Article

Identification of Structural Variants in Two Novel Genomes of Maize Inbred Lines Possibly Related to Glyphosate Tolerance

Medhat Mahmoud [1], Joanna Gracz-Bernaciak [2], Marek Żywicki [3], Wojciech Karłowski [3] , Tomasz Twardowski [2] and Agata Tyczewska [2],*

[1] Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; helmy.medhat@gmail.com

[2] Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznań, Poland; j.gracz@gmail.com (J.G.-B.); twardows@ibch.poznan.pl (T.T.)

[3] Department of Computational Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, Umultowska 89, 61-614 Poznań, Poland; marek.zywicki@amu.edu.pl (M.Ż.); Wojciech.Karlowski@amu.edu.pl (W.K.)

* Correspondence: agatat@ibch.poznan.pl; Tel.: +48-61-852-8503

check for updates

Abstract: To study genetic variations between genomes of plants that are naturally tolerant and sensitive to glyphosate, we used two *Zea mays* L. lines traditionally bred in Poland. To overcome the complexity of the maize genome, two sequencing technologies were employed: Illumina and Single Molecule Real-Time (SMRT) PacBio. Eleven thousand structural variants, 4 million SNPs and approximately 800 thousand indels differentiating the two genomes were identified. Detailed analyses allowed to identify 20 variations within the EPSPS gene, but all of them were predicted to have moderate or unknown effects on gene expression. Other genes of the shikimate pathway encoding bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase and chorismate synthase were altered by variants predicted to have a high impact on gene expression. Additionally, high-impact variants located within the genes involved in the active transport of glyphosate through the cell membrane encoding phosphate transporters as well as multidrug and toxic compound extrusion have been identified.

Keywords: maize; glyphosate; herbicide resistance; stress responses; genome sequencing; large structural variants; SNPs; indels

1. Introduction

Maize (*Zea mays* L.) originates from Mexico and therefore has adapted to a climate with high temperatures and high light intensity during the day and moderate temperatures at night [1]. Because of its high productivity, importance for food and feed production, and numerous industrial applications, maize has become one of the most important crop species worldwide; it is grown over a wide range of latitudes [2,3]. As a thermophilic plant, when grown in nonoptimal conditions, maize encounters environmental stresses and hence it is crucial to minimize the competition for water, light, and minerals [4] that occurs between maize seedlings and weeds growing in the field. The easiest way to eradicate weeds is the application of herbicides. One of the most popular and widely used (in the cultivation of maize and crops in general) herbicides of high economic value and effectiveness against a wide range of weeds is the nonselective foliar-applied Roundup® [5,6]. Glyphosate, the active ingredient in Roundup®, has its primary effect via the inhibition of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS; EC 2.5.1.19) [7], an enzyme involved in one of the reactions in the shikimate pathway,

which leads to the synthesis of aromatic amino acids—tryptophan, phenylalanine, and tyrosine. Exposure to glyphosate causes deficiencies in those amino acids and leads to the consequent inhibition of protein synthesis. Another important aftereffect is the disturbance of carbon flow related to the accumulation of useless shikimate, as well as the drainage of carbon into other essential pathways such as auxin, flavonoid, or quinone biosynthesis [8–14]. In the past decade, new proteomic [15] and transcriptomic [16–18] and metabolomic [19] methods have been applied to analyze glyphosate influence on plant metabolism. Despite these reports, we are still missing a holistic image of plant responses to glyphosate and, more generally, herbicide stress conditions.

Zea mays has a large genome of roughly 2.4 gigabases and approximately 40,000 genes, spread across 10 chromosomes (https://plants.ensembl.org/Zea_mays/Info/Annotation). Some of the chromosomes contain highly repetitive heterochromatic domains called "chromosomal knobs". The genome is transposon-rich and is composed of up to 85% repetitive sequences, such as transposable elements (TEs), and thus, its organization is highly complex [20,21]. Moreover, maize is one of the crop species with high genetic variation between different lines of the same species. These differences might range from 25% to 84% due to the disparities in TE content since the activation and/or loss of TEs has a great impact on genome structure and gene expression. Moreover, single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), and structural variants (SVs) such as copy number variations (CNVs), inversions and translocations also take part in generating substantial differences between maize genomes [22,23]. Numerous studies have shown that genome structure variations are associated with a wide range of plant phenotypic traits, resulting in metabolic fluctuations and different regulation of gene expression [24–27].

The first fully sequenced maize genome was B73, published in 2009 [21]. In the following years, full or partial genomic sequences of other varieties, such as Palomero Toluqueno, B104, CML247, DK105, EP1, F7, Mo17, PE0075, PH207, W22, and PI 566673, have been released, and the details are available in the MaizeGDB database (https://www.maizegdb.org/genome/assemblies_overview) [28]. Recently, a new version of the maize B73 reference genome was published (RefGen_v4) [22].

In this study, we analyzed the genome structures of two naturally inbred maize lines (S245 and S7957), which differ in resistance to glyphosate. To identify SVs, SNPs and indels which could be involved in gaining herbicide resistance in non-GMO maize, two different sequencing technologies have been applied (Illumina and SMRT PacBio).

2. Results

2.1. Verification of Sequencing Data

For the purpose of sequencing and to overcome the complexity of the two maize genomes two sequencing technologies: Illumina (two paired-end libraries with insert length of 400 bp and 500 bp generating short but highly accurate sequencing reads) and SMRT PacBio (two mate-pair libraries with insert length of 8 kb and 11 kb, longer but less precise reads) were used. To evaluate the obtained sequencing results, alignment rates and genome coverages for each library were calculated, based on maize B73 genome [21]. For Illumina libraries, the average coverage was high (>67×). Moreover, high alignment rates indicated that both lines are close to the reference genome of the maize B73 line. The examination of the cumulative distribution of genome coverage indicates that approximately 45% of the genome was covered with at least 50 Illumina reads (details are presented in Table 1). For PacBio reads, after correction with the HALC program, the average coverage for both lines was > 3×, with a slightly higher value observed in the sensitive line (3.59×, vs 3.06× in the tolerant line). Due to the low overall coverage of PacBio reads, the saturation of genic and intragenic regions of both genomes was compared. In genic regions, the coverage in the sensitive line was higher than the coverage in the tolerant line (accordingly, 78% vs 62%). Nevertheless, the minimum coverage of 2, which is necessary to identify the structural variants, was observed for the majority of genic and intergenic regions in both lines (55% and 70% of genic regions, 60% and 65% of intergenic regions of

tolerant and sensitive lines, respectively). We concluded that both obtained data sets (Illumina and PacBio) will allow the identification of structural variants between the lines.

Table 1. The alignment results for Illumina libraries in glyphosate-tolerant and glyphosate-sensitive *Zea mays* lines.

	Tolerant Line (S245)			Sensitive Line (S79757)		
Library Size	Millions of Reads	Genome Coverage	Alignment Rate	Millions of Reads	Genome Coverage	Alignment Rate
400 bp	816	37.5	97.1%	821	37.7	96.6%
500 bp	664	27.8	85.3%	615	29.1	97.0%
8 kb	129	3.7	86.4%	74	3.0	87.7%
11 kb	59	2.2	86.0%	116	2.6	70.0%

2.2. Identification of Large Structural Variants

Long PacBio reads, which are able to cover substantial chromosomal regions and directly reveal large deletions, insertions, duplications, or inversions [29], allowed for detection of structural differences between genomes of selected maize inbred lines. A total of 11,172 structural variants were identified representing 6062 insertions and 5110 deletions (only variants supported with 5 or more reads were investigated). Most variants were located in noncoding regions, such as upstream/downstream regions of the genes and intronic and intergenic regions. These sets of identified SVs were functionally annotated using the Ensembl variant effect predictor (VEP) to verify the probable influence of the identified changes on gene expression levels (Figure 1).
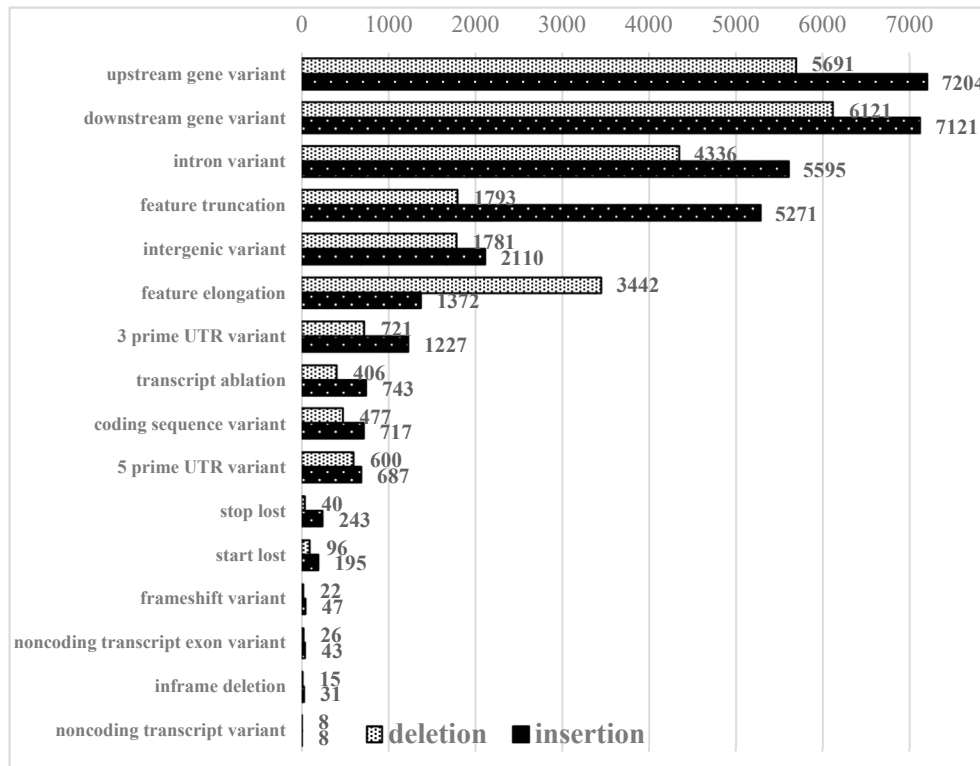

Figure 1. Predicted consequences of identified structural variants.

When analyzing such a narrow set of variants, the most frequent effect was transcript ablation (33.2%). Other predicted consequences of identified structural changes, such as coding sequence variation (12.32%), intron variant (9.25%), feature truncation (9.25%), or start/stop codon loss (8.41% and 8.18%, respectively) were less frequent. At the same time, only a few cases of transcript elongation or frameshift were identified. The vast majority of the variants were found to affect noncoding parts of the genome (upstream/downstream regions of the genes (22.16% and 22.75%, respectively), intronic (17.07%), and intergenic regions (6.69%)). Coding sequence variants, transcript ablation, frameshift variants and inframe deletions represented 2.05%, 1.97%, 12%, and 0.08%, respectively. Only those SVs with high possibility impact on phenotype were selected for further analysis.

2.3. Identification of Small Structural Changes

For the identification of smaller structural changes, such as SNPs and indels, in genomes of tolerant and sensitive maize lines, we used Illumina reads. After each dataset read was aligned to the reference genome of the maize B73 line, only changes differentiating between both lines were reported. In this way, we obtained a large number of variances (Table 2).

Considering that maize line B73 is sensitive to glyphosate [30] for a more accurate prediction, the pool of SNPs and indels was narrowed to those occurring only in the tolerant line. As a result, a comprehensive list of genetic variations observed exclusively in glyphosate-tolerant line has been obtained. This approach resulted in the identification of 4,068,829 SNPs and 729,866 indels (Table 2), among which 113,775 SNPs and 15,277 indels were located within the protein coding regions.

Table 2. Summary of identified small variants from Illumina reads.

Variant Group	SNPs	Indels
All	13,778,463	2,443,262
Tolerant line-specific	4,068,829	729,866
Located in coding sequence	113,775	15,277

To gain the insight into functional consequences of the identified variants, the annotation with Ensembl variant effect predictor (VEP) has been performed. The results for all and high impact consequences are given in Figure 2. The vast majority of the variants were found to affect noncoding (intergenic regions (25.13%), upstream/downstream regions of the genes (24.9% and 25.18%, respectively), and intronic (17.56%) parts of the genome.

2.4. Gene Ontology Analysis

Analysis of the distribution of gene ontology terms associated with genes influenced by high-impact large structural variants was performed in the next step. The obtained list of GO terms was summarized using REViGO. Among the genes affected by deletions, the most abundant GO terms in molecular function division were iron ion binding, methyltransferase activity, transferase activity and RNA binding, while in the biological process category, the most abundant terms were glycogen biosynthesis, protein K48-linked deubiquitination, carbohydrate metabolism and metal ion transport. It is worth noting that some of those genes were also annotated as a part of the response reaction to heat, nematode and oxidative stress such as Zm00001d037385, which encodes the adenine nucleotide alpha hydrolase-like superfamily protein or genes Zm00001d052001, Zm00001d010838 (ZIP zinc/iron transport family protein), and Zm00001d002426 (thylakoid lumenal 29 kDa protein, chloroplastic) [31–33]. On the other hand, among genes affected by insertions, the most abundant GO terms in molecular function were RNA binding, SUMO transferase activity and hydrolase activity. In the biological process category, protein sumoylation, cell growth, ATP hydrolysis and cellular response to extracellular stimulus were the most common GO terms. In the last category, two MYB transcription factors, Zm00001d029963 and Zm00001d037836, were identified.
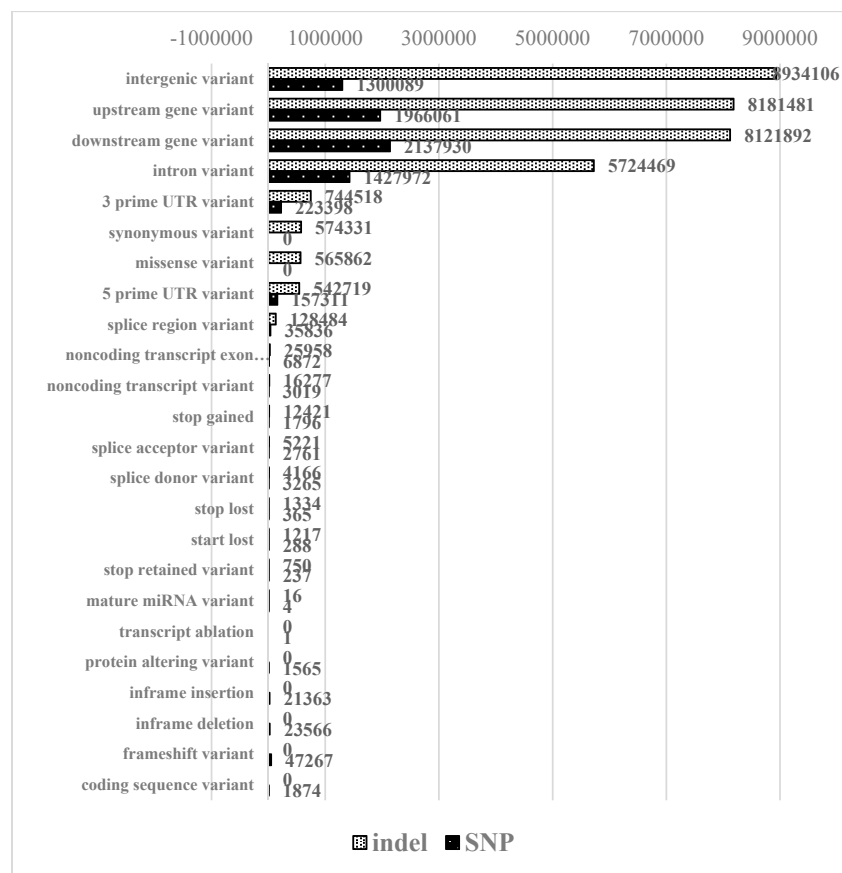
Figure 2. Predicted high impact consequences of identified indels and single nucleotide polymorphisms (SNPs).

Analysis of GO terms associated with high-impact SNPs and indels showed that a group of genes belonging to the RNA binding category was the most affected by those changes, similar to the results from the annotation of large insertions identified by PBSuite. The second most abundant GO term was methyltransferase activity. This group, in turn, was the most abundant GO term within genes affected by large deletions. Notably, in GO terms related to biological processes, the largest group of genes affected by SNPs and indels belonged to the category of DNA repair. These results suggest the existence of potential differences in the ability to maintain genome homeostasis in maize lines tolerant to glyphosate.

3. Discussion

The development and maturation of organisms is constantly influenced by environmental conditions that sometimes may exert a detrimental influence on them. Plants, due to their sedentary lifestyle, require short-term strategies to quickly and efficiently readapt their metabolism and have developed unique features in terms of habitat, growth and reproduction, often being a result of differences in their genomes [34].

Genetic variation, being the result of continuous changes ranging from single-nucleotide polymorphisms (SNPs), through indels, to large structural variants (SVs) like chromosomal rearrangements, provides raw material on which both natural selection and human act. Despite the advances in sequencing technologies, SVs continue to be largely unexplored elements of plant genomes. Understanding the type and size of SVs, their distribution among individuals and in population is critical for understanding the contributions of SVs to phenotypes, and the plants' ability to adapt to changing environments [25].

Since the basic challenge in genome studies is correlating genomic DNA variation with observed heritable phenotypes for our studies, we chose two maize inbred lines that differ in herbicide tolerance (Supplementary Figure S1), a variation that has been observed in the field (see Materials and methods). The differences recognized in both analyzed genomes and their functional diversity suggest a large genetic distance between tested maize inbred lines and that the vast majority of structural variants (SNPs and indels) are related to line-specific variability rather than to the glyphosate tolerance trait. Among all SNPs (13,778,463) and indels (2,443,262) detected in our study 4,068,829 (29.53%) SNPs and 729,866 (29.87%) indels were tolerant-line specific (Table 2) suggesting possible connection with glyphosate tolerance. Most highly represented changes caused by structural variants (for variants with a predicted high impact on gene expression) were transcript ablation (1149 events), coding sequence variant (426), followed by intron variation (320), feature truncation (320), stop loss (283), and 3' UTR variant (283) (Figure 1).

To recognize genome changes that are potentially responsible for the tolerant phenotype, we examined genes and pathways involved in glyphosate metabolism like the shikimate pathway (Supplementary Figure S2). First, we analyzed variants of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene as the best-known example of an enzyme inhibited by glyphosate [35–38]. To date several means of overcoming susceptibility to glyphosate have been described; for instance, amino acid substitution in the enzyme's active site [39–41], copy number variation of the gene and its overexpression [42]. In the case of the tolerant maize line studied herein no large structural variants, only several SNPs and indels within or in the vicinity of the EPSPS gene have been identified (Figure 3).
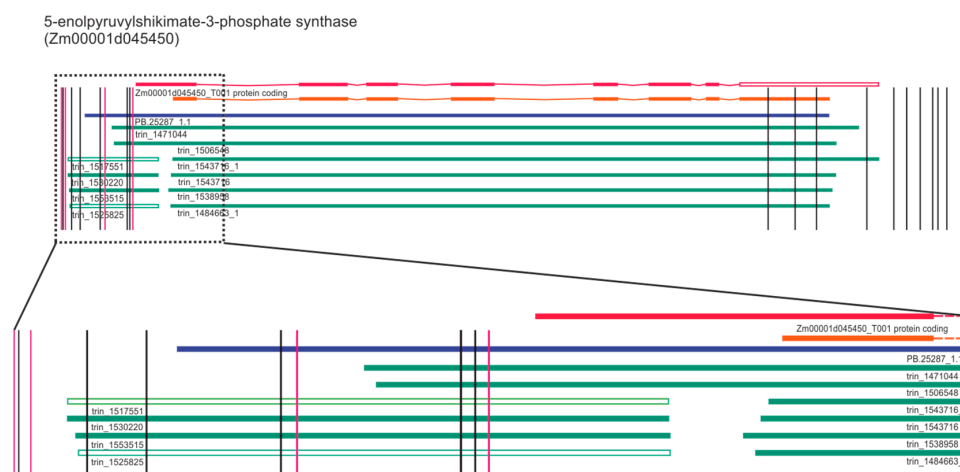

Figure 3. Location of identified SNPs (black) and indels (violet) on the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene.

In red, the EPSP gene structure according to the version 4 annotation of the maize genome, and in orange, the version 3 annotation have been given. Blue bars represent the publicly available full-length EPSPS transcript isoform identified with the iso-seq approach; green bars represent publicly available Trinity-assembled transcripts from RNA-seq data. The first part of the genes and transcripts, where the highest number of changes occurs, has been enlarged (indicated by black dashed line) to clearly see the SNPs and indels.

Importantly, all identified changes had, according to VEP, a moderate or modifier impact on gene expression, and none of them were located within the coding region. Previously it has been shown that a mutation in EPSPS coding sequence can provide resistance to glyphosate, as demonstrated in goosegrass, *Eleusine indica* and Pantoea sp. [35,43,44].

Two other genes coding enzymes involved in the shikimate pathway (bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase and chorismate synthase) [45] were also affected by detected variations (Table 3, Figure 4).

Table 3. Shikimate pathway genes affected by SNPs and indels.

Gene	Protein	Variation	Consequence
100272333	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase chloroplastic	SNPs, indels	Splice donor/frame shift
100381407	Chorismate synthase chloroplastic	SNPs, indels	Splice donor/frame shift
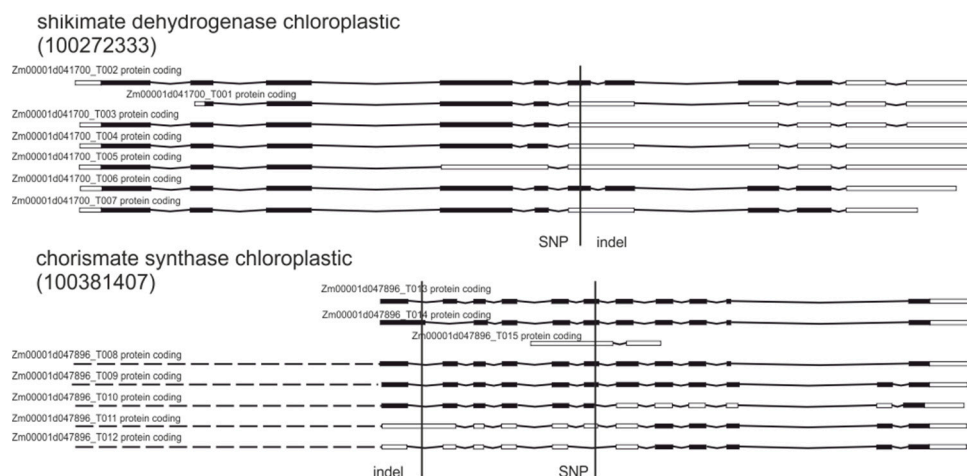
Figure 4. Genetic variations in genes encoding proteins involved in shikimate pathway.

For both tested genes (shikimate dehydrogenase chloroplastic and chorismate synthase chloroplastic) the impact on gene expression was high, affecting splice sites and resulting in frame shifts. If these changes lead to higher activity of both enzymes, it would be possible to at least partially compensate for the inhibitory effect of glyphosate on EPSPS by increasing EPSPS substrate synthesis and elevated intake of the EPSPS product.

Other identified structural variations that could result in increased tolerance to glyphosate were located in genes encoding phosphate transporters (Table 4, Figure 5).

Table 4. Genes encoding phosphate transporters that were affected by identified variations.

Gene	Protein	Variation	Consequence
Zm00001d051945	Phosphate transporter 2	SNP	Splice acceptor
Zm00001d018445	Phosphate transporter 3	indel	Splice donor
Zm00001d012747	Putative sugar phosphate/phosphate translocator	SNP, indel	Splice donor/Splice acceptor
Zm00001d021653	Glucose-6-phosphate/phosphate translocator 2	indel	Frame shift
100191756	Probable sugar phosphate/phosphate translocator	indel	Frame shift
Zm00001d011388	Putative sugar phosphate/phosphate translocator	indel	Frame shift

Previously it has been proposed that phosphate transporters 1 and 2 could participate in the active transport of glyphosate into plant cells [5,8,10,11,14,46]. Other identified transporters with SNP and indel changes (Table 4) participate in the distribution of inorganic phosphate, 3-phosphoglycerate, triose phosphates and, to a lesser extent, phosphoenolpyruvate (PEP), which is a substrate for the shikimate pathway that is directly affected by glyphosate inhibition of the EPSPS gene [47,48].
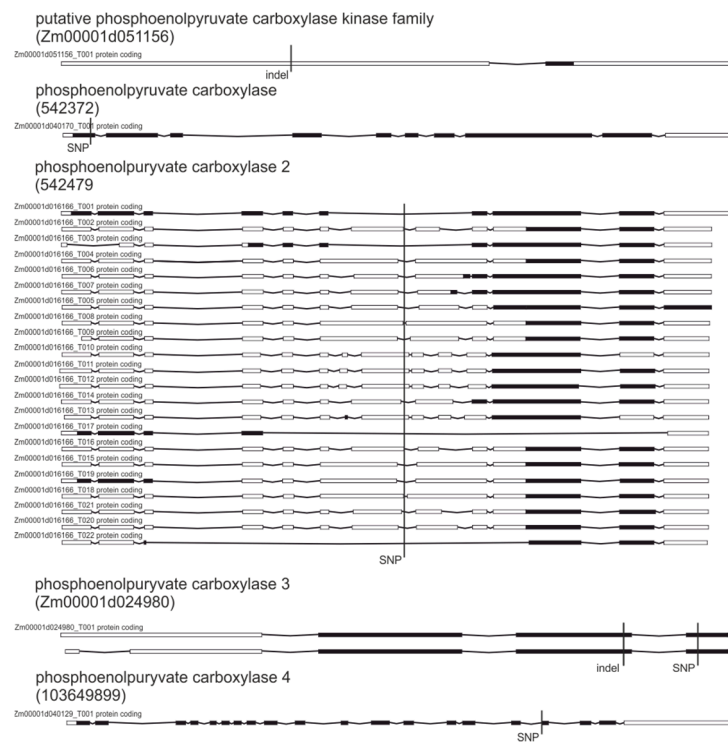
Figure 5. Genetic variations in genes encoding proteins related to lowering the availability of phosphoenolpyruvate for shikimate pathway.

Ten other genes affected by high-impact changes that are involved in PEP cellular availability are given in Table 5. Proteins encoded by those genes are important for maintaining appropriate concentrations of PEP in chloroplasts, where the shikimate pathway takes place [49–51].

Table 5. Genes associated with phosphoenolpyruvate availability affected by genetic variations between tested lines.

Gene	Protein	Variation	Consequence
100283648	Phosphoenolpyruvate/phosphate translocator 1 chloroplastic	SNP	Splice acceptor
103649694	Phosphoenolpyruvate/phosphate translocator 2 chloroplastic	indels	Frame shift
Zm00001d044715	Phosphoenolpyruvate/phosphate translocator 2 chloroplastic	indel	Frame shift
Zm00001d037659	Phosphoenolpyruvate/phosphate translocator 2 chloroplastic	SV insertion	Stop lost
542372	Phosphoenolpyruvate carboxylase	SNP	Stop gained
Zm00001d053453	Phosphoenolpyruvate carboxylase isoform 1	SNPs, indels	Splice acceptor/Splice donor
542479	Phosphoenolpyruvate carboxylase 2	SNP	Splice donor
Zm00001d024980	Phosphoenolpyruvate carboxylase 3	SNPs, indels	Stop gained/frame shift
103649899	Phosphoenolpyruvate carboxylase 4	SNP	Splice acceptor
Zm00001d051156	Putative phosphoenolpyruvate carboxylase kinase family protein	SNPs, indel	Stop gained/frame shift

The identified variations could potentially lead to stimulation of the shikimate pathway and at least partial compensation for herbicide action as observed changes cause splice acceptor changes, frame shifts, stop codon loss or gain (Figure 6).

In rice for instance, silencing of the PEPC gene lead to a 50%–60% increase in the activity of the shikimate pathway due to higher PEP availability [52]. On the other hand, increase of shikimic and protocatechuic (PCA) acids in nodules and leaves of nodulated lupine plants after herbicide application were explained by a diversion of most PEP into the shikimate pathway, depriving energy substrates to bacteroids to maintain nitrogen fixation [53].
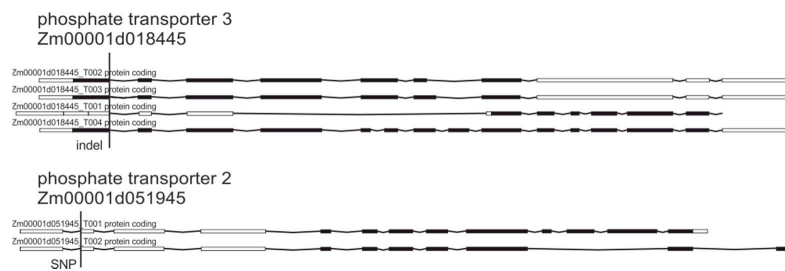

Figure 6. Genetic variations in phosphate transporter genes.

The next group of genes with high-impact structural variants that drew our attention in the context of herbicide stress responses were genes encoding proteins involved in multidrug and toxic compound extrusion (MATE, summarized in Table 6). This protein family is one of the most conservative and largest transporter families in plants and acts as membrane carriers of drugs and synthetic compounds, as well as organic acids, plant hormones and secondary metabolites [54–56].

Table 6. Multidrug and toxic compound extrusion genes affected by genetic variations.

Gene	Protein	Variation	Consequence
Zm00001d005080	Multidrug and toxic compound extrusion5	SNPs, indels	Splice acceptor/frame shift
Zm00001d013810	Multidrug and toxic compound extrusion2	indels, structural variant (SV) insertion	Frame shift/ start loss
100383875	Multidrug and toxic compound extrusion3	indel	Frame shift
100193278	Multidrug and toxic compound extrusion4	indel	Frame shift
Zm00001d035115	Multidrug and toxic compound extrusion1	indel	Stop gained

The MATEs gene family of cation antiporters also plays a large role in the exportation of toxins and other substrates and reportedly localize in plasma membranes [56,57]. Identified variants (that cause splice acceptor/donor change, frame shifts, stop codon loss or gain) may lead to the enhancement of glyphosate active transport and lowering of its cellular concentration (Figure 7).
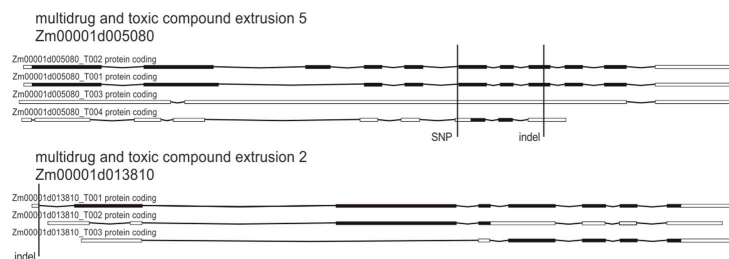

Figure 7. Genetic variations in selected genes encoding proteins involved in multidrug and toxic compound extrusion.

4. Conclusions

Harnessing the potential of genetic variations present within plant genomes and especially crop varieties, possible now due to technological and algorithmic advances, will undoubtedly forward our insight in some biological processes and may help in the improvement of agronomic crop species. In the face of increasing demands imposed by the impact of climate change and by a growing world population such improvement becomes necessity nowadays.

The presented results from the sequencing of two maize inbred lines give us insight into the molecular pathways that may be potentially involved in the glyphosate tolerance trait. The identified variations and changes are strong indicators for further biochemical analyses, especially those related to genes encoding three shikimate pathway enzymes (bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase and chorismate synthase). The described changes caused by genetic variations can lead to an increase in shikimate pathway efficiency and thus compensate for the decreased activity of the EPSPS caused by glyphosate. Additionally, genes encoding phosphoenolpyruvate carboxylase and its regulator, phosphoenolpyruvate carboxylase kinase, were also affected by identified variants of high impact on gene expression. The potential decrease in phosphoenolpyruvate carboxylase activity could lead to the increased availability of PEP, providing another level of EPSP inhibition compensation. Other mechanisms responsible for tolerance traits could be related to altered glyphosate transport via phosphate transporters or multidrug and toxic compound extrusion proteins. These mechanisms could modify the effective intracellular concentration of glyphosate and allow plants to thrive after Roundup® spraying. All identified differences in sequenced genomes shed a light on new potential ways that maize could respond to herbicidal stress conditions.

5. Materials and Methods

5.1. Plant Material and Sequencing

Maize seeds from inbred lines S245 (line tolerant to glyphosate, TL) and S79757 (line sensitive to glyphosate, SL) were grown in a greenhouse. Seeds from both lines were obtained from a local breeder (HR Smolice, Smolice, Poland). Field tests were used to verify the responses of the S245 and S7957 lines to glyphosate. When the highest concentration of herbicide (Round Up 360 SL) was used (300 g, 1.0 l/ha) the level of injuries to the tolerant maize variety was 40%, in contrast it reached 85% in the more sensitive line (K. Adamczewski, data not published) (Supplementary Figure S1).

Uniform seedlings from both lines were selected and grown at 22 °C, humidity 60% in the dark room. Seedlings were harvested after 7 days and immediately frozen in liquid nitrogen and stored at −80 °C until shipment. DNA isolation (done from a single seedling for each line), and sequencing were performed as outsourced services by the Fasteris company (Switzerland). Two sequencing technologies were used: Illumina libraries were prepared as two paired-end libraries with insert lengths of 400 bp and 500 bp and two mate-pair libraries with insert lengths of 8 kb and 11 kb; for long reads, Single Molecule Real-Time (SMRT) PacBio, RS II technology was used.

5.2. Preparation and Assessment of Sequencing Data

After removing adaptor sequences and sorting, Illumina reads were aligned using BWA-mem with default parameters and setting –M flag to mark split reads as secondary aligned. Coverage was calculated using the plotCoverage script from the deepTools package (https://github.com/fidelram/deepTools). For SMRT PacBio, 38 SMRT cells for the tolerant line and 40 SMRT cells for the sensitive line were sequenced. Reads were aligned using BWA [58] version 0.7.10 using -x PacBio with the following parameters: minimum seed length (-k17), where matches shorter than 17 are removed, max gap set to 40 where gaps longer than 40 are not found (-W40), (-r10) a key heuristic parameter for tuning the performance reseeding for an MEM longer than minimum seed length, matching score 2 (-A2), mismatch penalty 5 (-B5), gap opening penalty 2 (-O2), (-E1) setting gap penalty to

gap opening + minimum seed length * E), and clipping penalty 0 (-L0). After the assessment of correction tool performance [59], the HALC program was used for PacBio reads.

5.3. Calling Variance

GATK haplotypecaller (GATK HC) version 3.5 [60] was used to call variance between maize lines. The GATK best practice workflow was followed. Variants were filtered using hard filtering for SNPs, with suggested parameters, where SNPs matching any conditions will be marked FILTER and will not be considered in further analysis, and the remaining SNPs will be annotated as PASS (QD < 2.0).

5.4. Structure Variation Detection

Two methods of SV identification were employed: Sniffles [61] version 1.0.3 and PBSuite [62] version 15.8.24. In the first step, structural variants were detected with 2 minimum supporting reads, and the threshold was set on at least 5 reads. The comparison between variants detected with multiple approaches was performed using the BEDTools suite. The number of exact overlaps between corrected and uncorrected subreads identified SVs using the BEDTools intersect by setting -f 1.0 for 100% overlap and set–r flag, which requires the fraction of overlap to be reciprocal for both samples.

5.5. Functional Analysis

Only SNPs and SVs specific to the Roundup® tolerant line (S245) were further analyzed as potentially responsible for the tolerant phenotype. Changes in the S245 line were filtered out using SnpSift (http://snpeff.sourceforge.net/SnpSift.html), and then Variant Effect Predictor (VEP) was used to annotate selected SNPs. In the case of SVs, the gene ontology term accession associated with the SV affected genes was extracted using plant BioMart from Ensembl Plants genome browser (http://plants.ensembl.org/index.html) and then summarized using REViGO [63].

5.6. Sequence Deposition

All sequences have been deposited in ENA (European Nucleotide Archive) database under the following accession number PRJEB31400.

References

1. Jaenicke-Despres, V.; Buckler, E.; Smith, B.D.; Gilbert, M.T.; Cooper, A.; Doebley, J.; Paabo, S. Early allelic selection in maize as revealed by ancient DNA. *Science* **2003**, *302*, 1206–1208. [CrossRef]
2. Huang, C.; Sun, H.; Xu, D.; Chen, Q.; Liang, Y.; Wang, X.; Xu, G.; Tian, J.; Wang, C.; Li, D.; et al. ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E334–E341. [CrossRef] [PubMed]
3. Tyczewska, A.; Woźniak, E.; Gracz, J.; Kuczyński, J.; Twardowski, T. Towards food security: Current state and future prospects of agrobiotechnology. *Trends Biotechnol.* **2018**, *36*, 1219–1229. [CrossRef] [PubMed]
4. Craine, J.M.; Dybzinski, R. Mechanisms of plant competition for nutrients, water and light. *Funct. Ecol.* **2013**, *27*, 833–840. [CrossRef]

5.　Żywicki, M.; Gracz, J.; Karłowski, W.; Twardowski, T.; Tyczewska, A. Expression of miRNAs involved in phosphate homeostasis and senescence is altered in glyphosate-Treated maize. *Acta Physiol. Plant.* **2015**, *37*, 265. [CrossRef]

6.　Heap, I.; Duke, S.O. Overview of glyphosate-resistant weeds worldwide. *Pest Manag. Sci.* **2018**, *74*, 1040–1049. [CrossRef] [PubMed]

7.　Boocock, M.R.; Coggins, J.R. Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. *FEBS Lett.* **1983**, *154*, 127–133. [CrossRef]

8.　Denis, M.-H.; Delrot, S. Carrier-Mediated uptake of glyphosate in broad bean (Vicia faba) via a phosphate transporter. *Physiol. Plant.* **1993**, *87*, 569–575. [CrossRef]

9.　Geiger, D.R.; Kapitan, S.W.; Tucci, M.A. Glyphosate inhibits photosynthesis and allocation of carbon to starch in sugar beet leaves. *Plant Physiol.* **1986**, *82*, 468–472. [CrossRef]

10.　Hetherington, P.R.; Marshall, G.; Kirkwood, R.C.; Warner, J.M. Absorption and efflux of glyphosate by cell suspensions. *J. Exp. Bot.* **1998**, *49*, 527–533. [CrossRef]

11.　Morin, F.; Vera, V.; Nurit, F.; Tissut, M.; Marigo, G. Glyphosate uptake in Catharanthus roseus cells: Role of a phosphate transporter. *Pestic. Biochem. Physiol.* **1997**, *58*, 13–22. [CrossRef]

12.　Orcaray, L.; Zulet, A.; Zabalza, A.; Royuela, M. Impairment of carbon metabolism induced by the herbicide glyphosate. *J. Plant Physiol.* **2012**, *169*, 27–33. [CrossRef] [PubMed]

13.　Servaites, J.C.; Tucci, M.A.; Geiger, D.R. Glyphosate effects on carbon assimilation, ribulose bisphosphate carboxylase activity, and metabolite levels in sugar beet leaves. *Plant Physiol.* **1987**, *85*, 370–374. [CrossRef] [PubMed]

14.　Shaner, D.L. Role of translocation as a mechanism of resistance to glyphosate. *Weed Sci.* **2009**, *57*, 118–123. [CrossRef]

15.　Ahsan, N.; Lee, D.G.; Lee, K.W.; Alam, I.; Lee, S.H.; Bahk, J.D.; Lee, B.H. Glyphosate-Induced oxidative stress in rice leaves revealed by proteomic approach. *Plant Physiol. Biochem.* **2008**, *46*, 1062–1070. [CrossRef]

16.　Das, M.; Reichman, J.R.; Haberer, G.; Welzl, G.; Aceituno, F.F.; Mader, M.T.; Watrud, L.S.; Pfleeger, T.G.; Gutierrez, R.A.; Schaffner, A.R.; et al. A composite transcriptional signature differentiates responses towards closely related herbicides in Arabidopsis thaliana and Brassica napus. *Plant Mol. Biol.* **2010**, *72*, 545–556. [CrossRef]

17.　Unver, T.; Bakar, M.; Shearman, R.C.; Budak, H. Genome-Wide profiling and analysis of Festuca arundinacea miRNAs and transcriptomes in response to foliar glyphosate application. *Mol. Genet. Genom.* **2010**, *283*, 397–413. [CrossRef]

18.　Yu, W.; Zhang, R.; Li, R.; Guo, S. Isolation and characterization of glyphosate-Regulated genes in soybean seedlings. *Plant Sci.* **2007**, *172*, 497–504. [CrossRef]

19.　Trenkamp, S.; Eckes, P.; Busch, M.; Fernie, A.R. Temporally resolved GC-MS-based metabolic profiling of herbicide treated plants treated reveals that changes in polar primary metabolites alone can distinguish herbicides of differing mode of action. *Metabolomics* **2009**, *5*, 277–291. [CrossRef]

20.　Meyers, B.C.; Tingey, S.V.; Morgante, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genom. Res.* **2001**, *11*, 1660–1676. [CrossRef]

21.　Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [CrossRef] [PubMed]

22.　Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.; et al. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524–527. [CrossRef] [PubMed]

23.　McClintock, B. The significance of responses of the genome to challenge. *Science* **1984**, *226*, 792–801. [CrossRef]

24.　Huang, X.; Han, B. Natural variations and genome-Wide association studies in crop plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 531–551. [CrossRef] [PubMed]

25.　Zhou, Y.; Minio, A.; Massonnet, M.; Solares, E.; Lv, Y.; Beridze, T.; Cantu, D.; Gaut, B.S. The population genetics of structural variants in grapevine domestication. *Nat. Plants* **2019**, *5*, 965–979. [CrossRef] [PubMed]

26.　Zhang, X.; Chen, X.; Liang, P.; Tang, H. Cataloguing plant genome structural variations. *Curr. Issues Mol. Biol.* **2018**. [CrossRef] [PubMed]

27.　Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and exploiting pan-Genomics for crop improvement. *Mol. Plant.* **2019**, *12*, 156–169. [CrossRef]

28. Andorf, C.M.; Cannon, E.K.; Portwood, J.L.; Gardiner, J.M.; Harper, L.C.; Schaeffer, M.L.; Braun, B.L.; Campbell, D.A.; Vinnakota, A.G.; Sribalusu, V.V.; et al. MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Res.* **2015**, *44*, D1195–D1201. [CrossRef]

29. Rhoads, A.; Au, K.F. PacBio sequencing and its applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [CrossRef]

30. Forlani, G.; Racchi, M.L. Glyphosate tolerance in maize (Zea mays L.). 1. Differential response among inbred lines. *Euphytica* **1995**, *82*, 157–164.

31. Alexandrov, N.N.; Brover, V.V.; Freidin, S.; Troukhan, M.E.; Tatarinova, T.V.; Zhang, H.; Swaller, T.; Lu, Y.; Bouck, J.; Flavell, R.B.; et al. Insights into corn genes derived from large-Scale cDNA sequencing. *Plant Mol. Biol.* **2009**, *69*, 179–194. [CrossRef]

32. Soderlund, C.; Descour, A.; Kudrna, D.; Bomhoff, M.; Boyd, L.; Currie, J.; Angelova, A.; Collura, K.; Wissotski, M.; Ashley, E.; et al. Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. *PLoS Genet.* **2009**, *5*, e1000740. [CrossRef] [PubMed]

33. Vega-Arreguin, J.C.; Ibarra-Laclette, E.; Jimenez-Moraila, B.; Martinez, O.; Vielle-Calzada, J.P.; Herrera-Estrella, L.; Herrera-Estrella, A. Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genom.* **2009**, *10*, 299. [CrossRef] [PubMed]

34. Boyko, A.; Kovalchuk, I. Epigenetic control of plant stress response. *Environ. Mol. Mutagen.* **2008**, *49*, 61–72. [CrossRef] [PubMed]

35. Chen, J.; Huang, H.; Zhang, C.; Wei, S.; Huang, Z.; Chen, J.; Wang, X. Mutations and amplification of EPSPS gene confer resistance to glyphosate in goosegrass (Eleusine indica). *Planta* **2015**, *242*, 859–868. [CrossRef]

36. Funke, T.; Han, H.; Healy-Fried, M.L.; Fischer, M.; Schönbrunn, E. Molecular basis for the herbicide resistance of roundup ready crops. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13010–13015. [CrossRef] [PubMed]

37. Schönbrunn, E.; Eschenburg, S.; Shuttleworth, W.A.; Schloss, J.V.; Amrhein, N.; Evans, J.N.; Kabsch, W. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1376–1380. [CrossRef]

38. Rubin, J.L.; Gaines, G.C.; Jensen, R.A. Glyphosate inhibition of 5-enolpyruvylshikimate 3-phosphate synthase from suspension-Cultured cells of Nicotiana silvestris. *Plant Physiol.* **1984**, *75*, 839–845. [CrossRef]

39. Dong, Y.; Ng, E.; Lu, J.; Fenwick, T.; Tao, Y.; Bertain, S.; Sandoval, M.; Bermudez, E.; Hou, Z.; Patten, P.; et al. Desensitizing plant EPSP synthase to glyphosate: Optimized global sequence context accommodates a glycine-To-Alanine change in the active site. *J. Biol. Chem.* **2019**, *294*, 716–725. [CrossRef]

40. Zhou, M.; Xu, H.; Wei, X.; Ye, Z.; Wei, L.; Gong, W.; Wang, Y.; Zhu, Z. Identification of a glyphosate-Resistant mutant of rice 5-enolpyruvylshikimate 3-phosphate synthase using a directed evolution strategy. *Plant Physiol.* **2006**, *140*, 184–195. [CrossRef]

41. Kahrizi, D.; Salmanian, A.H.; Afshari, A.; Moieni, A.; Mousavi, A. Simultaneous substitution of Gly96 to Ala and Ala183 to Thr in 5-enolpyruvylshikimate-3-phosphate synthase gene of E. coli (k12) and transformation of rapeseed (Brassica napus L.) in order to make tolerance to glyphosate. *Plant Cell Rep.* **2007**, *26*, 95–104. [CrossRef] [PubMed]

42. Huang, Z.; Liu, Y.; Zhang, C.; Jiang, C.; Huang, H.; Wei, S. Molecular basis of natural tolerance to glyphosate in Convolvulus arvensis. *Sci. Rep.* **2019**, *9*, 8133. [CrossRef] [PubMed]

43. Chen, J.; Jiang, C.; Huang, H.; Wei, S.; Huang, Z.; Wang, H.; Zhao, D.; Zhang, C. Characterization of Eleusine indica with gene mutation or amplification in EPSPS to glyphosate. *Pestic Biochem. Physiol.* **2017**, *143*, 201–206. [CrossRef] [PubMed]

44. Liu, F.; Cao, Y.P. Improvement of glyphosate resistance through concurrent mutations in three amino acids of the Pantoea sp. 5-enolpyruvylshikimate-3-phosphate synthase. *J. Microbiol. Biotechnol.* **2018**, *28*, 1384–1390. [CrossRef] [PubMed]

45. Tzin, V.; Galili, G.; Aharoni, A. *Shikimate Pathway and Aromatic Amino Acid Biosynthesis*; eLS. John Wiley & Sons, Ltd.: Chichester, UK, 2012. [CrossRef]

46. Gomes, M.P.; Le Manac'h, S.G.; Moingt, M.; Smedbol, E.; Paquet, S.; Labrecque, M.; Lucotte, M.; Juneau, P. Impact of phosphate on glyphosate uptake and toxicity in willow. *J. Hazard. Mater.* **2016**, *304*, 269–279. [CrossRef] [PubMed]

47. Wang, D.; Lv, S.; Jiang, P.; Li, Y. Roles, regulation, and agricultural application of plant phosphate transporters. *Front. Plant Sci.* **2017**, *8*, 817. [CrossRef]

48. Nussaume, L.; Kanno, S.; Javot, H.; Marin, E.; Pochon, N.; Ayadi, A.; Nakanishi, T.M.; Thibaud, M. Phosphate import in plants: Focus on the PHT1 transporters. *Front. Plant Sci.* **2011**, *2*, 83. [CrossRef]

49. Shi, J.; Yi, K.; Liu, Y.; Xie, L.; Zhou, Z.; Chen, Y.; Hu, Z.; Zheng, T.; Liu, R.; Chen, Y.; et al. Phosphoenolpyruvate carboxylase in Arabidopsis leaves plays a crucial role in carbon and nitrogen metabolism. *Plant Physiology* **2015**, *167*, 671–681. [CrossRef]

50. O'Leary, B.; Rao, S.K.; Kim, J.; Plaxton, W.C. Bacterial-Type phosphoenolpyruvate carboxylase (PEPC) functions as a catalytic and regulatory subunit of the novel class-2 PEPC complex of vascular plants. *J. Biol. Chem.* **2009**, *284*, 24797–24805. [CrossRef]

51. Streatfield, S.J.; Weber, A.; Kinsman, E.A.; Häusler, R.E.; Li, J.; Post-Beittenmiller, D.; Kaiser, W.M.; Pyke, K.A.; Flügge, U.I.; Chory, J. The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-Dependent nuclear gene expression. *Plant Cell* **1999**, *11*, 1609–1621. [CrossRef]

52. Masumoto, C.; Miyazawa, S.; Ohkawa, H.; Fukuda, T.; Taniguchi, Y.; Murayama, S.; Kusano, M.; Saito, K.; Fukayama, H.; Miyao, M. Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5226–5231. [CrossRef] [PubMed]

53. De María, N.; Becerril, J.M.; García-Plazaola, J.I.; Hernandez, A.; De Felipe, M.R.; Fernandez-Pascual, M. New insights on glyphosate mode of action in nodular metabolism: Role of shikimate accumulation. *J. Agric Food Chem.* **2006**, *54*, 2621–2628. [CrossRef] [PubMed]

54. Lu, M. Structures of multidrug and toxic compound extrusion transporters and their mechanistic implications. *Channels (Austin)* **2016**, *10*, 88–100. [CrossRef]

55. Takanashi, K.; Shitan, N.; Yazaki, K. The multidrug and toxic compound extrusion (MATE) family in plants. *Plant Biotechnol.* **2014**, *31*, 417–430. [CrossRef]

56. Remy, E.; Duque, P. Beyond cellular detoxification: A plethora of physiological roles for MDR transporter homologs in plants. *Front Physiol.* **2014**, *5*, 201. [CrossRef] [PubMed]

57. Lu, P.; Magwanga, R.O.; Guo, X.; Kirungu, J.N.; Lu, H.; Cai, X.; Zhou, Z.; Wei, Y.; Wang, X.; Zhang, Z.; et al. Genome-Wide Analysis of Multidrug and Toxic Compound Extrusion (MATE) Family in Gossypium raimondii and Gossypium arboreum and Its Expression Analysis Under Salt, Cadmium, and Drought Stress. *G3 (Bethesda)* **2018**, *8*, 2483–2500. [CrossRef] [PubMed]

58. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.

59. Mahmoud, M.; Zywicki, M.; Twardowski, T.; Karlowski, W.M. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics* **2019**, *111*, 43–49. [CrossRef]

60. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-Generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]

61. Sedlazeck, F.J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M.C. Accurate detection of complex structural variations using single-Molecule sequencing. *Nat. Methods* **2018**, *15*, 461–468. [CrossRef]

62. English, A.C.; Salerno, W.J.; Reid, J.G. PBHoney: Identifying genomic variants via long-Read discordance and interrupted mapping. *BMC Bioinform.* **2014**, *15*, 180. [CrossRef] [PubMed]

63. Supek, F.; Bosnjak, M.; Skunca, N.; Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [CrossRef] [PubMed]