UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Agronomia Animali Alimenti Risorse naturali e Ambiente

TESI DI LAUREA MAGISTRALE

IN

SCIENZE E TECNOLOGIE AGRARIE

Genotyping by RAD sequencing analysis evaluated the genetic distinctiveness and relationship of chicory populations (*Chicorium intybus* var. *foliosum*, 2n=2x=18)

Supervisor: Prof. Gianni Barcaccia

Co-supervisors: Dott. Francesco Scariolo

Dott. Fabio Palumbo

Student: Giulia Agnoletto

ID number: 2023306

ACADEMIC YEAR 2021-2022

INDEX

ABSTRACT **1**

RIASSUNTO **3**

1. INTRODUCTION **4**

1.1. Aims of the study *6*

2. MATERIALS AND METHODS **8**

2.1. Plant Materials *8*

2.2. RADseq analysis and Data analysis *8*

2.3. Data analysis and genetic statistics *8*

2.4. BLASTn analysis and RADtags mapping *9*

3. RESULTS **11**

3.1. RADseq, Genetic similarity, and homozygosity estimates *11*

3.2. Genetic structure evaluation of the core collection *15*

3.3. BLASTn investigation for putative expressed SNP markers identification *16*

4. DISCUSSION **19**

5. CONCLUSIONS **20**

6. BIBLIOGRAPHY **22**

Abstract

Cichorium intybus L. is an important horticultural leafy vegetable cultivated worldwide, as in the countries of the Mediterranean basin as in the Americas, Oceania, Africa and Southern Asia, and its phenotypic variability is well represented by the large number of local biotypes, interspecific hybrids and cultivars. For the registration and protection of new commercial lines, the characterization of genetic diversity is crucial. In particular, population structure analysis, homozygosity estimates and genetic distinctiveness of commercial materials are essential for protecting them from theft. Regarding this, the Distinctness, Uniformity and Stability (DUS test) are the three major aspects considered in the constitution, registration, and protection of a new variety. To verify these parameters, beyond the classic morphological descriptors proposed by The International Union for the Protection of New Varieties of Plants (UPOV), several molecular tools and approaches can be adopted.

In this study, we investigated the genetic relationships of 94 chicory samples (*Chicorium intybus* var *foliosum*), belonging to four different experimental populations of chicory, using a *genotyping-by-sequencing* approach, based on the Restriction site-Associated DNA sequencing (RAD-seq) technology.

The sequencing data were filtered by removing the missing data and the multiallelic loci. A total of 2953 bi-allelic single nucleotide polymorphisms (SNPs) were retained and used to determine the homozygosity (82.9% on average), the mean genetic similarity in all pairwise comparisons (74.0% to 87.9%), to investigate the genetic relationships among the populations considered, and to reconstruct the genetic structure of the core collection. Moreover, a further investigation was performed to identify the SNPs contained in genic coding sequences (> 40%), which could be of future interest for the development of screening or selection essays. Overall, the RAD-Seq approach confirmed the distinctiveness of the analyzed populations. Future studies could take place starting from this study, leading to the development of a standardized, cheap and reliable analytical method able to provide useful information for chicory variety breeding, registration, and protection.

Riassunto

Cichorium intybus L. è un importante ortaggio a foglia coltivato in tutto il mondo, sia nei paesi del bacino del Mediterraneo che nelle Americhe, Oceania, Africa e Asia meridionale, e la sua variabilità fenotipica è ben rappresentata dal gran numero di biotipi locali, ibridi interspecifici e cultivar. Per la registrazione e la protezione di nuove linee commerciali, la caratterizzazione della diversità genetica è fondamentale. In particolare, l'analisi della struttura della popolazione, le stime di omozigosi e di distinguibilità genetica dei materiali commerciali sono aspetti essenziali per proteggere le varietà commerciali da furti o frodi. A questo proposito, la Distinguibilità, Uniformità e Stabilità (DUS test) sono i tre principali aspetti considerati nella costituzione, registrazione e protezione di una nuova varietà. Per verificare questi parametri, oltre ai classici descrittori morfologici proposti dall'Unione Internazionale per la Protezione delle Nuove Varietà di Piante (UPOV), possono essere adottati diversi strumenti e approcci molecolari.

In questo studio, abbiamo indagato le relazioni genetiche di 94 campioni di radicchio (*Chicorium intybus* var. *foliosum*), appartenenti a quattro diverse popolazioni sperimentali di radicchio, utilizzando un approccio di genotipizzazione tramite sequenziamento basato sulla tecnologia di sequenziamento del DNA associato al sito di restrizione (RAD- seq).

I dati di sequenziamento sono stati filtrati rimuovendo i dati mancanti e i loci multi-allelici. Un totale di 2953 polimorfismi bi-allelici a carico di singoli nucleotidi (SNPs) sono stati mantenuti e utilizzati per determinare l'omozigosi (82,9% in media), la similarità genetica media in tutti i confronti a coppie (dal 74,0% all'87,9%), per studiare le relazioni genetiche tra le popolazioni considerate, e per ricostruire la struttura genetica della collezione di campioni in esame. Inoltre, è stata condotta un'ulteriore indagine per identificare gli SNP contenuti nelle sequenze geniche codificanti (>40%), che potrebbero essere di futuro interesse per lo sviluppo di saggi di screening o di selezione. Nel complesso, l'approccio RAD-Seq ha confermato la particolarità delle popolazioni analizzate. A partire da questo studio potrebbero svolgersi studi futuri, che portino allo sviluppo di un metodo analitico standardizzato, rapido ed efficiente in grado di fornire informazioni affidabili per la selezione, la registrazione e la protezione delle varietà di radicchio.

1. Introduction

Chicory (*Cichorium intybus* L.) is a traditional European horticultural crop, and it is known as one of the most important crops cultivated in the region of Veneto [1]. It is a perennial herb from the *Cichorium* genus from the Asteraceae family [2].

The genus *Cichorium* consists of two important cultivated species: *Cichorium intybus,* or chicory (2n=18), and *Cichorium endivia,* or endive (2n=18). Moreover, other four wild species are known within this genus.

The first cultivated species is a diploid, perennial and self-incompatible (SI) leafy vegetable, while the second one is a diploid and annual crop, like the previous one, but it is a self-compatible species (SC).

Chicory domestication, as a vegetable crop, occurred in continental Europe, where it gradually differentiated into several cultivated varieties. The cultivation of this species is very ancient, being mentioned in many works from Virgil, Pliny and Theophrastus [3].

The name of this plant seems to be derived from Greek and Latin: "*Cichorium*" means field and "*intybus*" is partly derived from the Greek word "to cut", because of the leaves, and partly from the Latin "*tubus*" to indicate the hollow stem [4]. Noteworthy, other theories connect the origin of chicory's name to the Egyptian words *Kio* (= I) and *chorion* (= field) [5].

So far, chicory has been used in the food industry for the preparation of ready-to-eat salads, for teas and tea blends, for coffee supplementation, and as a source for inulin production, both for humans and in animal feed preparation. Some compounds present in this plant, such as polyphenols, inulin, oligofructose and sesquiterpene lactones may be considered as potential carriers of food functionality [6].

For its utilization in medicine, over 100 chemical compounds have been isolated and identified from this plant, the majority of which are accumulated in the root tissues.

These compounds are reported to have antimicrobial, anthelmintic, antimalarial, hepatoprotective, antidiabetic, gastroprotective, and anti-inflammatory properties [7], as well as many others that could be mentioned [8].

Chicory gradually underwent a process of naturalization in Europe. Although it cannot be considered an autochthonous species, it became part of the natural and agricultural European flora. Wild *C. intybus* covers a great portion of the European continent, and it has become a part of the traditional diet of local populations as an important ingredient of

typical local dishes. The cultural aspects of this crop might be both the consequence and the cause of the great differentiation that occurred in this species, which originated an ever-increasing number of cultivar groups, types, and populations that altogether comprise the horticultural landscape of the genus *Cichorium*. As said, this genus is important from a historical, cultural, agronomical, commercial, and scientific point of view due to its vastness of species, distribution, and cultivation around the world [9].

Chicory is a hardy plant and can endure extreme temperatures during both vegetative and reproductive growth, for this reason, it is cultivated in many countries and environments: it is widely distributed in Africa, temperate- Asia, tropical- Asia, Europe, Australia, Northern and Southern America. For what concerns the Veneto region, the origins of Radicchio (its common local name) are not clearly defined, although recent insights relate the origin of the Venetian biotypes to two distinct developmental pathways [10].

Chicory cultivars that are nowadays grown in the Veneto region are mostly derived from individuals having red leaves, which could be traced back to the biotype "Red of Treviso", firstly introduced in Veneto during the XV century [10].

Many types of radicchio are intensely cultivated nowadays, like the "Early" and "Late Red of Treviso", the "Red of Chioggia", the "Variegated of Castelfranco" and the "Red of Verona" [11].

Another division that can be done in separating the radicchio biotypes concerns two main groups: in the first one the pure chicory biotypes are grouped in such a way that include the two "Red of Treviso" and the "Red of Verona" biotypes, anciently developed by mass-selection of Late Red of Treviso plant materials; the second group, instead, includes the "Variegated of Castelfranco", originated from a controlled or spontaneous cross between the "Late Red of Treviso" and *Cichorium endivia* var. *latifolium* (endive), and the "Red of Chioggia", that was obtained by mass selection of the "Variegated of Castelfranco" biotype [9,10].

Differently from the past, in which plant breeding was performed by mass selection to promote the phenotype of interest, much progress has been achieved in the last few years in crops' genetic studies and in the use of molecular markers to improve their breeding strategies [12].

Marker-assisted breeding (MAB) approaches are nowadays widely adopted by seed firms when planning the development of new varieties. By determining the genetic similarity, homozygosity and relatedness of parental lines, as well as the distinctiveness, uniformity

and stability (DUS test) of the progenies, the development of stronger, more resistant and marketable cultivars has become easier, faster and more reliable.

Historically, commercial varieties were developed by recurrent mass selection based on phenotypic markers, but, in the last years, synthetic varieties have been constituted by breeders through inter-crossing or poly-crossing a considerable number of maternal individuals or clonal lines selected based on their morpho-phenological and agronomic traits, and, eventually, by performing progeny tests to assess their general combining ability [13]. During the past two decades, the agricultural scenery in the Mediterranean countries has radically changed for chicory cultivations, including radicchio biotypes, where subsistence mixed farming units have been converted into extensive farming systems growing mainly modern improved varieties instead of local varieties [13].

Currently, owing to the economic benefits, newly released varieties are mainly F1 hybrids developed by Italian or European seed firms through large-scale single crosses between inbred lines selected according to their specific combining ability and exploiting MAB strategies. Thus, radicchio's breeding programmes have significantly improved in recent years due to more efficient biotechnological tools that have been adopted, where synthetic varieties show higher distinctiveness, uniformity and stability for both agronomic and phenotypic traits [14,15]. In this regard, saturated linkage maps with DNA markers spanning the entire genome (approximately 2.6 Gb) are available for leaf chicory [16]. These maps are important considering that biotechnology and molecular genetics are largely utilised in breeding programmes of radicchio [16-20] as well as for most crops [21].

1.1. Aims of the study

This study was based on the characterization of 96 chicory samples belonging to four inbred populations by means of the restriction site-associated DNA sequencing (RAD-seq) approach.

Through this *genotyping-by-sequencing* (GBS) analysis, the genetic relationships and the genetic structure of the core collection were evaluated to verify the presence of potential crosses aimed at the maximisation of heterosis in the eventual progeny, and putative *single nucleotide polymorphisms* (SNP) to be adopted in future screening essays were investigated.

To maximize hybrids' uniformity and distinctiveness, the necessity of an informative molecular characterization of the parental lines is necessary. For this reason, among the genetic statistics that were calculated in this study, genetic similarity and homozygosity are

present, which are used to determine the uniformity and distinctiveness of the considered populations.

Given the high data throughput of the analytical methodology adopted, a parallel investigation of the possible location of the identified SNP markers was performed by means of a BLASTn analysis. For this aim, two phylogenetically related species' exomes were retrieved from international databases and used for the creation of a local database against which align the RAD-tags obtained by the RADseq analysis. This investigation was carried out to identify those markers that could putatively be within coding regions of the chicory genome, which could affect the plant phenotype and could, consequently, be adopted in marker-assisted selection (MAS) approaches in breeding programs of this crop.

2. Materials and methods

2.1. Plant Materials

For this experiment 96 samples of *Chicorium intybus* var. *foliosum* have been used, which were provided by a private seed company.

The genomic DNA samples have been extracted from 100 mg of fresh leaves using a DNeasy® 96 Plant Kit (Qiagen, Valencia, CA, USA) and following the procedure provided by the manufacturer. Both the quality and quantity of the genomic DNA samples were assessed by agarose gel electrophoresis (1% agarose/1× TAE gel containing 1× Sybr Safe DNA stain, (Life Technologies, Carlsbad, CA, USA) and a NanoDrop 2000c UV-Vis spectrophotometer (Thermo Fisher Scientific Inc., Pittsburgh, PA, USA), respectively.

2.2. RADseq analysis and Data analysis

The 96 samples were analysed by means of a restriction site-associated DNA sequencing (RAD-seq) approach. The RAD-based sequencing used is a well-established and powerful method used for recovering thousands of polymorphic loci across the genomes of many crop species. One microgram of DNA per individual was digested with the restriction enzyme *MseI* (New England Biolabs, Ipswich, MA, USA), following the procedure described by Stevanato [22].

For library preparation, digested DNA samples were diluted at a concentration of 3 ng/μL. Indexing, library preparation, sequencing and bioinformatic analyses were performed according to the protocol described by Stevanato et al [22].

Raw reads obtained through an Ion S5 sequencer (Thermo Fisher Scientific Inc., Waltham, MA, USA) were trimmed according to the restriction enzyme recognition motif. After quality assessment, all artifacts and Ns-containing reads were removed. Variants were called using Stacks v2.41 software [23].

SNPs were filtered to remove those meeting the following criteria: (1) SNPs with greater than 10% missing data, (2) SNPs with a sequence depth × 4, and (3) tri- and tetra-allelic SNPs.

2.3. Data analysis and genetic statistics

The obtained data were used for the construction of an unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on Rohlf's genetic similarity simple matching coefficient and a principal coordinate analysis (PCoA) centroid using NTSYS

software v2.21 [24]. Additionally, a Bayesian clustering algorithm implemented in STRUCTURE v.2.2 [25] was used to model the genetic structure of the chicory core collection. The number of founding groups ranged from 1 to 20, and 10 replicate simulations were conducted for each value of K based on a burn-in of 20,000 and a final run of 100,000 Markov chain Monte Carlo (MCMC) steps. STRUCTURE HARVESTER [26] was used to estimate the most likely value of K, and the results of membership were then plotted as a histogram using an Excel spreadsheet.

2.4. BLASTn analysis and RADtags mapping

We analyzed 94 samples of *Chicorium intybus* var *foliosum* and the reads with no missing data were used to identify those sequences most likely belonging to genomic coding sequences (CDSs).

Thus, the genomes of two phylogenetically closely related species to this genus, namely *Lactuca sativa* L. (GeneBank: GCA_002870075.2) and *Cynara cardunculus* L. (GeneBank: GCA_001531365.1), were considered.

The RAD-tags were aligned against both the *Lactuca sativa* and *Cynara cardunculus* CDS datasets using a local BLASTn (BLAST+ 2.11.0 package) with an E-value threshold $\leq 1.0\times10^{-10}$ and a percentage of identity ≥80%. This step was made to retrieve suitable information about the possibly expressed regions of the chicory genome, which could help future breeding programs in adopting marker-assisted selection (MAS) strategies.

3. Results

3.1. RADseq, Genetic similarity, and homozygosity estimates

From the RAD-seq analysis of the 96 chicory samples belonging to a core collection, two were not successfully sequenced because of poor DNA quality and were excluded from the following analyses for this reason.

The RAD sequencing produced a considerable number of raw reads that, after quality assessment and adapter trimming were used to create a catalog of consensus loci, then used for variant calling as a reference. A starting pool of 9351 SNPs was first identified. The RAD-tags (SNPs containing reads) considered in the following analyses were all 69bp long. Unfortunately, the raw data obtained from the sequencing of the 94 chicory accessions were lost due to a blackout that corrupted the server files containing the mentioned information. After the filtering step, in which all the sequences that presented at least one missing value in one sample were rejected, 2953 SNPs were retrieved in 2917 RAD-tags and used for the subsequent genetic statistics analyses.

Among the 2953 SNPs presenting no missing data in the entire core collection, the number of discriminative polymorphic sites among the most similar genotypes was 2 SNPs (0.0007 %), whereas the one calculated among the most dissimilar individuals was 1194 SNPs (40.43%), ~~bo''''th~~ considering heterozygous loci as non-discriminant.

Among the 94 sequenced samples, the genetic similarity (GS) was computed in all pair-wise comparisons, while the average GS was calculated within and among each population of the core collection, together with the respective standard error (SE).

Genetic similarity ranged overall from 47.21% to 98.56%, with an average value of 65.53%, whereas the mean GS within populations ranged from 82.33% (Pop1) to 91.05% (Pop4) and the GS calculated among them ranged from 54.08% (Pop3 vs. Pop4), as the lowest value, and 76.60% (Pop1 vs. Pop2), as the highest one. Standard errors were also computed that were always SE < 0.05% (Table 1).

The lowest genetic similarity values were observed in the comparisons between "Pop3" and the other populations, which were always below 57%, as reported in Table 1.

For each sample homozygosity was also estimated. The highest homozygosity was observed in sample "Pop4-21" (Ho = 94.48%), and the lowest was observed in sample "Pop1-03" (Ho = 66.71%). The mean homozygosity among the core collection was 82.90%. As for the GS, also the mean homozygosity was calculated within each of the four

populations that ranged from 74.04 ± 0.31% to 87.91 ± 0.13% ("Pop1" and "Pop4", respectively) (Table 1).

Table 1. Average pair-wise genetic similarity matrix and average homozygosity calculated within and among populations; standard error is also reported.

Homozygosis	Population ID	Mean Genetic Similarity			
74.04 ± 0.31%	**Pop1**	82.33 ± 0.04%			
83.67 ± 0.15%	**Pop2**	76.60 ± 0.01%	90.12 ± 0.01%		
82.83 ± 0.13%	**Pop3**	56.46 ± 0.00%	54.13 ± 0.00%	82.50 ± 0.01%	
87.91 ± 0.13%	**Pop4**	62.59 ± 0.01%	58.46 ± 0.00%	54.08 ± 0.00%	91.05 ± 0.01%
		Pop1	**Pop2**	**Pop3**	**Pop4**

From the GS matrix computed for the 94 chicory samples, a UPGMA dendrogram was created that grouped samples into four branches (from Pop1 to Pop4), which agreed with the populations' information given by the seed company that provided the samples of the core collection.

In the UPGMA, Pops 1, 2 and 4 were located in one main branch having an average within GS equal to 72.79%, while that computed in the comparison of this large group with Pop3 exhibited a genetic similarity that is on average equal to 54.66%.

The only exception in the exact matching of the UPGMA's clusters with expected populations interested sample Pop1-06 that was placed in the same branch containing Pop2's individuals (Figure 1, blue labelled arm in Pop2 cluster).

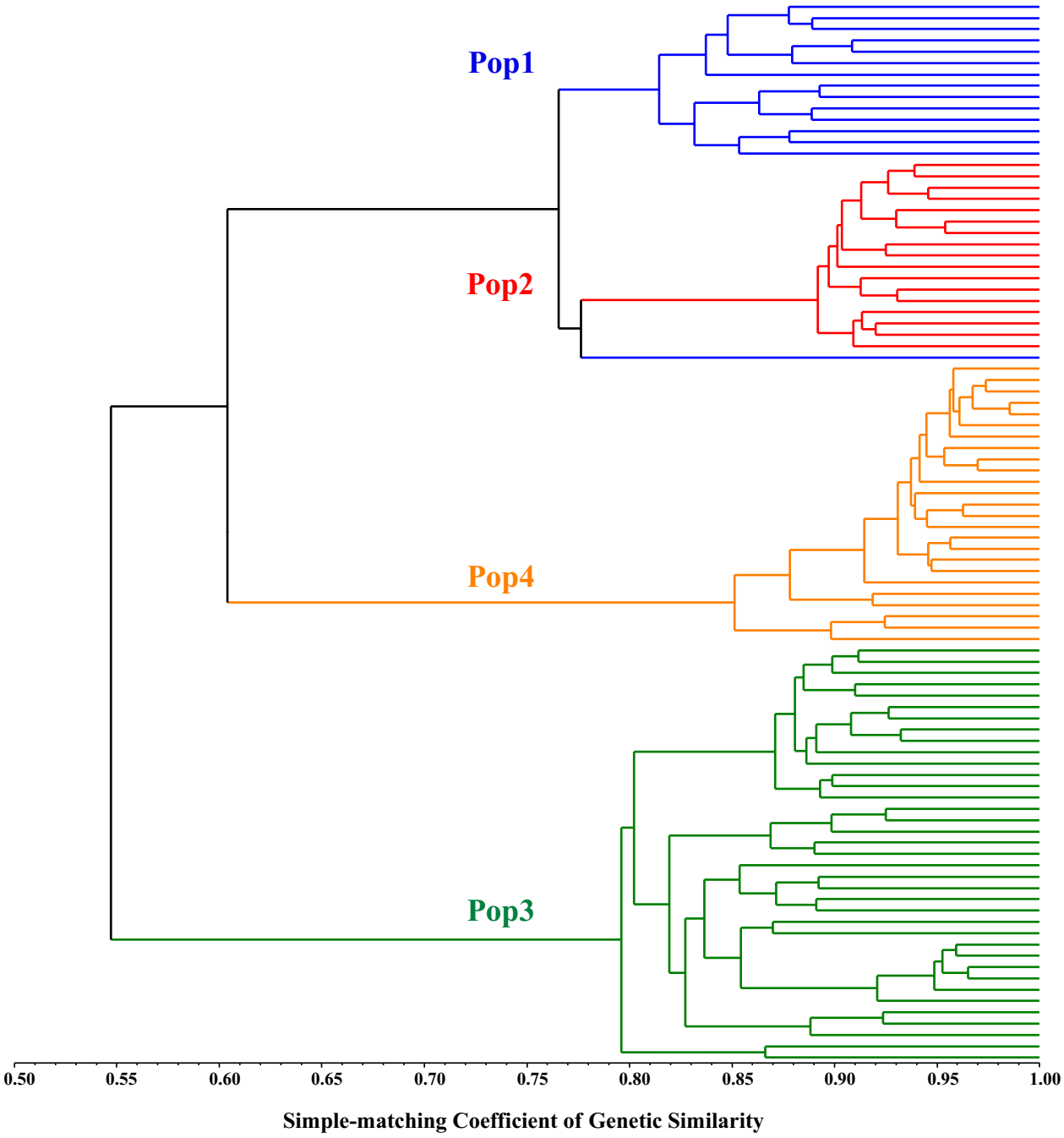

Figure 1. UPGMA dendrogram based on the pair-wise genetic similarity matrix highlighting four main "Clusters" for the no missing values containing dataset.

As the clustering in the dendrogram showed, the PCoA (Principal Coordinate Analysis) further divided the samples into four distinct populations. The PCoA clustered the samples in the chart depending on Dimension 1 and Dimension 2, which represented, respectively, 50.4% and 35.7% of the molecular variability, and 86.1% of the molecular variation in total. This analysis was based on the eigenvectors calculated starting from the genetic similarity matrix and highlighted that the 4 groups, previously identified in the UPGMA

dendrogram, once again agreed with the expected clustering based on the belonging population of each of the 94 samples of the core collection analysed.

All four populations found were clustered independently from the rest of the other groups. Pops 1,2, 3 and 4 formed unique groups with their respective individuals and were placed in different quadrants of the chart, except for Pop1 and Pop2 which groped close to each other, still remaining distinguishable (Figure 2).

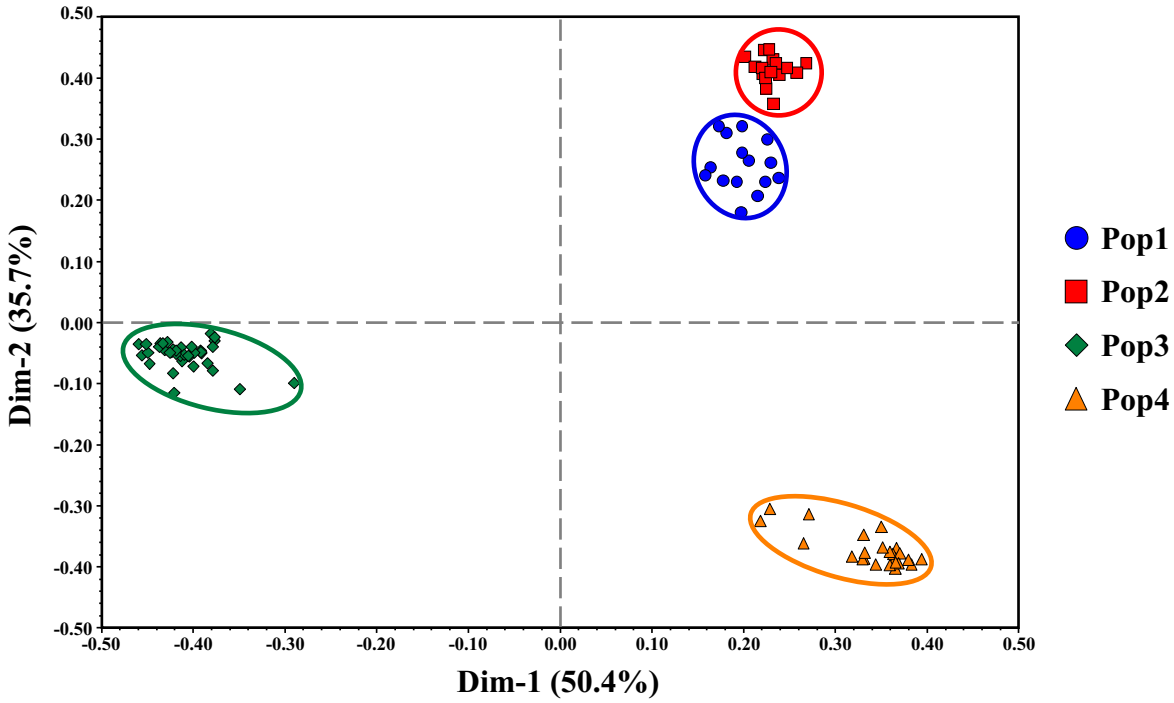

Figure 2. Principal Coordinate Analysis (PCoA) calculated starting from the genetic similarity matrix and highlighting the 4 mains groups (Pop1 to Pop4) identified for the 94 analysed samples of chicory.

3.2. Genetic structure evaluation of the core collection

As far as the genetic structure analysis of the chicory core collection was concerned, the results obtained from STRUCTURE software, then analysed using STRUCTURE harvester web software to estimate the most probable value of K [27], resulted in a value of K = 4 (ΔK = 40925.3) (Figure 3).

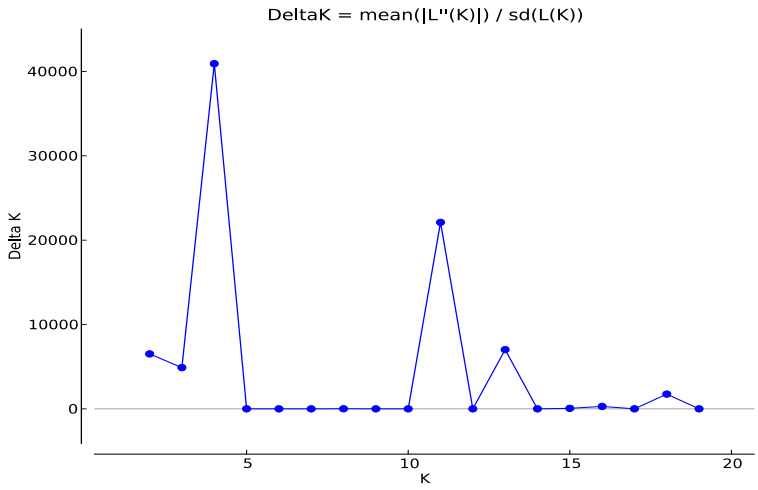

Figure 3. Graph representing the ΔK values resulting from Structure Harvester software analysis of the STRUCTURE software results.

Regarding the ΔK values (Figure 3) resulting from Structure Harvester software computation, it estimated the second most probable number of groups equal to K = 11. This said, after the results were plotted as a histogram, the observed clustering was completely in agreement with that for K=4 with the 7 newly identified groups presenting memberships < 5% (data not shown), and, for this reason, the result is not reported either discussed in the present study.

For K = 4, the memberships of the 94 samples with the respective hypothetical ancestral genotypes were plotted as a histogram and labelled with four different colours (Figure 4). Among the analysed populations, Pops 2, 3 and 4 had high memberships (> 95% on average) to one specific identified ancestral genotype, while Pop1 resulted to be equally admixed in its samples between a fourth ancestral genotype (Cluster-1) and that of Pop2 (Cluster-2), thus resulting with memberships ratios of 50:50 on average.

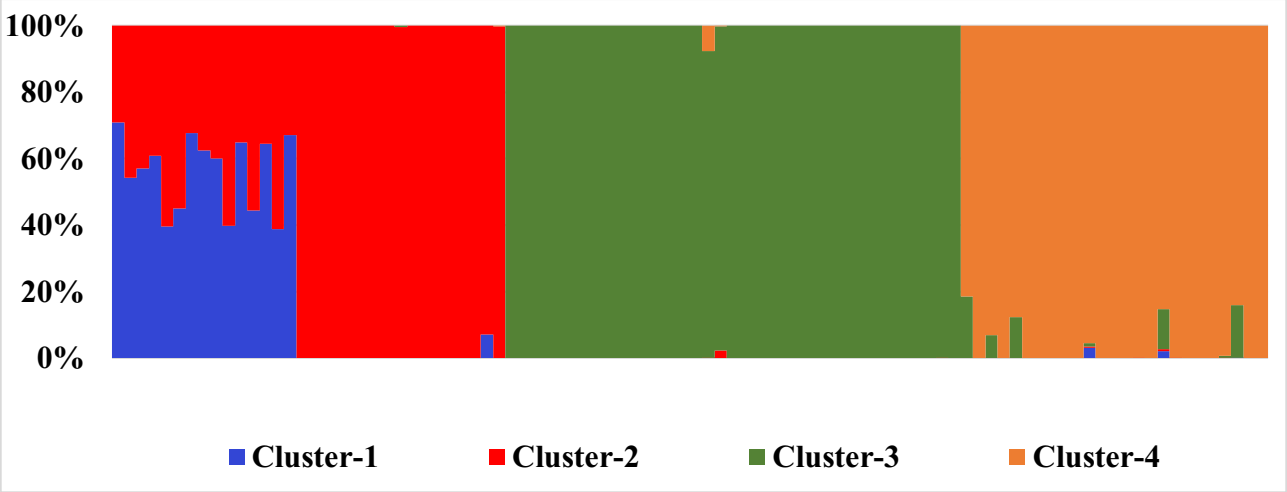

Figure 4. Genetic structure analysis of the radicchio core collection. Identified most likely value of K = 4.

3.3. BLASTn investigation for putative expressed SNP markers identification

For what concerned BLASTn analysis and RAD-tags mapping, the 94 samples of chicory were also used to identify those sequences most likely belonging to genomic coding sequences (CDSs) (Table 2). After the filtering steps, 2917 RAD-tags were aligned against the exome of both *L. sativa* and *C. cardunculus,* both species belonging to the family of Asteraceae. Concerning the lettuce exome, 1708 RAD-tags were mapped on 1710 CDS, whereas in the case of artichoke 1209 RAD-tags aligned on 1175 CDS.

In both crops, the 1710 and 1175 CDS were translated and produced the same number of proteins. In addition, the average percentage of identity (Avg ident) was respectively 90.01% with the exome of lettuce and 87.59% with the one of artichoke.

The calculated average nucleotide alignment length (Avg length) was equal to 66.46 bp and 66.44 bp in lettuce and artichoke, respectively. Considering the average E-value (Avg E-value), values were $1.73E^{-12}$ in lettuce and $3.09E^{-12}$ in artichoke, while the average bitscores (Avg Bs) were equal to 90.83 and 83.69, respectively. Furthermore, the average mismatches (Avg mis) resulted for the two investigations against both the related crops resulted to be 6.53 and 8.22 respectively, as the average of identical matches and average identity (Avg ident) were 59.79 and 58.16 (Table 2).

Table 2. Summary statistics of the BLASTN analysis of the RAD-Seq reads against the exomes of *L. sativa* and *C. cardunculus*. Average identity percentage (Avg ident), average length (Avg length), average E-value (Avg E-value), average bitscore (Avg Bs), average score (Avg score), average mismatches (Avg mis), average identity (Avg ident) and average positive positions values (Avg pp) were calculated.

BLASTn Result	RAD tags (n)	CDS	Proteins (n)	Avg ident (%)	Avg length (bp)	Avg E-value	Avg Bs	Avg score	Avg mis (n)	Avg ident (n)
Exome *L. sativa*	1708	1710	1710	90.01	66.46	$1.73\ E^{-12}$	90.83	99.43	6.53	59.79
Exome *C. cardunculus*	1209	1175	1175	87.59	66.44	$3.09E^{-12}$	83.69	91.45	8.22	58.16

4. Discussion

The RAD-Seq has become a more and more useful technology in the last decades, expanding its application from QTL mapping in crops [28-30] to gene mapping and marker-assisted breeding (MAB) [31,32].

Moreover, it has also been applied for the identification of new crop varieties, for the investigation of population structures [33] and in phylogenetics studies for the estimation of the evolutionary relationships among species, individuals, and genes [34].

The aim of the study was to characterize and discriminate a pool of 94 chicory accessions belonging to four different populations and, in the context of DUS testing, our findings may have great implications for new F1 hybrids constitution and new varieties registration.

Another goal of the study was the investigation of the genetic relationships among the populations considered and the reconstruction of the genetic structure of the core collection.

The RAD sequencing approach used in this research allowed to investigate the relatedness degree existing among the samples of the core collection thanks to the 2953 SNP sites with no missing data that were retrieved among 2917 RAD-tags, which were used for the subsequent genetic statistical analyses and population structure reconstruction.

From the analyses of the genetic variability and relationships, it was observed that the four chicory populations analysed were distinguishable and uniform within each of them. The genetic structure reconstruction of the core collection, in accordance with the GS estimates, the UPGMA clustering, and the PCoA grouping of samples, confirmed the distinctiveness and uniformity of the analysed populations. Moreover, the histogram in Figure 4 highlighted the relatedness between Pop1 and Pop2, as already expected due to the results obtained in the UPGMA and PCoA. From the combined results, it can be observed that Pop1 and Pop2 share a putative common ancestor (Cluster-2 in Figure 4), fact that can explain the close placing of their samples in the PCoA graph, the common branch of the UPGMA dendrogram, as well as the presence of Sample01-06 next to Pop2's accessions in Figure 1 (labelled in blue in the red branch). Noteworthy, Pop2 presents high membership with a single cluster, while Pop1 resulted admixed with a mean ratio of 50:50 with the cluster of Pop2. This result, together with the high GS among these two populations (average $GS_{Pop1/Pop2} = 76.60\%$), could indicate a common lineage, which makes their crossing unsuitable for the constitution of new F1 hybrids characterized by high heterosis (or heterotic vigor).

On the other hand, an interesting insight that is related to the GS estimates was the high dissimilarity of Pop3 when compared to Pops 1, 2 and 4. With a mean among-GS value equal to 54.66%, Pop3 resulted as the most genetically distant population of the core collection, which, in combination with its mean observed homozygosity (Ho_{Pop3} = 82.83%), makes it a potential parental line for the constitution of new F1 hybrid varieties if crossed with a highly uniform, homozygous and dissimilar population (e.g. Pop2 and Pop4).

For what concerned the BLASTn analysis, we considered the exomes of lettuce and artichoke for the identification of the SNP markers contained in the CDSs because of the absence of a chicory exome that is annotated and representative. In fact, the genome of chicory has been sequenced and published, but it has not been annotated yet, making it unsuitable for this analysis.

BLASTn results of the obtained RAD-tags against the two exome-based databases showed that 1708 and 1209 RAD-tags matched with the coding regions of lettuce (1710 CDS) and artichoke (1175 CDS) respectively, as expected from the phylogenetic relatedness of these two species with *C. intybus*. Noteworthy, also the low E-values observed from this analysis demonstrated the strong relatedness of these species belonging to the Asteraceae family. From these findings, the analytical method adopted, which was inspired by the work of Scariolo et al. [35], demonstrated its possible applicability to the genus *Chicorium* and its species and related taxa.

This said, these results highlighted the presence of several putative markers which could be of interest for the development of predictive screening essays aimed at the application of marker-assisted selection (MAS) strategies for planning crosses to constitute new varieties presenting phenotypic traits of interest. In the future, further studies should be carried out to verify the suitableness of the identified markers for their application in MAS essays for chicory, but these first insights provide a good starting point for this aim.

5. Conclusions

In conclusion, genotyping analysis by RAD-Seq approach demonstrated its usefulness in assessing the genetic identity and relationships of the given accessions of chicory. The discriminative ability of the molecular markers retrieved from the identified RAD-tags, the informativeness of the retrieved molecular data and their possible use in the bioinformatic analysis could help future projects aimed at developing new chicory varieties and their registration, as well as their traceability. Furthermore, the high data throughput of RAD-

seq could be applied for the further development of screening essays aimed at MAS approaches in this species, and not only at MAB strategies.

6. Bibliography

1. Simko, I.; Jia, M.Y.; Venkatesh, J.; Kang, B.C.; Weng, Y.Q.; Barcaccia, G.; Lanteri, S.; Bhattarai, G.; Foolad, M.R. Genomics and Marker-Assisted Improvement of Vegetable Crops. *Crit Rev Plant Sci* **2021**, *40*, 303-365, doi:10.1080/07352689.2021.1941605.

2. Bremer, K. *Asteraceae: Cladistics and Classification*; Timber Press: Portland, Oregon, 1994.

3. Candolle, A. *Origin of cultivated plants*; Litres: 2021.

4. Al-Snafi, A.E. Medical importance of *Cichorium intybus*–A review. *IOSR Journal of Pharmacy* **2016**, *6*, 41-56.

5. Kiers, A.M. Endive, chicory, and their wild relatives. A systematic and phylogenetic study of *Cichorium* (Asteraceae). *Gorteria-Supplement* **2000**, *5*, 1-77.

6. Perovic, J.; Tumbas Saponjac, V.; Kojic, J.; Krulj, J.; Moreno, D.A.; Garcia-Viguera, C.; Bodroza-Solarov, M.; Ilic, N. Chicory (*Cichorium intybus* L.) as a food ingredient - Nutritional composition, bioactivity, safety, and health claims: A review. *Food Chem* **2021**, *336*, 127676, doi:10.1016/j.foodchem.2020.127676.

7. Wang, F.X.; Deng, A.J.; Li, M.; Wei, J.F.; Qin, H.L.; Wang, A.P. (3S)-1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid from *Cichorium endivia* L. induces apoptosis of human colorectal cancer HCT-8 cells. *Molecules* **2012**, *18*, 418-429, doi:10.3390/molecules18010418.

8. Street, R.A.; Sidana, J.; Prinsloo, G. *Cichorium intybus*: Traditional Uses, Phytochemistry, Pharmacology, and Toxicology. *Evid Based Complement Alternat Med* **2013**, *2013*, 579319, doi:10.1155/2013/579319.

9. Barcaccia, G.; Ghedina, A.; Lucchin, M. Current Advances in Genomics and Breeding of Leaf Chicory (*Cichorium intybus* L.). *Agriculture* **2016**, *6*, 50, doi:10.3390/agriculture6040050.

10. Basso, A.; Scariolo, F.; Negrisolo, E.; Lucchin, M.; Barcaccia, G. Molecular Relationships and Genetic Diversity Analysis of Venetian Radicchio (Leaf Chicory, *Cichorium intybus* subsp. *intybus* var. *sylvestre*, 2n = 2x = 18) Biotypes. *Diversity* **2022**, *14*, doi:10.3390/d14030175.

11. Lucchin, M.; Varotto, S.; Barcaccia, G.; Parrini, P. Chicory and Endive. In *Vegetables I: asteraceae, brassicaceae chenopodi-caceae*, Prohens, J., Nuez, F., Eds. Springer: New York, NY, 2008; Vol. 1, pp. 3-48.

12. Hasan, N.; Choudhary, S.; Naaz, N.; Sharma, N.; Laskar, R.A. Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J Genet Eng Biotechnol* **2021**, *19*, 128, doi:10.1186/s43141-021-00231-1.

13. Barcaccia, G.; Lucchin, M.; Parrini, P. Characterization of a flint maize (Zea mays var. indurata) Italian landrace, II. Genetic diversity and relatedness assessed by SSR and Inter-SSR molecular markers. *Genet Resour Crop Ev* **2003**, *50*, 253-271, doi:10.1023/A:1023539901316.

14. Barcaccia, G.; Pallottini, L.; Soattin, M.; Lazzarin, R.; Parrini, P.; Lucchin, M. Genomic DNA fingerprints as a tool for identifying cultivated types of radicchio (*Cichorium intybus* L.) from Veneto, Italy. *Plant Breed.* **2003**, *122*, 178-183, doi:10.1046/j.1439-0523.2003.00786.x.

15. Patella, A.; Scariolo, F.; Palumbo, F.; Barcaccia, G. Genetic Structure of Cultivated Varieties of Radicchio (*Cichorium intybus* L.): A Comparison between F1 Hybrids and Synthetics. *Plants* **2019**, *8*, 213, doi:10.3390/plants8070213.

16. Galla, G.; Ghedina, A.; Tiozzo, S.C.; Barcaccia, G. Toward a First High-quality Genome Draft for Marker-assisted Breeding in Leaf Chicory, Radicchio (*Cichorium intybus* L.). In *Plant Genomics*, IntechOpen: 2016; 10.5772/61747.

17. Cadalen, T.; Mörchen, M.; Blassiau, C.; Clabaut, A.; Scheer, I.; Hilbert, J.L.; Hendriks, T.; Quillet, M.C. Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Molecular Breeding* **2010**, *25*, 699-722, doi:10.1007/s11032-009-9369-5.

18. Gonthier, L.; Blassiau, C.; Morchen, M.; Cadalen, T.; Poiret, M.; Hendriks, T.; Quillet, M.C. High-density genetic maps for loci involved in nuclear male sterility (NMS1) and sporophytic self-incompatibility (S-locus) in chicory (*Cichorium intybus* L., Asteraceae). *Theor. Appl. Genet.* **2013**, *126*, 2103-2121, doi:10.1007/s00122-013-2122-9.

19. Muys, C.; Thienpont, C.N.; Dauchot, N.; Maudoux, O.; Draye, X.; Van Cutsem, P. Integration of AFLPs, SSRs and SNPs markers into a new genetic map of industrial chicory (*Cichorium intybus* L. var. *sativum*). *Plant Breed.* **2014**, *133*, 130-137, doi:10.1111/pbr.12113.

20. Palumbo, F.; Qi, P.; Pinto, V.B.; Devos, K.M.; Barcaccia, G. Construction of the First SNP-Based Linkage Map Using Genotyping-by-Sequencing and Mapping of the Male-Sterility Gene in Leaf Chicory. *Front. Plant Sci.* **2019**, *10*, 276, doi:10.3389/fpls.2019.00276.

21. Singh, B.D.; Singh, A.K. *Marker-Assisted Plant Breeding: Principles and Practices*; Springer New Delhi, India, 2015; 10.1007/978-81-322-2316-0.

22. Stevanato, P.; Broccanello, C.; Biscarini, F.; Del Corvo, M.; Sablok, G.; Panella, L.; Stella, A.; Concheri, G. High-Throughput RAD-SNP Genotyping for Characterization of Sugar Beet Genotypes. *Plant Molecular Biology Reporter* **2014**, *32*, 691-696, doi:10.1007/s11105-013-0685-x.

23. Rochette, N.C.; Rivera-Colon, A.G.; Catchen, J.M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* **2019**, *28*, 4737-4754, doi:10.1111/mec.15253.

24. Rohlf, F.J. *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System Version 2.1: Owner manual*; Exeter Publishing: Setauket, New York, 2000.

25. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945-959, doi:10.1093/genetics/155.2.945.

26. Earl, D.A.; Vonholdt, B.M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **2011**, *4*, 359-361, doi:10.1007/s12686-011-9548-7.

27. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **2005**, *14*, 2611-2620, doi:10.1111/j.1365-294X.2005.02553.x.

28. Wang, S.Z.; Wang, B.C.; Liu, J.; Ren, J.; Huang, X.X.; Zhou, G.L.; Wang, A.H. Novel polymorphic EST-based microsatellite markers characterized in lettuce (*Lactuca sativa*). *Biologia* **2017**, *72*, 1300-1305, doi:10.1515/biolog-2017-0154.

29. Zhang, F.; Kang, J.; Long, R.; Yu, L.X.; Wang, Z.; Zhao, Z.; Zhang, T.; Yang, Q. High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biol* **2019**, *19*, 165, doi:10.1186/s12870-019-1770-6.

30. Wang, L.; Conteh, B.; Fang, L.; Xia, Q.; Nian, H. QTL mapping for soybean (*Glycine max* L.) leaf chlorophyll-content traits in a genotyped RIL population by using RAD-seq based high-density linkage map. *BMC Genomics* **2020**, *21*, 739, doi:10.1186/s12864-020-07150-4.

31. Wu, K.; Liu, H.; Yang, M.; Tao, Y.; Ma, H.; Wu, W.; Zuo, Y.; Zhao, Y. High-density genetic map construction and QTLs analysis of grain yield-related traits in sesame (*Sesamum indicum* L.) based on RAD-Seq techonology. *BMC Plant Biol* **2014**, *14*, 274, doi:10.1186/s12870-014-0274-7.

32. Peng, Y.; Hu, Y.; Mao, B.; Xiang, H.; Shao, Y.; Pan, Y.; Sheng, X.; Li, Y.; Ni, X.; Xia, Y., et al. Genetic analysis for rice grain quality traits in the YVB stable variant line using RAD-seq. *Mol Genet Genomics* **2016**, *291*, 297-307, doi:10.1007/s00438-015-1104-9.

33. Feng, J.; Zhao, S.; Li, M.; Zhang, C.; Qu, H.; Li, Q.; Li, J.; Lin, Y.; Pu, Z. Genome-wide genetic diversity detection and population structure analysis in sweetpotato (*Ipomoea batatas*) using RAD-seq. *Genomics* **2020**, *112*, 1978-1987, doi:10.1016/j.ygeno.2019.11.010.

34. Liu, L.; Jin, X.; Chen, N.; Li, X.; Li, P.; Fu, C. Phylogeny of *Morella rubra* and Its Relatives (Myricaceae) and Genetic Resources of Chinese Bayberry Using RAD Sequencing. *PLoS One* **2015**, *10*, e0139840, doi:10.1371/journal.pone.0139840.

35. Scariolo, F.; Palumbo, F.; Vannozzi, A.; Sacilotto, G.B.; Gazzola, M.; Barcaccia, G. Genotyping Analysis by RAD-Seq Reads Is Useful to Assess the Genetic Identity and Relationships of Breeding Lines in Lavender Species Aimed at Managing Plant Variety Protection. *Genes (Basel)* **2021**, *12*, doi:10.3390/genes12111656.