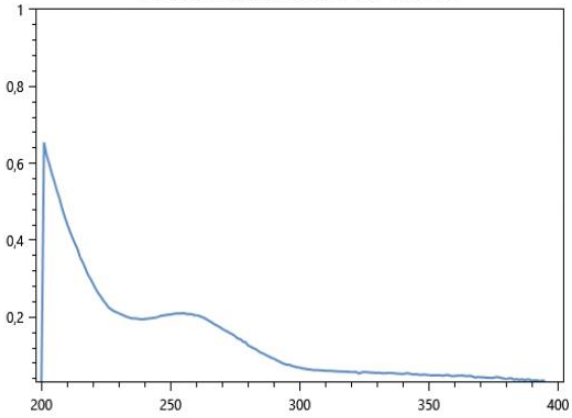
C2: Absorbance Scan 200-400 nm



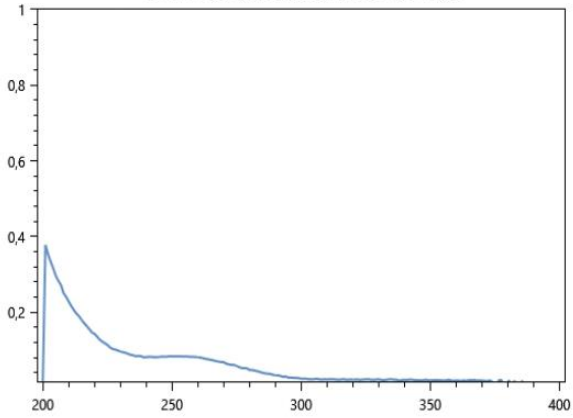
Report

TECAN.

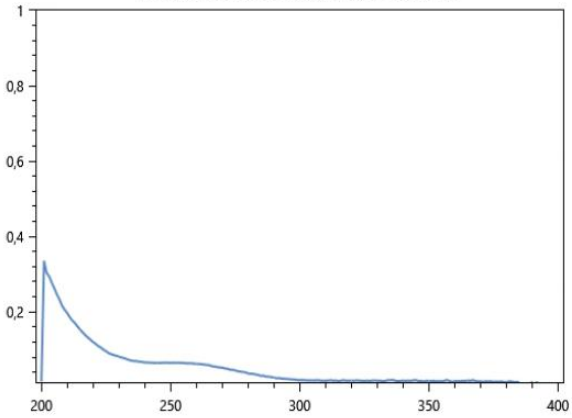
D1: Absorbance Scan 200-400 nm



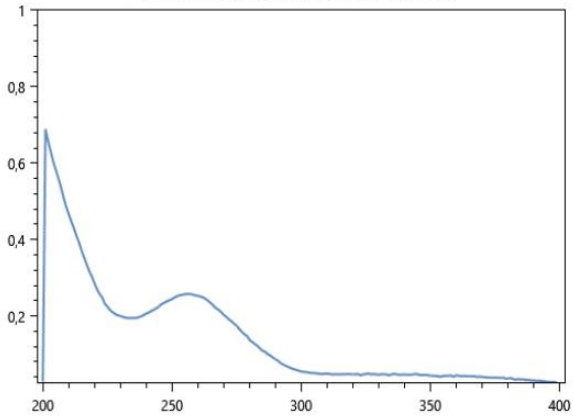
D2: Absorbance Scan 200-400 nm



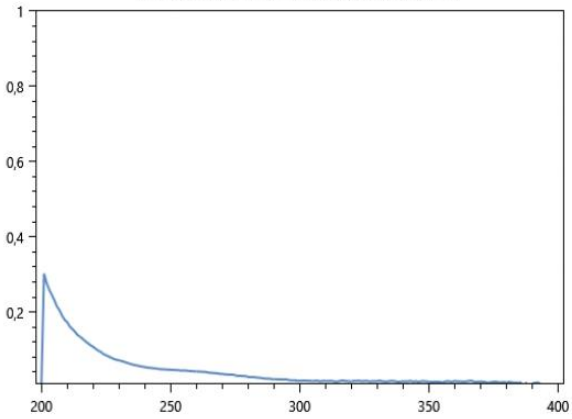
E1: Absorbance Scan 200-400 nm



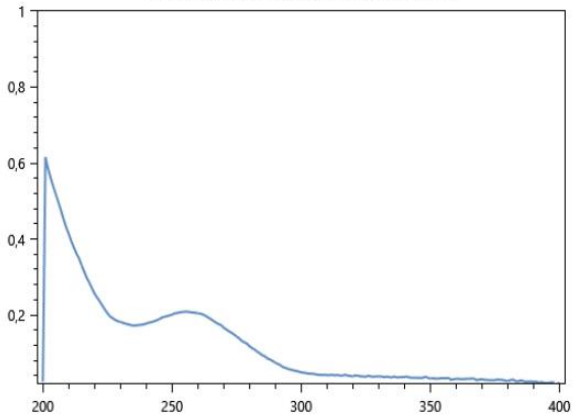
E2: Absorbance Scan 200-400 nm



F1: Absorbance Scan 200-400 nm

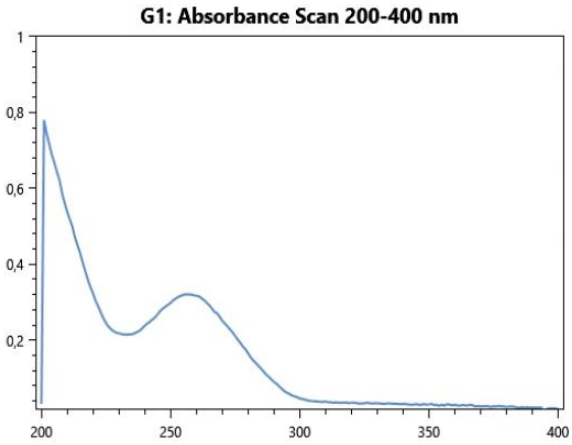


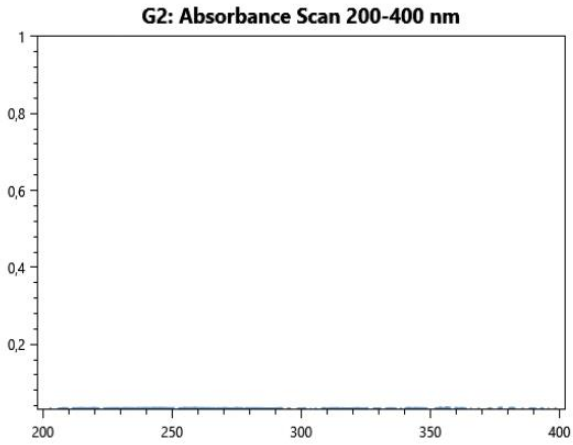
F2: Absorbance Scan 200-400 nm

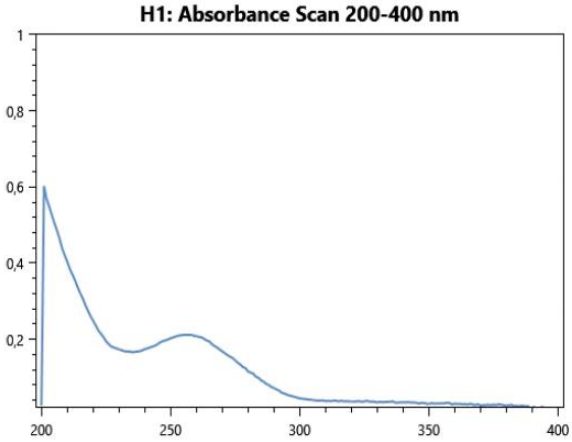


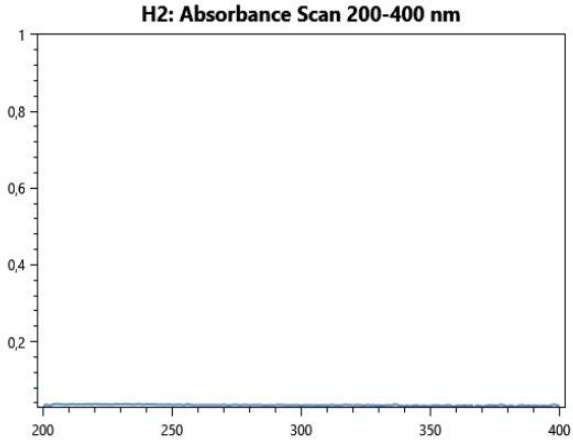
Report

TECAN.







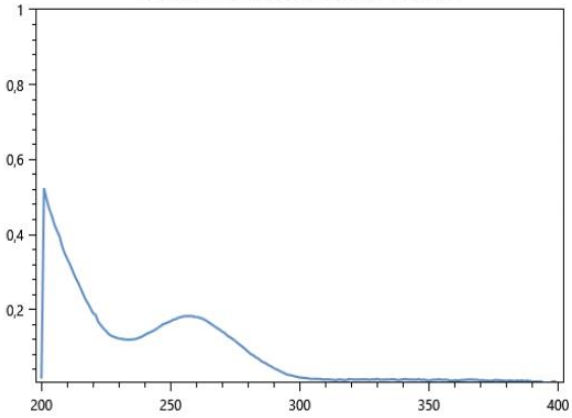


Well	260	230	280	ng/µl	Ratio 260/230	Ratio 260/280
A1	0,1677	0,1093	0,0763	167,70	1,53	2,20
A2	0,0247	0,0488	0,0102	24,73	0,51	2,42
B1	0,0687	0,0664	0,0297	68,70	1,03	2,31
B2	0,0261	0,0484	0,0105	26,12	0,54	2,48
C1	0,0009	0,0013	0,0000	0,87	0,70	-24,72
C2	0,0005	0,0007	-0,0001	0,54	0,82	-8,98
D1	0,0002	0,0000	-0,0003	0,22	-12,58	-0,65
D2	0,0010	0,0015	0,0005	1,00	0,67	1,96
E1	0,0008	0,0013	0,0002	0,77	0,58	3,48
E2	0,0008	0,0014	0,0003	0,81	0,58	2,72
F1	0,0009	0,0013	0,0000	0,89	0,70	-46,21
F2	0,0010	0,0018	0,0001	1,05	0,58	14,28
G1	0,0011	0,0022	0,0004	1,09	0,49	3,01
G2	0,0004	0,0009	-0,0001	0,39	0,43	-7,61
H1	0,0016	0,0031	0,0005	1,56	0,50	3,18
H2	0,0013	0,0030	0,0002	1,29	0,44	6,07

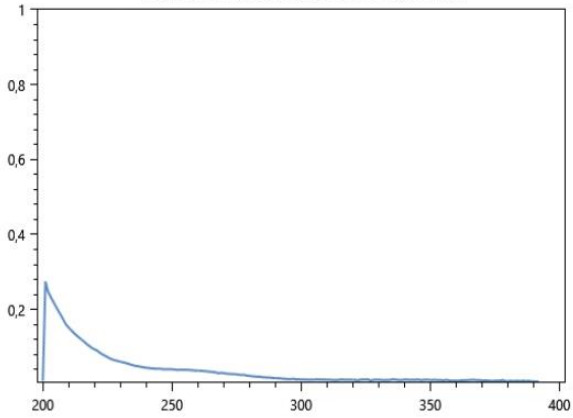
Report

TECAN.

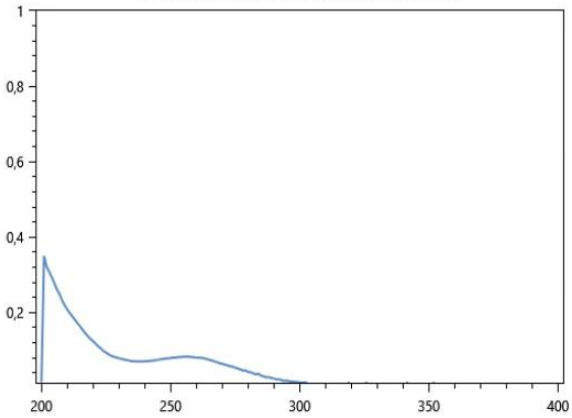
A1: Absorbance Scan 200-400 nm



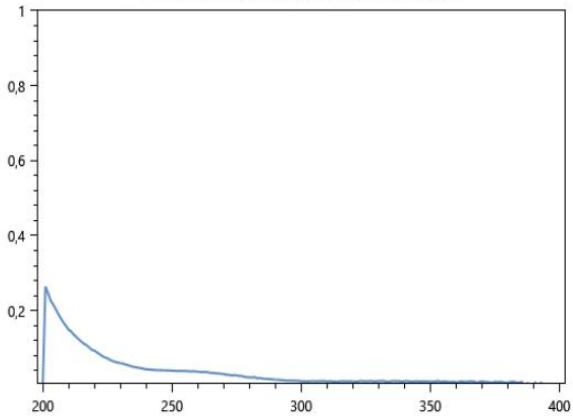
A2: Absorbance Scan 200-400 nm



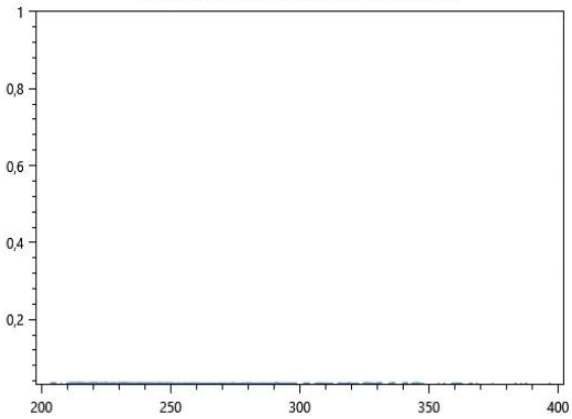
B1: Absorbance Scan 200-400 nm



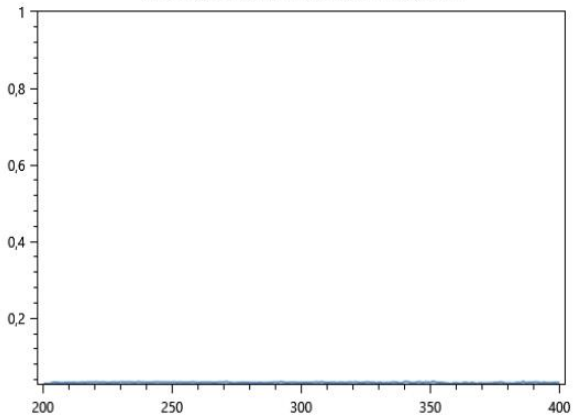
B2: Absorbance Scan 200-400 nm



C1: Absorbance Scan 200-400 nm



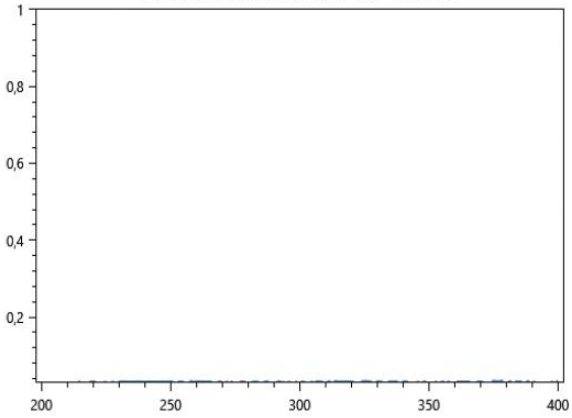
C2: Absorbance Scan 200-400 nm



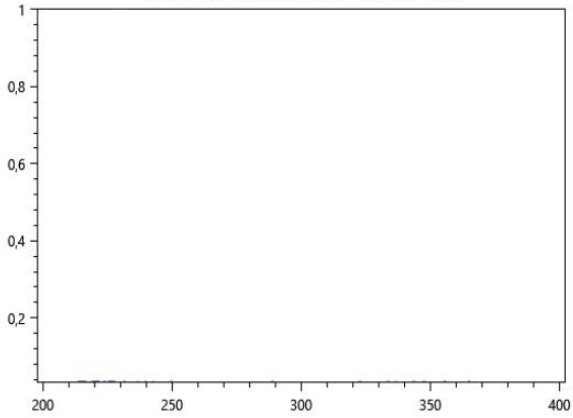
Report

TECAN.

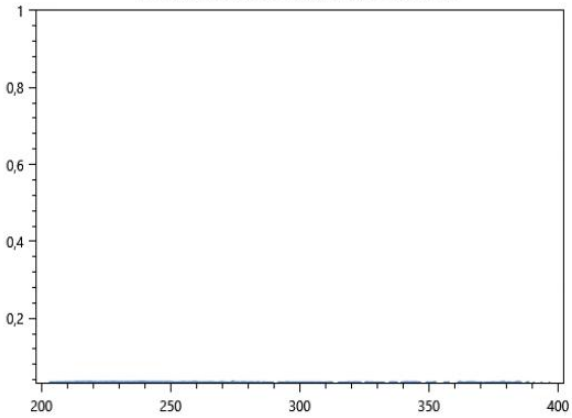
D1: Absorbance Scan 200-400 nm



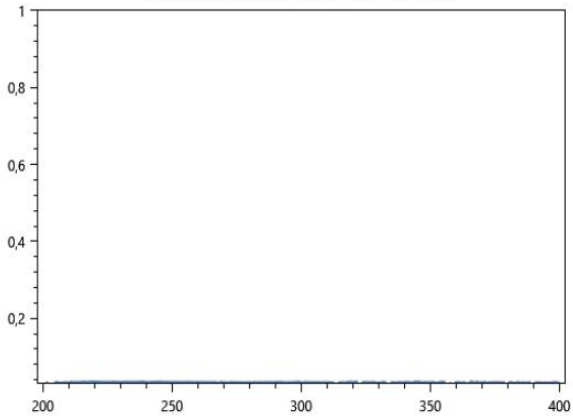
D2: Absorbance Scan 200-400 nm



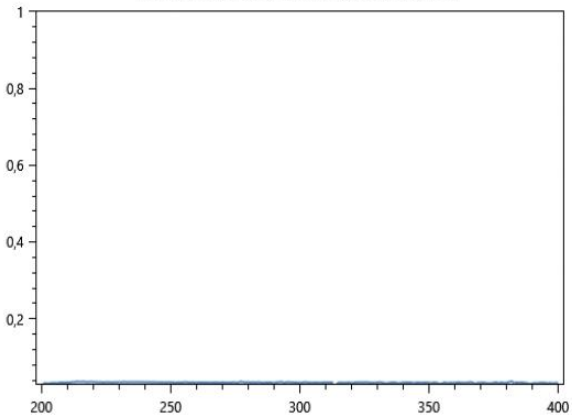
E1: Absorbance Scan 200-400 nm



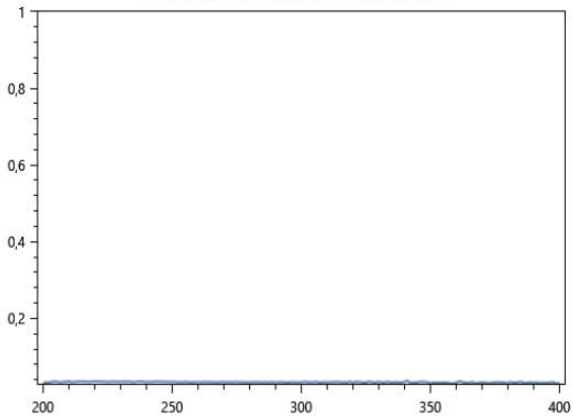
E2: Absorbance Scan 200-400 nm



F1: Absorbance Scan 200-400 nm

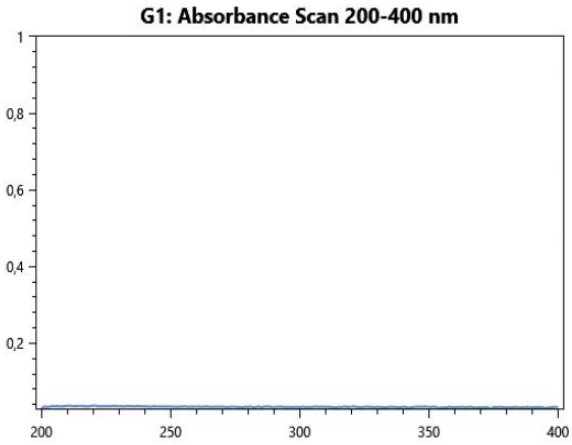


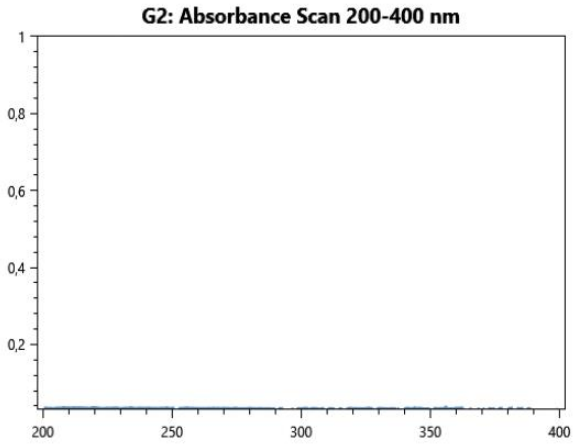
F2: Absorbance Scan 200-400 nm

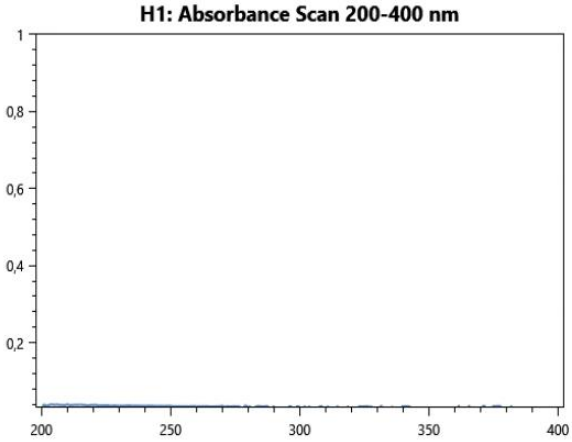


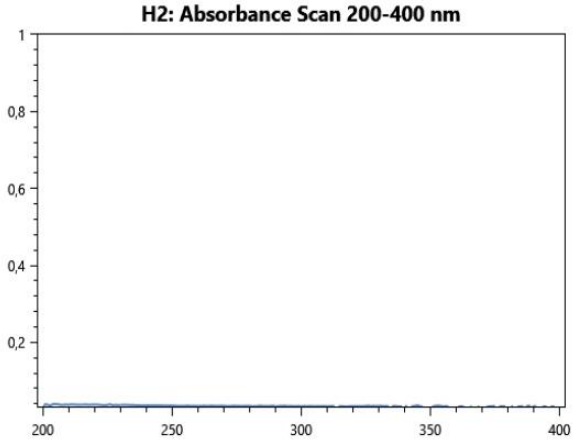
Report

TECAN.









Annex 2: DNA extraction and purification protocol

agap	Instruction Atelier de marquage moléculaire INRA **Extraction d'acides nucléiques à partir de lots de graines de tournesol.** Purification phénol/chloroforme, billes magnétiques et robot KingFisher	PEX-EXT-004GT 15/11/2022 Page 1 sur 7

Purification d'acide nucléique total à partir de lots de graines de tournesol

Avec purification au phénol/chloroforme et billes magnétiques CHEMAGEN, à l'aide du robot KingFisher

Historique des versions

Référence : PEX-EXT-0040		**Gestionnaire : Responsable qualité**
Version	**date de version**	Historique des modifications
Version 00	15/11/2022	Sylvain Santoni

Etat du document : A Contrôler

Confidentialité : Public

Sommaire

1. Objectif /Principe

L'objectif de ce protocole est d'isoler rapidement et de manière fiable de l'ADN génomique de haute qualité à partir de tissus végétaux. Ce protocole utilise la technologie de « binding » des acides nucléiques sur billes magnétiques et est donc dédié à l'utilisation conjointe d'une robotique adaptée (KingFisher Flex, instruction « **INS-EXT-005-Kingfisher-Flex.pdf** »).

Ce protocole a été spécifiquement mis au point pour purifier l'ADN végétal et minimiser la co-purification avec des polysaccharides et des polyphénols de la cellule. Le tampon de lyse cellulaire utilise un détergent universel. Les lipides, en quantité importantes dans les graines de tournesol, sont éliminés au cours d'une extraction à l'aide de solvants organiques phénol/chloroforme. Enfin, les acides nucléiques totaux, ADN et ARN sont purifiés à l'aide des billes métalliques coatées avec de la silice (groupements Silanol) (Billes Perkin-Elmer Chemagen CMG 252-A). L'adsorption des acides nucléique est réalisée par déshydratation en présence d'alcool.

agap	Instruction Atelier de marquage moléculaire INRA **Extraction d'acides nucléiques à partir de lots de graines de tournesol.** **Purification phénol/chloroforme, billes magnétiques et robot KingFisher**	PEX-EXT-004GT 15/11/2022 Page 2 sur 7

Consignes de sécurité
Si nécessaire, décrire les précautions

- A la manipulation de produits dangereux
  - Ethanol
  - Isopropanol
  - Phénol
  - Chloroforme
  - Betamercapto-ethanol

- A l'utilisation de l'équipement (KingFisher Flex, voir **INS-EXT-005-Kingfisher-Flex.pdf**)

2. Matériel utilisé/nécessaire

- Centrifugeuse (3000 à 5000g)
- Incubateur-agitateur (65°C)
- Robot de purification des acides nucléiques (KingFisher Flex)
- Matériel de laboratoire classique, dont tubes plastiques en polypropylène Sarsdedt de 50 et 13 ml.

agap	Instruction Atelier de marquage moléculaire INRA **Extraction d'acides nucléiques à partir de lots de graines de tournesol.** **Purification phénol/chloroforme, billes magnétiques et robot KingFisher**	PEX-EXT-004GT 15/11/2022 Page 3 sur 7

3. Protocole expérimental

BROYAGE

Les graines fraiches sont broyées par lots de 60 graines.

Les graines sont broyées manuellement, dans un mortier avec un pillon.
Une petite quantité d'azote liquide est utilisée pour rendre cassante les graines et faciliter le broyage.

La poudre/pate obtenue, qui contient les coques et les tissus adipeux broyés, est transférée dans un tube Sarsdedt de 50 ml.

EXTRACTION

La poudre est reprise par 15 ml de tampon d'extraction chaud.

TAMPON D'EXTRACTION : *200mM Tris pH = 8.0, 50mM EDTA, 500mM NaCl, 1.25 % SDS, 1% PVP 40000 + 1 ml/100ml Betamercapto-ethanol, 1% final (à ajouter extemporanément)*
Le tampon est préchauffé vers 50°C.

Agiter par inversion et au Vortex.
Incuber à 65°C pendant 30 min dans une étuve (tube posé à plat) sous agitation douce, 60rpm.

DEPROTEINISATION
Travailler sous sorbonne pour éviter de respirer les vapeurs d'acide acétique.

Incuber 5 min dans la glace.
Ajouter 7.5 ml d'acétate de K (5M/3M) froid (dans la glace). Agiter doucement par inversion puis homogénéiser au Vortex.
Incuber 5 min dans la glace.
Centrifuger 10 min à 4500 rpm à 4°C (Rotor Swing-out).

Prélever 5 ml de lysat clarifié pour chaque tube. 2 ou 3 fractions de lysats sont mélangées en fonction des échantillons.
Pour les différents échantillons mélangés, ajouter 1 volume (10 ou 15 ml) d'isopropanol.
Travailler sous sorbonne pour éviter de respirer les vapeurs d'isopropanol
Laisser 1h à -20°C

La manipulation peut être stoppée à cette étape en laissant les échantillons à -20°C durant la nuit

Centrifuger 10 min à 4500 rpm à 4°C (Rotor Swing-out), éliminer le surnageant, essorer le bord du tube.

Reprendre les culots dans 5 ml d'eau UP. Agiter fortement pour re-suspendre les culots

Elimination des lipides, extraction aux solvants organiques, PHENOL/CHLOROFORME

Ajouter 1 volume (5 ml) de phénol/Chloroforme (1/1)

Mélanger par retournement 5 min à T° amb.
Centrifuger 5 min à 4500 rpm à 4°C (Rotor Swing-out).

Transférer 2,5 ml de la phase aqueuse dans un autre tube de 13 ml.

Ajouter 1 volume (2,5 ml) de Chloroforme/Alcool isoamylique (24/1)
Mélanger 5 min par retournement à T° amb.
Centrifuger 5 min à 4500 rpm à 4°C (Rotor Swing-out).

Transférer 1,5 ml de la phase aqueuse dans une plaque Deepwell 24 du robot KingFisher

PURIFICATION des ADN sur Billes magnétique en utilisant un robot KingFisher

Préparation de la plaque de binding :

- Ajouter 150 ul de NaCl 5M
- 60 µl de billes magnétiques Chemagic CMG-252-A
- 0,7 ml de Chlorure de Guanidium 7.8M
- 2 ml d'isopropanol

Remarque : à cette étape, il est possible de laisser l'échantillon à 4°C pendant quelques heures (temps du repas ou d'une nuit par exemple)

Préparation des plaques de lavage pour le robot :

Wash 1 : 1 plaque 24 puits deepwell contenant 3600 ul de tampon AMMLAV/E
Wash 2 : 1 plaque 24 puits deepwell contenant 3600 ul de tampon AMMLAV/E
Wash 3 : 1 plaque 24 puits deepwell contenant 3600 ul de tampon Ethanol 75%
Wash 4 : 1 plaque 24 puits deepwell contenant 3600 ul de tampon Ethanol 75%

Elution : 1 plaque 24 puits deepwell contenant 250 ul de TE 1X

A partir de cette étape c'est le robot KingFisher qui prend en charge la purification.

Dépôt des plaques sur robot KingFisher en utilisant le programme adapté (voir le protocole d'utilisation du robot)

Récupérer la plaque d'élution à la fin du programme. Poser la plaque d'élution sur une plaque aimantée. Transférer les ADN dans les tubes Eppendorf de 2 ml en prenant le moins de billes restantes possible. Les ADN sont ensuite conservés à -20°C ou utilisés selon besoin.

Procéder à l'élimination des déchets et au lavage des plaques.

4. Réactifs

Réactifs chimiques, éléments de sécurité

	Formule chimique	Toxicité	Protection	Phrase de risque	Phrase de conseil de prudence
Tris (Trisma Base, Tris[hydroxymethyl]aminomethane)	$C_4H_{11}NO_3$	irritant	cutanée	R36/37/38	S26-36
EDTA (Ethylenediaminetetraacetic Acid)	$C_{10}H_{14}N_2O_8Na_2 2 H_20$	irritant	cutanée	R36	S26
Chlorure de sodium (sodium chloride)	NaCl		cutanée		
SDS (Sodium Dodecyl Sulfate, Sodium lauryl sulfate)	$C_{12}H_{25}O_4SNa$	irritant	Cutanée, respiratoire (si poudre)	R22-36/37/38	S26-36
Betamercapto-ethanol, 2-mercaptoethanol	HSCH2CH2OH	Toxique Nocif Irritant	Cutanée, respiratoire, oculaire	R20/22-24-34-51/53	S26-36/37/39-45-61
Acetate de potassium (potassium acetate)	CH_3COOK	irritant	Cutanée, respiratoire	R36/37/38	S26-36
Acide acétique glacial (glacial acetic acid)	CH_3COOH	irritant	Cutanée, respiratoire, oculaire	R10-35	S26-36/37/39-45
Polyvinylpyrrolidone PVP (PVP 40000, PVT 40T)	$[C_6H_9NO]_n$	Irritant	Cutanée, respiratoire		S22-24/25
Isopropanol (Alcool isopropylique, isopropyl alcohol)	[CH3]2CHOH		Cutanée, respiratoire	R11-36-67	S7-16-24/25-26
Phénol (hydroxybenzene, benzenol)	C6H5OH	Toxique	Cutanée, respiratoire, oculaire	R24/25-34	S26-28-36/37/39-45
Chloroforme (chloroform, trichloromethane)	CHCl3	Nocif irritant	Cutanée, respiratoire, oculaire	R22-38-40-48/20/22	S36/37
Alcool isoamylique	C5H12O/CH3CH(CH3)CH2CH2OH	Nocif irritant	Cutanée, respiratoire, oculaire	R: 10-20	S: (2-)24/25
Ethanol (alcool ethylique, Ethyl alcool)	CH_3CH_2OH		cutanée	R11-20/21/22-36/37/38	S7-16-24/25/26

agap	Instruction Atelier de marquage moléculaire INRA **Extraction d'acides nucléiques à partir de lots de graines de tournesol.** **Purification phénol/chloroforme, billes magnétiques et robot KingFisher**	PEX-EXT-004GT 15/11/2022 Page 6 sur 7

PREPARATION DES SOLUTIONS

Tampons d'extraction

Tampon d'extraction SDS :
200mM Tris pH = 8.0, 50mM EDTA, 500mM NaCl, 1.25 % SDS, 1% PVP 40000

Préparé à partir de solutions stock disponibles (cf fiches de préparation)

SOLUTIONS/Poudres	100 ml	200 ml	300 ml	400ml
TRIS 1M PH = 8	20ml	40	60	80
EDTA 0,5M	10ml	20	30	40
NACL 2.5M	20ml	40	60	80
SDS 20%	6,25ml	12,5	19	25
PVP 40000	1 g	2 g	3 g	4 g
NA BISULFITE	1 g	2 g	3 g	4 g
H_2O UP	43.5ml	87.5ml	131ml	175

Ajouter le **Betamercapto-ethanol** au dernier moment.

Acétate de K (3M K et 5M Ac)
Pour 100 ml, mélanger 60 ml de KAc 5M, 11.5 ml d'acide acétique glacial (37%) et 28.5 ml d'H_2O.

Phénol/Chloroforme/Alcool Isoamylique (25/24/1) : solution tamponnée à pH 8, prête à l'emploi. Produit Sigma ref 77617-500 Ml)

Chloroforme:
Dans un récipient en verre, tamponner 4 volumes de chloroforme avec 1 volume d'une solution de Tris 10 mM pH 8. N'utiliser que la phase solvant pour le mélange avec l'alcool isoamylique.

Chloroforme/Alcool Isoamylique (24/1) : mélanger 24 volumes de chloroforme tamponné avec 1 volume d'alcool isoamylique

Chlorure de Guanidium 7,8 M : Dissoudre 37.25 gr de chlorure de guanidium (produit Sigma ref G-3272) dans l'eau jusqu'à un volume final de 50 mL de solution (faire le mélange dans un tube Sarstedt ou Falcon de 50 mL). Vérifier, si possible, l'indice de réfraction au réfractomètre : n = 1.449.

Billes magnétiques Chemagic (Perkin Elmer), ref. CMG-252-A.

Tampons de lavage

Wash 1 et 2 : **AMMLAV/E**

Faire une solution aqueuse **AMMLav** (Acétate de potassium 160 mM, Tris HCl ph 8 22,5 mM, EDTA 0,1 mM)

Solutions	100 ml	500 ml	1 litre
Tris 1M pH=8	2.25 ml	11.25 ml	22.5 ml

EDTA 0.5M	20 ul	100 ul	200 ul
Acétate de Potassium 5M	3.2 ml	16 ml	32 ml
H2O UP	Qsp 100 ml	Qsp 500 ml	Qsp 1 000 ml

AMMLAV/E : mélanger 100 ml d'AMMLAV avec 170 ml d'éthanol à 96%

Wash 3 et 4 : **éthanol 75 %**
Ethanol à 75%. En respectant les proportions de la Table de Gay-Lussac :
100 ml d'éthanol à 96% + 31 ml d'eau.

Tampon d'élution TE1X : Tris 10 mM pH 8, EDTA 1 mM.

Eau UP : eau ultra pure ou eau MilliQ : eau de laboratoire purifiée par le système MilliQ de Millipore. Test de conductivité 18MΩ/cm.

Nettoyage, élimination des déchets

Les tubes plastiques de 50 et 13 ml et la plaque DeepWell de binding contenant les résidus d'extraction sont éliminés en poubelle spécifique.
Les solutions contenant du chlorure de guanidium et de l'éthanol sont versées dans le bidon de récupération spécifique prévu à cet usage.
Les autres solutions peuvent être éliminées directement à l'évier.

Les plaques DeepWell de lavage, récupérables sont rincées à l'eau, à l'eau osmosée et sont autoclavées avant re-emploi.
Les plaques d'élution sont éliminées en poubelle spécifique.

Annex 4: Table of the FST calculated between bulks (yellow cells are statistically different)

	B1			B2			B3			B4A			B4B			B5		
	CLPlus	CL_545	CL_569	CLPlus	CL_545	CL_569	CLPlus	CL_545	CL_569	CLPlus	CL_545	CL_569	CLPlus	CL_545	CL_569	CLPlus	CL_545	CL_569
B1																		
B2	0,0016	0,0031	0															
B3	**0,0809**	0,0024	**0,0666**	**0,0909**	0	**0,0666**												
B4A	**0,0106**	**0,0104**	**0,1674**	**0,0176**	0,0025	**0,1674**	**0,044**	0,003	**0,036**									
B4B	**0,3413**	0,0046	**0,0443**	**0,88**	0,0002	**0,0443**	**0,136**	0	0,004	**0,287**	0,001	**0,06**						
B5	**0,0568**	**0,0231**	**0,0092**	**0,0663**	**0,0122**	**0,0092**	0,003	**0,013**	**0,035**	**0,025**	0,005	**0,126**	**0,176**	**0,01**	**0,018**			
B6	0	**0,0231**	**0,0185**	0,0013	**0,0122**	**0,0185**	**0,082**	**0,013**	**0,022**	**0,011**	0,005	**0,104**	**0,343**	**0,01**	**0,008**	**0,058**	0	0,002

Annex 5: Table of allelic frequencies for each bulk

	CL-Plus (mut_319)				CL(mut_545)				CL(mut_569)				CL(mut_545_569)			
	Nb_R1	Nb_R2	Nb_R_mapping	Allele_Frequency	Nb_R1	Nb_R2	Nb_R_mapping	Allele_Frequency	Nb_R1	Nb_R2	Nb_R_mapping	Allele_Frequency	Nb_R1	Nb_R2	Nb_R_mapping	Allele_Frequency
B1_T1	162	46	17922	1,16%	773	155	17698	5,24%	168	54	17721	1,25%	1	2	17721	0,02%
B1_T2	166	30	18627	1,05%	669	129	18310	4,36%	134	56	18322	1,04%	1	0	18310	0,01%
B2_T1	45	10	19495	0,28%	495	105	19460	3,08%	263	82	19479	1,77%	1	0	19460	0,01%
B2_T2	74	25	20836	0,48%	403	80	20642	2,34%	47	20	20675	0,32%	1	0	20642	0,00%
B3_T1	3519	603	20705	**19,91%**	745	128	20481	4,26%	2715	825	20507	**17,26%**	26	4	20481	0,15%
B3_T2	2818	491	21285	**15,55%**	291	53	21118	1,63%	2255	653	21134	**13,76%**	8	1	21118	0,04%
B4A_T1	711	101	23719	3,42%	21	10	23598	0,13%	4992	1366	23627	**26,91%**	3	1	23598	0,02%
B4A_T2	1067	174	21545	5,76%	446	95	21444	2,52%	6038	1670	21469	**35,90%**	19	6	21444	0,12%
B4B_T1	8001	1349	16902	**55,32%**	231	42	16745	1,63%	1525	503	16756	**12,10%**	12	1	16745	0,08%
B4B_T2	8394	1462	19318	**51,02%**	449	112	19166	2,93%	1578	471	19176	**10,69%**	14	4	19166	0,09%
B5_T1	1694	309	19029	**10,53%**	21	6	18813	0,14%	670	218	18832	4,72%	0	1	18813	0,01%
B5_T2	2785	443	19696	**16,39%**	23	7	19494	0,15%	602	201	19530	4,11%	1	0	19494	0,01%
B6_T1	273	60	18434	1,81%	10	7	18211	0,09%	742	244	18231	5,41%	1	0	18211	0,01%
B6_T2	17	21	18861	0,20%	14	5	18696	0,10%	826	251	18717	5,75%	1	0	18696	0,01%
TN1	20	10	18439	0,16%	17	14	18325	0,17%	23	17	18346	0,22%	0	0	18325	0,00%
TN2	107	30	21107	0,65%	18	10	21043	0,13%	50	22	21074	0,34%	0	0	21043	0,00%
TP1	16	10	17699	0,15%	14	6	17601	0,11%	472	140	17642	3,47%	0	0	17601	0,00%
TP2	47	20	19226	0,35%	39	13	19162	0,27%	323	119	19178	2,30%	0	0	19162	0,00%
T1_2007	20	30	19073	0,26%	28	11	18702	0,21%	20	28	18733	0,26%	0	0	18702	0,00%
T2_2007	10	16	15247	0,17%	9	8	15266	0,11%	15	19	15295	0,22%	0	0	15266	0,00%
H201	0	0	25	0,00%	0	0	25	0,00%	1	0	25	4,00%	0	0	25	0,00%
H202	1	0	9	11,11%	0	0	7	0,00%	0	0	7	0,00%	0	0	7	0,00%