



Article

# Genomic Variability Survey in *Ilex aquifolium* L., with Reference to Four Insular Populations from Eastern Europe

Ciprian Valentin Mihali <sup>1,2,\*,†</sup>, Alexandru Eugeniu Mizeranschi <sup>1,3,†</sup>, Daniela Elena Ilie <sup>1</sup>, Ludovic-Toma Cziszter <sup>4</sup>, Radu Ionel Neamț <sup>1</sup>, Andreea Ștefania Anton <sup>1</sup>, Endre Mathe <sup>2,5</sup>, Bence Pecsenye <sup>6</sup>, Viviane Beatrice Bota <sup>2,7,8</sup> and Violeta Turcuș <sup>2,8</sup>

- <sup>1</sup> Research and Development Station for Bovine Arad, 310059 Arad, Romania; alexandru.mizeranschi@scdcbarad.ro (A.E.M.); daniela.ilie@scdcbarad.ro (D.E.I.); radu.neamt@scdcbarad.ro (R.I.N.); andreea.anton@scdcbarad.ro (A.Ș.A.)
  - <sup>2</sup> Faculty of Medicine, “Vasile Goldiș” Western University from Arad, 310025 Arad, Romania; mathe.endre@uvvg.ro (E.M.); bota.viviene@uvvg.ro (V.B.B.); turcus.violeta@uvvg.ro (V.T.)
  - <sup>3</sup> Institute for Advanced Environmental Research, West University of Timisoara, 300086 Timisoara, Romania
  - <sup>4</sup> Faculty of Bioengineering of Animal Resources, University of Life Sciences “King Mihai I” from Timisoara, 300645 Timisoara, Romania; ludoviccziszter@usvt.ro
  - <sup>5</sup> Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary
  - <sup>6</sup> Institute of Nutrition Science, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary; pecsenye.bence@agr.unideb.hu
  - <sup>7</sup> Faculty of Biology, “Alexandru Ioan Cuza” University of Iași, 700505 Iași, Romania
  - <sup>8</sup> Centre for Mountain Economy (CE-MONT), National Institute for Economic Research “Costin C. Kirițescu”, Romanian Academy, 725700 Suceava, Romania
- \* Correspondence: ciprian.mihali@scdcbarad.ro  
† These authors contributed equally to the study.



**Citation:** Mihali, C.V.; Mizeranschi, A.E.; Ilie, D.E.; Cziszter, L.-T.; Neamț, R.I.; Anton, A.Ș.; Mathe, E.; Pecsenye, B.; Bota, V.B.; Turcuș, V. Genomic Variability Survey in *Ilex aquifolium* L., with Reference to Four Insular Populations from Eastern Europe. *Int. J. Mol. Sci.* **2024**, *25*, 13593. <https://doi.org/10.3390/ijms252413593>

Academic Editor: Pedro Martínez-Gómez

Received: 20 November 2024

Revised: 13 December 2024

Accepted: 17 December 2024

Published: 19 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Cosmopolitan in the western areas of Europe as well as on other continents, the *Ilex* genus is interesting for its genetic, phenotypic, and biogeographic variabilities. Its insular/local distribution, according to existing data on the periphery of the central and southern European areas, represents a suitable case study with reference to the adaptive plasticity or acclimatization of the *Ilex aquifolium* L. species to new climatic conditions. The aim of the present study was to analyze the genetic variability at the genome level in four insular populations of *Ilex aquifolium* L., i.e., in three spontaneous populations from Romania (RO), Serbia (SR), and Bulgaria (BG) and a cultivated population from Hungary (HU). According to the obtained results, the most genetically similar populations among the four considered in this study were those from SR and RO. Genetic variation overlapped genes that were generally associated with metabolic regulation/transport factors, water, and abiotic stress factors. The analysis of single-nucleotide polymorphisms (SNPs) at the levels of the chloroplast and mitochondrion, from the point of view of their distributions at the gene level, identified two clusters: one that includes the native populations (BG, SR, and RO) and a second one including the cultured population from HU.

**Keywords:** *Ilex aquifolium* L.; genome analysis; short reads; Eastern Europe; Romanian insular population

## 1. Introduction

*Ilex aquifolium* L. is a dioecious, woody, evergreen species shrub or tree with an average height of up to 10 m, belonging to the Aquifoliaceae family [1]. More than 660 subspecies are described within the genus *Ilex* [2]. Currently, the Aquifoliaceae family is placed in the order Aquifoliales. Initially, the classification of the genus was achieved through studies of floral unit–cyme analysis (years 1949/50) [3,4] and, later, by molecular phylogeny studies (in the 2000s) [5].

The species/subspecies of the genus *Ilex* have diverse uses, being sources for the production of some tea drinks consumed in South America (yerba mate—*I. paraguariensis* A.St.-Hil.—and guayusa—*I. guayusa* Loes., with caffeine contents), China ('kudingcha'—bitter spike tea—*I. latifolia* Thunb., and *I. kaushue* S.Y. Hu, without caffeine or other methylxanthines [6]), and the United States (yaupon—*I. vomitoria* Sol. ex Aiton, with caffeine, and *I. opaca* Aiton—Appalachian tea, without caffeine) [7]. They are sources in the production of medicines (active principles used in traditional Chinese medicine) with protective effects on the cardiocirculatory system, as well as anti-tumor, antioxidant, anti-inflammatory, lipid metabolism regulatory, anti-microbial, and anti-diabetic effects [8]; similar studies are also presented in the cases of South American species [9]. *Ilex aquifolium* L. has an ornamental importance in urban landscaping and is socio-culturally associated with the celebration of some holidays [10].

The species of the *Ilex* genus spread from the tropics to the temperate zones, the area covering the American continents (the native residence of the genus *Ilex* is in South America, and the native or invasive residence—invasive in the Pacific Northwest wildlands—is in North America, *Ilex aquifolium* L.) [11], North Africa (Tunisia and Morocco), Central East Asia, Australia, and New Zealand [12]. It is present in almost all of Europe, originating in the Mediterranean basin, as in the *Ilex*–*Taxus* complex [13]. This species is mentioned in the 2013 IUCN Red List of Threatened Species, listed as a least-concern species. In agreement with the IUCN, on the European continent, *Ilex aquifolium* L. is present as a native resident species in the Scandinavian countries, the British Isles, from Western Europe descending to the South in Spain, and on the Balkan peninsula, in Serbia, Bulgaria, and Greece. The same list mentions that in Hungary, *Ilex aquifolium* L. exists as a cultivated species, while in Romania, it has an uncertain status regarding the native residence [14]. At the same time, other scientific sources mention the native uncertain residence of the Romanian insular population or the likelihood that it is a native resident species in Romania [15].

The first records about the *Ilex aquifolium* L. presence in the territory of Romania, near the village of Zimbru, were presented in 1893 as a part of an inventory study of forest species, where *Ilex aquifolium* L. was mentioned together with other woody species, such as *Acer platanoides* L., *A. pseudoplatanus* L., *Fagus sylvatica* L., and *A. tegmentosum* Maxim [16,17]. During the 90's *Ilex aquifolium* L. was presented in Romania as a protected species [18]. As a consequence, in 1999, it was preserved within the Dosul Laurului Nature Reserve (0.33 km<sup>2</sup>), with GPS coordinates 46°23055.500 N, 22°22050.900 E, located in the Codru-Moma Mountains, Zimbru population, Gurahonț Commune, Arad County [19].

There are studies of molecular analysis and the phylogeny of the *Ilex*-genus species at the levels of the chloroplast (*matK*, *rbcL*, *atpB-rbcL*, *trnL-F*, *psbA-trnH*, and *rpl32trnL*) and nuclear sequences (*ITS* and *nepGS*) [20]. Genomic analysis and de novo reconstruction, with annotations of the chloroplast genome based on seven different species belonging to the *Ilex* genus from China (not including *Ilex aquifolium*) [21], as well as phylogenetic analysis between species from the Hong Kong region were carried out [22]. The analysis/construction of a complete chloroplast genome and phylogenetic analysis studies were carried out on an *Ilex* 'Beryl' hybrid between *Ilex cornuta* and *Ilex latifolia* [23] and on *Ilex* 'Tall Boy', a hybrid between *Ilex aquifolium* and *Ilex latifolia*, respectively [24]. In DNA methylation studies (on *Ilex aquifolium* L.) based on methylation-sensitive amplified polymorphisms at the level of the entire genome, a decreasing methylation tendency was observed between leaves without thorns and those with thorns, with a correlation between the DNA methylation mechanism and phenotypic plasticity [25].

A popular option for genomic DNA analysis is represented by SNPs (single-nucleotide polymorphisms). Genomic analyses by intraspecific SNP markers were performed in different studies with the aim of discriminating between different cultivars in *Lactuca sativa* L. [26], in biodiversity phylogeny studies between different yeast strains [27], or in association studies between genetic structure and productive morphological characteristics in species of economic interest [28]. The literature also mentions studies of the phylogeny and distribution of SNPs correlated with their effects at the gene level in cultivated trees

(*Prunus* sp.) [29] or in economic species with importance in perennial forest tree resistance to various diseases [30] by NGS (next-generation sequencing). Currently, no research studies have been performed in which SNPs were used (through NGS) as a tool in population genetic analysis on the species *Ilex aquifolium* L.

Studies that “started” from leaf material using long reads in NGS technology (PacBio CLR and HiC sequencing) produced the de novo construction of new genomes (in the high quality–deep coverage range from 40× to 230×) in plants belonging to *Eucalyptus* and *Arachis* species [31]; in addition, using WGS (whole-genome sequencing), evolutionary history analyses in endangered plant species (*Acer catalpifolium* **Rehd.**) [32] were performed.

WGS analyses by short reads are used to obtain fragments with a high reading accuracy with the aim of constructing/reconstructing DNA/RNA sequences or cellular organelle genomes (chloroplasts and mitochondria) [33]; using WGS technology, in combinatory reading mode (long reads combined with short DNA reads), the outputs are analyzed in phylogeny/biodiversity studies in different species [34], plant breeding [35], or diagnosis [36]. WGS projects by NGS (short/long reads) in the *Ilex* genus were mostly carried out in species from Asia/South America, areas corresponding to speciation/diversification points [37]. Present species in genomic analysis by NGS are *Ilex asprella* (online data at the chromosomal assembly level) [38], *Ilex chinensis* **Sims.** (at the scaffold assembly level, without online publication of the assembly but which can be obtained upon request from the authors) [39], *Ilex polyneura* L. (online data for assembled chromosomes with annotations) [37], and *Ilex paraguariensis* **A.St.-Hil.** (online data at the scaffold assembly level) [40].

Currently, there are no present studies with reference to entire-genome sequencing by NGS (long or short reads) in the *Ilex aquifolium* L. species. The present study had several aims, the first of which was the entire-genome sequencing of the species *Ilex aquifolium* L. through NGS by short reads, the most representative and cosmopolitan species of the genus present in Europe. The WGS was performed on four DNA samples belonging to four different island locations, Romania (RO), Hungary (HU), Serbia (SR), and Bulgaria (BG). The second purpose of the WGS of *Ilex aquifolium* L. was to identify the genomic affiliation of the Romanian insular population of *Ilex aquifolium* L. from the residency status point of view, as a native or as a newly introduced and cultivated species from somewhere else and acclimatized in Romania. We comparatively analyzed the characteristics of the whole-genome (SNPs) between the RO population with two recognized resident native populations (BG and SR) and with one cultured population (HU). As such, we set out to identify if the RO population is oriented toward a native population or a cultivated one from the point of view of its genomic structure. It should also be mentioned that these three locations selected for the *Ilex aquifolium* L. sample collection (HU, SR, and BG) are the closest geographically to the Romanian insular location from the Zimbru population in the Codru-Moma Mountains, Romania.

## 2. Results

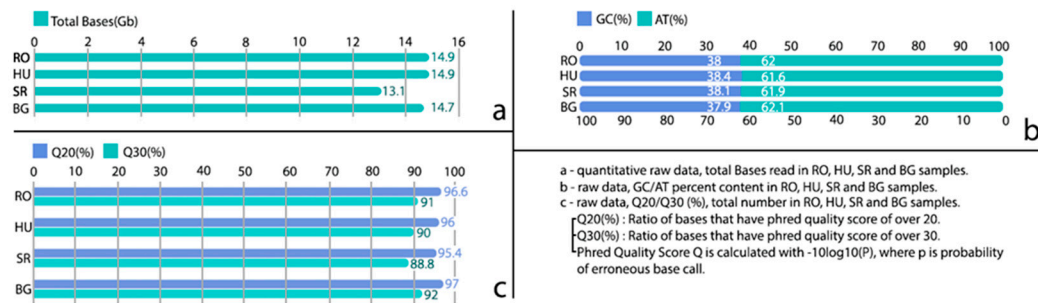
The results of the present study are divided into two categories. The first set of results were obtained by NGS sequencing and subsequent QC for the four locations from which the sampling was made. The second set of results are presented with reference to the following: (1) genomic SNP identification across all the populations of *Ilex aquifolium*, using the *Ilex asprella* (**Hook. and Arn.**) **Champ. ex Benth.** genome as a reference; (2) identification of positions and comparison of SNPs in genes from the four populations, at the chloroplast level, with *Ilex aquifolium* L. as a reference; (3) identification of the positions and comparison of SNPs in genes from the four populations at the level of the mitochondria, with reference to *Ilex pubescens* **Hook. and Arn.**; (4) quality metrics for the primary assemblies and the alignment of genomes between the studied populations, with reference to the *Ilex asprella* (**Hook. and Arn.**) **Champ. ex Benth.** genome.

## 2.1. Preliminary Results

### NGS Raw Data and QC Results in All Four *Ilex aquifolium* L. Populations

The NGS data, with gross total numbers of 14,877,967,184 bases for the sample from the RO population, 14,877,794,440 bases for the HU population, 13,117,326,210 bases for SR, and 14,713,914,442 bases for the BG population, were generated. Total numbers of 98,529,584 reads for the RO population, 98,528,440 in the HU population, 86,869,710 in SR, and 97,443,142 reads for the BG population were generated.

The GC and AT percentages had similar relative values in the four populations, with Q20 (%) and Q30 (%) base quality scores also having similar values (Figure 1).



**Figure 1.** NGS raw-data statistics for all four locations (RO, HU, SR, and BG). (a) quantitative raw data, total bases read in RO, HU, SR and BG samples; (b) raw data, GC/AT percent content in RO, HU, SR and BG samples; (c) raw data, Q20/Q30(%) total number in RO, HU, SR and BG samples.

High-quality reads were obtained at all four locations, with Q30 values between 89.4% and 92.4%. The percentages of the reads were in the range 98.4–99.0%, and well-expected values for the CG content after filtering (range: 37.8–38.3%) were obtained.

We used a modular tool to aggregate results from bioinformatics analyses (quality control, GC content, raw reads, and sequence alignments) across all the samples (RO, HU, SR, and BG). We summarized a single statistical report; more details can be found in the Supplementary Material summarized in Table S1.

## 2.2. Processed Results

### 2.2.1. SNP Analysis

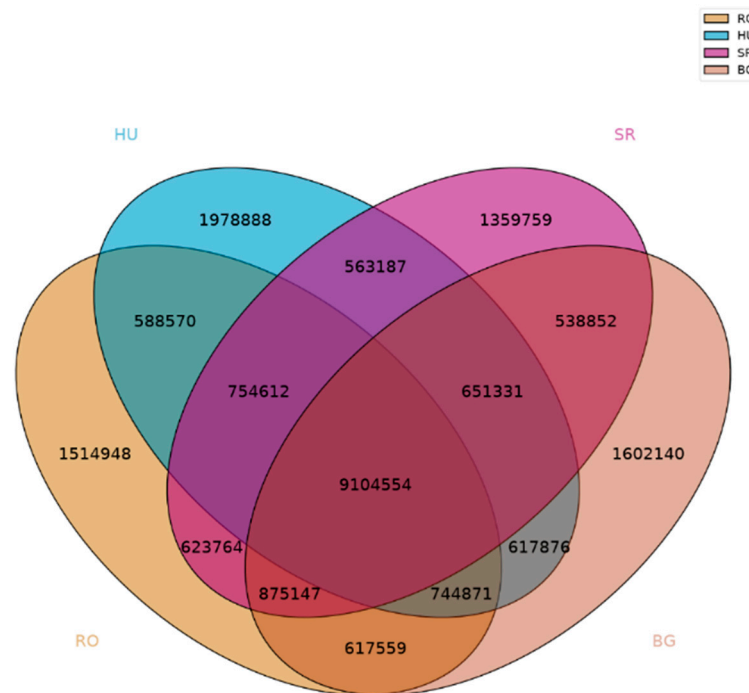
We performed the genomic SNP identification across all the populations of *Ilex aquifolium* L. (RO, HU, SR, and BG), using the *Ilex asprella* (Hook. and Arn.) Champ. ex Benth. genome (CUHK\_Ilex\_v2.1) as a reference. A similar ratio of transitions vs. transversions in SNPs was present. Details can be found below in Table 1.

**Table 1.** Transition (Ts) vs. transversion (Tv) ratios in SNPs at all four locations.

Sample	Vars	SNPs	Indels	Ts/Tv
BG	14,752,330	14,752,330	0	1.72
HU	15,003,889	15,003,889	0	1.72
RO	14,824,025	14,824,025	0	1.72
SR	14,471,206	14,471,206	0	1.72

Vars, variants; Indels, insertions-deletions; RO, Romania; HU, Hungary; SR, Serbia; BG, Bulgaria.

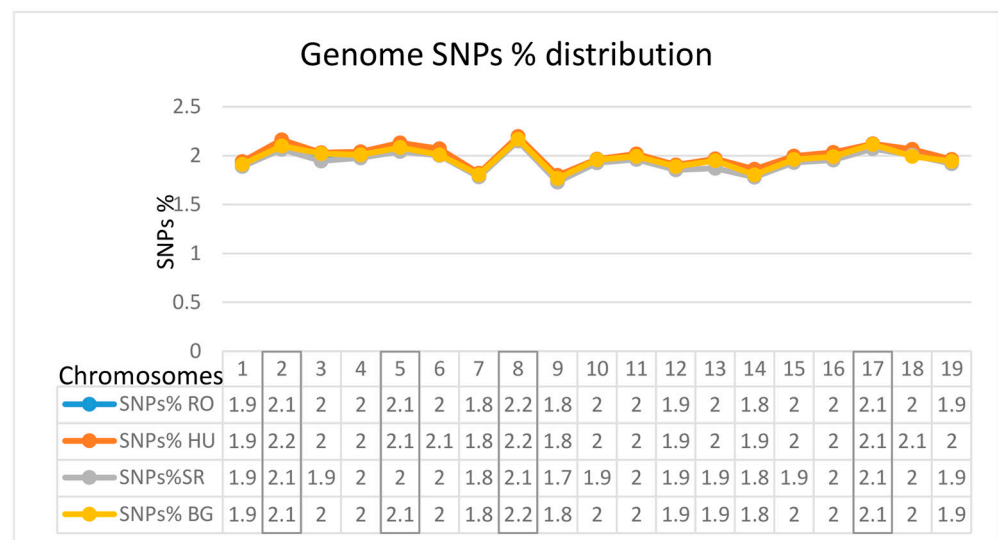
For SNP comparison in all four datasets (RO, HU, SR, and BG), a Venn diagram was generated for visualization (Figure 2). Sets of Venn diagrams were generated for paired locations, such as HU-BG, HU-SR, RO-BG, RO-HU, RO-SR, and SR-BG. For more details, see the Supplementary Material Figures S1–S6.



**Figure 2.** Venn diagram depiction of SNP intersections in all four datasets (RO/HU/SR/BG).

### 2.2.2. Chromosome-Level Distributions of SNPs Across the Four Locations

Higher-percentage values of SNPs were obtained for chromosomes 2, 5, 8, and 17 according to the reference genome, *Ilex asprella* (Hook. and Arn.) Champ. and Benth. genome (CUHK\_Ilex\_v2.1) (Figure 3).



**Figure 3.** The chromosomal SNP-percentage-level representations with highest values at all four locations (RO, HU, SR, and BG).

### 2.2.3. Identification of SNP Positions in Genes at the Four Locations at the Levels of Chloroplasts and Mitochondria

According to our analysis at all four locations, we identified a cumulative total of 14 chloroplast genes that presented SNPs. Of these, six genes (*psbA*, *rpoB*, *clpP*, *rpl14*, *rpl16*, and *ndhD*) were present at all four locations. Three of these genes were highlighted with a higher number of SNPs: *clpP*, *rpl14*, and *rpl16*. Also, a differentiation into two categories

of SNPs as the number per gene identified was observed, with similar values for samples from spontaneous populations compared to the cultivated population (Table 2).

**Table 2.** Counts of SNP positions in genes at the four locations at the chloroplastidial level.

Chr	Start	End	Gene	RO	HU	SR	BG
NC_068798.1	450	1512	<i>psbA</i>	3	4	6	5
NC_068798.1	8184	8370	<i>psbK</i>	1	0	1	1
NC_068798.1	12,575	13,831	<i>atpF</i>	1	0	1	1
NC_068798.1	21,792	24,612	<i>rpoC1</i>	3	0	2	2
NC_068798.1	24,638	27,851	<i>rpoB</i>	3	1	3	3
NC_068798.1	69,183	69,257	<i>trnP-UGG</i>	0	0	1	0
NC_068798.1	72,652	74,715	<i>clpP</i>	13	16	13	13
NC_068798.1	80,637	81,651	<i>rpoA</i>	0	0	1	0
NC_068798.1	83,439	83,808	<i>rpl14</i>	12	9	19	19
NC_068798.1	83,935	85,265	<i>rpl16</i>	20	6	28	27
NC_068798.1	117,349	118,315	<i>ccsA</i>	1	0	1	1
NC_068798.1	118,573	120,151	<i>ndhD</i>	1	1	1	1
NC_068798.1	122,601	124,838	<i>ndhA</i>	1	0	1	1
NC_068798.1	126,765	132,456	<i>ycf1</i>	1	0	2	1

Chr, chromosome; RO, Romania; HU, Hungary; SR, Serbia; BG, Bulgaria.

At all four locations, we identified a higher cumulative number of genes with SNPs in mitochondria than in chloroplasts, namely, in a total of 17 genes. Of these, 12 genes (*rrn18*, *nad4*, *atp4*, *nad2*, *cob*, *nad7*, *atp1*, *nad1*, *matR*, *rrn26*, and *atp9*) were present at all four locations. Of these, the *nad2* and *rrn26* genes presented higher numbers of SNPs. An equal number of SNPs were identified in seven genes. Here, too, the differentiation of SNP values into two categories is observed, correlated with the population status in terms of the origin: native or cultivated (Table 3).

**Table 3.** Counts of SNP positions in genes at the four locations at the mitochondrial level.

Chr	Start	End	Gene	RO	HU	SR	BG
NC_045078.1	37,841	39,772	<i>rrn18</i>	1	3	1	1
NC_045078.1	43,064	43,138	<i>trnW-CCA</i>	0	1	1	1
NC_045078.1	56,375	58,624	<i>cox2</i>	2	0	2	2
NC_045078.1	68,915	77,567	<i>nad4</i>	3	3	3	3
NC_045078.1	158,633	159,212	<i>atp4</i>	1	1	1	1
NC_045078.1	174,865	178,373	<i>rps3</i>	0	1	0	0
NC_045078.1	216,298	221,012	<i>nad2</i>	4	2	4	4
NC_045078.1	249,304	250,486	<i>cob</i>	1	1	1	1
NC_045078.1	252,738	253,002	<i>atp9</i>	0	1	0	0
NC_045078.1	276,579	282,510	<i>nad7</i>	1	2	1	1
NC_045078.1	295,025	295,634	<i>rps1</i>	0	1	0	0
NC_045078.1	300,823	302,353	<i>atp1</i>	2	2	2	2
NC_045078.1	438,334	439,971	<i>nad1</i>	1	2	1	1
NC_045078.1	471,524	475,330	<i>nad1</i>	2	2	2	2

**Table 3.** *Cont.*

Chr	Start	End	Gene	RO	HU	SR	BG
NC_045078.1	472,644	474,615	<i>matR</i>	1	1	1	1
NC_045078.1	483,961	487,154	<i>rrn26</i>	4	2	4	4
NC_045078.1	503,822	504,068	<i>atp9</i>	1	1	1	1

Chr, chromosome; RO, Romania; HU, Hungary; SR, Serbia; BG, Bulgaria.

#### 2.2.4. Quality Metrics for the Primary Assemblies

The representative statistical data generated with reference to the genomic alignments among the four locations in the HTML report are presented in Table 4.

**Table 4.** Genome assembly statistics at all four locations.

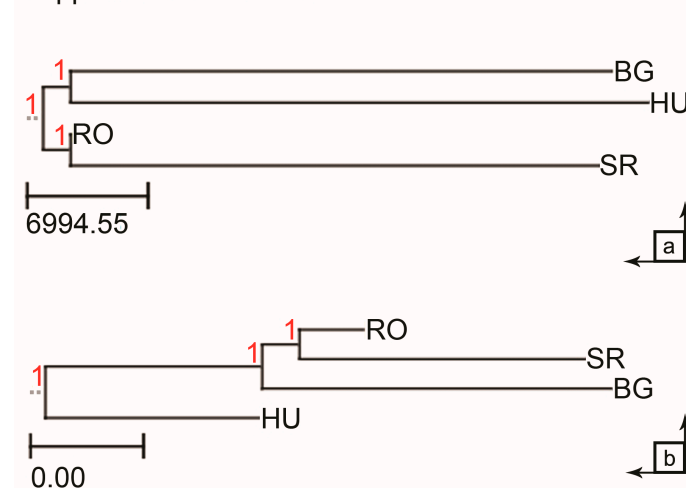
QUAST Stat	Location			
	RO	HU	SR	BG
All contigs	96,602	102,971	112,689	93,492
Total length (bp)	604,400,875	603,264,456	582,153,373	599,835,904
Longest contig (bp)	147,155	109,510	111,060	135,597
N50	12,964	12,102	10,112	13,200
N90	96,602	102,971	112,689	93,492

All the statistics are based on contigs of a length of at least 500 bp.

#### 2.2.5. Modeling and Graphical Representation of Genomic Alignments

According to the genomic analysis between the individuals at the four locations, we generated phylogenetic trees that showed the existing interrelationships between them. According to the analysis, a common origin is identified for the RO-SR clade (in the case of the breakpoint distance) and a clade with a common ancestor among the three spontaneous populations (in the case of the substitution distance). Within this clade, there is a sub-clade with a common ancestor between RO and SR. In both genomic analyses, the common ancestor between the populations from RO and SR is identified, as shown in Figure 4.

Support values are shown in red.



**Figure 4.** Phylogenetic trees based on (a) breakpoint distance; (b) substitution distance—for all four samples.

### 3. Discussion

The events with importance in evolution and that lead to the emergence of genetic diversity are mutations, genetic drift, gene flow, and natural selection, an important role being played by the interaction between the environment and genotypes, with reference to genetic

diversity [41,42]. The quantification of these events with importance in evolution (implicitly in genetic diversity) can be performed by analyzing the allelic variation, the proportion of polymorphic loci, the average allelic value/locus, the polymorphism/rate/frequency of the polymorphism at the SNP level, etc.

### 3.1. SNP Population Structure

In the present study, the analysis of the SNPs at the genome level in samples of *Ilex aquifolium* L. identified the existing differences among the local populations in RO, SR, and BG (from spontaneous flora) and HU (cultivated). The HU (cultivated) population has a higher number of SNPs (with approximately half-a-million SNPs/genome) compared to those in the other three populations (RO, SR, and BG). Similar intra-population studies have shown that plant-genome-based SNP numbers, in compared populations (cultivated vs. spontaneous), have higher levels of polymorphism in cultivated populations compared to those in native ones [43].

Venn diagrams showing the SNPs found among the four locations reveal the fact that the RO location of *Ilex aquifolium* L. has the highest number of SNPs in common with the SR location (11,358,077) and decreasing numbers in common with the BG (11,342,131) and HU (11,192,607) locations. Given the highest level of the SNP-based similarity between RO and SR, we believe that the location in RO, with an extremely compact, non-dispersed location, has a connection or a common ancestor with the population of SR. The population of RO, because of the extremely compact area (Supplementary Materials Figure S7), does not show individuals scattered in the forest area, as would be expected. Also, the RO population is found in only one very limited location, which leads us to hypothesize the existence of a possible anthropic factor responsible for the presence of the *Ilex aquifolium* L. species at this location in RO. From the point of view of the origin of the insular area in the RO population, we propose that it is likely to have a common ancestor with the local population from SR according to the number of SNPs found in common during the SNP analysis at the genome level.

### 3.2. The Distributions of SNPs at the Chromosomal Level

Based on the analyses that we carried out, the largest genomic differences were found on chromosomes 2, 5, 8, and 17, as shown in Figure 3. Sequencing studies and chromosomal analysis in *Arabidopsis thaliana* (L.) Heynh. [44], on chromosome 2, have shown that the genes/proteins expressed by this chromosome have regulatory functions of signal transduction, of transport, and related to the binding of proteins. Similarly, for chromosome 5 [45], the proteins expressed play roles in metabolism (21.1%), transcription (18.6%), and defense (11.9%). On chromosome 8 in rice [46], genetic regions (between RM6999 and RM22529) are involved in the rate of photosynthesis and are important in the hydraulic conductance of plants, root length, root function, surface area, and stomatal conductance. SNPs with locations on chromosome 17 may be important in relation to the *HaSPL* genes, which play crucial roles in the flowering process, stem/root tissue development, as well as in the plant's responses to abiotic stress factors [47].

### 3.3. Identification of SNP Positions on Genes at the Four Locations at the Level of the Chloroplast and Mitochondrion

#### 3.3.1. Chloroplastidial SNPs

Higher SNP counts were identified at the level of some chloroplastidial genes, *rpl14/16*—chloroplast genes encoding a ribosomal protein (L14/16), which are constituents of the large ribosomal subunit. High SNP numbers in these genes may indicate genetic variation affecting ribosomal function, which could impact plant growth, development, or stress responses. Mutations in *clpP*—the proteolytic subunit of the ATP-dependent Clp protease—may lead to altered protein turnover, degradation, or repair. Disruptions in ClpP function could impact chloroplast homeostasis, photosynthesis, and overall plant health. Regarding comparative values between the spontaneous flora locations (RO, SR, and BG) and the cultivated location (HU), the presence of SNPs was observed in the genes *psbK*

(Photosystem II reaction center protein K), *rpoC1* ( $\beta$ -subunit of RNA polymerase), *ccsA* (involved in cytochrome c biogenesis within the chloroplast), *ndhA* (NADH dehydrogenase subunit A), and *ycf1* (translocon at the inner envelope membrane of chloroplasts 214 and *ycf1.2*) and the absence of SNPs in the cultivated population. High SNP values for *psbK* might affect PSII function, electron transport, or stability; an altered *rpoC1* function could affect the overall plant development and stress tolerance. For *ccsA*, high SNP values might disrupt cytochrome c assembly or function; for *ndhA*, SNPs in this gene may affect electron transport and energy conversion. Altered *ndhA* function could impact plant growth and stress responses. High SNP values for *ycf1* might influence plastid biogenesis, protein translation, or RNA processing. Implications could range from altered chloroplast structure to overall plant fitness. Based on chloroplast gene mutations, changes are observed that intervene in energy metabolism correlated with genes involved in plant growth and responses to abiotic stresses.

### 3.3.2. Mitochondrial SNPs

From the analysis of the seventeen mitochondrial genes, three main directions were identified.

In the first instance (with *Ilex pubescens* Hook. and Arn as a reference) two clusters are individualized among all four populations, with relatively uniform SNP values among the native flora (RO, SR, and BG) locations compared to the cultivated location (HU).

In the second case, similar values for the RO-SR-BG cluster were evident for genes such as *rrn18* (mitochondrial 18S ribosomal RNA, a component of the 30S small subunit of mitochondrial ribosomes), *cox2* (cytochrome c oxidase subunit 2), *rps3* (nuclear ribosomal protein S3, implicated in the assembly of the ribosomal small subunit), *nad2* (NADH dehydrogenase subunit 2), *atp9* (subunit 9 of mitochondrial F<sub>0</sub>-ATPase), *nad7* (NADH dehydrogenase subunit 7), *rps1* (nuclear ribosomal protein S1), *nad1* (NADH dehydrogenase subunit 1), and *rrn26* (mitochondrial 26S ribosomal RNA protein).

In the third case, the genes can be clustered into a group in relation to ribosomal function and protein biosynthesis, as well as a second group that contains genes with importance in energy metabolism/electron transport chain–ATP synthase.

Thus, in the ribosomal function group, the following genes are identified: SNPs in *rrn18/26* can impact ribosomal function, which, in turn, affects protein synthesis within the cell; SNPs in *rps3* can influence protein synthesis and cellular growth; and SNPs in *rps1* can impact protein synthesis and cellular processes.

From the electron transport chain/ATP synthase group: SNPs in *cox2* may alter the efficiency of oxidative phosphorylation; SNPs in *nad2/7/1* may impact energy metabolism; and SNPs in *atp9* can affect cellular energy levels. Also, similar SNP values were highlighted regardless of the location. These values were identified for the genes *nad4* (NADH dehydrogenase subunit 4), *atp4* (subunit 4 of mitochondrial F<sub>0</sub>F<sub>1</sub> ATPase), *cob* (cytochrome b), *atp1* (ATP synthase subunit alpha), *nad1* (NADH dehydrogenase subunit 1), *matR* (reverse transcriptase/RNA maturase protein, involved in splicing), and *atp9* (subunit 9 of mitochondrial F<sub>0</sub>-ATPase).

The genetic variations in these genes and their SNPs contribute to plant adaptation, evolution, and responses to environmental changes, so these results could be used to study plant diversity and improve crop resilience [48,49].

At the locations of Mala-Reka SR (1033 m altitude), Borino BG (1143 m altitude), and Zimbru RO (258 m altitude), the *Ilex aquifolium* L. species is located in the forest mountain area, with low temperatures correlated with intense ultraviolet radiation, which can lead to adaptive gene changes at the mitochondrial level [50]. Saline soils and water stress affect the energy balance of plants. Mitochondrial genes may respond in order to maintain cellular homeostasis.

### 3.4. Quality Metrics of Genome Assemblies

We generated several quality metrics for the country-specific genome assemblies. These metrics are reported in Table 4 and are based on contigs of lengths longer than or equal to

500 bp. In general, the RO, HU, and BG assemblies were of similar quality, with the SR assembly slightly lagging behind. The RO assembly had the longest total length, at 604 Mbp distributed in 96,602 contigs, while SR had only 582 Mbp in total, for a total of 112,689 contigs. The longest contig was found in the RO assembly, at 147,155 bp. In terms of contiguity, the N50 metric, which represents the length (in bp) of the shortest contig at 50% of the total assembly length, had the highest value for the BG assembly (at 13,200 bp), with RO coming in at a close second, with an N50 value of 12,964 bp. Similar trends were observed for the BG and RO assemblies in terms of the N90 metric, which assesses contiguity based on 90% of the assembly. The corresponding values were 2798 for BG and 2690 for RO. The SR assembly had the lowest contiguity, with an N50 value of 10,112 and an N90 value of 2230.

### 3.5. Genomic Alignments

In this study, our goal was not to achieve a pan-genome, as was achieved, for example, in cultivated plantain species [51,52], but to understand the diversity or genomic proximity of *Ilex aquifolium* L. among the four locations. Also, we focused on the possible origin of the population in RO and its likelihood to be either a native or a cultivated variety of the species. Figure 4a shows the structural variation evolution, in the case of the four genomes, based on genomic rearrangements over time, which indicates a common origin at the genome level between the locations RO and SR and between BG and HU. According to the substitution distance phylogenetic tree from Figure 4b, the comparative analysis was followed by genomic alignment among the four locations, having as a criterion the number of SNPs existing at each location. As in the previous case (from Figure 4a), here, the results also indicate a common origin of the populations of RO and SR and these two having a common origin with the population of BG. The obvious difference is given by the fact that the cultivated population of *Ilex aquifolium* L. (HU) is on a separate branch of the phylogenetic tree compared to those at the locations with native populations.

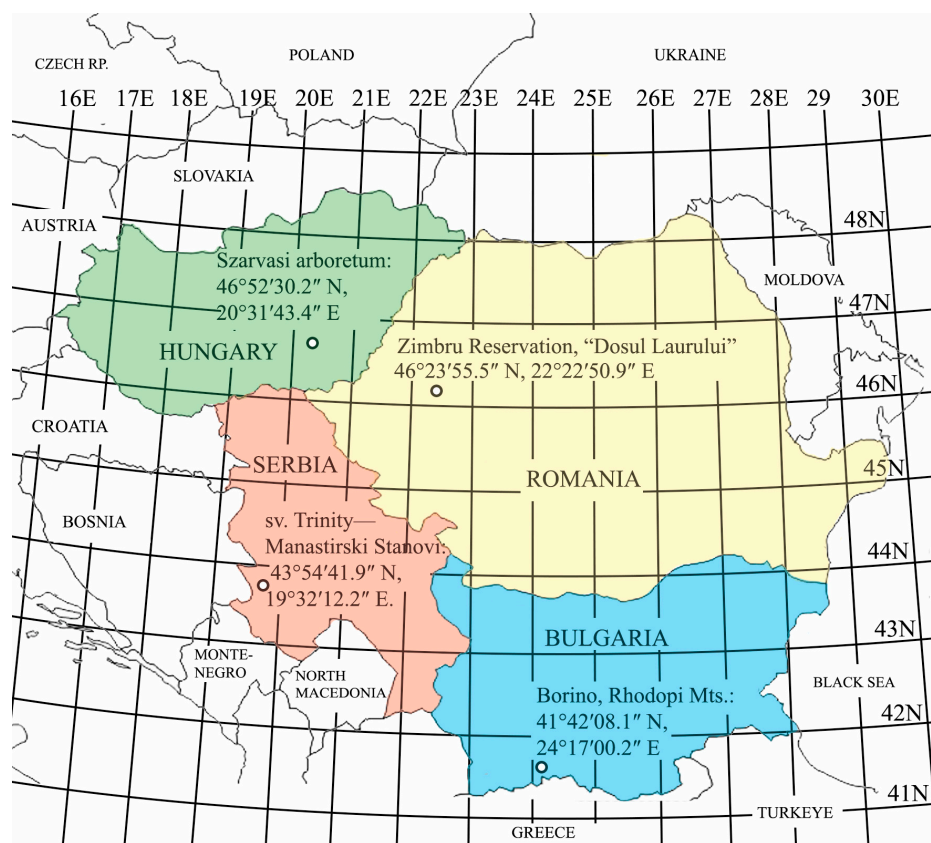
### 3.6. Phenotypic Traits in Relation to Genomic Variability

The present study shows the genomic variability among the four locations (RO, SR, BG, and HU), indicating a possible common origin between the RO and SR locations. In our previous studies on the morphologies and phytonutrient profiles in the same populations of *Ilex aquifolium* L., it emerged that the population in HU is close to that in SR [53]. By comparing the RO and SR locations with the HU location, differences in habitat, plant habitus, and morphology can be observed. In the HU population (in a maintained park area), the location is at a lower altitude than the two locations RO and SR, which, over time, could have produced morphological adaptation changes correlated with changing climatic factors in comparison to those of the populations from RO and SR, which are located in forest areas. Even so, both genomic and morphological changes show a differentiation into two clusters, i.e., the native species populations versus the cultivated one from HU. One possible reason for this is the abiotic factors that are relatively controlled in terms of the water regime and nutrient access for individuals in the HU population.

## 4. Materials and Methods

### 4.1. Plant Harvesting

All the harvested plants used in this study belong to the *Ilex aquifolium* L. species. The GPS coordinates of the collection points are the following: Romania, Zimbru Reservation, "Dosul Laurului", GPS coordinates of 46°23'55.5" N, 22°22'50.9" E; Hungary, Szarvasi Arboretum, 46°52'30.2" N, 20°31'43.4" E; Bulgaria, Borino, Rhodope Mountains, 41°42'08.1" N, 24°17'00.2" E; and Serbia, in the middle of the Mala Reka region, Sv. Trinity—Manastirski Stanovi, 43°54'41.9" N, 19°32'12.2" E, as shown in Figure 5. At each collection point, 5 shoots, with the approximate age of 1 year, were harvested from female plants, in the 11 AM–2 PM interval during November 2022. The collected samples were wrapped in aluminum foil to avoid contamination and kept at a temperature of −80 °C before their processing.



**Figure 5.** Harvesting points of *Ilex aquifolium* L. samples. Romania, Zimbru Reservation, “Dosul Laurului”, GPS coordinates of  $46^{\circ}23'55.5''$  N,  $22^{\circ}22'50.9''$  E; Hungary, Szarvasi Arboretum,  $46^{\circ}52'30.2''$  N,  $20^{\circ}31'43.4''$  E; Bulgaria, Borino, Rhodope Mountains,  $41^{\circ}42'08.1''$  N,  $24^{\circ}17'00.2''$  E; and Serbia, in the middle of the Mala Reka region, Sv. Trinity—Manastirski Stanovi,  $43^{\circ}54'41.9''$  N,  $19^{\circ}32'12.2''$  E.

#### 4.2. DNA Extraction and Sequencing

The genomic DNA (gDNA) of *Ilex aquifolium* L. was extracted from an amount of 50 mg of fresh leaves preserved on dry ice (4 extractions, separated for each location) followed by WGS for all four locations.

Genomic DNA extractions for all the samples were performed using DNeasy plant mini purification kit from Qiagen (S.C. Omnivet Impex SRL/OmniGen, Bucharest, Romania). The DNA quality control (QC) method—DNA quantification was performed with Quant-iT™ PicoGreen™ dsDNA assay kits (Thermo Fisher Scientific Inc., BERD Trading Ltd., Bucharest, Romania) and dsDNA reagents from Invitrogen™, using Victor 3 fluorometry (Perkin-Elmer, Lisboa, Portugal). The DNA condition assessment was performed using the gel electrophoresis method. The DNA size check for DNA fragments <1 kb was performed with a 2100 Bioanalyzer and a DNA 1000–7500 chip for normal PCR products.

For NGS analysis, a Truseq PCR-free (350 bp short reads) library preparation kit was used. An Illumina system (Sequencer/Novaseq 6000), a read length of 151, a paired-end (2 × 150 bp) read type, and 12 Gb/sample (at each location) were used. The sequencing was performed by Macrogen Ltd., Seoul, Republic of Korea.

Library QC Method—The size of the PCR-enriched fragments was measured by checking the template size distribution on an Agilent Technologies 2100 Bioanalyzer using a DNA 1000 chip. The library quantity check was performed using an Illumina library.

Considering the size of the *Ilex asprella* (Hook. and Arn.) Champ. ex Benth. genome is 835 Mb, the sequencing read depth coverage was estimated to be at least 12×. The raw-genomic-archived FastQ sequences resulted in NGS for all four locations and are indexed in the NCBI SRA under BioProject accession number PRJNA941469.

The experiments of the gDNA extractions and NGS services for WGS—short reads—were performed by MacroGen Europe and its affiliate MacroGen Seoul, Republic of Korea. The sample shipment was made on dry ice.

#### 4.3. WGS Analysis

The raw NGS data files, RO\_1.fastq.gz, RO\_2.fastq.gz, HU\_1.fastq.gz, HU\_2.fastq.gz, SR\_1.fastq.gz, SR\_2.fastq.gz, BG\_1.fastq.gz, and BG\_2.fastq.gz, were deposited in the NCBI SRA database under BioProject accession number PRJNA941469.

The raw sequenced reads were checked with FastQC v. 0.11.9 [54], and all the quality control tests were passed by the four samples. Next, decontamination was performed using Kraken2 v. 2.1.2 [55] on a prebuilt database available online, denoted as “PlusPF”, and built on 14.03.2023 from RefSeq sequences of archaea, bacteria, viruses, plasmids, humans, protists, and fungi [56]. Decontamination was then performed using the following settings: “--memory-mapping --paired --gzip-compressed --use-names --minimum-hit-groups 4 --confidence 0.1”. Classified reads were discarded, and unclassified reads were retained and used for all the further analyses.

MaSuRCA (Maryland Super Read Cabog Assembler) v. 4.1.0 [57] was used to assemble the decontaminated reads into contigs for each of the 4 samples. QUAST v. 5.2.0 [58] was then used to compute various quality metrics for the primary assemblies and RepeatModeler and were used to identify and mask repetitive sequences, respectively. The latter two software packages were obtained via the TETools Docker image, v. dfam/tetools 1.8 [59]. For the first joint variant-calling analysis, we used the Ilex *Asprella* assembly, version CUHK\_Ilex\_v2.1 (NCBI ID GCA\_023539305.1), as a reference. The Sarek pipeline, v. 3.1.2, from nf-core [60] was used to identify small variants (SNPs and indels) for each of the four samples relative to the Ilex *asprella* reference. Some of the main steps implemented in the pipeline relied on the following software packages: BWA v. 0.7.17-r1188 [61], for mapping reads to the reference via the bwa mem command; GATK v. 4.3.0.0 [62], for marking duplicates and identifying variants via the MarkDuplicates and HaplotypeCaller subroutines, respectively; Samtools v. 1.16.1 and Bcftools v. 1.16 [63], for quality control and various maintenance operations on mapped reads and identified variants, respectively; and MultiQC v. 1.13 [64], for integrating various statistics and results from the different pipeline stages. The Nextflow framework, v. 23.04.0 [65], was responsible for running the pipeline. The result was a consolidated VCF file, with variants called in the four NGS samples.

For the chromosome-level distributions of SNPs across the four locations—*Ilex asprella* chromosomes (based on the *Ilex asprella* (Hook. and Arn.) Champ. ex Benth. genome CUHK\_Ilex\_v2.1 as a reference)—the SNP counts for each individual chromosome were computed using basic Unix tools, such as grep, sort, uniq, and sed. In the next step, the percentages of SNPs obtained for all 19 chromosomes at the four locations were calculated as the ratio between the length of the chromosomes (expressed in nucleotides) and the number of SNPs obtained in the previous analysis, see the Supplementary Material (Table S2).

For the second joint variant-calling analysis, we used the *Ilex aquifolium* L. chloroplast assembly (NCBI ID NC\_068798.1) as a reference, based on the same pipeline as previously described. After obtaining the VCF file, specific variants were retained, which overlapped the 86 annotated genes (as of 01.02.2023) for this chloroplast sequence [66].

Finally, for the third joint variant-calling analysis, we acquired the *Ilex pubescens* Hook. & Arn. mitochondrion sequence from NCBI (accession NC\_045078.1) and used it as a reference. As in the chloroplast analysis, we first selected all the biallelic SNPs, after which, for a separate analysis, we retained only those that overlapped the 54 annotated genes (as of 01.02.2023) found in NCBI [67].

In order to generate Venn and UpSet plots, biallelic SNPs were extracted using Bcftools for the following three scenarios:

- SNPs of RO, HU, SR, and BG, relative to the *Ilex asprella* (Hook. & Arn.) Champ. ex Benth. genome;
- SNPs of the four samples, relative to the *Ilex aquifolium* L. chloroplast sequence;



## 5. Conclusions

The study of genomic variability through comparative analysis at the genome level within the four populations (three native populations (RO, SR, and BG) and one cultivated population (HU)) identified correlations based on whether they were phylogenetically close or distant. The highest number of SNPs in common was found between the RO and SR locations. SNPs identified in all four populations showed higher frequencies for 4 chromosomes (2, 5, 8, and 17), this being in correlation with cellular regulatory functions, transport of metabolites/water, and responses of the plants to abiotic stress factors. Similar observations for SNPs' positions on chloroplast and mitochondrial genes were found in all four populations. Finally, according to the genome alignments, the common origin of the SR-RO populations is reconfirmed.

In accordance with the data presented in this study, at the genome level within the populations of SR, RO, BG, and HU of the species *Ilex aquifolium* L., we propose that the origin of the population in RO is common with that of the population in SR and different from that of the population of the cultivated variant in HU.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms252413593/s1>.

**Author Contributions:** Conceptualization, C.V.M.; methodology, A.E.M.; software, A.E.M.; validation, C.V.M. and A.E.M.; formal analysis, C.V.M., D.E.I., V.T., R.I.N., L.-T.C. and V.B.B.; investigation, C.V.M. and A.E.M.; resources, A.Ş.A., B.P. and R.I.N.; data curation, R.I.N., L.-T.C., A.Ş.A. and E.M.; writing—original draft preparation, C.V.M. and A.E.M.; writing—review and editing, D.E.I., B.P., V.B.B., V.T. and R.I.N.; visualization, C.V.M., B.P., D.E.I., E.M. and A.E.M.; supervision, C.V.M. and A.E.M.; project administration, C.V.M. and A.E.M.; funding acquisition, V.T., C.V.M., A.E.M., A.Ş.A., E.M., V.B.B. and L.-T.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Links to publicly archived datasets: BioProject Accession PRJNA941469, <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA941469> (accessed on 6 June 2024). Links to resources: BG 1 ILLUMINA (Illumina NovaSeq 6000) run: 48.7 M spots, 14.7 G bases, 4.5 Gb downloads, Accession SRX19590379; SR 1 ILLUMINA (Illumina NovaSeq 6000) run: 43.4 M spots, 13.1 G bases, 4.3 Gb downloads, Accession SRX19590378; HU 1 ILLUMINA (Illumina NovaSeq 6000) run: 49.3 M spots, 14.9 G bases, 4.7 Gb downloads, Accession SRX19590377; RO 1 ILLUMINA (Illumina NovaSeq 6000) run: 49.3 M spots, 14.9 G bases, 4.6 Gb downloads, Accession SRX19590376 (All the data links were accessed on 7 March 2023).

**Acknowledgments:** We would like to thank Márta Somlyai at the Szarvas Arboretum and Goran T. Anačkov at the University of Novi Sad, Department of Biology and Ecology, for providing information about the analyzed insular populations of *Ilex aquifolium* L. This research was partially supported by MOISE infrastructure, grant number 240/2020, ID 911 POC/398/1/1, financed by the European structural funds and Romanian government funds (hpc.uvt.ro).

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

## References

1. Guerrero Hue, N.; Caudullo, G.; de Rigo, D. *Ilex aquifolium* in Europe: Distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species*; San-Miguel Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A., Eds.; Publications Office of the EU: Luxembourg, 2016; p. e011fbc+.
2. Yao, X.; Song, Y.; Yang, J.-B.; Tan, Y.-H.; Corlett, R.T. Phylogeny and biogeography of the hollies (*Ilex* L., Aquifoliaceae). *J. Syst. Evol.* **2021**, *59*, 73–82. [[CrossRef](#)]
3. HU, S.-Y. The genus *Ilex* in China. *J. Arnold Arbor. Harv. Univ.* **1949**, *30*, 233–344, 348–387. [[CrossRef](#)]
4. HU, S.-Y. The genus *Ilex* in China. *J. Arnold Arbor. Harv. Univ.* **1950**, *31*, 39–80, 214–240, 241–263. [[CrossRef](#)]

5. Savolainen, V.; Chase, M.W.; Hoot, S.B.; Morton, C.M.; Soltis, D.E.; Bayer, C.; Fay, M.F.; De Bruijn, A.Y.; Sullivan, S.; Qiu, Y.-L. Phylogenetics of Flowering Plants Based on Combined Analysis of Plastid *atpB* and *rbcL* Gene Sequences. *Syst. Biol.* **2000**, *49*, 306–362. [CrossRef]
6. Negrin, A.; Long, C.; Motley, T.J.; Kennelly, E.J. LC-MS metabolomics and chemotaxonomy of caffeine-containing holly (*Ilex*) species and related taxa in the Aquifoliaceae. *J. Agric. Food Chem.* **2019**, *67*, 5687–5699. [CrossRef]
7. Yao, X.; Zhang, F.; Corlett, R.T. Utilization of the Hollies (*Ilex* L. spp.): A Review. *Forests* **2022**, *13*, 94. [CrossRef]
8. Yi, F.; Zhao, X.; Peng, Y.; Xiao, P. Genus *Ilex* L.: Phytochemistry, ethnopharmacology, and pharmacology. *Chin. Herb. Med.* **2016**, *8*, 209–230.
9. Nouredine, T.; El Hussein, Z.; Nehme, A.; Massih, R.A. Antibacterial activity of *Ilex paraguariensis* (Yerba Mate) against Gram-positive and Gram-negative bacteria. *J. Infect. Dev. Ctries.* **2018**, *12*, 712–719. [CrossRef]
10. Peterken, G.F.; Lloyd, P.S. *Ilex aquifolium* L. *J. Ecol.* **1967**, *55*, 841–858. [CrossRef]
11. David, L.S.; Elliott, D.C.; David, M.C.; Santiago, L. Pictures of an Invasion: English Holly (*Ilex aquifolium*) in a Semi-Natural Pacific Northwest Forest. *Northwest Sci.* **2014**, *88*, 75–93.
12. Available online: <https://www.gbif.org/species/5414222> (accessed on 15 January 2023).
13. Nimis, P.L.; Bolognini, G. Quantitative phytogeography of the Italian Beech Forests. *Vegetatio* **1993**, *109*, 125–143. [CrossRef]
14. Available online: <https://www.iucnredlist.org/species/202963/2758279> (accessed on 15 January 2023).
15. Available online: <https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:83051-1> (accessed on 10 January 2023).
16. Bedö, A.; Horváth, S. *Erdészeti Lapok, az Országos Erdészeti Egyesület, Közlönye, Erdő és földbirtokosok, erdészeti ügyekkel foglalkozó és erdőtisztek számára. Harminczkettedik Évfolya; Az Országgyűlési Ertesítő kö es könyvnyomdája részv: Budapest, Hungary, 1893; pp. I–XII.*
17. Simonkai, L. *Aradvármegye és Arad Szabad királyi város, Természetrzaji Leírása, Monographiája; Kiadja a Monographia-Bizottság: Arad, Romania, 1893.*
18. Muica, E.C.; Popova-Cucu, A. The composition and conservation of Romania’s plant cover. *GeoJournal* **1993**, *29*, 9–18. [CrossRef]
19. UNEP-WCMC and IUCN. Protected Planet: Dosul Laurului in Romania. The World Database on Protected Areas WDPA: 14574/The Global Database on Protected Areas Management Effectiveness. GD-PAME, UNEP-WCMC/IUCN, Cambridge, UK. 1999. Available online: [www.protectedplanet.net](http://www.protectedplanet.net) (accessed on 3 June 2018).
20. Fan, Y.; Sun, L.E.; Hao, D.-C.; Peng, Y.; Han, F.-M.; Xiao, P.-G. Complex phylogenetic placement of *Ilex* species (Aquifoliaceae): A case study of molecular phylogeny. *Pak. J. Bot.* **2017**, *49*, 215–225.
21. Yao, X.; Tan, Y.H.; Liu, Y.Y.; Song, Y.; Yang, J.B.; Corlett, R.T. Chloroplast genome structure in *Ilex* (Aquifoliaceae). *Sci. Res.* **2016**, *6*, 28559. [CrossRef]
22. Kong, B.L.H.; Park, H.S.; Lau, T.W.D.; Zhixiu, L.; Tae-Jin, Y.; Pang-Chui, S. Comparative analysis and phylogenetic investigation of Hong Kong *Ilex* chloroplast genomes. *Sci. Rep.* **2021**, *11*, 5153. [CrossRef]
23. Zhang, F.; Chen, H.; Zhou, Y.; Li, N.; Chong, X.; Li, Y.; Lu, X.; Wang, C. The complete chloroplast genome sequence and phylogenetic analysis of *Ilex* ‘Beryl’, a hybrid of *Ilex cornuta* × *Ilex latifolia* (Aquifoliaceae). *Mitochondrial DNA B Resour.* **2021**, *6*, 227–228. [CrossRef]
24. Fan, Z.; Yanwei, Z.; Hong, C.; Naiwei, L.; Chuanyong, W.; Xiaoqing, L.; Yunlong, L. The complete chloroplast genome of *Ilex* ‘Tall Boy’, *Ilex aquifolium* × *Ilex latifolia* (Aquifoliaceae). *Mitochondrial DNA Part B* **2021**, *6*, 229–230.
25. Carlos, M.H.; Pilar, B. Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Bot. J. Linn. Soc.* **2013**, *171*, 441–452.
26. Jee-Soo, P.; Min-Young, K.; Eun-Jo, S.; JongHee, O.; Kyoung-In, S.; Kyung, S.K.; Sung-Chur, S.; Sang-Min, C.; Younghoon, P.; Gung Pyo, L.; et al. Genome-wide core sets of SNP markers and Fluidigm assays for rapid and effective genotypic identification of Korean cultivars of lettuce (*Lactuca sativa* L.). *Hortic. Res.* **2022**, *9*, uhac119. [CrossRef]
27. Ben-Ari, G.; Zenvirth, D.; Sherman, A.; Simchen, G.; Lavi, U.; Hillel, J. Application of SNPs for assessing biodiversity and phylogeny among yeast strains. *Heredity* **2005**, *95*, 493–501. [CrossRef] [PubMed]
28. Anjan, H.; Rakesh, K.; Chandan, S.; Sauren, D. Genome-wide SNP discovery from Darjeeling tea cultivars—Their functional impacts and application toward population structure and trait associations. *Genomics* **2021**, *113 Pt 1*, 66–78.
29. Guajardo, V.; Solís, S.; Almada, R.; Saski, C.; Gasic, K.; Moreno, M.Á. Genome-wide SNP identification in *Prunus* rootstocks germplasm collections using Genotyping-by-Sequencing: Phylogenetic analysis, distribution of SNPs and prediction of their effect on gene function. *Sci. Rep.* **2020**, *10*, 1467. [CrossRef]
30. Younessi-Hamzekhanlu, M.; Gailing, O. Genome-Wide SNP Markers Accelerate Perennial Forest Tree Breeding Rate for Disease Resistance through Marker-Assisted and Genome-Wide Selection. *Int. J. Mol. Sci.* **2022**, *23*, 12315. [CrossRef] [PubMed]
31. Driguez, P.; Bougouffa, S.; Carty, K.; Alexander, P.; Kamel, J.; Muppala, R.; Richard, S.; Ming, S.C.; Yoshinori, F.; Luca, E. LeafGo: Leaf to Genome, a quick workflow to produce high-quality de novo plant genomes using long-read sequencing technology. *Genome Biol.* **2021**, *22*, 256. [CrossRef]
32. Yu, T.; Hu, Y.; Zhang, Y.; Zhao, R.; Yan, X.; Dayananda, B.; Wang, J.; Jiao, Y.; Li, J.; Yi, X. Whole-Genome Sequencing of *Acer catalpifolium* Reveals Evolutionary History of Endangered Species. *Genome Biol. Evol.* **2021**, *13*, evab271. [CrossRef] [PubMed]
33. Zhang, F.; Li, W.; Gao, C.W.; Dan, Z.; Li-zhi, G. Deciphering tea tree chloroplast and mitochondrial genomes of *Camellia sinensis* var. *assamica*. *Sci. Data* **2019**, *6*, 209. [CrossRef]

34. Liu, G.-Q.; Lian, L.; Wang, W. The Molecular Phylogeny of Land Plants: Progress and Future Prospects. *Diversity* **2022**, *14*, 782. [CrossRef]
35. Wang, X.; Jiao, Y.; Ma, S.; Yang, J.-T.; Wang, Z.-X. Whole-Genome Sequencing: An Effective Strategy for Insertion Information Analysis of Foreign Genes in Transgenic Plants. *Front. Plant Sci.* **2020**, *11*, 573871. [CrossRef]
36. Jones, S.; Baizan-Edge, A.; MacFarlane, S.; Torrance, L. Viral Diagnostics in Plants Using Next Generation Sequencing: Computational Analysis in Practice. *Front. Plant Sci.* **2017**, *8*, 1770. [CrossRef]
37. Xin, Y.; Zhiqiang, L.; Yu, S.; Xiaodi, H.; Richard, T.C. A chromosome-scale genome assembly for the holly (*Ilex polyneura*) provides insights into genomic adaptations to elevation in Southwest China. *Hortic. Res.* **2022**, *9*, uhab049.
38. Kong, B.L.; Nong, W.; Wong, K.H.; Law, S.T.; So, W.L.; Chan, J.J.; Zhang, J.; Lau, T.D.; Hui, J.H.; Shaw, P.C. Chromosomal level genome of *Ilex asprella* and insight into antiviral triterpenoid pathway. *Genomics* **2022**, *114*, 110366. [CrossRef] [PubMed]
39. Zhou, P.; Li, J.; Huang, J.; Li, F.; Zhang, Q.; Zhang, M. Genome Survey Sequencing and Genetic Background Characterization of *Ilex chinensis* Sims (Aquifoliaceae) Based on Next-Generation Sequencing. *Plants* **2022**, *11*, 3322. [CrossRef]
40. Paiva, D.; Cascales, J.; Rosetti, M.; Scherer, R.; Gauchat, M.E.; Gottlieb, A. Unraveling the genetic complexity of a cultivated breeding population of “yerba mate” (*Ilex paraguariensis* St. Hil.). *An. Acad. Bras. Ciências* **2020**, *92*, e20190113. [CrossRef] [PubMed]
41. Sork, V.L. Gene flow and natural selection shape spatial patterns of genes in tree populations: Implications for evolutionary processes and applications. *Evol. Appl.* **2015**, *9*, 291–310. [CrossRef]
42. Booy, G.; Hendriks, R.J.J.; Smulders, M.J.M.; Van Groenendael, J.M.; Vosman, B. Genetic Diversity and the Survival of Populations. *Plant Biol.* **2000**, *2*, 379–395. [CrossRef]
43. Wang, D.; Li, X.; Zhang, Y. Comparative Study of Genetic Structure and Genetic Diversity between Wild and Cultivated Populations of *Taxus cuspidata*, Northeast China. *Phyton-Int. J. Exp. Bot.* **2024**, *93*, 355–369. [CrossRef]
44. Lin, X.; Kaul, S.; Rounsley, S.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.; Bowman, C.L.; Barnstead, M.; et al. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **1999**, *402*, 761–768. [CrossRef]
45. Tabata, S.; Kaneko, T.; Nakamura, Y.; Kotani, H.; Kato, T.; Asamizu, E.; Miyajima, N.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; et al. Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 823–826.
46. Adachi, S.; Tsuru, Y.; Nito, N.; Murata, K.; Yamamoto, T.; Ebitani, T.; Ookawa, T.; Hirasawa, T. Identification and characterization of genomic regions on chromosomes 4 and 8 that control the rate of photosynthesis in rice leaves. *J. Exp. Bot.* **2011**, *62*, 1927–1938. [CrossRef]
47. Jadhao, K.R.; Kale, S.S.; Chavan, N.S.; Janjal, P.H. Genome-wide analysis of the SPL transcription factor family and its response to water stress in sunflower (*Helianthus annuus*). *Cell Stress Chaperones* **2023**, *28*, 943–958. [CrossRef]
48. Androsiuk, P.; Paukszto, Ł.; Jastrzębski, J.P.; Milarska, S.E.; Okorski, A.; Pyszczółkowska, A. Molecular Diversity and Phylogeny Reconstruction of Genus *Colobanthus* (Caryophyllaceae) Based on Mitochondrial Gene Sequences. *Genes* **2022**, *13*, 1060. [CrossRef] [PubMed]
49. Niu, Y.; Lu, Y.; Song, W.; He, X.; Liu, Z.; Zheng, C.; Wang, S.; Shi, C.; Liu, J. Assembly and comparative analysis of the complete mitochondrial genome of three *Macadamia* species (*M. integrifolia*, *M. ternifolia* and *M. tetraphylla*). *PLoS ONE* **2022**, *17*, e0263545. [CrossRef]
50. Kang, N.; Hu, H. Adaptive evidence of mitochondrial genes in Pteromalidae and Eulophidae (Hymenoptera: Chalcidoidea). *PLoS ONE* **2023**, *18*, e0294687. [CrossRef]
51. RPAN. Available online: <https://cgm.sjtu.edu.cn/3kricedb> (accessed on 13 September 2023).
52. Sunflower Genome Database, Species: *Helianthus Annuus* (Sunflower). Available online: <https://www.sunflowergenome.org> (accessed on 13 September 2023).
53. Mihali, C.V.; Petrescu, C.M.; Ciolacu-Ladasiu, C.F.; Mathe, E.; Popescu, C.; Bota, V.; Mizeranschi, A.E.; Ilie, D.E.; Neamț, R.I.; Turcus, V. Assessing Phenotypic Variability in Some Eastern European Insular Populations of the Climatic Relict *Ilex aquifolium* L. *Plants* **2022**, *11*, 2022. [CrossRef]
54. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 20 March 2023).
55. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef]
56. Available online: [https://genome-index.s3.amazonaws.com/kraken/k2\\_pluspf\\_20230314.tar.gz](https://genome-index.s3.amazonaws.com/kraken/k2_pluspf_20230314.tar.gz) (accessed on 20 March 2023).
57. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [CrossRef] [PubMed]
58. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
59. Lerat, E.; Fablet, M.; Modolo, L.; Lopez-Maestre, H.; Vieira, C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* **2017**, *45*, e17. [CrossRef] [PubMed]
60. Ewels, P.A.; Peltzer, A.; Fillinger, S.; Patel, H.; Alneberg, J.; Wilm, A.; Garcia, M.U.; Di Tommaso, P.; Nahnsen, S. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **2020**, *38*, 276–278. [CrossRef]
61. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
62. Van der Auwera, G.A.; O’Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*; O’Reilly Media: Sebastopol, CA, USA, 2020.

63. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)]
64. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)] [[PubMed](#)]
65. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [[CrossRef](#)]
66. Available online: [https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore\\_gene&from\\_uid=2327861972](https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore_gene&from_uid=2327861972) (accessed on 24 April 2023).
67. Available online: [https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore\\_gene&from\\_uid=1770691604](https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore_gene&from_uid=1770691604) (accessed on 28 April 2023).
68. Available online: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1708-7> (accessed on 7 September 2023).
69. Hickey, G.; Monlong, J.; Ebler, J.; Novak, A.M.; Eizenga, J.M.; Gao, Y.; Marschall, T.; Li, H.; Paten, B. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat. Biotechnol.* **2024**, *42*, 663–673. [[CrossRef](#)]
70. Kolmogorov, M.; Raney, B.; Paten, B.; Pham, S. Ragout—A reference-assisted assembly tool for bacterial genomes. *Bioinformatics* **2014**, *30*, i302–i309. [[CrossRef](#)]
71. Hubisz, M.J.; Pollard, K.S.; Siepel, A. PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **2011**, *12*, 41–51. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.