

Article

# The Molecular Determination of Hybridity and Homozygosity Estimates in Breeding Populations of Lettuce (*Lactuca sativa* L.)

Alice Patella, Fabio Palumbo \*, Giulio Galla  and Gianni Barcaccia 

Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, 35020 Legnaro PD, Italy; alice.patella@phd.unipd.it (A.P.); giulio.galla@unipd.it (G.G.); gianni.barcaccia@unipd.it (G.B.)

\* Correspondence: fabio.palumbo@unipd.it

Received: 23 September 2019; Accepted: 7 November 2019; Published: 9 November 2019



**Abstract:** The development of new varieties of horticultural crops benefits from the integration of conventional and molecular marker-assisted breeding schemes in order to combine phenotyping and genotyping information. In this study, a selected panel of 16 microsatellite markers were used in different steps of a breeding programme of lettuce (*Lactuca sativa* L.,  $2n = 18$ ). Molecular markers were first used to genotype 71 putative parental lines and to plan 89 controlled crosses designed to maximise recombination potentials. The resulting 871 progeny plants were then molecularly screened, and their marker allele profiles were compared with the profiles expected based on the parental lines. The average cross-pollination success rate was  $68 \pm 33\%$ , so 602 F1 hybrids were completely identified. Unexpected genotypes were detected in 5% of cases, consistent with this species' spontaneous out-pollination rate. Finally, in a later step of the breeding programme, 47 different F3 progenies, selected by phenotyping for a number of morphological descriptors, were characterised in terms of their observed homozygosity and within-population genetic uniformity and stability. Ten of these populations had a median homozygosity above 90% and a median genetic similarity above 95% and are, therefore, particularly suitable for pre-commercial trials. In conclusion, this study shows the synergistic effects and advantages of conventional and molecular methods of selection applied in different steps of a breeding programme aimed at developing new varieties of lettuce.

**Keywords:** pure lines; F1 hybrids; microsatellite markers; marker-assisted breeding; crop improvement; varieties

## 1. Introduction

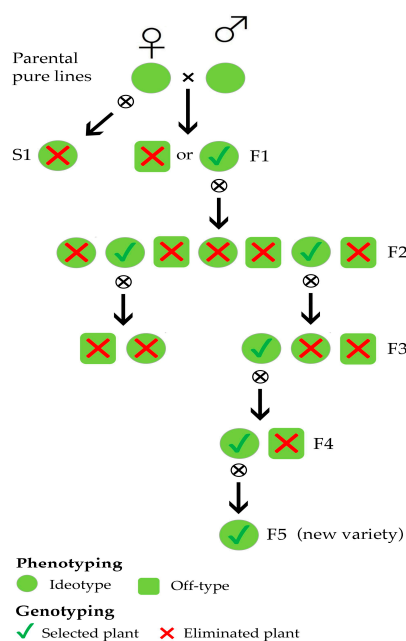
Lettuce (*Lactuca sativa* L.) is a self-pollinating leafy vegetable species ( $2n = 2x = 18$ ) of the Asteraceae family. It is cultivated on a large scale throughout the world for consumption as a fresh vegetable on its own or in combination with other ready-to-eat vegetable products [1]. Its growing economic importance has led seed companies to regularly develop new varieties with ever higher agronomic traits. However, breeding programmes are highly limited by the reproductive system of this species. The flower structure of lettuce determines a reproductive strategy known as cleistogamy, in which anther dehiscence and subsequent pollination take place before flower opening, resulting in a very high rate of self-pollination, very often equal to or close to 100% [2]. According to recent estimates, out-cross rates are limited to 1–6% [3]. These reproductive barriers mean that in natural conditions the species spontaneously constitutes pure lines, characterised by phenotypic uniformity and genotypic stability, due to their very high homozygosity. In conventional breeding programmes, developing segregating and recombinant F2 populations traditionally requires crosses to be hand pollinated

while self-pollination is prevented by emasculating the flowers. The most popular emasculation and hand-pollination technique is that described by Olivier [4]. Known as the “wash method”, it involves hand-spraying the inflorescence with water during pistil emergence to remove the pollen attached to the female part of the flower. The inflorescence is then left to dry for a short period, after which it is rubbed with a ripe flower of the pollinating variety [5]. A slightly different, but also widely used, technique is the “clip-and-wash method”, which involves clipping the tip of the corolla before spraying with water. This guarantees more efficient pollen removal and cross rates close to 100% from the subsequent manual pollination [2]. However, these breeding techniques are time-consuming and technically highly demanding, and are only really effective if coupled with molecular analyses aimed at screening progeny plants and assessing their hybridity.

In recent years, many seed firms have begun using molecular markers to carry out assisted selection schemes and to speed up varietal development programmes [3]. Simple Sequence Repeat (SSR) markers are, so far, the most commonly used markers for these purposes [6–8] as they are codominant, have high reproducibility and multi-allelism, and can be detected at any stage of plant development, without being influenced by the environment [9]. There are a considerable number of SSR markers for lettuce in the literature [10]. Truco, et al. [11] produced an integrated genome map from 7 different linkage maps, which included 130 SSR loci organised in 9 linkage groups. Rauscher and Simko [10] augmented this genomic map with 54 genomic SSR and 52 EST-SSR (Expressed Sequence Tag) loci. Finally, with the publication of the *L. sativa* genome draft [12], tens of thousands of new SSR regions have become available for testing and use.

Given the availability of markers in lettuce, Marker-Assisted Selection (MAS) has started to be adopted in plant breeding programmes for various purposes, including identification of resistance genes [13,14] or Quantitative Trait Loci (QTLs) of phytopathogens [15,16], the study of QTLs controlling complex traits [17,18], and investigation of the genetic identity and purity of either experimental or commercial lines [19]. On the other hand, very few attempts have been made to prove the efficiency of molecular markers in Marker-Assisted Breeding (MAB) activities, where the genotypic background is molecularly investigated to complement traditional phenotypic selection [20].

In this work, SSR markers were used in three different steps of a conventional breeding scheme aimed at developing new varieties characterised by distinctiveness, uniformity, and stability (Figure 1).



**Figure 1.** Simplified overview of a lettuce breeding scheme in which selection is based on both plant phenotyping and genotyping.

We first examined the genetic background of a number of superior pure lines in order to plan experimental matings to produce F1 hybrids and then derived F2 progenies manifesting morphological variability as a result of genetic segregation and recombination (Figure 1). Each offspring in the F1 generation was analysed to distinguish the individuals resulting from planned out-crosses from those resulting from accidental selfing (Figure 1). After genotyping, the S1 individuals were discarded, and the F1 individuals were self-pollinated. In the F3 generations (Figure 1); experimental populations, previously selected according to their morphological traits, were also characterised by molecular markers due to the need to assess their stability and uniformity in order to run pre-commercial trials.

## 2. Materials and Methods

### 2.1. Plant Materials and Breeding Techniques

Plant materials were developed and provided by Blumen Group SpA, Italy and belonged to five different lettuce cultivar types (Table S1). Seventy-one parental lines (germplasm composed of experimental, pre-commercial and commercial lines) were involved in 89 combinations of crosses, in which each progeny consisted of 6–12 individuals (871 progeny samples). Parental lines were grown in the spring of 2015, and the 89 programmed crosses were carried out in the summer using the clip-and-wash method [2]. This involved making an incision in the calyx and corolla and washing the anthers in the early morning before the pollen grains could settle naturally on the outermost stigmatic surface of pistils. The plants were then manually pollinated by rubbing anthers of the pollen donor on the stigma of the seed parent. For each planned cross, a bulk of 4/5 flowers from a pollinator parent was used to pollinate as many flowers of a seed plant. Seeds were collected from the seed plant and sown in early autumn for genotyping selection and agronomic evaluation (spring 2016).

Finally, to assess the uniformity of the 47 experimental lines, previously chosen for morpho-phenological traits and pathogen resistances, 940 samples belonging to the 47 F3 populations (labelled 1 to 47) were collected in the spring of 2018. Each experimental line comprised 20 individuals.

### 2.2. DNA Isolation

A total of 100 mg of fresh leaves was collected from each of the 1882 lettuce samples (71 parents, 871 progeny and 940 F3) and ground to a fine powder using Tissue Lyser II (Qiagen, Valencia, CA, USA). Genomic DNA (gDNA) was extracted with the Dneasy<sup>®</sup> 96 Plant Kit (Qiagen), according to the manufacturer's protocols. After extraction, the integrity of the gDNA was assessed by electrophoresis on 1% (*w/v*) agarose gel stained with SYBR Safe<sup>®</sup> 1 × DNA Gel Stain (Life Technologies, Carlsbad, CA, USA) in Tris-Acetate-EDTA (TAE) running buffer. Both the yield and purity of the extracted gDNA samples were evaluated using a NanoDrop 2000c UV-Vis Spectrophotometer (Thermo Scientific, Carlsbad, CA, USA). Following DNA quantification, all DNA samples were diluted to a final concentration of 20 ng/μL.

### 2.3. Primer Design and Testing of SSR Marker Amplification

Sixteen SSR marker loci were selected from those available in the scientific literature [10,21], according to (i) chromosomal location; (ii) polymorphism rate, expressed as PIC (Polymorphism Information Content); (iii) allele size range; (iv) annealing temperature of the locus-specific primers. Amplifications were performed according to the method previously described by Schuelke [22], with some minor modifications. Briefly, three primers were used to amplify each microsatellite locus: a pair of locus-specific primers, one with an oligonucleotide tail at the 5' end (M13, PAN-1, PAN-2 or PAN-3, Table S2), and a third universal primer complementary to the tail and labelled with a fluorescent dye (6-FAM, VIC, NED, or PET). Primer pair efficiency was tested *in silico* using the PRaTo [23] web-tool and were organised in three multiplex reactions, as shown in Table 1.

**Table 1.** Microsatellite loci information. For each primer pair, the original simple sequence repeat (SSR) name, ID used in this work, linkage group [10,21], SSR motif, primer sequences (PAN1, PAN2, PAN3, or M13 tails at the 5' end are indicated in square brackets; for further details see Table S2), dye and the multiplex to which the SSR marker locus belongs is shown.

Marker Name	ID	LG	Motif		Primer Sequence	Dye	Multiplex
LSSA27-2 [10]	Lsat1	1	(AC) <sub>7</sub>	For Rev	[M13]CACACTACCCACCCAACACG ACCCCTTCGCTTCTTCTT	6-FAM	1
SML-045 [21]	Lsat2	2	(AAG) <sub>9/12</sub>	For Rev	ACAAAACCGTTTCACCCAAA [M13]AGCCCTGCCTTTCAGGAT	6-FAM	1
LSSB54 [10]	Lsat3	8	(GT) <sub>10</sub>	For Rev	[PAN1]CTTGAGAGTGCTTGGAGAGGAT CACATACAACAAGACAAGTCCCA	VIC	1
LSSA05 [10]	Lsat4	8	(TC) <sub>18</sub>	For Rev	AGAACAACGGTAGCTTGTTAAATTG [PAN1]ATCGTCGGTTAATCTTCGTCG	VIC	1
LSSA04 [10]	Lsat5	4	(TC) <sub>14</sub>	For Rev	[PAN2]AAGGAAAGGAAGGGTTGACTTGT TTGGTGAAGAAAAGAGAGAGTTT	NED	1
LSSA11 [10]	Lsat6	3	(CT) <sub>20</sub>	For Rev	[PAN2]ACTCCCACTATCCTCTTTGCAT GCCACATTCTAATCTTGTC	NED	1
LSSA14 [10]	Lsat7	9	(AG) <sub>18</sub>	For Rev	[PAN3]TGATGACTCCAGTCTTAGATACCA AGTCCCGACTATCAGTCTCA	PET	1
LSSB09 [10]	Lsat8	2	(TG) <sub>8</sub>	For Rev	AGAATGAGAAGGATGAAATGGCTG [M13]AAACACCTTTAGCATCAAAATACCC	6-FAM	2
SML-029 [21]	Lsat9	9	(GAG) <sub>7/8</sub>	For Rev	[M13]AGCCCAGAAGAGCGTGATTA TGCAGGGCTCCTTGATCTAC	6-FAM	2
LSSB17-1 [10]	Lsat10	7	(GT) <sub>11</sub>	For Rev	ACTAGGGCTCTAATACAACCTTGT [PAN1]TTGGCTTACAGTTATGGATTAAATG	VIC	2
LSSA17 [21]	Lsat11	3	(AG) <sub>21</sub>	For Rev	[PAN1]AATGTGCGTGAGAGTTTCCTTT CAAGAAGGCAGTGATGAAGTTG	VIC	2
LSSA12 [10]	Lsat12	5	(GT) <sub>11</sub>	For Rev	[PAN2]ACAAGGCCCAATCCTTTTCT TCGAAAATTTGGAGAGAGTTTCTT	NED	2
LSSA15 [10]	Lsat13	1	(AC) <sub>11</sub>	For Rev	GCCCAACCCAAGAAGAGGAG [PAN3]TGGAGAGGAGTGGAGAGTGTT	PET	2
LSSA28-1 [10]	Lsat14	4	(GA) <sub>28</sub>	For Rev	TTCATCTCTCTCCTCCTTCAGC [M13]ATCCCCATTGTCCTCCC	6-FAM	3
LSSA21-1 [10]	Lsat15	8	(TC) <sub>19</sub>	For Rev	[PAN2]TTGTACCCAGTTGTCCAAACAG CAGATTGTTGCAGATTCTTCG	NED	3
LSSB68 [10]	Lsat16	na	(CT) <sub>20</sub>	For Rev	GTCTGTGTGGTTTTGGT [PAN3]TGTGGTGGAGTGTGATT	PET	3

The 16 primer pairs were first tested individually (singleplex reactions) using three randomly chosen lettuce gDNA to evaluate primer efficiency and to check the correspondence between expected and actual size of the bands; they were then evaluated in multiplex PCRs to assess possible primer interactions.

All amplification reactions (both singleplex and multiplex) were performed in a 10 µL reaction volume containing 1× Platinum<sup>®</sup> Multiplex PCR Master Mix (Thermo Scientific), 10% GC Enhancer (Thermo Scientific), 0.25 µM of non-tailed primer, 0.75 µM of tailed primer, 0.50 µM of fluorophore-labelled primer (universal primer) and 20 ng of genomic DNA. Thermal cycling conditions were as follows for multiplex 1 and 2: 94 °C for 5 minutes followed by 6 cycles of 94 °C for 30 seconds, 61 °C for 30 seconds, 72 °C for 45 seconds, with a 1 °C annealing temperature stepdown per cycle (from 61 °C to 56 °C). The annealing temperature for the following 35 cycles was set at 56 °C, with denaturation and extension phases as above and a final extension at 60 °C for 30 minutes. The multiplex 3 thermal cycling conditions were instead: 94 °C for 5 minutes followed by 6 cycles of 94 °C for 30 seconds, 56 °C for 30 seconds, 72 °C for 45 seconds, with a 1 °C annealing temperature stepdown per cycle (from 56 °C to 51 °C). The annealing temperature for the following 35 cycles was set at 51 °C with denaturation and extension phases as above and a final extension at 60 °C for 30 minutes. All amplifications were performed in a GeneAmp<sup>®</sup> PCR 9700 thermal cycler (Applied Biosystems, Carlsbad, CA, USA). PCR products were first checked on gel electrophoresis (2% Ultrapure<sup>™</sup> Agarose in TAE 1×, SYBR Safe<sup>®</sup> 1×, Life Technologies) then run on capillary electrophoresis with ABI 3730 DNA Analyzer (Applied Biosystem), using LIZ500 as the molecular weight standard. The size of each peak was determined with the Peak Scanner 1.0 software (Applied Biosystems).

#### 2.4. Genotyping and Data Analysis

The 71 potential parents were genotyped at 16 SSR loci and statistical analyses were performed using NTSYS (Numerical Taxonomy and Multivariate Analysis System) version 2.2 (Exeter Software) [24]. Rohlf's (or the simple matching) coefficient was used to calculate pairwise genetic similarity (GS) in all possible comparisons and to construct a genetic similarity matrix, according to the formula:

$$GS_{ij} = m/(m + n) \quad (1)$$

where "i" and "j" are two different individuals, while "m" and "n" represent the number of matching and non-matching attributes, respectively. An unweighted pair group method with an arithmetic mean (UPGMA) dendrogram and a Principal Coordinates Analysis (PCoA) of parental lines were carried out using the Jaccard coefficient in the PAST software v. 3.14 with 10,000 bootstrap repetitions [25]. The genetic structure of the lines was modelled using a Bayesian clustering algorithm implemented in STRUCTURE v. 2.2 [26]. Since no *a priori* knowledge of the origin of the populations under study was available, the admixture model and then the correlated allele frequencies model were used. Ten replicate simulations were conducted for each value of K, with the number of founding groups ranging from 1 to 8, using a burn-in of 200,000 and 1,000,000 iterations. The most likely K Estimates were determined using the method described by Evanno et al. [26]. Estimates of membership were plotted as a histogram in an Excel spreadsheet. Finally, observed homozygosity ( $H_o$ ) was determined with the POPGENE software [27].

The 89 subsequent crosses were planned according to the following criteria: (i) high genetic dissimilarity values among parents within the same lettuce cultivar type and between them, (ii) availability of informative loci able to distinguish between the resulting offspring and individuals resulting from accidental self-pollinated. Only homozygous loci for different alleles were considered informative, whereas heterozygous loci were taken into account only if the origin of the parental alleles could be clearly discerned in the progenies. The resulting offspring (871 samples) were then screened, with the analysis restricted to those SSR loci which had previously proven to be informative for hybrid detection. This made it possible to determine whether individuals belonging to a given F1 population originated from cross-pollination or self-pollination. The successful crosses (S.C.) rate of 89 was calculated as follows:

$$S.C. = (No F1 \times 100)/(No Tot - No U.G.) \quad (2)$$

where "No F1" is the number of hybrid individuals, "No Tot" is the number of all individuals in a progeny population (No tot = No F1 + No U.G. + No SP) and "No U.G." is the number of unexpected genotypes deriving from unplanned crosses.

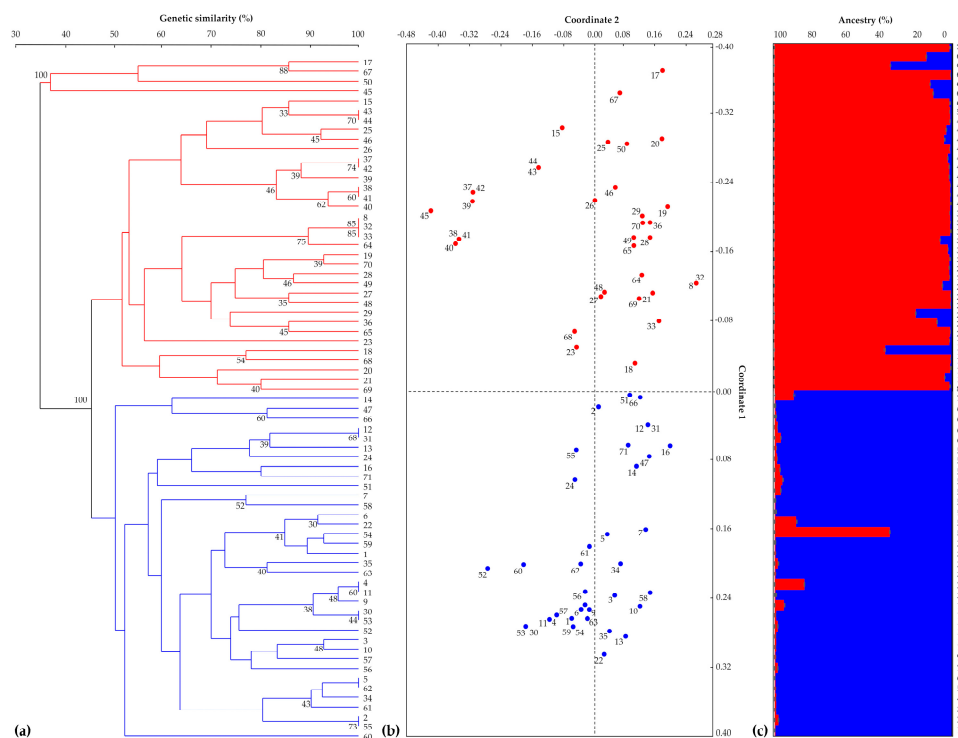
Finally, 940 samples from the 47 F3 populations were genotyped using the previously-described panel of SSR markers. The POPGENE software [27] was used to compute the mean values of observed homozygosity for each population (3), where n is the total number of samples). In addition, the median of genetic similarity between the 47 lines was calculated using Rohlf's coefficient, which was designed for codominant molecular markers [28,29]. Comparison of genetic similarity among ten selected populations was instead calculated using the Jaccard coefficient, in accordance with the literature [30]. Genetic similarity matrices were generated in the NTSYS software [24].

$$\bar{H}_o = \sum_n H_o/n \quad (3)$$

### 3. Results

#### 3.1. Parental Lines

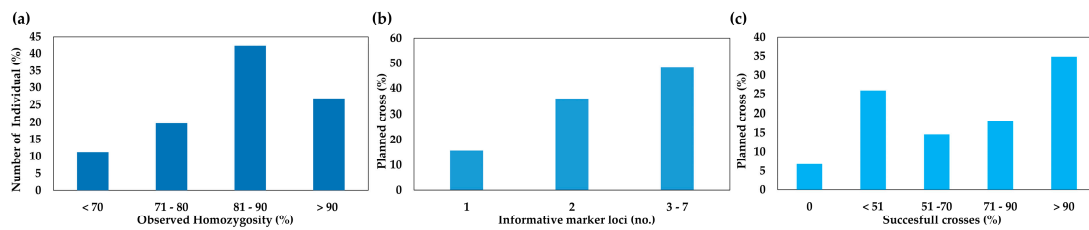
The 16 SSR markers, organised in three multi-locus PCRs, were used firstly to amplify and score the 71 parental lines. Fourteen of the 16 SSR markers proved to be polymorphic among plant accessions. The similarity matrix constructed using Rohlf's coefficient revealed genetic similarity values ranging from 53% to 100% (Figure S1). The resulting unweighted pair group method with an arithmetic mean (UPGMA) dendrogram showed the samples clustering into two main sub-groups. Eighteen parental lines were not fully distinguishable, while the remaining 53 had unique genotypic profiles. The first principal coordinate from the PCoA accounted for 22% of the total variation and divided the samples into two groups, analogous to the clustering in the tree. The second principal coordinate accounted for 12% of the total variation. These results were confirmed by investigation of the genetic structure of the 71 parental lettuce lines based on allele frequencies; the best estimate of population size was  $K = 2$  (Figure S2), such that the samples were grouped into two genetically distinct clusters (Figure 2). The lettuce cultivar types were reported in Table S1, but they did not correspond to different clusters in the UPGMA tree.



**Figure 2.** (a) Genetic similarity-based unweighted pair group method with an arithmetic mean (UPGMA) dendrogram of 71 parental lines calculated using the Jaccard coefficient. Bootstrap estimates  $\geq 30\%$  are reported next to the nodes (red and blue branches indicate the two clusters identified). (b) Principal coordinate analysis (PCoA). The 71 samples are shown in red or blue according to the clustering shown in the UPGMA tree. (c) The population genetic structure of the 71 lines as estimated by STRUCTURE. Each sample is represented by a vertical histogram partitioned into  $K = 2$  coloured segments (red or blue, in accordance with (a,b)) representing the estimated membership. The proportion of subgroup membership (%) is reported on the ordinate axis, and the identification number of each accession is reported below each histogram.

The mean observed homozygosity was 82%, with a minimum value of 69% and a maximum of 100%. It is worth noting that 19 of the 71 parental lines (27%) had observed homozygosity values greater than 90%, while 30 of the 71 (42%) had a medium-high observed homozygosity ( $H_o$ ) between

81% and 90%. Fourteen of the 71 parental lines (20%) had observed homozygosity ranging from 71% to 80%, and only 8 individuals had values lower than 70% (Figure 3a and Figure S1).



**Figure 3.** (a) Observed homozygosity of 71 lettuce parental individuals belonging to as many pure lines. (b) Histogram of discriminating loci in 89 cross combinations (in percentages). (c) Histogram of the percentages of pollination success in 89 programmed lettuce crosses.

### 3.2. Determination of Hybridity

Using a combination of genotypic and phenotypic data, 89 cross combinations were planned (Table S3). Before proceeding, we also checked the availability of informative loci able to distinguish between offspring resulting from out-cross and those obtained by accidental self-pollination. Screening identified 1 discriminant locus in 16% of cases, 2 informative loci in 36% of cases, and 3 to 7 informative molecular markers in 48% of the crosses (Figure 3b). The three most informative loci were *Lsat3*, *Lsat7*, and *Lsat6*, while *Lsat4* and *Lsat13* were monomorphic in almost all parental groups. It is worth noting that the *Lsat8* marker was in a heterozygous state in all but four parental genotypes (7, 45, 58, and 60).

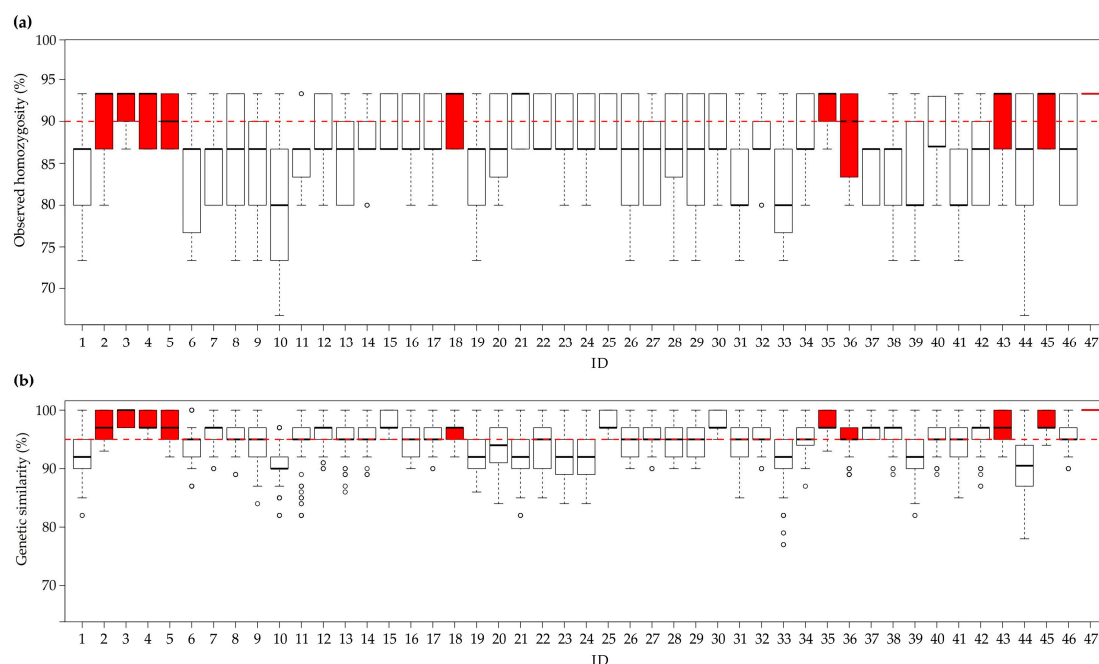
We were able to take advantage of these informative loci to calculate the success rate of each cross. In 30% of manual pollinations (27 out of 89), a success rate of 100% was achieved (i.e., all the offspring were hybrids), whereas in 18% of crosses (16 out of 89), the S.C. ranged from 71% to 90%. A hybridity rate fluctuating between 51% and 70% was reported in 15% of cases (13 out of 89), while 26% of the crosses produced fewer than 50% hybrids each. Finally, in only 7% of crosses (6 out of 89) were all the offspring the result of self-pollination (Figure 3c and Table S3). Overall, the mean hybridization rate (the average number of hybrids per crosses) was  $68 \pm 33\%$ , and out of a total of 871 individuals, 602 (69%) were hybrids, and 556 were derived from programmed crosses. The remaining 46 individuals (5% of the total) had a unexpected genotypes (U.G.) compared with their putative parents (Table S3).

### 3.3. Lettuce Breeding Populations

The 47 F3 experimental lines were genotyped using the same set of 16 SSR loci as for the previous analyses. The homozygosity estimates of all samples (940) ranged from 67% to 93% (Figure 4a). Ten experimental populations had a median observed homozygosity  $\geq 90\%$ . Outliers—with homozygosity values consistently deviating from the median—were present in only three experimental populations (11, 14, and 32).

The median genetic similarity observed within each line was always greater than 90% (Figure 4b), and 37 experimental populations had a median genetic similarity  $\geq 95\%$ . Outliers were present in 21 of the 47 lines (Figure 4b).

After assembling the data, we found 10 breeding populations, belonging to butterhead type (Table S4), to have  $H_o$  values  $\geq 90\%$ , and a median genetic similarity  $\geq 95\%$ ; the box-plots of these populations are labelled in red in Figure 4a,b. Finally, in the genetic similarity matrix calculated from all pairwise comparisons of these ten populations, the Jaccard's index ranged from  $44\% \pm 3\%$  (between populations 3 and 18) to  $96 \pm 5\%$  (between populations 45 and 47, Figure S3). Moreover, the populations called 45 and 47 were constituted starting from the same parents ( $2 \times 6$ , Table S4).



**Figure 4.** Statistics relating to the observed homozygosity and genetic similarity among lines. (a) Box-plot of the median observed homozygosity (in percentages) in each of the 47 populations. The red dotted line represents the homozygosity threshold set at 90%. (b) Box plot of the median genetic similarity in each experimental population (in percentages). The red dotted line represents the genetic similarity threshold set at 95%. The red box-plots represent the ten best experimental populations (observed homozygosity  $\geq 90\%$  and genetic similarity values  $\geq 95\%$ ). The second and third quartiles are marked inside the square and are divided by a bold bar (median). Dots show outlier samples.

#### 4. Discussion

The last decade has seen major advances in the acquisition of knowledge concerning the genetics of lettuce and, in particular, the development of molecular markers [1,11,21]. This has facilitated marker-assisted selection programmes, especially those aimed at countering the onset of new diseases. For example, several studies have dealt with identifying the QTLs associated with biotic and abiotic stress resistance [17,31,32]. Molecular markers have also been extensively used to assess genetic variation and relationships in lettuce germplasm [19] and to identify possible duplicate varieties [33]. However, although the benefits derived from exploitation of these molecular tools have also been discussed in marker-assisted breeding programmes [34] and demonstrated in several species [7,35], there are only a few studies on this type in lettuce [20]. The aim of our work, therefore, was to integrate conventional and biotechnological methods in three different steps of a breeding programme to show that this strategy is also effective in *L. sativa* (Figure 1). This is of pivotal importance if we consider the economic impact of lettuce (the world production of lettuce and chicory in 2017 was 26.8 million tons [36]) and the need to regularly develop new varieties.

Commercial lettuce varieties are usually characterised by pure lines due to the autogamous nature of this species. In order to introduce variability, manual pollination is usually carried out to cross genetically stable parent lines with agronomic traits of interest. Progeny selection is a crucial step, but despite the efficiency of some emasculation and hand pollination methods developed over the years [2], a major problem—distinguishing unequivocally and rapidly F1 individuals from self-pollinated progeny—still remains. The use of molecular assays to quickly and accurately screen progeny makes it possible to overcome most of the conventional breeding limits in this species.

In this context, our SSR-based analysis has (i) facilitated selection of the best parents to cross in order to maximise the variability of the progeny both within the same cultivar type and among



them, (ii) allowed accurate evaluation of the resulting offspring, and (iii) sped up the screening of experimental F3 lines for their stability and uniformity.

The first part of our work focused on pre-screening 71 parental lines for crossing with the aim of maximising the gains obtained from each out-pollination within cultivar type and, in some cases, among them. As expected, the similarity matrix and the unweighted pair group method with an arithmetic mean (UPGMA) dendrogram showed varying levels of similarity among the different parental genotypes. Parental germplasm appeared to divide into two different groups, as revealed by the Principal Coordinates Analysis (PCoA) results and particularly by the genetic structure analysis. However, samples did not separate in UPGMA tree and PCoA according to the cultivar type, but we may assume that increasing the number of markers it could be possible to clarify this clustering. Although 53 parental lines were found to be fully distinguishable, with similarity values ranging from 53% to 98% and characterised by a unique genotypic profile, it was impossible to identify unequivocally the remaining 18. This is not surprising if we consider that some of the parental lines were closely related. We may speculate that increasing the number of SSRs would allow us to address these remaining issues. Given the aim of this study, these data were useful to avoid crosses between parents with 100% similarity. To introduce variability according to the phenotype and the lowest similarity values, we carried out 89 crossing combinations. Another aspect that needs to be considered when planning crosses is the stability of the parental line in terms of homozygosity. In our study, the median observed homozygosity of the parental lines was lower than expected (82%), especially in light of the strictly autogamous reproductive system of lettuce [37]. Overall, the fact that only one individual in four had homozygosity values greater than 90% showed that some of these lines were not entirely stable. However, it must be borne in mind that, although the observed homozygosity was not optimal, some of these lines, experimental lines, were chosen to produce F1 partly because they displayed resistance to multiple pathogens and had interesting phenotypic traits.

Before proceeding with hand pollination, in order to distinguish between F1 individuals resulting from cross-pollination and those resulting from self-pollination, we first examined the informative loci among the parental lines used in the crosses. Only homozygous loci for different alleles in parental lines were considered informative. Our analysis showed at least 2 informative loci in 84% of the programmed crosses. It is worth pointing out that restricting the analysis to the informative loci brought us considerable savings in terms of time and costs.

Overall, the molecular determination of hybridity was successful: F1 individuals represented at least 51% of the offspring in 67% of the manual crosses, and 100% of the offspring in 30% of the crosses (100% success rate), in agreement with the estimates originally reported by the developer of the pollination technique [2]. Unexpected genotypes (U.G.) were identified in 5% of the individual progeny. In these cases, the progeny genotypes appeared to diverge from what would be expected given the parents. This percentage is consistent with the spontaneous or undesired occurrences of cross-pollination (1–6%) reported in the literature for this species [3], mainly due to pollinator insects. However, we cannot exclude human error during manual pollination or seed collection.

Finally, at an advanced step of the breeding programme, we genetically assessed 47 different experimental F3 populations (940 samples), previously selected for their morpho-phenotypic traits and resistances of interest (Figure 1). Interestingly, the findings in terms of both homozygosity and intra-line similarity were very good. This would suggest that in strictly autogamous species, such as lettuce, three cycles of self-pollination may already be sufficient to reach desired outcomes in terms of genetic uniformity and homozygosity. This also confirms that the use of molecular markers could speed up the process by making it possible to select the best individuals on the basis of their genotype, thereby reducing the number of generations needed to develop new varieties. The ten experimental populations with the highest homozygosity estimates ( $\geq 90\%$ ) and the highest intra-genetic similarity values ( $\geq 95\%$ ) were considered suitable for pre-commercial trials (red box plot, Figure 4). However, a pairwise comparison of two of them (identified as 45 and 47) showed them to be genetically too similar ( $96\% \pm 5\%$  genetic similarity, Figure S3), in agreement with phenotypic data and

their common origin (Table S4), to be registered and marketed as distinct varieties. According to the most recent guidelines concerning the protection of new plant varieties, the similarity threshold to define two lettuce varieties as distinct is set at 96% [30]. The next step will be to integrate molecular data and morphological observations in order to select the best genotypes (positive selection) for evaluation as pre-commercial varieties. In particular, the eligible genotypes will be self-pollinated to multiply the seed so that their agronomic performance can be compared in different locations and periods of the year, and with the best commercial varieties already on the market.

For the remaining experimental populations (white box blot, Figure 4), an attempt could be made to increase their genetic uniformity through negative selection to remove the most genetically divergent individuals (i.e., outlier samples). Moreover, if necessary, the remaining genotypes can undergo a further selfing cycle aimed at reaching optimum values of homozygosity.

## 5. Conclusions

The results of this study demonstrate the advantages of mutual integration of traditional and biotechnological methods and show the added value that molecular markers can give to breeding programmes. We used microsatellite markers in three different steps of a conventional lettuce breeding program (see Figure 1) and demonstrated, firstly, the efficiency of SSR markers not only in selecting the best parental plants for crossing based on their observed homozygosity and dissimilarity values, but also in screening the resulting F1 progeny to distinguish between the offspring resulting from cross-pollination and those resulting from self-pollination. Furthermore, using the same SSR panel, we were able to act downstream of the breeding scheme to assess the uniformity of some pre-commercial cultivars. Our molecular assay could therefore also be used by seed firms to assess newly developed varieties for distinctiveness, uniformity and stability (DUS test), three major requirements for registering plant materials [6]. Finally, molecular characterisation of a new variety could also be used to register it in national or international varietal catalogues. In fact, the genotype or molecular profile of a registered variety can be crucial in solving cases of fraudulent practices, and in curbing plagiarism and unfair free-riding on the original plant breeder's time and investment [30].

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/11/916/s1>: Figure S1: Pairwise genetic similarity matrix of the 71 individuals analysed (in percentages) based on Rohlf's genetic similarity coefficient. High genetic similarity values are labelled in green, the low values in red, and intermediate values are coloured on a scale from green to red. The observed homozygosity values of the 71 putative parental lines are reported to the left of each ID name. Figure S2: Definition of the subgroup number of parental lines based on the SSR marker dataset. Mean  $\Delta K$  is calculated as  $|L''(K)|/(SD(L(K)))$ , following Evanno et al. [23]. The blue line represents the  $\Delta K$  values; Figure S3: Pairwise genetic similarity matrix of ten selected populations (in percentages) based on the Jaccard coefficient. The high genetic similarity values are labelled in green, the low values in red, and intermediate values are coloured on a scale from green to red; Table S1: Lettuce parental lines information, including ID of accessions, cultivar type of materials and subpopulation classification based on STRUCTURE analysis (1 = blue and 2 = red); Table S2: SSR primer tails and dyes. List of the primer tails used with their sequences and corresponding dyes; Table S3: Lettuce plant material information, including ID of accessions used in the crosses, total number of plants per programmed cross, number of informative marker loci, hybrid plants, selfed plants and unexpected genotypes, and the mean hybridisation values (in percentages) for all the programmed crosses; Table S4: Information about ten F3 selected populations, including ID of parental lines used in the crosses.

**Author Contributions:** Conceptualization, F.P. and G.B.; formal analysis, A.P. and G.G.; investigation, A.P. and G.G.; methodology, F.P. and G.B.; resources, A.P.; software, A.P. and G.G.; supervision, G.B.; Writing—Original Draft, A.P. and F.P.; Writing—Review and Editing, F.P. and G.B.

**Funding:** This project was financially supported by Blumen Group SpA (Bologna, Italy), within a Research Contract stipulated with DAFNAE, University of Padova (Italy): "Development of new varieties in horticultural species using molecular marker-assisted breeding methods" PI: Gianni Barcaccia.

**Acknowledgments:** This research was carried out in partial fulfillment of the Ph.D. Program of Alice Patella by taking advantage of a Doctoral Research Fellowship provided by Blumen Group SpA (Bologna, Italy).

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

## References

1. Wang, S.; Wang, B.; Liu, J.; Ren, J.; Huang, X.; Zhou, G.; Wang, A. Novel polymorphic EST-based microsatellite markers characterized in lettuce (*Lactuca sativa*). *Biologia* **2017**, *72*, 1300–1305. [[CrossRef](#)]
2. Nagata, R. Clip-and-wash Method of Emasculation for Lettuce. *HortScience* **1992**, *27*, 907–908. [[CrossRef](#)]
3. Barcaccia, G.; Falcinelli, M. *Genetica e Genomica*; Liguori Editore, S.r.l.: Naples, Italy, 2006; Volume 3.
4. Oliver, G.W. New methods of plant breeding. *J. Hered.* **1910**, *1*, 21–30. [[CrossRef](#)]
5. Prohens-Tomás, J.; Nuez, F. *Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae*; Springer Science & Business Media: Berlin, Germany, 2007; Volume 1.
6. 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. [[CrossRef](#)] [[PubMed](#)]
7. Palumbo, F.; Galla, G.; Vitulo, N.; Barcaccia, G. First draft genome sequencing of fennel (*Foeniculum vulgare* Mill.): Identification of simple sequence repeats and their application in marker-assisted breeding. *Mol. Breed.* **2018**, *38*, 122. [[CrossRef](#)]
8. Jha, N.; Jacob, S.; Nepolean, T.; Jain, S.; Kumar, M. SSR markers based DNA fingerprinting and its utility in testing purity of eggplant hybrid seeds. *Qual. Assur. Saf. Crop. Foods* **2016**, *8*, 333–338. [[CrossRef](#)]
9. Singh, D.; Singh, C.K.; Tomar, R.S.S.; Taunk, J.; Singh, R.; Maurya, S.; Chaturvedi, A.K.; Pal, M.; Singh, R.; Dubey, S.K. Molecular Assortment of Lens Species with Different Adaptations to Drought Conditions Using SSR Markers. *PLoS ONE* **2016**, *11*, e0147213. [[CrossRef](#)] [[PubMed](#)]
10. Rauscher, G.; Simko, I. Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. *BMC Plant Biol.* **2013**, *13*, 11. [[CrossRef](#)]
11. Truco, M.J.; Antonise, R.; Lavelle, D.; Ochoa, O.; Kozik, A.; Witsenboer, H.; Fort, S.B.; Jeuken, M.J.W.; Kesseli, R.V.; Lindhout, P.; et al. A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor. Appl. Genet.* **2007**, *115*, 735–746. [[CrossRef](#)]
12. Reyes-Chin-Wo, S.; Wang, Z.; Yang, X.; Kozik, A.; Arikait, S.; Song, C.; Xia, L.; Froenicke, L.; Lavelle, D.O.; Truco, M.-J.; et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* **2017**, *8*, 14953. [[CrossRef](#)]
13. Bull, C.T.; Goldman, P.H.; Hayes, R.; Madden, L.V.; Koike, S.T.; Ryder, E. Genetic Diversity of Lettuce for Resistance to Bacterial Leaf Spot Caused by *Xanthomonas campestris* pv. *vitiensis*. *Plant Health Prog.* **2007**, *8*, 11. [[CrossRef](#)]
14. Giesbers, A.K.J.; Pelgrom, A.J.E.; Visser, R.G.F.; Niks, R.E.; Ackerveken, G.V.D.; Jeuken, M.J.W. Effector-mediated discovery of a novel resistance gene against *Bremia lactucae* in a nonhost lettuce species. *New Phytol.* **2017**, *216*, 915–926. [[CrossRef](#)] [[PubMed](#)]
15. Macias-González, M.; Truco, M.J.; Bertier, L.D.; Jenni, S.; Simko, I.; Hayes, R.J.; Michelmore, R.W. Genetic architecture of tipburn resistance in lettuce. *Theor. Appl. Genet.* **2019**, *132*, 2209–2222. [[CrossRef](#)] [[PubMed](#)]
16. Mamo, B.E.; Hayes, R.J.; Truco, M.J.; Puri, K.D.; Michelmore, R.W.; Subbarao, K.V.; Simko, I. The genetics of resistance to lettuce drop (*Sclerotinia* spp.) in lettuce in a recombinant inbred line population from Reine des Glaces × Eruption. *Theor. Appl. Genet.* **2019**, *132*, 2439–2460. [[CrossRef](#)] [[PubMed](#)]
17. Hartman, Y.; Hooftman, D.A.P.; Uwimana, B.; Schranz, M.E.; Van De Wiel, C.C.M.; Smulders, M.J.M.; Visser, R.G.F.; Michelmore, R.W.; Van Tienderen, P.H. Abiotic stress QTL in lettuce crop–wild hybrids: Comparing greenhouse and field experiments. *Ecol. Evol.* **2014**, *4*, 2395–2409. [[CrossRef](#)] [[PubMed](#)]
18. Su, W.; Tao, R.; Liu, W.; Yu, C.; Yue, Z.; He, S.; Lavelle, D.; Zhang, W.; Zhang, L.; An, G.; et al. Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* **2019**. [[CrossRef](#)]
19. El-Esawi, M.A. Molecular Genetic Markers for Assessing the Genetic Variation and Relationships in *Lactuca* Germplasm. *Annu. Res. Rev. Biol.* **2015**, *8*, 1–13. [[CrossRef](#)]
20. Hayes, R.; Simko, I. Breeding lettuce for improved fresh-cut processing. *Acta Hort.* **2016**, *1141*, 65–76. [[CrossRef](#)]
21. Simko, I. Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J. Hered.* **2009**, *100*, 256–262. [[CrossRef](#)]

22. Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234. [[CrossRef](#)]
23. Nonis, A.; Scortegagna, M.; Nonis, A.; Ruperti, B. PRaTo: A web-tool to select optimal primer pairs for qPCR. *Biochem. Biophys. Res. Commun.* **2011**, *415*, 707–708. [[CrossRef](#)] [[PubMed](#)]
24. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [[PubMed](#)]
25. Hammer, Ø.; Harper, D.A.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.
26. 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. [[CrossRef](#)]
27. Yeh, F.C.; Yang, R.-C.; Boyle, T. *POPGENE Version 1.31*; University of Alberta: Edmonton, AB, Canada, 1999.
28. Palumbo, F.; Galla, G.; Martinez-Bello, L.; Barcaccia, G. Venetian Local Corn (*Zea mays* L.) Germplasm: Disclosing the Genetic Anatomy of Old Landraces Suited for Typical Cornmeal Mush Production. *Diversity* **2017**, *9*, 32. [[CrossRef](#)]
29. Feng, J.Y.; Li, M.; Zhao, S.; Zhang, C.; Yang, S.T.; Qiao, S.; Tan, W.F.; Qu, H.J.; Wang, D.Y.; Pu, Z.G. Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species *I-trifida* using RAD-seq. *BMC Plant Biol.* **2018**, *18*, 181. [[CrossRef](#)]
30. Lawson, C. *Plant Breeder's Rights and Essentially Derived Varieties: Still Searching for Workable Solutions*; European Intellectual Property Review 499: Griffith University Law School Research: Nathan, QLD, Australia, 2016; pp. 16–17.
31. Hand, P.; Kift, N.; McClement, S.; Lynn, J.; Grube, R.; Schut, J.; Van der Arend, A.; Pink, D. Progress towards mapping QTLs for pest and disease resistance in lettuce. In Proceedings of the Eucarpia Leafy Vegetables Conference, Centre for Genetic Resources, Wageningen, The Netherlands, 22–24 May 2003; pp. 31–35.
32. Šimko, I.; Pechenick, D.; McHale, L.; Truco, M.; Ochoa, O.; Michelmore, R.; Scheffler, B. Development of Molecular Markers for Marker-Assisted Selection of Dieback Disease Resistance in Lettuce (*Lactuca sativa*). *Acta Hort.* **2010**, *859*, 401–408. [[CrossRef](#)]
33. Sochor, M.; Jemelková, M.; Doležalová, I. Phenotyping and SSR markers as a tool for identification of duplicates in lettuce germplasm. *Czech J. Genet. Plant Breed.* **2019**, *55*, 110–119. [[CrossRef](#)]
34. Bhat, J.A.; Shivaraj, S.M.; Ali, S.; Mir, Z.A.; Islam, A.; Deshmukh, R. Genomic Resources and Omics-Assisted Breeding Approaches for Pulse Crop Improvement. In *Pulse Improvement*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 13–55.
35. Phing Lau, W.C.; Latif, M.A.; Rafii, M.Y.; Ismail, M.R.; Puteh, A. Advances to improve the eating and cooking qualities of rice by marker-assisted breeding. *Crit. Rev. Biotechnol.* **2016**, *36*, 87–98. [[CrossRef](#)]
36. FAO Site. Available online: <http://www.fao.org/faostat/en/#data/QC> (accessed on 11 September 2019).
37. Uwimana, B.; Smulders, M.J.M.; Hooftman, D.A.P.; Hartman, Y.; Van Tienderen, P.H.; Jansen, J.; McHale, L.K.; Michelmore, R.W.; Van De Wiel, C.C.M.; Visser, R.G.F. Hybridization between crops and wild relatives: The contribution of cultivated lettuce to the vigour of crop-wild hybrids under drought, salinity and nutrient deficiency conditions. *Theor. Appl. Genet.* **2012**, *125*, 1097–1111. [[CrossRef](#)]

