agronomy

MDPI

Article

Optimized In Vitro Restriction Digestion Protocol for Preparing Maize and Barley ddRAD-Seq Libraries

Marta Puchta-Jasińska, Paulina Bolc, Urszula Piechota and Maja Boczkowska *

Plant Breeding and Acclimatization Institute—National Research Institute, Radzików, 05-870 Błonie, Poland; m.puchta@ihar.edu.pl (M.P.-J.); p.bolc@ihar.edu.pl (P.B.); u.piechota@ihar.edu.pl (U.P.)
* Correspondence: m.boczkowska@ihar.edu.pl

Abstract: In recent years, high-throughput sequencing methods have become increasingly popular in molecular biology laboratories, mainly due to the relatively low cost of small, benchtop platforms, the simplicity of library preparation, and the low price per unit of information. Sequencing huge and complex genomes, such as cereal genomes, remains challenging and may not always be necessary. Therefore, several techniques have been developed to sequence a reduced representation of the genome. The most flexible and widely used of these is ddRAD-Seq, which uses a pair of restriction enzymes to generate a pool of DNA fragments. The aim of this study was to validate in vitro the efficacy of different combinations of restriction enzymes for ddRAD-Seq library construction in barley and maize. Eleven pairs of restriction enzymes were selected and tested to determine the concentrations of fragments with the expected length range and to select suitable pairs for sampling the genomes of these two cereals using ddRAD-Seq. For the selected pairs, i.e., PstI—MspI and HindIII—FspBI for barley and maize, respectively, libraries were prepared for NGS sequencing on Illumina MiSeq. Sequencing confirmed the suitability of the selected enzymes to perform ddRAD-Seq in different genotypes. The results presented can be used for extensive research on these important cereal species.

Keywords: ddRAD-Seq; restriction enzyme; maize; barley; next-generation sequencing

Citation: Puchta-Jasińska, M.; Bolc, P.; Piechota, U.; Boczkowska, M. Optimized In Vitro Restriction Digestion Protocol for Preparing Maize and Barley ddRAD-Seq Libraries. *Agronomy* **2023**, *13*, 2956. https://doi.org/10.3390/agronomy13122956

Academic Editors: Nikolaos Nikoloudakis and Angelos Kyratzis

Received: 8 November 2023
Revised: 27 November 2023
Accepted: 28 November 2023
Published: 30 November 2023

1. Introduction

In the nearly 70 years since James Watson and Francis Crick discovered the structure of DNA [1], several different technologies have been developed to study the sequence of nucleotides found in DNA molecules. The first attempts at DNA sequencing date back to the 1950s, while the late 20th century witnessed intensive advances in sequencing methods, leading to the establishment of three generations of sequencing methods. However, whole genome sequencing of every specimen in the population of interest is still time-consuming, costly, and often unnecessary for many applications. The focus has, therefore, been on developing methods for reduced representation genome sequencing (RRS) of genomes. This approach has been extremely successful and has revolutionized research in evolutionary and conservation genetics and molecular ecology over the past decade [2]. These methods allow reproducible pools of sequenced DNA fragments to be obtained from consecutive samples analyzed and provide information on single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs). Depending on the approach, reproducibility has been achieved by targeted sequencing of PCR products, library enrichment by probe hybridization or restriction enzyme (RE) fragmentation [3–5].

Technological advances over the past 15 years have made sequencing a routine and widely available analysis in molecular biology, and a milestone has been the invention of parallel, high-throughput next-generation sequencing (NGS) technology. The beginning of the 21st century exhibited a swift advancement in sequencing methods. Second-generation sequencing technologies, such as Illumina, have facilitated the assembly of over 200 plant

genomes. Genomic development has enabled comprehensive analysis of plant genomes and examination of the genetic foundation of agronomic characteristics [6]. Genome sequencing provides an opportunity to gain insight into the process of domestication, whereby human selection has resulted in the cultivation of plants that are more suitable for agriculture than for survival in natural environments. The comparison of genome sequences among plant species forms the basis for determining evolutionary relationships. The genome of *Oryza sativa*, commonly known as rice, was the first crop to undergo sequencing. This decision was made due to its importance as a major crop with a genome that is relatively small in size. The sequencing of rice subsequently served as a model for the sequencing of other cereal and grass genomes. *Brachypodium distachyon*, a model grass genome, was similarly sequenced and is relevant to the sequencing of the wheat genome [7].

Restriction enzyme-based methods use one or two restriction enzymes to cleave genomic DNA in the initial step of library preparation [8]. One of these methods is double-digest restriction site-associated DNA sequencing (ddRAD-Seq), which uses two different restriction enzymes (RE) to cut the DNA [9]. The first RE is a rare cutter, i.e., an RE with a rare restriction recognition site of 6–8 bp in length, which is used to generate long fragments. The second is a common cutter, i.e., an RE with a much more common four bp restriction site, whose role is to reduce the length of the resulting fragments. This eliminates random shearing and end repair of genomic DNA and minimizes the number of fragments bound by the same restriction sites at both ends. In addition, the method uses precise size selection of genomic fragments, providing greater control over the proportion of fragments represented in the resulting library [9]. As a result, ddRAD-Seq allows the generation of a library consisting of only a subset of hundreds to hundreds of thousands of DNA fragments generated after the double restriction cut. In addition, precise and reproducible size selection increases the ability to sequence the same regions from multiple samples and reduces the risk of sequencing adjacent regions. This is reflected in the minimization of the required sequencing depth for optimal coverage. The percentage of missing loci is also reduced compared to other methods, such as restriction site-associated DNA sequencing (RADseq) [9–11]. The ddRAD-Seq technology can be successfully applied to the analysis of both species with known and unknown genome sequences [11,12]. The reference genome sequence is an advantage right from the start, i.e., it offers the possibility to perform in silico restriction analysis to select REs on this basis [13]. For species with an unknown genome sequence, it is important to optimize the protocol, especially regarding the selection of REs [14]. Therefore, ddRAD-Seq is highly customizable in terms of the total number of loci, based on the RE selection and the range of fragment sizes selected. It allows for the production of a large number of fragments, thus enabling the identification of variants at tens of thousands of orthologous loci [15,16]. Consequently, this method is used for genetic mapping, genetic diversity detection, population genomics, and molecular evolution studies [17–21]. Another application of ddRAD-Seq can be the detection of methylation levels using methylation-sensitive restriction enzymes, as described by Marconi et al. [22] and Dimond and Roberts [23].

The aim of this study was to experimentally confirm the efficacy of restriction enzymes used in constructing ddRAD-Seq libraries to evaluate genetic diversity in maize (*Zea mays* L.) and barley (*Hordeum vulgare* L.). In parallel, the results obtained by in silico restriction analysis were verified. Barley and maize are among the most important crops in terms of acreage and protein sources for both humans and livestock. As these species are of interest to scientists and breeders, it is crucial to efficiently break down their huge genomes, i.e., about 5.3 Gbp for barley and about 2.4 Gbp for maize, into fragments that are fractionally accurate for case- or gene-specific ddRAD-Seq library construction. Presenting the efficiency of different RE combinations is crucial, especially for low-budget laboratories that cannot afford an additional assay before implementing the ddRAD-Seq methodology. Based on this, the optimal set of enzymes can be selected depending on the expected coverage and depth of sequencing.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

The material for the study were barley (*Hordeum vulgare* sp. *vulgare* L.) accessions from the National Centre for Plant Genetic Resources, that is, the Polish Gene Bank and maize (*Zea mays* L.) inbred lines obtained from DSc Elżbieta Kochańska-Czembor (Plant Breeding and Acclimatization Institute-NRI, Poland) (Table 1). The maize genotypes utilized in constructing the ddRad-Seq library were outlined in Czembor et al., 2019 [24].

Table 1. List of materials used in the construction of the ddRAD-Seq libraries.

***Zea mays* Inbred Lines**	***Hordeum vulgare* Accessions**
104N	40004
11IDT	40068
21IDT	40073
23UN	40215
24S/I	40552
30IDT *	40554
32SSS	41265
39L	41268 *
43F	41269
44S/I	41274
45SSS	41279
49L	41282
4F	41283
54D	41284
58F	41365
67D	41430
96F	41434
	41435
	41525
	41526
	41527
	41528
	41529
	41530

* The line/accession used for restriction enzyme effectiveness analysis.

Detailed information on all barley accessions can be found in the EGISET database (https://wyszukiwarka.ihar.edu.pl/pl, accessed date: 27 October 2023). The materials used in the study's first phase included a maize line, 30IDT, with moderate resistance to ear rot caused by *Fusarium verticillioides* and a barley accession numbered 41268. This landrace was collected during an expedition from south-eastern Poland and introduced into the long-term storage in 1978. It was found to be resistant to steam rust, net blotch, and scald, and identified as *Hordeum vulgare* var. *nutans*.

DNA was isolated from the first leaf using the CTAB protocol followed by treatment with RNase A (10 mg/mL) (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C for 30 min [25]. The isolated genetic material was dissolved in DNAse-free water, quantified using the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). and qualified by agarose gel electrophoresis analysis. Samples were diluted to 20 ng/µL.

2.2. Double Digestion with Restriction Enzymes of Genomic DNA

To optimize the digestion with REs, three rare and four common cutters were selected (Table 2). Several variants were tested. A 20 µL reaction containing 200 ng of genomic DNA, 2× CutSmart Buffer, 8 U Enzyme I, 8 U Enzyme II, and 7.2 µL water was prepared. The reaction was run at 37 °C for 2 h and cooled down at 4 °C. The following combinations of restriction enzymes were tested: EcoRI—FspBI, EcoRI—HaeIII, EcoRI—MspI, HindIII—FspBI, HindIII—HaeIII, HindIII—MseI, HindIII—MspI, PstI—FspBI, PstI—HaeIII, PstI—MseI,

and PstI—MspI. The enzymes were selected based on available literature data [11,13,19] and in silico analysis results.

Table 2. Restriction enzymes used in the study.

Enzyme I (Rare Cutter)		Enzyme II (Common Cutter)	
Enzyme	**Cut Site**	**Enzyme**	**Cut Site1**
EcoRI *	5′...G | AATTC...3′ 3′...CTTAA | G...5′	HaeIII ***	5′...GG | CC...3′ 3′...CC | GG...5′
HindIII	5′...A | AGCTT...3′ 3′...TTCGA | A...5′	FspBI	5′...C | TAG...3′ 3′...GAT | C...5′
PstI	5′...CTGCA | G...3′ 3′...G | ACGTC...5′	MseI	5′...T | TAA...3′ 3′...AAT | T...5′
		MspI **	5′...C | CGG...3′ 3′...GGC | C...5′

All enzymes were manufactured by New England Biolabs, Ipswich, MA, USA; * Blocked by some combinations of overlapping CpG methylation; ** Blocked by hemi and fully methylated mCCGG and mCmCGG sequences; *** Blocked by GGm5CC methylation.

2.3. Evaluation of Restriction Enzymes Efficiency

The products of enzymatic digestion with the above combinations of restriction enzymes were analyzed using an Agilent 2100 Bioanalyzer with a High Sensitivity DNA kit. Within the 300–600 bp range, the concentration, molarity, and percentage of the resulting fragment pool were calculated for each sample.

2.4. Preparation of Adaptors

Oligonucleotide-forming adaptors were annealed prior to library construction to form the duplex of the P1 adaptor. The oligonucleotides P1.1/P1.2 and P2.1/P2.2 were designed in pairs for each barcode, respectively, to form their duplex. The reactions were carried out according to Table 3. Mixtures of P1 and P2 adaptors were incubated in a thermocycler at 97.5 °C for 2.5 min and then cooled to 21 °C at a rate of no more than 3 °C per minute, held at 21 °C for 10 min, and stored at 4 °C. A diagram of adaptor ligation is shown in Figure 1.

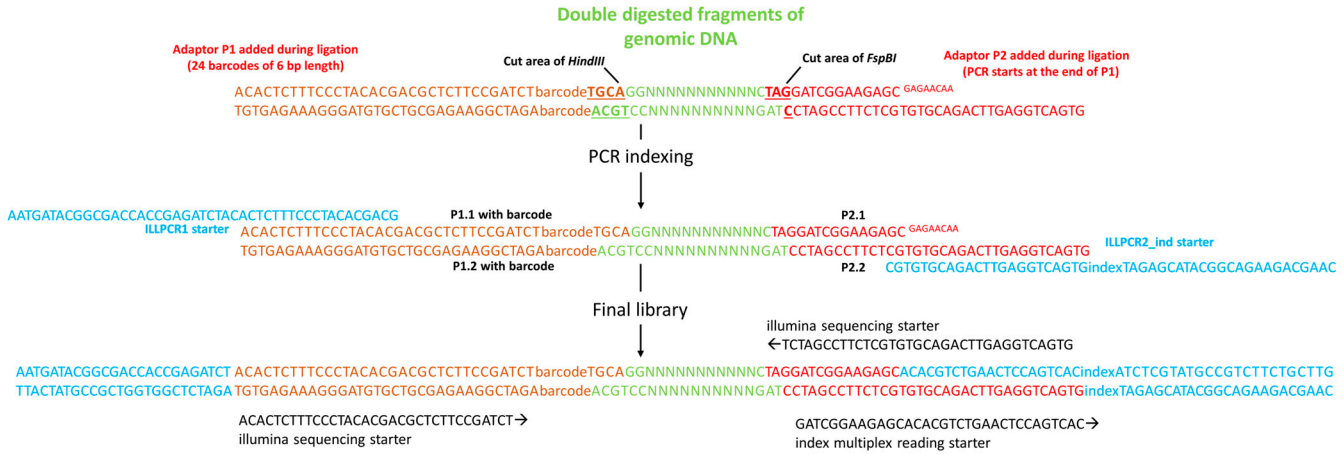

Figure 1. Scheme of double digestion of genomic DNA and ligation of adaptors.

Table 3. Composition of the adaptors duplex.

P1 Adaptor		P2 Adaptor	
Ingredient	**Volume (µL)**	**Ingredient**	**Volume (µL)**
P1.1 (100 µM)	10	P2.1 (100 µM)	10
P1.2 (100 µM)	10	P2.2 (100 µM)	10
water	80	water	80
final volume	100	final volume	100

2.5. Library Construction and Sequencing

In the next step, fragments resulting from PstI—MspI and HindIII—FspBI digestion for barley and maize genomic DNA, respectively, were ligated with adaptors compatible with the restriction site. A ligation reaction mix was prepared according to Table 4. One µL of P1 adaptor was added directly to 20 µL of digested DNA, and then the mixture was added to the ligation mix. Different adapters (P1) were used for each accession (Table S1), and then 9 µL of mix was added. The reaction was incubated at 16 °C for 2 h and then held at 4 °C. A diagram of ddRAD-Seq library preparation can be found in Figure S1.

Table 4. Composition of the adaptors' ligation reaction.

Ingredient	**Volume (µL)**
Cut Smart Buffer (10×)	1
T4 DNA Ligase (400 U/µL)	0.5
10 mM ATP	3
10 mM P2 adaptor	1
water	3.5
final volume	9

DNA fragments were then purified from the post-reaction components using 1× Ampure XP beads (Beckman Coulter, Pasadena, CA, USA), and the purified product was suspended in 20 µL of water. The purified library was amplified with 5 µM Illumina PCR1, PCR2 primers, and 25 µL NebNext Ultra II Q5 master mix. The reaction was denatured at 98 °C for 30 s. This was followed by 16 cycles of denaturation at 98 °C for 10 s and combined with annealing at 65 °C for 75 s. The next step was a final extension at 65 °C for 5 min and a final incubation at 4 °C. For this purpose, a PCR mix was prepared.

The library was again purified with 1× Ampure XP beads. The purified product was then suspended in 20 µL of water. Size selection was then performed using a 1.5% DF cassette with Marker K Pippin Prep. The prepared libraries were analyzed qualitatively and quantitatively using the Qubit fluorimeter (ThermoFisher Scientific) with the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) and the Bioanalyzer 2100 automated electrophoresis system with the High Sensitivity DNA Kit (Agilent, Santa Clara, CA, USA). For normalization, the library was diluted to 4nM and then denatured with 0.2N NaOH. After normalization, the libraries were sequenced on MiSeq (Illumina, San Diego, CA, USA) at 2 × 150 bp paired-end using the MiSeq Reagent Kit v3 (600 cycles). Raw reads were qualitatively evaluated using FastQC software v 0.11.9 [26].

2.6. Bioinformatic Analysis

In silico analysis was performed using the web-based tool ddgRADer [27]. Reference sequences of the barley (MorexV3_pseudomolecules_assembly) and maize (Zm-B73-REFERENCE-NAM-5.0) genomes, downloaded from the EnsemblPlants database release 55 (http://plants.ensembl.org/index.html, accessed date: 25 May 2022), were used. The same sequencing parameters were used for both species, i.e., read length to be sequenced −150 bp, pair end, sequencing yield −1 × 10^6, and desired depth −5× [28]. The number

of SNPs was estimated for fragments in the length range of 300–600 bp. SNP calling was performed as described by Czembor et al. [24].

3. Results

3.1. Efficacy of Restriction Enzymes

For barley and maize, RE efficacy was evaluated separately. A fragment fraction of 300–700 bp is often used to construct ddRAD-Seq libraries, but it can be modified depending on the case, enzyme, species, etc. [29,30]. In general, the size of fragments in the library should not be less than 200 bp to avoid overlapping of paired-end sequences, which can hinder bioinformatic analysis [31]. However, to avoid degradation of base quality in Illumina paired-end sequencing, fragments should also not exceed 800 bp [32]; a selection of fragments in the range of 300 to 600 bp was used here. It is worth noting here that the parameters obtained from automated electrophoresis show the aggregate presence of all types of fragments that can be generated during digestion with a pair of REs, i.e., those with the same restriction sites specific for one of the REs at both ends, two different restriction sites, and fragments that were mechanically broken off during DNA handling. In silico analysis, on the other hand, only shows the presence of fragments with different restriction sites at both ends. Such fragments are indeed sequenced later.

3.1.1. Maize

Restriction of maize DNA using seven RE in 11 combinations produced a diverse range of fragment concentrations and molarities. The distribution of fragments obtained by 11 pairs of RE is shown in Figure 2. Based on the results of automated electrophoresis, the concentration, molarity, and percentage of the desired fraction of fragments were determined (Table 5). The highest values of molarity, concentration, and percentage of fragments of appropriate length were observed for a HindIII—FspBI, i.e., above 124,398 pmol/L 48,290 pg/µL and 46%, respectively.

Table 5. Results of in vitro and in silico enzymatic digestion of the maize genome for fragments in the range of 300–600 bp.

Enzymes Combination	In Vitro Concentration (pg/µL)	In Vitro Molarity (pmol/L)	Percentage of Useful Fragments (%)	In Silico Number of Fragments
PstI—MspI	0	0	0	160,965
PstI—HaeIII	8998	17,158	25.80	144,142
EcoRI—MspI	1615	6770	6.83	166,259
HindIII—MspI	2269	8094	14.75	263,417
HindIII—FspBI	48,290	124,398	46.02	270,679
HindIII—MseI	21,941	92,976	20.38	214,230
EcoRI—HaeIII	22,388	95,413	20.78	175,882
PstI—MseI	457	1998	20.51	195,383
HindIII—HaeIII	10,662	45,698	17.32	222,274
PstI—FspBI	42,471	101,132	39.20	208,475
EcoRI—FspBI	20,920	91,359	27.50	207,162

Acceptable fragmentation results were also obtained for the PstI—FspBI. Low fragment concentrations in the range of 300–600 bp were observed after using PstI—MseI, EcoRI—MspI, and HindIII—MspI RE pairs. This means that no fragments of this length were obtained.

In silico analysis showed that the most fragments in the expected range of 300–600 bp should be obtained using HindIII—FspBI. This was confirmed by in vitro analysis. In contrast, for RE pairs such as PstI—MseI and HindIII—MspI, significant discrepancies were observed between the wet analysis performed in the laboratory and the computer

simulation. Fragmentation efficiency with PstI—MseI was unsatisfactory for the sample tested here, although in silico results indicated a better yield.

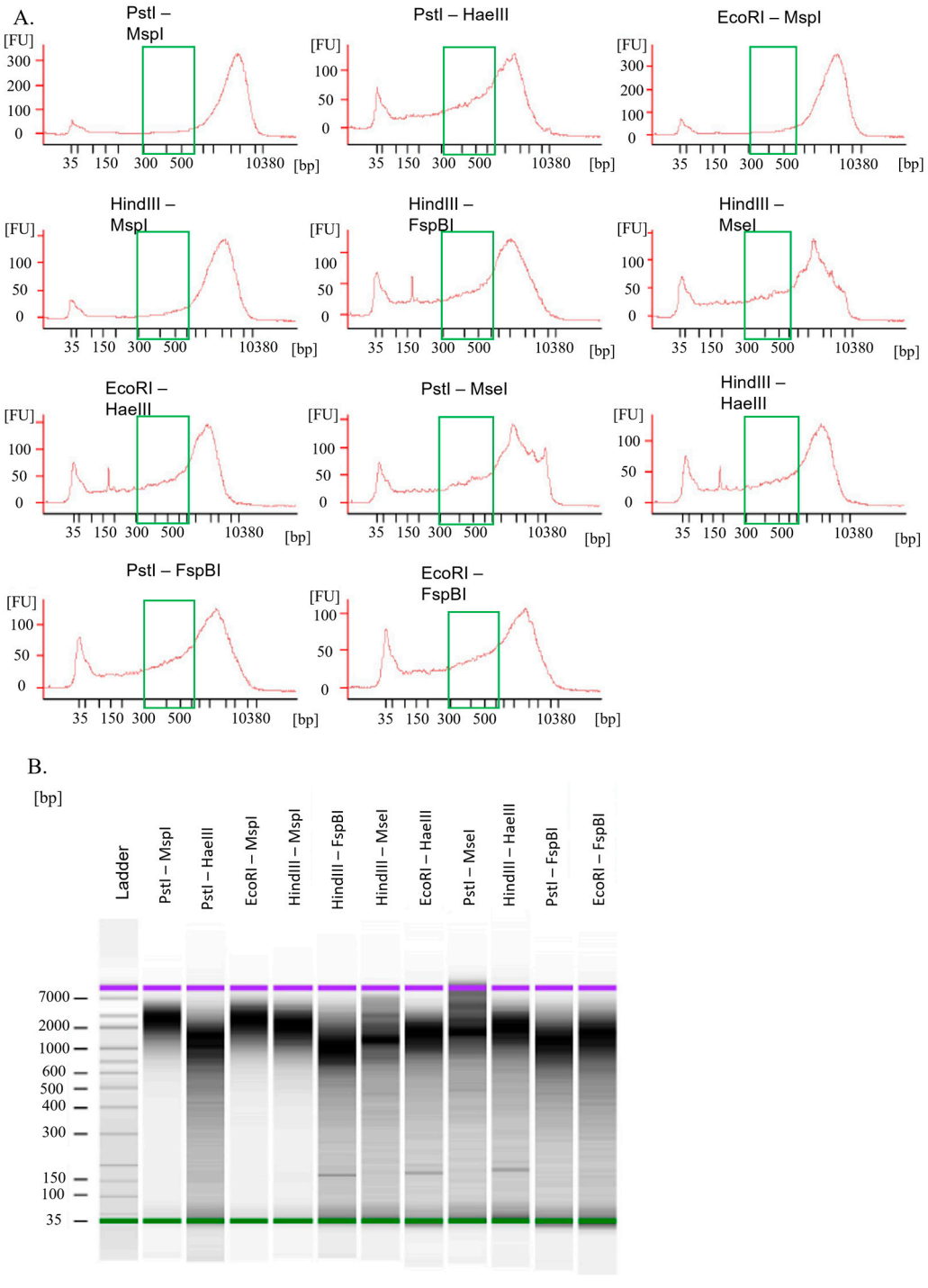

Figure 2. Results of enzymatic digestion of maize DNA. Automated electrophoresis was performed on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit. (**A**) Agilent Bioanalyzer chromatograms of DNA fragment size (FU-Arbitrary fluorescence units); (**B**) Agilent 2100 Bioanalyzer gel-like image of DNA after automated electrophoresis.

3.1.2. Barley

The effectiveness of the different RE combinations was also investigated for barley DNA. Different sizes and amounts of fragments were also observed in the chromatograms

resulting from automated on-chip gel electrophoresis (Figure 3). The highest concentration and molarity of fragments in the range of 300–600 bp occurred for EcoRI—FspBI (Table 6), while the lowest efficiency was shown by the EcoRI—MspI pair. The highest proportion of fragments with the required length was obtained in the sample treated with PstI—MspI. Meanwhile, the results of the in silico analysis indicated that the HindIII—MseI RE pair should be the most efficient in digesting the barley genome into 300–600 bp fragments. In silico treatment of the reference genome with PstI—MspI RE was identified as the least effective.

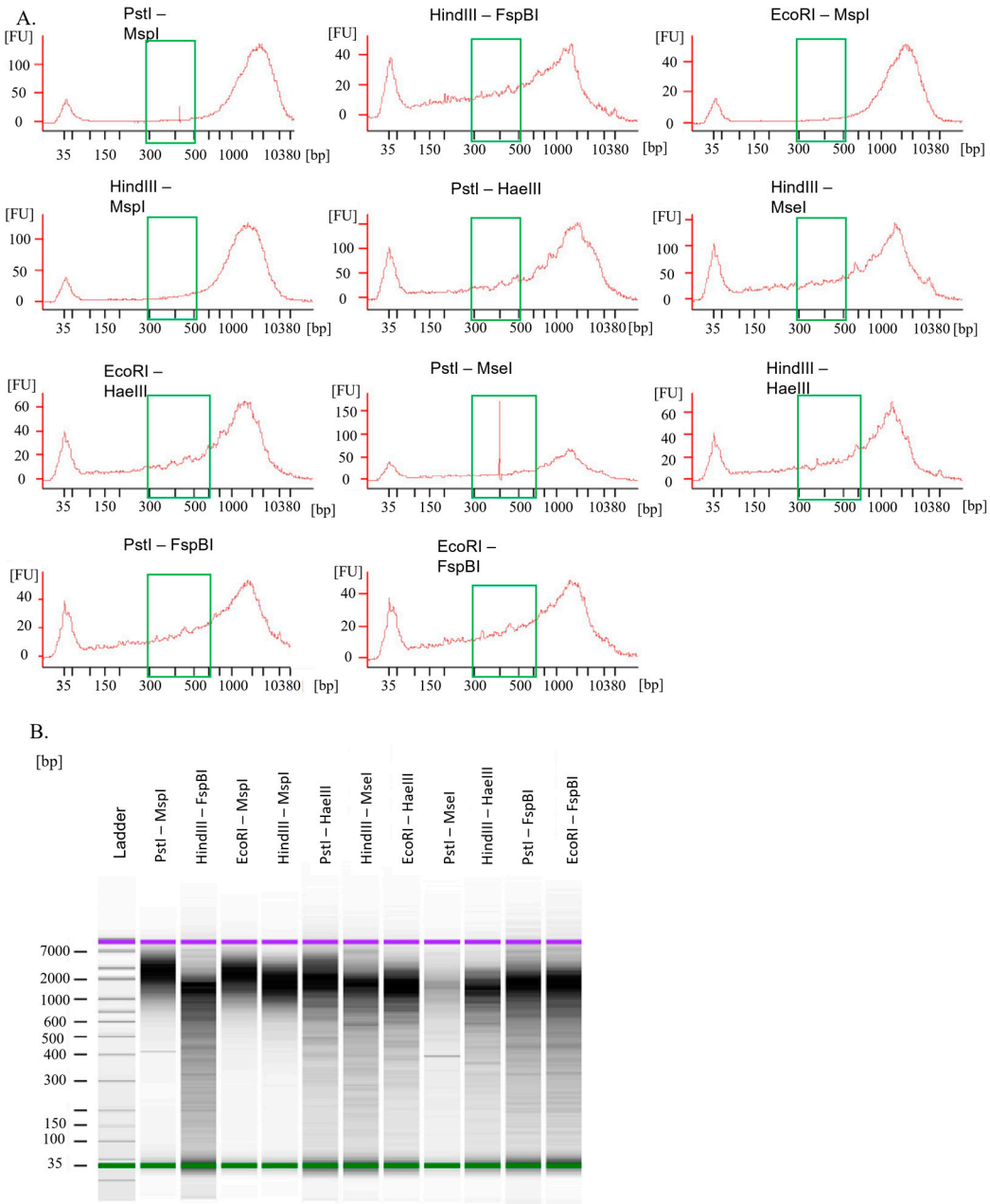

Figure 3. Results of enzymatic digestion of barley DNA. Automated on-chip electrophoresis was performed on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit. (**A**) Agilent Bioanalyzer chromatograms of DNA fragment size (FU-Arbitrary fluorescence units); (**B**) Agilent 2100 Bioanalyzer gel-like image of DNA after automated electrophoresis.

Table 6. Results of in vitro and in silico enzymatic digestion of the barley genome for fragments in the range of 300–600 bp.

Enzymes Combination	Concentration (pg/µL)	In Vitro Molarity (pmol/L)	Percentage of Useful Fragments (%)	In Silico Number of Fragments
PstI—MspI	6658	30,134	24.93	250,144
HindIII—FspBI	6644	25,115	23.50	645,928
EcoRI—MspI	804	2583	1.92	312,531
HindIII—MspI	2667	9313	11.14	566,198
PstI—HaeIII	4230	17,709	10.08	453,778
HindIII—MseI	8694	37,430	19.62	744,406
EcoRI—HaeIII	8107	33,563	17.85	553,525
PstI—MseI	7427	32,204	14.85	524,379
HindIII—HaeIII	1780	7030	12.36	712,117
PstI—FspBI	4869	21,357	22.01	288,032
EcoRI—FspBI	10,202	43,327	21.13	321,946

3.2. NGS Sequencing

Based on the results of the efficiency analysis of the RE pairs and the literature data, PstI—MspI and HindIII—FspBI for barley and maize, respectively, were selected for further testing. To verify that the selected restriction systems were effective regardless of the genotype tested, the number of DNA samples was increased to 17 for wheat and 24 for barley. From the resulting fragment mixtures after restriction digestion, a total of 41 ddRAD-Seq libraries were prepared and subjected to NGS sequencing. After sequencing, a total of 17,648,717 raw reads were obtained for maize and 23,876,096 for barley. The Phred score for all reads from the ddRAD-Seq libraries was above 30 for the entire length of the fragments (Figure 4). Bioinformatic analysis and SNP calling revealed that approximately 50,500 fragments with allele frequencies $\geq$0.05 and at least 5× [28] coverage were obtained for barley. For maize, 10,500 such fragments were obtained.

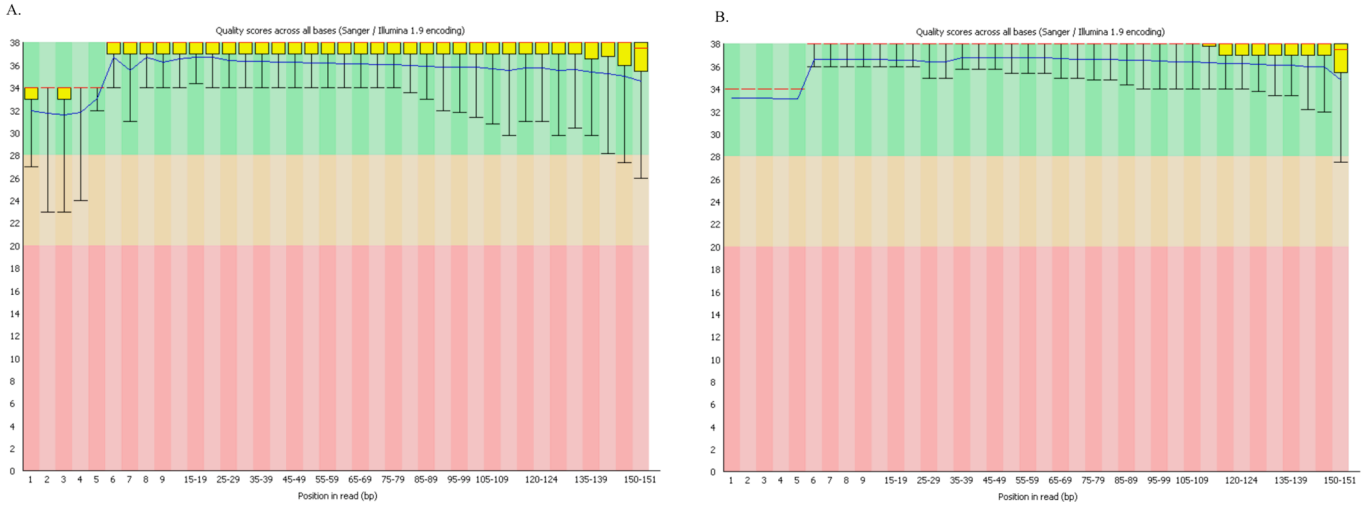

Figure 4. Quality of reads obtained from sequencing ddRAD-Seq maize (**A**) and barley libraries (**B**).

4. Discussion

The ddRAD-Seq method can be an essential and simple answer to the needs of scientists and breeders. Due to its high reproducibility and low cost per sample, the method can be an ideal analytical solution for teams with tighter budgets and struggling to analyze large plant genomes [9]. The use of a pair of RE during genome sampling, followed by adaptor ligation and magnetic bead purification, generates reproducible genomic targets [33]. The ddRAD-Seq method allows the sequencing of the same loci within the list

of sequencing samples of complex genomes with a high rate of repeatable fragments and accurate mapping [34]. This approach increases the possibility for easy and cost-effective development of molecular markers.

Although the original ddRAD-Seq protocol developed by Peterson et al. [9] used EcoRI—MspI, the method needs to be optimized for each species and/or application, especially in terms of RE selection, as genomes vary in size, complexity, and density of restriction sites. The appropriate pair of RE allows efficient sequencing by constructing libraries that contain precise DNA length fragments and cover the appropriate portion of the genome. This optimization should be prepared for the species, DNA fragment range, sequencing platform, and strategy [11]. In general, the goal is to generate an appropriate number of sequencable fragments, according to the downstream analysis of interest.

In silico analysis is often used to pre-identify the optimal combination of RE for genome fragmentation in the ddRAD-Seq protocol. Peterson et al. [9] performed an in silico restriction analysis for 17 species and 5 pairs of RE. Maize was one of the species analyzed. The researchers estimated that EcoRI—MspI digestion would yield 30,000 fragments. Yang et al. [33] indicated the potential presence of about 120,000 fragments for the same REs. The in silico analysis presented here predicted more than 165,000 fragments. Similar inconsistencies appeared in the abundance of fragments obtained by PstI—MspI restriction. According to Yang et al. [33], this combination should generate a yield of about 130,000 fragments, while here, it was estimated to yield about 160,000 fragments. Differences in the in silico estimation of the number of fragments are due to the analysis of different maize genome assemblies, variable sequencing parameters, different length ranges of the fragments analyzed, and the use of different bioinformatics tools. Therefore, the results of the in silico analysis should be treated with appropriate caution, and the possibility that the results of the in vitro analysis may not agree with the estimate should be considered. This is illustrated by the comparison of parameters obtained during the automated electrophoresis of maize DNA. Here, the results clearly indicated that the efficiency of EcoRI—MspI digestion was very low for the expected fragment length, and the efficiency of PstI—MspI was even below the detection threshold of the Agilent Bioanalyzer 2100 with the High Sensitivity DNA chip loaded. However, there is evidence of a high agreement between in silico estimated and laboratory-analyzed results [9,35]. Based on both types of results, further analysis of the maize lines was performed using HindIII—FspBI and satisfactory NGS results were obtained. Other RE combinations that have been successfully used in reduced representation protocols for maize genome sequencing can also be found in the literature, i.e., SphI—MluCI [36].

Furthermore, in barley, different teams have used different combinations of RE, i.e., PstI—MseI and PstI—MspI, to effectively reduce genome complexity [37,38]. Here, the concentration and molarity of these two combinations were similar in the desired length range. However, the percentage of 300–600 bp fragments was higher for PstI—MspI, while the in silico analysis indicated more than double the number of fragments for PstI—MseI. The PstI—FspBI combination, which, according to in silico estimation, should generate a comparable number of fragments to PstI—MspI in in vitro analysis, was characterized by a significantly lower digestion efficiency. Thus, discrepancies between laboratory analysis and computer simulations can be observed, as in the case of maize. High efficiencies were observed for EcoRI—FspBI and HindIII—MseI. However, considering the size of the barley genome (5.3 Gbp), the planned sequencing depth (1.25 million reads/sample) and the aim of the study (diversity analysis of the landrace collection), the most efficient RE combinations were discarded in order to obtain a higher fraction of fragments with at least a five-fold sequence coverage. Finally, the combination PstI—MspI was chosen because of the possibility of comparing the results with those obtained by Milner et al. for 22,626 accessions from the gene bank [37].

As can be seen, the selection of RE for library construction is not a trivial matter. Different teams have used different pairs of RE to analyze even the same species. The efficiency of the restriction digestion in generating the appropriate fragment length is

one of the most important conditions for success in constructing a library containing only a representation of the genome. However, the selection of RE for analysis should also be guided by the availability of data from the experimental work by other teams. Using the same RE greatly increases the chances of obtaining compatible results from different experiments, facilitating co-analysis and inference. The lower number of fragments generated by this RE pair also leads to a reduction in sequencing costs. A lower total number of reads per sample is required to achieve adequate coverage of the resulting fragments.

Another factor to consider is the methylation sensitivity of the RE and its consequences. The PstI—MspI RE pair used in the barley analysis consists of both a methylation-sensitive rare-cutting RE and a common-cutting RE. As the results show, this arrangement has certain advantages, such as a combination resulted in greater uniformity in the depth of reads across loci and provided a higher quality of acquired genetic information compared to pairs containing no methylation-sensitive RE as well as those containing only one methylation-sensitive RE [39]. In addition, the use of a methylation-sensitive RE pair increases the correlation between SNP and gene density. This may be a derivative of the methylation sensitivity itself, which increases the pool of under-methylated regions of the genome in the pool of sequenced fragments and their association with a higher gene density [39,40]. However, the methylation sensitivity of the RE also has drawbacks. In maize, all RE combinations containing MspI showed low efficiency in generating fragments in the length range analyzed. MspI is completely blocked by hemi- and fully methylated mCCGG and mCmCGG sequences. Thus, a highly methylated genome may cause a low cutting efficiency and significantly reduce the number of fragments useful for ddRAD-Seq analysis of maize. However, attention to the efficiency of generating the required fragment length remains critical. If an enzyme pair does not generate fragments of the appropriate length, as in the case of PstI—MseI in maize, it does not matter what the longer/shorter fragments contain and how often they occur, as they will be discarded and not sequenced anyway due to the fragment length selection step in the ddRAD-Seq protocol.

The inconsistency of in silico and experimental genome restriction analysis can arise from several factors. In silico tools use algorithms to predict restriction sites by recognizing restriction enzyme sequence motifs. However, the accuracy of these algorithms is limited, particularly when dealing with complex or repetitive DNA sequences [41]. The accuracy of in silico restriction analysis depends on the quality of the genome sequence used as input. If the genome sequence is incomplete or inaccurate, it may result in the misidentification or omission of true restriction sites [41]. The substantial genetic distance between the reference and the genotype(s) being examined could also contribute to the observed discrepancies [27]. This could lead to the existence of polymorphisms in restriction sites, as well as the emergence of new restriction sites [42,43]. Moreover, natural adaptations to specific environmental conditions may lead to structural changes. Such changes may also arise from natural and artificial introgression of genome fragments from another species [44,45]. Additionally, mobile genetic elements, both in quantity and type, significantly contribute to these modifications of the genome [46]. DNA modifications, such as methylation, can modify recognition sequences for restriction enzymes. These modifications can hinder cleavage at forecasted restriction sites, thereby resulting in inconsistencies between in silico and experimental outcomes [27,47]. The discrepancy between the in silico predictions and the experimental results may also be attributed to varying experimental conditions, such as temperature, pH, and reaction time, which could impact the activity of the restriction enzyme. Under conditions that are far from optimal for REs activity, they can cleave sequences that are similar but not identical to their canonical recognition sequences. This altered specificity has been termed “star activity”. Star sites are associated with a canonical recognition site, but usually differ by one or more bases. This results in non-specific fragments which presence was not considered in the in silico prediction [48]. Initial DNA quality is also a crucial determinant for sequencing reduced representative fractions of the genome. The number of fragments obtained, as well as the depth of sequencing, significantly decrease only after significant genomic DNA degradation. This effect

is observed alongside a decrease in the number of SNPs [49]. Another factor, that could cause discrepancies between laboratory analysis and computer simulations is the presence of non-target DNA, i.e., residuals of microbial DNA. Restriction enzymes may recognize and cleave such sequences and can lead to the generation of unexpected fragments [50]. The final outcome of these factors could be an inadequate initial estimation of the number of fragments to be included in the library, resulting in a different number of reads and sequencing depth than expected. In instances with low sequencing depth, heterozygous loci might be miscategorized as sequencing errors during the analysis of results. In the context of research on genetic variation or population structure, this may result in a falsely elevated level of homozygosity [51]. This could potentially suggest the presence of a bottleneck phenomenon that did not actually occur. Insufficient fragments of the appropriate length may result in too few reads, leading to errors in downstream applications such as gene mapping or, in the worst case, making the analysis of the sequences obtained completely impossible. In addition, selecting an inappropriate enzyme pair might lead to inadequate or partial digestion of DNA, significantly increasing the likelihood of poorly reproducible results [52].

The implementation of a ddRAD-Seq methodology, in which restriction enzymes are selected to produce a large number of fragments in the range of 300–600 bp, allows the generation of an effective ddRAD-Seq library. By sequencing such libraries, substantial amounts of data are obtained, which then allows for an in-depth analysis of the variability of SNP types. The outcomes derived from the optimization process, which helped identify the optimal ERs for maize and barley, were deployed during the study conducted by Czembor et al. in 2019 [24]. A study on the use of specific REs in barley is being prepared for publication.

5. Conclusions

For optimal sequencing results, the restriction REs used in ddRAD-Seq analysis should be matched to the genome under study. This is a key step in ddRAD-Seq library construction that affects the number of fragments generated for sequencing. The in silico analyses used to assess the number of genomic fragments resulting from enzymatic digestion may differ from the in vitro results of experimental studies. There is no universal pair of restriction REs that can be recommended for ddRAD-Seq library construction for all species to produce an efficient number of fragments. Therefore, to minimize the inconsistencies between in silico and experimental genome restriction analysis results, it is important to use high-quality genome sequences, carefully control experimental conditions, and validate in silico predictions with experimental data whenever possible.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13122956/s1, Figure S1: Diagram of ddRAD-Seq library preparation; Table S1: List of oligonucleotides used to construct barley and maize ddRAD-Seq libraries.

Author Contributions: Conceptualization, M.P.-J., P.B., U.P. and M.B.; methodology, M.P.-J., P.B. and U.P.; investigation, M.P.-J., P.B. and U.P.; writing—original draft preparation, M.P.-J., P.B. and U.P.; writing—review and editing, M.B. and M.P.-J.; visualization, M.P.-J.; supervision, M.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture and Rural Development of the Republic of Poland.

Data Availability Statement: The row data of DArTseq SNP used in this study are openly available on the platform Center for Open Science at https://osf.io/akrd8/ (accessed on 27 February 2023); (https://doi.org/10.17605/OSF.IO/AKRD8).

Acknowledgments: The authors would like to thank Elżbieta Kochańska-Czembor for providing the material for the study. They would also like to thank Jerzy Henryk Czembor for the opportunity to perform the analyses.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Watson, J.D.; Crick, F.H. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738. [CrossRef] [PubMed]
2. Choquet, M. Customize and get the most out of your reduced-representation sequencing experiment with the new simulation software RADinitio. *Mol. Ecol. Resour.* **2021**, *21*, 351–354. [CrossRef] [PubMed]
3. Bybee, S.M.; Bracken-Grissom, H.; Haynes, B.D.; Hermansen, R.A.; Byers, R.L.; Clement, M.J.; Udall, J.A.; Wilcox, E.R.; Crandall, K.A. Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biol. Evol.* **2011**, *3*, 1312–1323. [CrossRef] [PubMed]
4. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef] [PubMed]
5. Myllykangas, S.; Natsoulis, G.; Bell, J.M.; Ji, H.P. Targeted sequencing library preparation by genomic DNA circularization. *BMC Biotechnol.* **2011**, *11*, 122. [CrossRef] [PubMed]
6. Pourkheirandish, M.; Golicz, A.A.; Bhalla, P.L.; Singh, M.B. Global role of crop genomics in the face of climate change. *Front. Plant Sci.* **2020**, *11*, 922. [CrossRef]
7. Henry, R.J. Progress in plant genome sequencing. *Appl. Biosci.* **2022**, *1*, 113–128. [CrossRef]
8. Jiang, Z.; Wang, H.; Michal, J.J.; Zhou, X.; Liu, B.; Woods, L.C.S.; Fuchs, R.A. Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development. *Int. J. Biol. Sci.* **2016**, *12*, 100. [CrossRef]
9. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135. [CrossRef]
10. Timm, H.; Weigand, H.; Weiss, M.; Leese, F.; Rahmann, S. DDRAGE: A data set generator to evaluate ddRADseq analysis software. *Mol. Ecol. Resour.* **2018**, *18*, 681–690. [CrossRef]
11. Aguirre, N.C.; Filippi, C.V.; Zaina, G.; Rivas, J.G.; Acuña, C.V.; Villalba, P.V.; García, M.N.; González, S.; Rivarola, M.; Martínez, M.C. Optimizing ddRADseq in non-model species: A case study in *Eucalyptus dunnii* maiden. *Agronomy* **2019**, *9*, 484. [CrossRef]
12. Wright, B.R.; Farquharson, K.A.; McLennan, E.A.; Belov, K.; Hogg, C.J.; Grueber, C.E. A demonstration of conservation genomics for threatened species management. *Mol. Ecol. Resour.* **2020**, *20*, 1526–1541. [CrossRef] [PubMed]
13. Chafin, T.K.; Martin, B.T.; Mussmann, S.M.; Douglas, M.R.; Douglas, M.E. FRAGMATIC: In silico locus prediction and its utility in optimizing ddRADseq projects. *Conserv. Genet. Resour.* **2018**, *10*, 325–328. [CrossRef]
14. Cumer, T.; Pouchon, C.; Boyer, F.; Yannic, G.; Rioux, D.; Bonin, A.; Capblancq, T. Double-digest RAD-sequencing: Do pre- and post-sequencing protocol parameters impact biological results? *Mol. Genet. Genom.* **2021**, *296*, 457–471. [CrossRef] [PubMed]
15. Shortt, J.A.; Card, D.C.; Schield, D.R.; Liu, Y.; Zhong, B.; Castoe, T.A.; Carlton, E.J.; Pollock, D.D. Whole genome amplification and reduced-representation genome sequencing of *Schistosoma japonicum* miracidia. *PLoS Neglected Trop. Dis.* **2017**, *11*, e0005292. [CrossRef] [PubMed]
16. Adamo, M.; Blandino, M.; Capo, L.; Ravetto Enri, S.; Fusconi, A.; Lonati, M.; Mucciarelli, M. A ddRADseq Survey of the Genetic Diversity of Rye (*Secale cereale* L.) Landraces from the Western Alps Reveals the Progressive Reduction of the Local Gene Pool. *Plants* **2021**, *10*, 2415. [CrossRef]
17. Gilman, I.S.; Tank, D.C. Species tree estimation using ddRADseq data from historical specimens confirms the monophyly of highly disjunct species of Chloropyron (Orobanchaceae). *Syst. Bot.* **2018**, *43*, 701–708. [CrossRef]
18. Zhang, Y.M.; Vitone, T.R.; Storer, C.G.; Payton, A.C.; Dunn, R.R.; Hulcr, J.; McDaniel, S.F.; Lucky, A. From pavement to population genomics: Characterizing a long-established non-native ant in North America through citizen science and ddRADseq. *Front. Ecol. Evol.* **2019**, *7*, 453. [CrossRef]
19. Ivanov, V.; Marusik, Y.; Pétillon, J.; Mutanen, M. Relevance of ddRADseq method for species and population delimitation of closely related and widely distributed wolf spiders (Araneae, Lycosidae). *Sci. Rep.* **2021**, *11*, 2177. [CrossRef]
20. Nazareno, A.G.; Bemmels, J.B.; Dick, C.W.; Lohmann, L.G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Mol. Ecol. Resour.* **2017**, *17*, 1136–1147. [CrossRef]
21. O'Connell, K.A.; Oaks, J.R.; Hamidy, A.; Kurniawan, N.; Smith, E.N.; Fujita, M.K. Congruent demographic responses to Pleistocene geological processes in Sumatran parachuting frogs: A comparison of target-capture and ddRADseq for population genomic analyses. *bioRxiv* **2019**, 843664. [CrossRef]
22. Marconi, G.; Capomaccio, S.; Comino, C.; Acquadro, A.; Portis, E.; Porceddu, A.; Albertini, E. Methylation content sensitive enzyme ddRAD (MCSeEd): A reference-free, whole genome profiling system to address cytosine/adenine methylation changes. *Sci. Rep.* **2019**, *9*, 14864. [CrossRef] [PubMed]
23. Dimond, J.L.; Roberts, S.B. Convergence of DNA methylation profiles in a novel environment in the reef coral *Porites astreoides*. *bioRxiv* **2019**, 747840. [CrossRef]
24. Czembor, E.; Waśkiewicz, A.; Piechota, U.; Puchta, M.; Czembor, J.H.; Stępień, Ł. Differences in ear rot resistance and Fusarium verticillioides-produced fumonisin contamination between Polish currently and historically used maize inbred lines. *Front. Microbiol.* **2019**, *10*, 449. [CrossRef] [PubMed]
25. Murray, M.; Thompson, W. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [CrossRef] [PubMed]
26. Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **2017**, *33*, 3137–3139. [CrossRef]

27. Lajmi, A.; Glinka, F.; Privman, E. Optimising ddRAD sequencing for population genomic studies with ddgRADer. *bioRxiv* **2022**. [CrossRef]
28. Durbin, R.M. An integrated map of genetic variation from 1092 human genomes. *Nature* **2012**, *491*, 56.
29. Yu, H.; You, X.; Li, J.; Liu, H.; Meng, Z.; Xiao, L.; Zhang, H.; Lin, H.-R.; Zhang, Y.; Shi, Q. Genome-wide mapping of growth-related quantitative trait loci in orange-spotted grouper (*Epinephelus coioides*) using double digest restriction-site associated DNA sequencing (ddRADseq). *Int. J. Mol. Sci.* **2016**, *17*, 501. [CrossRef]
30. Ulaszewski, B.; Meger, J.; Burczyk, J. Comparative analysis of SNP discovery and genotyping in *Fagus sylvatica* L. and *Quercus robur* L. using RADseq, GBS, and ddRAD methods. *Forests* **2021**, *12*, 222. [CrossRef]
31. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [CrossRef] [PubMed]
32. Tan, G.; Opitz, L.; Schlapbach, R.; Rehrauer, H. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Sci. Rep.* **2019**, *9*, 2856. [CrossRef] [PubMed]
33. Yang, G.-Q.; Chen, Y.-M.; Wang, J.-P.; Guo, C.; Zhao, L.; Wang, X.-Y.; Guo, Y.; Li, L.; Li, D.-Z.; Guo, Z.-H. Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods* **2016**, *12*, 39. [CrossRef]
34. Shirasawa, K.; Hirakawa, H.; Isobe, S. Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Res.* **2016**, *23*, 145–153. [CrossRef] [PubMed]
35. Aballay, M.M.; Aguirre, N.C.; Filippi, C.V.; Valentini, G.H.; Sánchez, G. Fine-tuning the performance of ddRAD-seq in the peach genome. *Sci. Rep.* **2021**, *11*, 6298. [CrossRef] [PubMed]
36. Srii, V.S.; Nethra, N.; Umarani, K.; Lohithaswa, H.; Shadakshari, Y.; Prasad, S.R. SNP genotyping of maize (Zea mays) hybrids and parental inbred lines for genetic purity testing using double digest restriction site-associated DNA sequencing. *Seed Sci. Technol.* **2021**, *49*, 193–206. [CrossRef]
37. Milner, S.G.; Jost, M.; Taketa, S.; Mazón, E.R.; Himmelbach, A.; Oppermann, M.; Weise, S.; Knüpffer, H.; Basterrechea, M.; König, P. Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* **2019**, *51*, 319–326. [CrossRef]
38. Liu, H.; Bayer, M.; Druka, A.; Russell, J.R.; Hackett, C.A.; Poland, J.; Ramsay, L.; Hedley, P.E.; Waugh, R. An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. *BMC Genom.* **2014**, *15*, 104. [CrossRef]
39. Pootakham, W.; Sonthirod, C.; Naktang, C.; Jomchai, N.; Sangsrakru, D.; Tangphatsornruang, S. Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: A case study in oil palm (*Elaeis guineensis*). *Mol. Breed.* **2016**, *36*, 154. [CrossRef]
40. De Donato, M.; Peters, S.O.; Mitchell, S.E.; Hussain, T.; Imumorin, I.G. Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE* **2013**, *8*, e62137. [CrossRef]
41. Lepais, O.; Weir, J.T. Sim RAD: An R package for simulation-based prediction of the number of loci expected in RAD seq and similar genotyping by sequencing approaches. *Mol. Ecol. Resour.* **2014**, *14*, 1314–1321. [CrossRef] [PubMed]
42. Verdu, C.F.; Guichoux, E.; Quevauvillers, S.; De Thier, O.; Laizet, Y.H.; Delcamp, A.; Gévaudant, F.; Monty, A.; Porté, A.J.; Lejeune, P. Dealing with paralogy in RAD seq data: In silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecol. Evol.* **2016**, *6*, 7323–7333. [CrossRef] [PubMed]
43. Vendrami, D.L.; Forcada, J.; Hoffman, J.I. Experimental validation of in silico predicted RAD locus frequencies using genomic resources and short read data from a model marine mammal. *BMC Genom.* **2019**, *20*, 72. [CrossRef] [PubMed]
44. Eaton, D.A.; Ree, R.H. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Syst. Biol.* **2013**, *62*, 689–706. [CrossRef]
45. Lambert, S.M.; Streicher, J.W.; Fisher-Reid, M.C.; Mendez de la Cruz, F.R.; Martínez-Méndez, N.; García-Vázquez, U.O.; Nieto Montes de Oca, A.; Wiens, J.J. Inferring introgression using RADseq and DFOIL: Power and pitfalls revealed in a case study of spiny lizards (Sceloporus). *Mol. Ecol. Resour.* **2019**, *19*, 818–837. [CrossRef]
46. Zhou, S.; Wei, F.; Nguyen, J.; Bechner, M.; Potamousis, K.; Goldstein, S.; Pape, L.; Mehan, M.R.; Churas, C.; Pasternak, S. A single molecule scaffold for the maize genome. *PLoS Genet.* **2009**, *5*, e1000711. [CrossRef]
47. López, A.; Carreras, C.; Pascual, M.; Pegueroles, C. Evaluating restriction enzyme selection for reduced representation sequencing in conservation genomics. *Mol. Ecol. Resour.* **2023**. *Online ahead of print*. [CrossRef]
48. Robinson, D.; Walsh, P.R.; Bonventre, J.A. Restriction endonucleases. In *Molecular Biology Problem Solver: A Laboratory Guide*; Wiley-Liss, Inc.: Hoboken, NJ, USA, 2001; pp. 225–266.
49. Graham, C.F.; Glenn, T.C.; McArthur, A.G.; Boreham, D.R.; Kieran, T.; Lance, S.; Manzon, R.G.; Martino, J.A.; Pierson, T.; Rogers, S.M. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADS eq). *Mol. Ecol. Resour.* **2015**, *15*, 1304–1315. [CrossRef]
50. Piffanelli, P.; Droc, G.; Mieulet, D.; Lanau, N.; Bès, M.; Bourgeois, E.; Rouvière, C.; Gavory, F.; Cruaud, C.; Ghesquière, A. Large-scale characterization of Tos17 insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **2007**, *65*, 587–601. [CrossRef]

51. Arnold, B.; Corbett-Detig, R.B.; Hartl, D.; Bomblies, K. RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol. Ecol.* **2013**, *22*, 3179–3190. [CrossRef]
52. Nguyen, T.K.; Yu, J.; Choi, H.-W.; In, B.-C.; Lim, J.-H. Optimization of genotyping-by-sequencing (GBS) in chrysanthemums: Selecting proper restriction enzymes for GBS library construction. *Hortic. Sci. Technol.* **2018**, *36*, 108–114. [CrossRef]