forests

MDPI

Article

Kinship Analysis and Pedigree Reconstruction of a Natural Regenerated Cork Oak (*Quercus suber*) Population

Bruna Mendes [1,†], Teresa Sampaio [2,†], Marta A. Antunes [1,‡], Hugo Magalhães [1,§], Filipe Costa e Silva [2], Carla Borges [3], Fernanda Simões [3], Ana Usié [1,4,*], Maria Helena Almeida [2] and António Marcos Ramos [1,4]

1 Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo (CEBAL)/Instituto Politécnico de Beja (IPBeja), 7801-908 Beja, Portugal; bruna.mendes@cebal.pt (B.M.); fc48389@alunos.fc.ul.pt (M.A.A.); hugo.magalhaes@hhu.de (H.M.); marcos.ramos@cebal.pt (A.M.R.)
2 Centro de Estudos Florestais (CEF), Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal; tsampaio@isa.ulisboa.pt (T.S.); filipecs@isa.ulisboa.pt (F.C.e.S.); nica@isa.ulisboa.pt (M.H.A.)
3 Instituto Nacional de Investigação Agrária e Veterinária (INIAV), 2780-157 Oeiras, Portugal; carla.borges@iniav.pt (C.B.); fernanda.simoes@iniav.pt (F.S.)
4 Mediterranean Institute for Agriculture, Environment and Development (MED), Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo (CEBAL), 7801-908 Beja, Portugal
* Correspondence: ana.usie@cebal.pt
† These authors contributed equally to this work.
‡ Current address: Centre for Ecology, Evolution and Environmental Changes (cE3c), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal.
§ Current address: Medical Faculty, Institute for Medical Biometry and Bioinformatics, Heinrich Heine University, 40225 Düsseldorf, Germany.

Citation: Mendes, B.; Sampaio, T.; Antunes, M.A.; Magalhães, H.; Costa e Silva, F.; Borges, C.; Simões, F.; Usié, A.; Almeida, M.H.; Ramos, A.M. Kinship Analysis and Pedigree Reconstruction of a Natural Regenerated Cork Oak (*Quercus suber*) Population. *Forests* **2022**, *13*, 226. https://doi.org/10.3390/f13020226

Academic Editors: María-Dolores Rey, Jesus V. Jorrin Novo and María Ángeles Castillejo

Received: 13 December 2021
Accepted: 31 January 2022
Published: 2 February 2022

Abstract: Cork oak (*Quercus suber* L.) is a valuable forest species in the western Mediterranean Basin due to its ecological value and the production of cork (a renewable natural material). Cork quality depends on the genetic background and cork oak environment, which has long been recognized. As no cork oak genetic trials with pedigree information were available, the inference of the genetic relatedness between individuals from molecular markers can potentially be applied to natural populations. This work aimed to investigate the potential of performing kinship prediction and pedigree reconstruction by SNP genotyping a natural cork oak population. A total of 494 trees located in Portugal were genotyped with 8K SNPs. The raw SNP set was filtered differently, producing four SNP sets that were further filtered by missing data, genotype frequency, and minor allele frequency. For each set, an identity by descent (IBD) matrix was generated to perform the relationship prediction, revealing from 22,114 to 23,859 relationships. Familial categories from the first to the third degree were able to be assigned. The feasibility of SNP genotyping for future studies on the kinship analysis and pedigree reconstruction of cork oak populations was demonstrated. The information produced may be used in further breeding and conservation programs for cork oak.

Keywords: cork oak; pedigree reconstruction; SNP; kinship

1. Introduction

Cork oak (*Quercus suber* L.) is a valuable forest species in the western Mediterranean Basin. Cork oak covers 23% of the national forest area in Portugal, being the dominant species of the "montado" agroforestry system ("dehesa" in Spain). In addition to its important ecological value, "montado" combines cork production with extensive agriculture or pasture and livestock. Portugal and Spain are the main cork producers worldwide (50 and 31% of total world cork production, respectively) [1]. Cork is a truly sustainable product as it is renewable and biodegradable. Cork harvesting occurs every nine years, by an environmentally friendly process, during which not a single tree is cut down. Cork stripping must be conducted when the phellogen, which has the property of self-regeneration after damage

or the peeling-off of the cork layer, is active in late spring and early summer [2,3]. It is this property that allows successive cork stripping from the same tree and sustainable cork production during a tree's life. Cork is a source of income in rural areas as raw material for the cork industry although dependent each year from cork production and quality [4]. However, over recent decades, cork oak woodlands have been under pressure due to a combination of biotic, abiotic and anthropogenic factors [5,6], which currently compromise the sustainability of the cork industry.

Cork oak is an evergreen long-lived tree (200–250 years), growing to a height of 15–20 m. Stem diameter at breast height can reach more than 300 cm. The earliest flowering and fructification occur at around 15–20 years of age, producing both annual and biennial acorns [7]. Cork oak is wind pollinated, having separate male and female flowers on the same tree with the initiation of the growth cycle in April or May, dependent on weather conditions. After flowering, pollinated flowers cease to grow in periods of water scarcity. Fertilization takes place in late August and acorn development is complete during autumn.

Benefiting from the European Union-funded programs for the reforestation of abandoned former agricultural lands, cork oak has been one of the most planted forest species in the Iberian Peninsula, both by direct sowing and by planting. The option of regenerating cork oak is influenced not only by the different socio-economic and ecological conditions but is also framed by the landscape perspective.

Defining and optimizing breeding strategies will contribute to the sustainable management of valuable forest species, such as cork oak, but require knowledge regarding the genetic control (i.e., heritability) of quantitative traits with economic relevance. The dependence of cork quality on genetic makeup and the environment in which the cork oak grows has long been recognized [7–9]. However, until now, the amount this is determined by genetic control is unknown. In populations with available pedigree information, heritability may be estimated by comparing the phenotypic variation within and between family groups. This can be achieved using a pedigree-based relationship matrix from a family-structured population, relying on data from progeny trials evaluated across sites and years. Although the cork oak progeny trials established in 1998 [10] are a valuable tool to obtain estimations of cork oak genetic parameters [11], trees in these trials are still too young to allow reproduction and cork assessments. In fact, industrial valuable cork can only be obtained from the third extraction cycle onwards, usually from trees over 40 years old.

Since the 1990s, methods using molecular markers have been used to infer genetic relatedness between individuals and to indirectly estimate quantitative genetic parameters, such as heritability values, from the regression of phenotypic similarity on the marker-based co-ancestry [12–17]. These approaches have received attention as they can potentially be applied directly to natural populations. However, in natural cork oak woodlands, unknown genealogical information combined with a long life cycle, overlapping generations, long reproductive cycle with a lengthy juvenile phase, and complex reproductive biology with self-incompatibility and high degree of heterozygoty [18] do not allow a traditional estimation of genetic parameters.

Therefore, kinship prediction and accurate pedigree inference methods are extremely valuable tools for breeding programs in plants. To achieve this type of information, it is important to develop methods for feasible relationship predictions and pedigree reconstructions. Several molecular studies for relatedness prediction were performed primarily with microsatellites that focused, for example, on the reconstruction of a pedigree to estimate heritability values in radiata pine [13], or improvement of the accuracy of genetic parameters and breeding values [19,20]. However, these studies analyzed populations for which there was some pedigree information. To our knowledge, no such studies have ever been carried out on cork oak.

Over the last decade, the fast development of high-throughput sequencing technologies greatly accelerated the sequencing of genomes for many species, which in turn opened up a new genomics era. This brought along a significant increase in our ability to iden-

tify SNPs, since they are more abundant and cost efficient. The recent sequencing of the cork oak genome [21] allowed the identification of large SNP datasets for this species, which then became available for use in different studies focusing on cork oak genomics research. However, for many genomics studies it is still unclear which filtering rules should be applied for the filtering of these SNP datasets, since studies with different goals may require specific approaches for SNP filtering. This need is further reinforced in kinship and pedigree studies, which may require specific filtering strategies, when compared, for example, with genome-wide association studies. The goal of this study was to perform a kinship analysis and investigate to what extent a pedigree of a natural regenerating cork oak population could be determined, using a set of over eight thousand SNPs genotyped by high-throughput sequencing technology, and filtered with four different approaches.

2. Materials and Methods

2.1. Cork Oak Population Analyzed

The study was conducted at a cork oak stand located in the Setúbal region in southern Portugal. The stand has approximately 440 ha of forested area, with over 300 ha covered by cork oak stands. These stands originated from natural regeneration and are around 100 years old. The altitude ranges between 2 and 25 m a.s.l and the orography is mainly flat, with slopes below 5%. The studied area is characterized by non-wet Psalmytic Regosols and Entic Podzols [22], of predominantly sandy type textures. The climate is of Mediterranean type, with hot and dry summers and wet mild winters. The long-term (1971–2000) mean annual temperature is 16.2 °C, and the mean annual precipitation is 715.9 mm. Precipitation occurs predominantly from autumn to early spring (October–April).

A total of 535 trees (Figure 1) were randomly selected from an area of around 7 ha, which corresponds to a density of 76 trees/ha. All trees were georeferenced and identified in the field with a unique identification tag. The individual tree age and microenvironmental conditions that caused heterogeneity in soil or water availability are unknown factors for this study site.

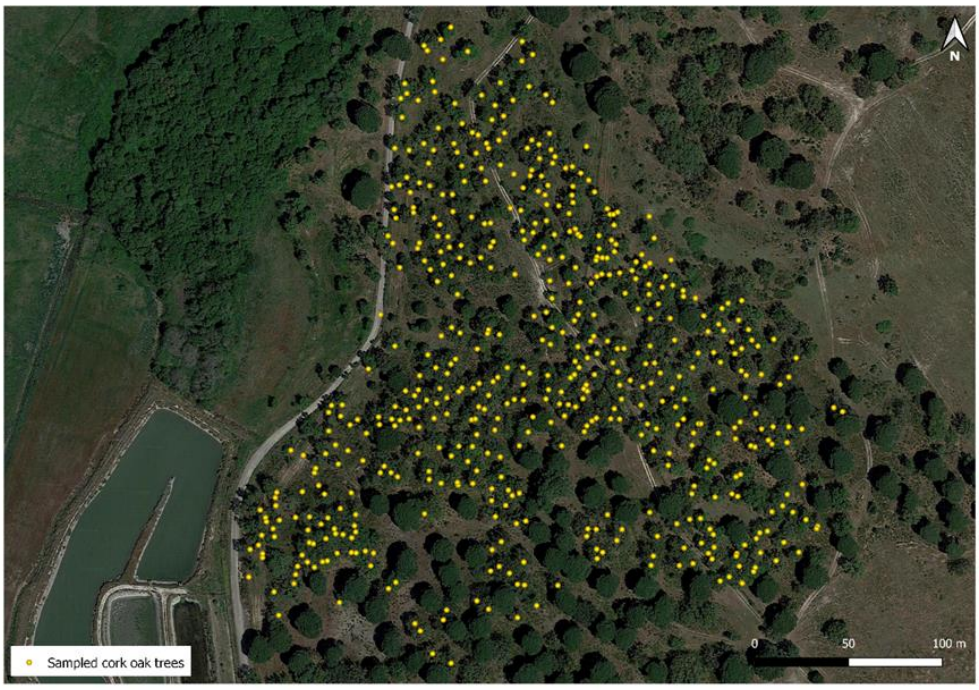

Figure 1. Map of the cork oak population analyzed. The trees included in the study are indicated with a yellow dot.

2.2. DNA Extraction

DNA was extracted from the leaf samples collected in each tree. A total of 200 mg of leaf tissue cut into small pieces and dried under vacuum for 15 min was used. The dried material was then reduced to powder using SpeedMill PLUS (Analytik Jena) (Analytic Jena) in innuSPEED Lysis Tube P (Analytic Jena). DNA extraction proceeded using the innuPREP Plant DNA Kit (Analytik Jena AG, Berlin, Germany) according to the manufacturer's protocol. DNA quality was checked on 0.8% agarose gel, and the DNA concentration was estimated using a NanoDrop ND2000 spectrophotometer (Thermo Scientific, Massachusetts, MA, USA). After checking for DNA quality, a total of 41 samples had to be discarded, due to insufficient DNA quality for downstream genotyping procedures.

2.3. SNP Genotyping and Calling

A total of 494 samples were genotyped using a novel technology based on target enrichment coupled with high-throughput sequencing, labeled as SeqSNP, which is available through LGC Biosearch Technologies. Genotyping was performed for a set of 8411 SNPs, selected from a larger cork oak SNP dataset generated with whole-genome resequencing data produced for 26 individuals. Sequencing was performed using the Illumina NextSeq 500 platform, with single-end reads that were 75 base pairs long.

SNP identification was performed using a bioinformatics pipeline that involved mapping the reads to the cork oak genome using Bowtie2 with default parameters [23]. Then, the mapped reads were sorted by coordinates and indexed using Samtools v.1.4 [24]. Lastly, Freebayes v1.02 [25] was used to perform the targeted SNP calling with the following parameters: –min-base quality 20; –min-supporting-allele-sum 10; –read-mismatch-limit 4; –min-coverage 4; –min-alternate-count 2; –report-genotype-likelihood-max; –exclude-unobserved-genotypes; –genotype-qualities; –ploidy 2; –min-alternate-fraction 0.166666666667; –report-monomorphic; –no-mnps; –no-complex; –mismatch-base-quality-threshold 10 [26]. Once we obtained the raw set of SNPs, individuals with more than 90% of missing SNPs genotypes were discarded. Then, the SNPs were filtered using several criteria which included minimum read coverage per SNP genotype (DP $\geq$ 8), the removal of SNPs with more than two alleles and indels, a minimum SNP quality (SNPQ) of 30 and a minimum genotype quality (GTQ) of 10.

In addition to this first set of SNPs, another set of SNPs was produced by applying the same procedures described above plus a threshold of maximum missing data (maxMD) of 50%, using VCFtools v.0.1.17 [26], and minimum genotype frequency (minGTF) of 2%. Then, each of these SNP sets was filtered using MAF (minor allele frequency) values of 1% and 5% to evaluate the effect on IBD estimation (and further relationship prediction) of using different MAF thresholds. Therefore, a total of four different SNP sets were generated:

SNP Set 1) DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, MAF—1%;

SNP Set 2) DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, MAF—5%;

SNP Set 3) DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, maxMD—50%, minGTF—2%, MAF—1%;

SNP Set 4) DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, maxMD—50%, minGTF—2%, MAF—5%.

Each of the SNP sets was then converted into the PLINK [25] file format (.map and .ped) using VCFtools with the "—plink" option.

2.4. IBD Estimation

IBD values were determined using the method-of-moments estimation implemented in PLINK v1.9 [27]. For IBD estimation, the PLINK flat files were used as input to PLINK using the "—genome" option. IBD estimation was performed for each of the four SNP sets generated, first with MAF of 1%, the default value, and then with MAF of 5%, using the "—maf 0.05" option. In the end, a total of four IBD matrices were generated, one for each set of SNPs. The IBD matrices were filtered with a PI_HAT (proportion of IBD estimated by PLINK) value of 0.1 to identify possibly related and unrelated pairs. We considered the

pairs with a PI_HAT equal or above 0.1 to be related, while the pairs with a PI_HAT below 0.1 were considered to be unrelated, in accordance with the expected mean of IBD portions for familial relationship categories outlined by Blouin and colleagues [28].

2.5. Pedigree Reconstruction

PRIMUS v.19 [29] was used for the identification of possible kinships and pedigree reconstruction. Since no information was available regarding the age of each cork oak tree or possible relationships between them, only the previously generated IBD matrixes, without PI_HAT filtering, were provided to PRIMUS using the default likelihood cut-off and the "—plink" option. Additionally, after using all the 493 trees for pedigree reconstructions, four subsets were created, one for each IBD matrix.

3. Results and Discussion

The application of different filtering parameters in the initial SNP output resulted in four sets of SNPs used for IBD estimation, each containing different numbers of SNPs. After checking for individuals with a high number of missing genotypes, one individual was discarded from the dataset. From the initial set of 8411 SNPs selected for genotyping, a total of 6016 SNPs were kept, after filtering for genotype quality, deep coverage per SNP genotype and SNP quality, and removing non-biallelic SNPs and indels. This represented a loss of 2225 SNPs from the initial target number of SNPs to be genotyped. To our knowledge, this was the first high-throughput SNP genotyping study on cork oak. SNP losses in genotyping were somewhat expected as the cork oak genome is highly heterozygous and assay optimization can be challenging to achieve.

The first SNP set comprising 6016 SNPs was filtered using MAF thresholds of 1% and 5%, which resulted in new SNP sets that included 6016 SNPs (MAF—1%) and 4752 SNPs (MAF—5%). Two additional SNP sets were produced by using two more filtering parameters, namely a maximum missing data value, per SNP, of 50% and a minimum genotype frequency of 2%, which resulted in a decrease in the total number of SNPs to 5093. Once again, the number of SNPs remained equal when the MAF threshold of 1% was applied and decreased to a total of 4362 SNPs with a MAF of 5%.

The observed decrease in the number of SNPs kept when the minimum number of individuals per SNP was applied could possibly be due to technical reasons, such as the presence of additional SNPs in the vicinity of the SNP targeted for genotyping, possibly affecting the efficiency of the SeqSNP methodology in some individuals. Moreover, the additional decrease observed when a minimum genotype frequency of 2% was applied may reflect the allelic and genotypic differences that may have occurred when the genotype frequencies observed in 26 samples (the ones that were used in the initial SNP discovery effort) were extrapolated for a population of 493 individuals, originating from a different geographic region than that of the initial set of 26 cork oak trees.

A total of four IBD matrices were estimated with PLINK, one for each SNP set, and then filtered with a PI_HAT value of 0.1 to predict how many possible relationships could be detected in the dataset comprising 493 individuals, considering the pairs (dyads) with a PI_HAT equal or above 0.1 to be possibly related and the ones with a PI_HAT below 0.1 to be unrelated. These results are summarized in Table 1.

Table 1. Number of related and unrelated dyads in four distinct IBD matrices.

	Nr of SNPs	Related Dyads (PI_HAT $\geq$ 0.1)	Unrelated Dyads (PI_HAT < 0.1)
IBD matrix 1	6016	12,334	108,945
IBD matrix 2	4752	11,882	109,397
IBD matrix 3	5093	9587	111,695
IBD matrix 4	4362	9208	112,071

SNP filtering criteria for each IBD matrix:

1. DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, MAF—1%;
2. DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, MAF—5%;
3. DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, maxMD—50%, minGTF—2%, MAF—1%;
4. DP $\geq$ 8, SNPQ $\geq$ 30, GTQ $\geq$ 10, bi-allelic only, no indels, maxMD—50%, minGTF—2%, MAF—5%;

The IBD matrix 1, generated with the larger set of 6016 SNPs, resulted in a higher number of possible relationships, with a total of 12,334 related dyads in a universe of 121,279 dyads. On the other hand, the IBD matrix 4, generated with a smaller set that included 4362 SNPs, resulted in a total of 9208 dyads that were possibly related, which was the lowest number of predicted relationships among the four IBD matrices. The number of related dyads showed a general trend of increasing along with the number of SNPs present in each set, even though an exception to this rule was also observed, since IBD matrix 2 displayed a higher number of related dyads when compared with IBD matrix 3, even though it had 249 fewer SNPs. Applying filters that removed SNPs for which more than 50% of the individuals had a missing genotype, together with discarding SNPs with genotype frequency less than 2% (IBD matrices 3 and 4), had a clear impact on decreasing the number of related dyads. These thresholds are frequently used to filter SNP datasets, for example, produced with SNP arrays containing tens, or even hundreds, of thousands of SNPs, prior to being used in several analyses, such as population genetics or genome-wide association studies. For pedigree reconstruction, however, SNPs genotyped in a small number of individuals from the overall population still seem to carry a sufficient amount of information useful for estimating relationships between the individuals. This means that SNPs that would likely be removed in the context of a genome-wide association study, for example, should be kept in the datasets used in studies like the one reported here. Therefore, perhaps different filtering approaches will be necessary to filter SNP datasets used in the context of studies focusing on kinship analysis and pedigree reconstruction.

Regarding the relationship categories, all IBD matrices showed evidence of the existence of identical genotypes and dyads related in first degree (Figure 2). Dyads possibly related to the second- and third-degree were also detected, however the fact that they are overlapped did not allow their differentiation.

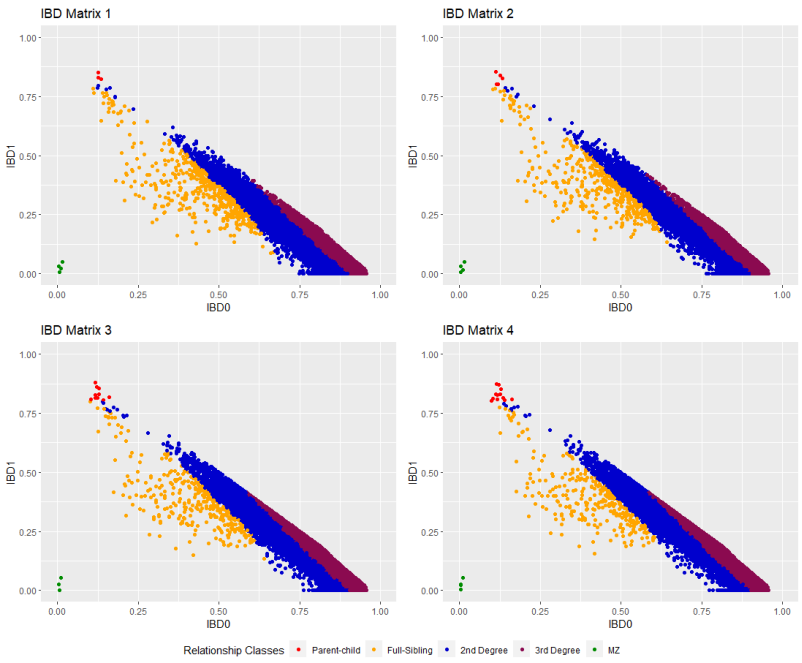

Figure 2. Distribution of relationship categories determined for the four IBD matrices.

Considering the population being analyzed in this study, a self-regeneration cork oak population for which no family information was available, PRIMUS was the best software tool available, since it does not have the limitation of other pedigree-reconstruction methods that require information about age and other population related data. The results showed that PRIMUS predicted more relationships than previously anticipated in Table 1, with filtering of the IBD matrices using expected mean of IBD portions for familial relationship categories, using the approach described by Blouin and colleagues [28] (Table 2).

Table 2. Kinship assignments obtained for each IBD matrix, indicating the 1st-, 2nd- and 3rd-degree relationships, distantly related individuals and identical genotypes (MZ).

	IBD Matrix 1	IBD Matrix 2	IBD Matrix 3	IBD Matrix 4
1st Degree	502	436	330	306
Parent–child	3	5	10	12
Full sibling	499	431	320	294
2nd Degree	10,362	9502	6400	5866
3rd Degree	10,887	13,078	13,195	13,534
Distant	609	839	2275	2404
MZ	4	4	4	4
Total Related	22,364	23,859	22,204	22,114
Unrelated	98,914	97,419	99,074	99,164

The highest number of related dyads and familial classes assigned was detected for IBD matrix 2, with a total of 23,859 dyads. Although, when comparing the IBD matrices, IBD matrix 1 had the largest number of dyads related to the first and second degrees, with 502 and 10,887 dyads, respectively, and the lowest number of dyads related to the third degree and distant related, with 10,887 and 609 dyads, respectively (Table 2). Therefore, this increased number of SNPs was associated with a higher number of related dyads in the first and second degrees, even though some of the SNPs had characteristics that would likely have them discarded from datasets to be used in studies with different goals, such as genome-wide association studies.

The first-degree relationships include parent–child (PC) and full sibling (FS) relationships. The largest number of PC relationships was identified using IBD matrix 4, with a total of 12 possible parents. The increase in PC relationships using IBD matrix 4 was linked to a decrease in the number of SNPs. Parent–child relationship assignments can potentially be performed more efficiently with a lower number of SNPs, in comparison with the second- and third-degree relationships. Nevertheless, future studies will be needed to further our understanding of the impact of the number of SNPs used to estimate familial relationships in cork oak.

The results observed for IBD matrices 3 and 4 were also quite similar for the first degree and third degree relationships (Table 2). However, matrix 3 showed a higher number of second degree relationships, when compared with matrix 4, while the latter displayed more distant related dyads. The filtering rules applied to produce the SNP sets used in each matrix were the same except for the MAF value.

In general, keeping SNPs with higher amounts of missing data in the dataset (IBD matrices 1 and 2) was associated with the identification of more first- and second-degree relationships, and with less distant relationships. Thus, even though the total number of SNPs influenced the ability to detect relationships, the type of SNPs that were used also had a clear impact on the type of relationships that were detected.

A total of 34,851 different relationships were predicted with the analyses of all IBD matrices, of which 12,093 relationships were commonly predicted (same pairs of individuals with the same type of relationship) by all four IBD matrices (290 dyads related to the first

degree, 5330 in second degree and 6466 in third degree). These results are illustrated in Figure 3. Once again, the SNP set used to determine each IBD matrix had a clear effect, since only 34.70% of all relationships determined were shared by all matrices. Moreover, for IBD matrix 1, which had the largest number of SNPs, a total of 2295 unique relationships were detected, which indicated that the increased number of SNPs used carried a sufficient volume of information for the identification of these unique relationships.

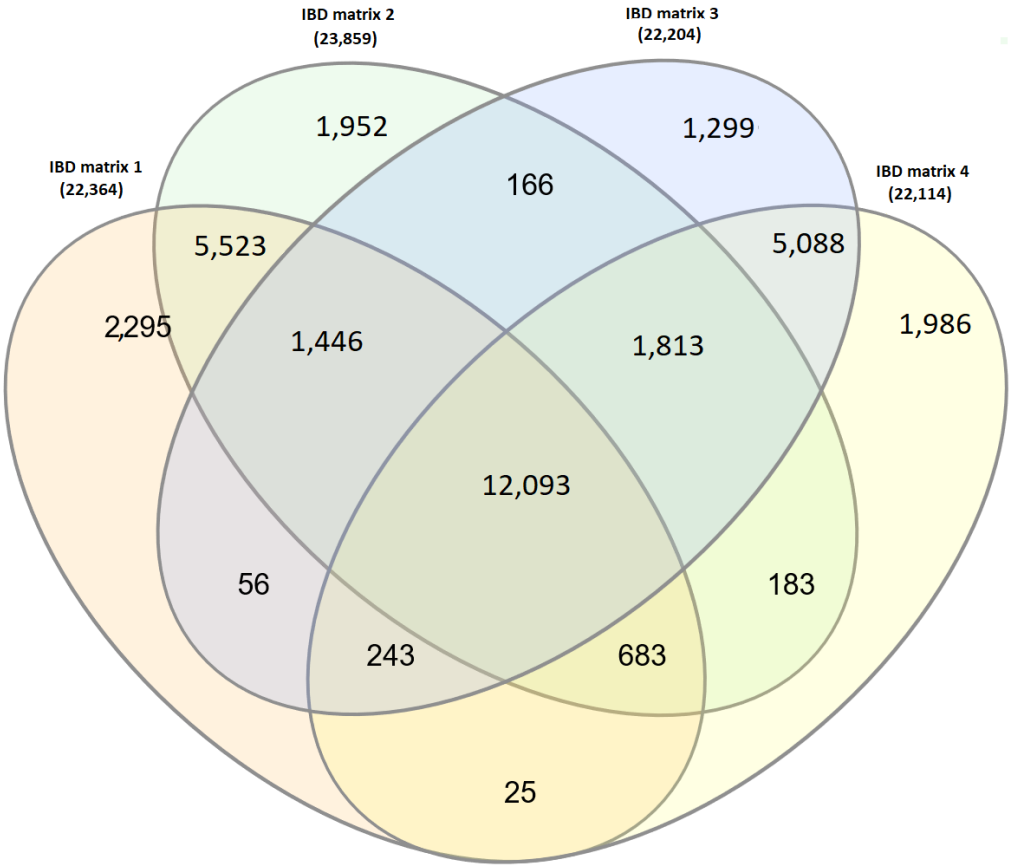

Figure 3. Venn diagram representing the relationships shared between the four IBD matrices. The relationships identified in a single IBD matrix are also represented.

In this study, we analyzed a considerable number of trees derived from a natural regenerating cork oak population, whose breeding and reproductive process occurred without human intervention. The large complexity present in the population was most likely the reason why it was not possible to establish a complete population structure and fully reconstruct the pedigree for all trees. For example, in this population, a tree can simultaneously be a half-sibling of another tree, and the parent, or even offspring, of another tree. This was evidenced by the average number of relationships per sample, which ranged from 83 to 97 among all IBD matrices. Moreover, the largest number of relationships assigned to an individual was 425, using IBD matrix 1 as the reference, and a total of 32 trees had at least 200 relationships. The average number of relationships established remained high even for the IBD matrices where fewer SNPs were used (Figure 4).

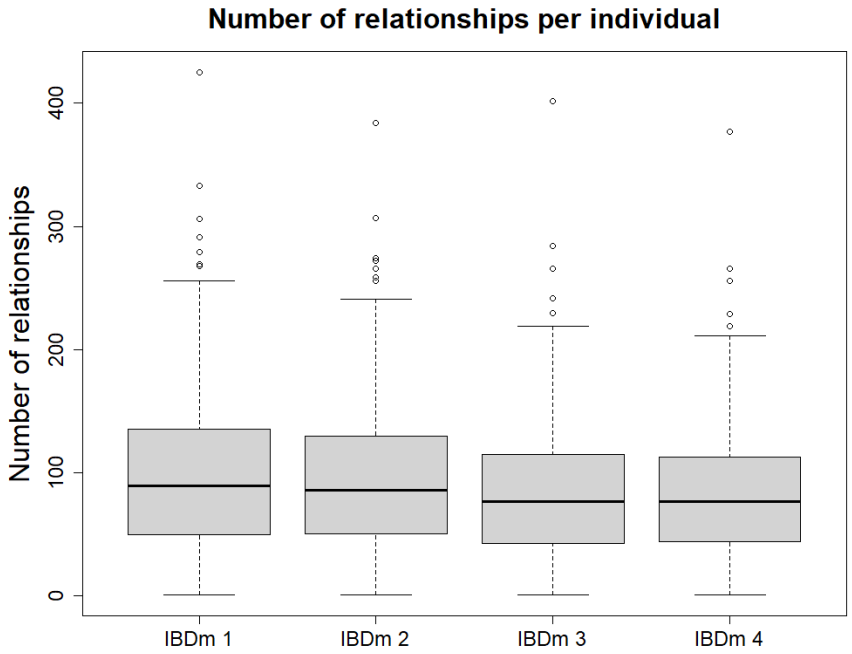

Figure 4. Boxplot of the number of relationships per individual for the 4 IBD matrices.

The high number of relationships added a significant amount of complexity to the PRIMUS pedigree reconstruction, which was not originally designed to analyze populations with such complexity and in the complete absence of any familial information. In order to circumvent this problem, a subset of the dyads only related only to first degree for each IBD matrix was created, which was subsequently used for pedigree reconstruction.

As expected, the number of individuals and the type of relationships varied according to the IBD matrix used. Nevertheless, due to the complexity of the data associated with the lack of information, only smaller families were identified with all four IBD matrices. These families were then enriched manually with kinships in second-degree dyads that were not included within the families identified. The pedigrees for these families were then determined (Figures 5 and 6)

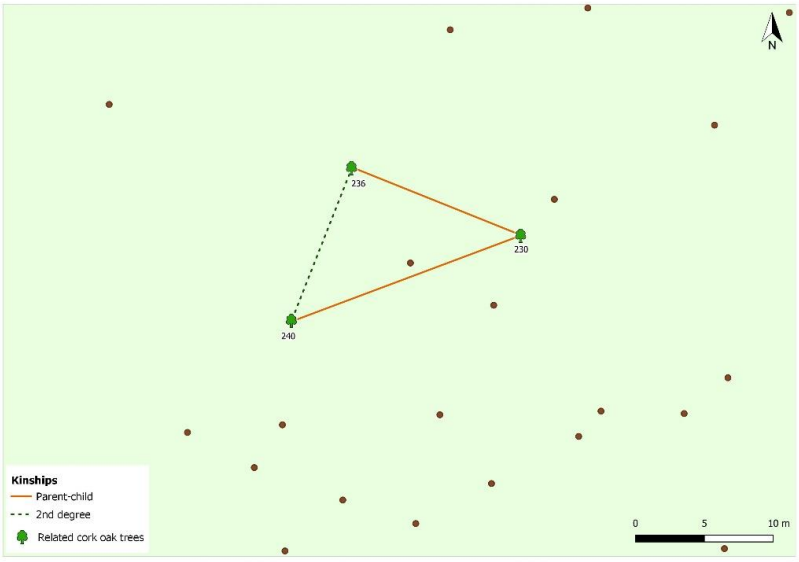

Figure 5. Spatial distribution of the reconstructed pedigree for a cork oak family comprising three individuals. The parent–child and second-degree relationships are illustrated.

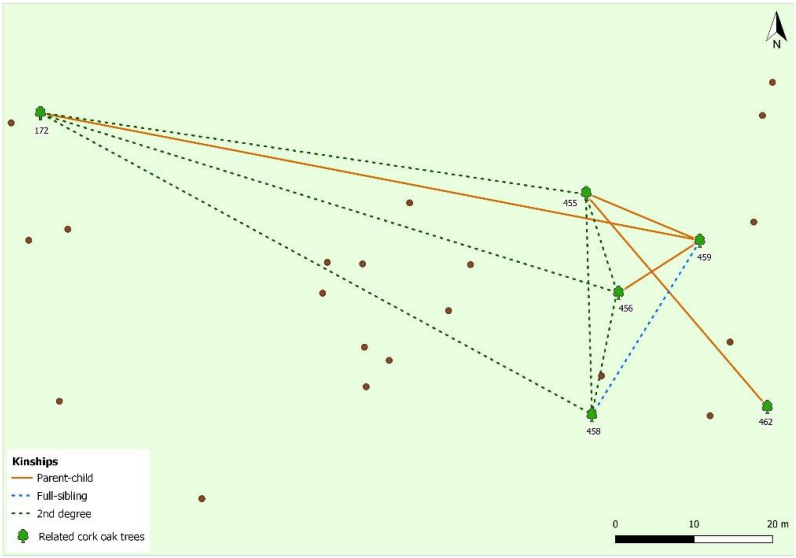

Figure 6. Spatial distribution of the reconstructed pedigree for a cork oak family comprising six individuals. The parent–child, full-sibling and second-degree relationships are illustrated.

These examples indicate that smaller pedigrees, at the family level, can be reconstructed from this type of analysis. However, it will always be very challenging to determine full pedigrees, considering that natural regeneration cork oak populations have characteristics that cannot be controlled, which will hinder the ability to determine whole pedigrees. For cork oak, the main vector for pollination is the wind, which is the reason why in these populations pollination cannot be controlled, which subsequently increases the complexity of pollination possibilities. Pollen can travel considerable distances and successfully pollinate the cork oak trees located in the stand analyzed in this study. In these cases, it will be impossible to determine the male parent of the cork oak tree. In addition, the cork oak stand analyzed in this study is more than 100 years old, a time window that opens the possibility that some of the trees have already been removed as a result of disease or died from natural processes.

Finally, a total of four pairs of identical genotype oaks were commonly identified using each IBD matrix. Identical genotype oaks were located very close to each other, either originating from early sprouts (several decades ago) or multi-seeded acorns resulting from mitotic reproduction of the initial zygote [30–32]. Animals such as squirrels could excise the acorns from the soil (or a seed from it) and move them from their original locations to a close spot, explaining the existing distance between each identical genotype oak [33–35]. However, when comparing the identical genotypes, which have long been thought to be genetically "identical", small differences were found; they were not 100% identical. To our knowledge, this is the first work reporting cork oak trees with almost identical genotypes, although it was observed that double-seeded cork oak acorns germinate two viable plants (Ramos, personal communication). Identical genotypes in cork oak trees or germinated plants still need further studies in order to clarify into what extent identical genotypes may co-exist in the species.

4. Conclusions

This study documents the first time a set comprising thousands of SNPs was genotyped in a large natural regenerated cork oak population, aiming to investigate if a kinship analysis and pedigree reconstruction would be feasible. The successful identification of kinships and establishment of some families' pedigree indicate the potential of this approach for future studies. Pedigree information can be a valuable tool for future management strategies of cork oak populations, including future cork oak breeding schemes. However, a full pedigree reconstruction exclusively based on geographical and genotype information

may prove to be difficult for large populations having convoluted familial relationships and complex reproductive and seed dispersal systems. The inclusion of additional information, such as the age of each tree or any known familial relationships, would be helpful in improving the accuracy and extent of pedigree restoration.

We have proved that, by using a high number of single-nucleotide markers, it is possible to predict pedigrees from cork oak trees from a natural cork oak stand resulting only from natural regeneration over a period of 100 years. Knowing only the georeference from each tree and their respective genotypes, it was possible to predict and identify several kinship relationships for small families. The high number of relationships found for each individual is a balance of tree age, reproductive stag and seed production (thousands of seeds by tree). However, in cork oak, natural regeneration is rare [1]. Seedling establishment is deficient and tree recruitment is often not sufficient to compensate for natural or induced mortality [36,37], being a major concern for forest management and cork producers.

Cork oak stand management based on the species genetic variability assumes particular importance due to the economic impact of the reduction in cork production by cork oak trees. Pedigree reconstruction in cork oak stands may be seen as a new tool contributing to the improvement of knowledge in this field. However, additional research is needed in order to assess its feasibility for cork oak species.

Another potential application of pedigree reconstruction using genome span molecular markers is the breed without breeding approaches. These approaches for cork oak may take advantage of wind pollination systems that minimize full-siblings crossing, the absence of self-pollination, natural progeny estimation and genetic evaluation at the landscape level, emphasizing adaptive traits and their respective interaction with environmental conditions. Molecular markers data may be complemented by phenotype characteristics and tested with performance prediction models. Lstibůrek [38] concluded that breed without breeding strategies provides an effective and economically feasible method to breed outcrossing forest tree species. We, therefore, forecast the utility of these methods for cork oak stand management and future breeding and conservation based on molecular-based pedigree reconstruction.

Author Contributions: A.M.R. and M.H.A. conceived and designed the study; F.S. and C.B. performed the laboratorial experiments; B.M., M.A.A., H.M. and A.U. performed bioinformatics analyses of the data; A.M.R. coordinated the bioinformatics analyses; B.M., T.S., F.C.e.S., A.U., M.H.A. and A.M.R. interpreted the results; B.M., T.S., A.U., F.S. and A.M.R. prepared the manuscript; M.A., F.S., F.C.e.S., A.U. and M.H.A. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FCT—Foundation for Science and Technology, Portugal, under the projects CEF (UIDB/00239/2020) and MED (UIDB/05183/2020). Funding for these research activities was also provided by Amorim Florestal. Teresa Sampaio was funded by the Portuguese Foundation for Science and Technology through a doctoral grant under the SUSFOR Doctoral Programme (PD/BD/52402/2013).

Data Availability Statement: The list of SNPs genotyped is openly available. These data can be found here: https://bitbucket.org/agbi_cebal/cork-oak-snps/.

Acknowledgments: The authors thank Joaquim Gomes and Daniel Gaspar for their support in the field work.

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

References

1. *APCOR's Cork Yearbook 19/20*; Portuguese Cork Association: Santa Maria de Lamas, Portugal, 2020; p. 128. Available online: https://www.apcor.pt/wp-content/uploads/2019/12/boletim_estatistico_apcor_2019.pdf (accessed on 20 January 2022).
2. Pereira, H. *Cork: Biology, Production and Uses*; Elsevier: Amsterdam, The Netherlands, 2011.
3. Caritat, A.; Gutiérrez, E.; Molinas, M. Influence of weather on cork ring width. *Tree Physiol.* **2000**, *20*, 893–900. [CrossRef] [PubMed]

4. Bugalho, M.; Plieninger, T.; Aronson, J.; Ellatifi, M.; Crespo, D.G. Open woodlands: A diversity of uses (and overuses). In *Cork Oak Woodlands on the Edge: Conservation, Adaptive Management and Restoration*, 1st ed.; Aronson, J., Pereira, J.S., Pausas, Eds.; Island Press: New York, NY, USA, 1999; pp. 33–47.
5. Costa, A.; Pereira, H.; Madeira, M. Analysis of spatial patterns of oak decline in cork oak woodlands in Mediterranean conditions. *Ann. For. Sci.* **2010**, *67*, 204. [CrossRef]
6. Camilo-Alves, C.; Clara, M.; Ribeiro, N. Decline of Mediterranean oak trees and its association with Phytophthora cinnamomi: A review. *Eur. J. For. Res.* **2013**, *132*, 411–432. [CrossRef]
7. Natividade, J.V. Cortiças: Contribuição para o estudio do melhoramento da qualidade. *J. Oliveira Jr.* **1934**, *1*, 1–143.
8. Pereira, H. Chemical composition and variability of cork from *Quercus suber* L. *Wood Sci. Technol.* **1988**, *22*, 211–218. [CrossRef]
9. Conde, E.; Cadahía, E.; García-Vallejo, M.C.; Fernández de Simón, B. Polyphenolic composition of *Quercus suber* cork from different Spanish provenances. *J. Agric. Food Chem.* **1998**, *46*, 3166–3171. [CrossRef]
10. Varela, M.C. The EUFORGEN Quercus suber Network and the research projects for the evaluation of genetic variability of cork oak. In Proceedings of the Mediterranean Oaks Network, Report of the First Meeting, Antalya, Turkey, 12–14 October 2000; International Plant Genetic Resources Institute: Rome, Italy, 2000; Volume 472, p. 6.
11. Sampaio, T.; Gonçalves, E.; Faria, C.; Almeida, M.H. Genetic variation among and within *Quercus suber* L. populations in survival, growth, vigor and plant architecture traits. *For. Ecol. Manag.* **2021**, *483*, 118715. [CrossRef]
12. Coltman, D.W. Testing marker-based estimates of heritability in the wild. *Mol. Ecol.* **2005**, *14*, 2593–2599. [CrossRef]
13. Kumar, S.; Richardson, T.E. Inferring relatedness and heritability using molecular markers in radiate pine. *Mol. Breed.* **2005**, *15*, 55–64. [CrossRef]
14. Lynch, M.; Ritland, K. Estimation of pairwise relatedness with molecular markers. *Genetics* **1999**, *152*, 1753–1766. [CrossRef]
15. Ritland, K. A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **1996**, *50*, 1062–1073. [CrossRef]
16. Ritland, K. Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* **2000**, *9*, 1195–1204. [CrossRef]
17. Thomas, S.C. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2005**, *360*, 1457–1467. [CrossRef]
18. Álvarez, R.; Alonso, P.; Cortizo, M.; Celestino, C.; Hernández, I.; Toribio, M.; Ordás, R.J. Genetic transformation of selected mature cork oak (*Quercus suber* L.) trees. *Plant Cell Rep.* **2004**, *23*, 218–223. [CrossRef]
19. Doerksen, T.K.; Herbinger, C.M. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genet. Genomes* **2010**, *6*, 591–600. [CrossRef]
20. Vidal, M.; Plomion, C.; Harvengt, L.; Raffin, A.; Boury, C.; Bouffier, L. Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. *Tree Genet. Genomes* **2015**, *11*, 105. [CrossRef]
21. Ramos, A.M.; Usié, A.; Barbosa, P.; Barros, P.M.; Capote, T.; Chaves, I.; Simões, F.; Abreu, I.; Carrasquinho, I.; Faro, C.; et al. The draft genome sequence of cork oak. *Sci. Data* **2018**, *5*, 1–12. [CrossRef]
22. IUSS Working Group WRB. *World Reference Base for Soil Resources*, 2nd ed.; World Soil Resources Reports 103; FAO: Rome, Italy, 2006; p. 103.
23. Langmead, B.; Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
24. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
25. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907. (q-bio.GN).
26. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
27. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
28. Blouin, M.S. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol. Evol.* **2003**, *18*, 503–511. [CrossRef]
29. Staples, J.; Qiao, D.; Cho, M.H.; Silverman, E.K.; Nickerson, D.A.; Below, J.E.; University of Washington Center for Mendelian Genomics. PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet.* **2014**, *95*, 553–564. [CrossRef]
30. Garrison, W.J.; Augspurger, C.K. Double- and single-seeded acorns of bur oak (*Quercus macrocarpa*): Frequency and some ecological consequences. *Bull. Torrey Bot. Club* **1983**, *110*, 154–160. [CrossRef]
31. McEuen, A.B.; Steele, M.A. Atypical acorns appear to allow seed escape after apical notching by squirrels. *Am. Midl. Nat.* **2005**, *154*, 450–458. [CrossRef]
32. Costa, A.; Pereira, H. Montados e sobreirais: Uma espécie, duas perspectivas. In *Os Montados. Muito para além das Árvores. Árvores e Florestas de Portugal*, 1st ed.; Sande Silva, J., Ed.; Público, Comunicação Social SA & Fundação Luso-Americana para o Desenvolvimento: Lisboa, Portugal, 2007; pp. 17–37.
33. Zhang, M.; Dong, Z.; Yi, X.; Bartlow, A.W. Acorns containing deeper plumule survive better: How white oaks counter embryo excision by rodents. *Ecol. Evol.* **2014**, *4*, 59–66. [CrossRef]

34. Steele, M.A.; Turner, G.; Smallwood, P.D.; Wolff, J.O.; Radillo, J. Cache management by small mammals: Experimental evidence for the significance of acorn-embryo excision. *J. Mammal.* **2001**, *82*, 35–42. [CrossRef]
35. Xiang, J.; Li, X.; Yi, X. One acorn produces two seedlings in Chinese cork oak *Quercus variabilis*. *Plant Signal. Behav.* **2019**, *14*, e1654817. [CrossRef] [PubMed]
36. Acácio, V.; Holmgren, M.; Moreira, F.; Mohren, G.M.J. Oak persistence in Mediterranean landscapes: The combined role of management, topography, and wildfires. *Ecol. Soc.* **2010**, *15*. [CrossRef]
37. Zavala, M.A.; Zamora, R.; Pulido, F.; Blanco, J.A.; Imbert, J.B.; Marañón, T.; Castillo, F.J.; Valladares, F. Nuevas perspectivas en la conservación, restauración y gestión sostenible del bosque mediterráneo. In *Ecología del Bosque Mediterráneo en un Mundo Cambiante*, 1st ed.; Valladares, F., Egraf, S.A., Eds.; Ministerio de Medio Ambiente: Madrid, Spain, 2004; pp. 509–529.
38. Lstibůrek, M.; El-Kassaby, Y.A.; Skrøppa, T.; Hodge, G.R.; Sønstebø, J.H.; Steffenrem, A. Dynamic gene-resource landscape management of Norway spruce: Combining utilization and conservation. *Front. Plant Sci.* **2017**, *8*, 1810. [CrossRef] [PubMed]