BMC Plant Biology

RESEARCH Open Access

Hazelnut allergome overview and *Cor a* gRNAs identification

Ciro Gianmaria Amoroso[1] and Giuseppe Andolfo[1*]

Abstract

Background *Corylus* species (hazelnuts) are a valuable source of nutrients and are widely consumed worldwide. Nevertheless, *Corylus avellana* (Cor a) contains 13 allergens (*Cor a 1*, *Cor a 2*, *Cor a 6*, *Cor a 8*, *Cor a 9*, *Cor a 10*, *Cor a 11*, *Cor a 12*, *Cor a 13*, *Cor a 14*, *Cor a 15*, *Cor a 16*, and *Cor a TLP*) that have been deposited into the official database (WHO/IUIS) for allergen nomenclature. The recent availability of several *Corylus* genomes provided opportunities to explore allergome variability, and thus to develop hypoallergenic varieties using modern biotech approaches. Certainly, the identification of CRISPR-Cas9 guide RNA (gRNA) is a pivotal step in achieving this goal. User-friendly web tools include limited reference genomes to design CRISPR-Cas9 gRNAs, while bioinformatic software for custom analysis require advanced command-line skills.

Results This work explored the evolutionary trajectories of allergenic Cor a homologs in *C. avellana*, *C. americana*, *C. heterophylla*, and *C. mandshurica* genome assemblies. 52 Cor a orthologs were found in the analyzed species, and a recent tandem duplication of *Cor a 1* was found in *C. americana*. Three new gene models were predicted in *C. avellana* and *C. mandshurica* for *Cor a 16* and *Cor a 10*. Additionally, we identified 56 Cor a isoallergens, of which ten Cor a isoforms. Furthermore, phylogenetic analysis sheds light on the evolutionary dynamics of three hazelnut allergens revealing the evolutionary complexity of *Cor a 1*, *Cor a 2*, and *Cor a TLP* within the *Corylus* genus. A list of multiple gRNAs designed for the CRISPR-Cas9 system was provided for the singular and multiple silencing of Cor a homologs in each *Corylus* genome.

Conclusions This study enhances our knowledge on the evolutionary path of Cor a allergens among *Corylus* species and provides highly accurate on-target guides targeting hazelnut allergome.

Keywords *Corylus* Sp., Genome editing, Cor a protein, Phylogeny, Isoallergens, Isoforms, Multiplexing

*Correspondence:
Giuseppe Andolfo
giuseppe.andolfo@unina.it
[1]Department of Agricultural Sciences, University of Naples 'Federico II', Portici, Italy

Background

Hazelnuts, scientifically known as *Corylus* spp., belong to the family Betulaceae and are a source of valuable nutrients, which makes them one of the most nutritive among nuts [1]. Because of its nutritional properties, the worldwide economic relevance of hazelnuts reaches 2.3 billion (USD) of exports per year, with European countries considered the main importing countries [2]. Lamentably, hazelnuts also represent the main culprit of nut allergy in Europe, which could be justified by their high consumption rate [3]. The National Institute of Allergy and Infectious Diseases (NIAID) describes food allergy (FA) as an IgE-mediated response of the immune system to specific food proteins [4]. This process triggers the release of potent inflammatory mediators from mast cells and basophils in sensitive subjects, causing itching, hives, swelling, and, in severe cases, anaphylactic shock [5]. The allergens found in the *Corylus* genus have been extensively researched in *C. avellana*, leading to the identification and characterization of several hazelnut proteins [6]. So far, hazelnut allergens found in *C. avellana*, are *Cor a 1* (PR-10), *Cor a 2* (Profilin), *Cor a 6* (Isoflavone reductase), *Cor a 8* (Non-specific lipid transfer protein, nsLTP), *Cor a 9* (11 S globulin), *Cor a 10* (Heat shock protein, HSP), *Cor a 11* and *Cor a 16* (7 S globulins), *Cor a 12*, *Cor a 13* and *Cor a 15* (Oleosins), *Cor a 14* (2 S albumin), which have been included in the Allergen nomenclature (WHO/IUIS) database (https://allergen.org/), and *Cor a TLP* (thaumatin like-protein) (not included). *Cor a 1* is related to mild local hypersensitivity reactions and is heat labile [7], while the risk of systemic reactions increases with seed storage proteins (SSPs) *Cor a 9*, *Cor a 14*, and *Cor a 11*, and the lipid transfer protein (LTP) *Cor a 8* [8–11]. Otherwise, *Cor a 2*, *Cor a 6*, and *Cor a 10* are most abundant in pollen, and it is thought that sensitization to these proteins primarily occurs at the mucosal membranes of the respiratory system, leading to localized symptoms [12]. *Cor a 12* and *Cor a 13*, classified as oleosins, present an unclear clinical significance, while *Cor a 15* has been classified as a new allergen in pediatric patients [13]. More recently, *Cor a 16* was classified as a new hazelnut allergen [14]. Moreover, *Cor a TLP* described by Palacín et al. [15], showed cross-reactivity with sera from fruit-allergic patients, which present high sequence similarity with allergens *Mal d 2* (apple) and *Pru p 2* (peach) [12]. Modern biotechnological approaches such as CRISPR-based genome editing have revolutionized the field of plant breeding, offering precise and efficient tools for targeted modifications. The most popular CRISPR system utilizes the Cas9 nucleases and requires the design of highly efficient guide RNAs (gRNAs) to successfully generate mutated plants [16]. However, so far, CRISPR-Cas9 in hazelnut is impeded by the lack of efficient regeneration protocols and user-friendly online tools integrating hazelnut genomes for gRNA design [17]. Indeed, currently available tools for custom gRNA design require bioinformatics knowledge and command-line skills. Therefore, the design of efficient gRNAs could provide a further step toward the obtainment of hypoallergenic hazelnut crops [18]. Furthermore, the sequencing of several *Corylus* genomes in recent years [12, 19–24] allows the exploration of the variability of allergenic gene families and the identification of efficient on-target gRNAs. The task to specifically predict the gRNAs of Cor a genes directly within *Corylus* genomes of studied taxa has not yet been addressed. This work encompasses the comprehensive identification of hazelnut allergens in four different *Corylus* species, aiming to investigate the evolution events that occurred along the *Corylus* genus speciation. Moreover, we leverage computational analysis for the generation of efficient gRNAs for all 13 Cor a genes and relative orthologs and paralogs for singular and simultaneous targeting of allergenic isoforms.

Methods

Taxa data set

To investigate the evolution of Cor a genes along the *Corylus* plant species, we used the genomic data of four taxa (*Corylus avellana*, *Corylus americana*, *Corylus heterophylla*, and *Corylus mandshurica*) [12, 22–24]. The genome sequence, gene model annotation, and protein sequences were downloaded from three (Phytozome, GigaDB, and GenBank) public databases (Supplementary Table 1). Moreover, for comparative purposes, we also added 12 well-characterized reference Cor a genes (*Cor a 1*, *Cor a 2*, *Cor a 6*, *Cor a 8*, *Cor a 9*, *Cor a 10*, *Cor a 11*, *Cor a 12*, *Cor a 13*, *Cor a 14*, *Cor a 15* and *Cor a 16* (Supplementary Table 2) retrieved from the WHO/IUIS Allergen Nomenclature Database (http://allergen.org, accessed on 1 February 2024), and *Cor a TLP* [12].

Identification and manual curation of *Cor a* paralogs and orthologs

In order to retrieve predicted Cor a genes, we scanned the *C. avellana*, *C. americana*, *C. heterophylla*, and *C. mandshurica* proteome data set (Supplementary Table 1) with a well-known algorithm for inferring orthologs based on all-versus-all sequence comparison with BLAST [25], implemented in InParanoid (v4) [26] with default parameters (Supplementary Tables 3–5). We used a confidence score threshold of 1 to directly estimate gene–gene correspondences between the Cor a proteins. Cor a protein domain architecture was annotated using InterProScan [27] and Conserved Domain search [28] with default parameters. *Corylus* orthogroups were identified by Orthofinder v2.3.3 [29], using the default settings. To improve the gene prediction procedure for *Cor*

a 10 (GenBank ID: AJ295617) and *Cor a 16* (GenBank ID: ON060881) gene loci in *C. mandshurica* and in *C. avellana* genome assemblies, we used the AUGUSTUS-PPX (Protein Profile eXtension) module implemented in Augustus (v2.5.5) pipeline, with Arabidopsis parameters [30]. AUGUSTUS-PPX analysis used amino acidic profiles of conserved, ungapped blocks from the full-length protein alignments of *Cor a 16* and *Cor a 10* orthologs. The predicted genes encoded proteins with a confidence score > 0.

Multiple sequence alignment and phylogenetic analysis

Sequence similarities were determined by performing a ClustalW (v2) [31] multiple alignments with default settings, using the full-length protein sequences of annotated Cor a proteins as input. Phylogenetic analysis was performed by using newly identified Cor a homologs. The dataset was completed with the following proteins: *Cor a 1*, *Cor a 2*, *Cor a 6*, *Cor a 8*, *Cor a 9*, *Cor a 10*, *Cor a 11*, *Cor a 12*, *Cor a 13*, *Cor a 14*, *Cor a 15*, *Cor a 16* and *Cor a TPL* (Supplementary Table 2). Evolutionary relationships between Cor a proteins were inferred using the maximum likelihood method based on the Whelan and Goldman model [32], using the MEGA6 software (http://www.megasoftware.net, accessed on 1 February 2024) [33]. The model with the lowest BIC (Bayesian Information Criterion) score was considered to describe the substitution pattern. The bootstrap consensus tree, obtained from 100 replicates, was taken to represent the Cor a family phylogenetic history [34].

Estimation of average evolutionary divergence

The evolutionary divergence (π) was estimated as number of base substitutions per site from averaging over all sequence pairs. Standard error was obtained by a bootstrap procedure (100 replicates). Analyses were conducted using the Maximum Composite Likelihood model [35]. All positions containing gaps and missing data were eliminated. Selective pressures acting on the *Cor a TLP* genes were investigated by determining the nonsynonymous to synonymous nucleotide substitution (dN/dS) indicated as ω. Tests were conducted to estimate the evolution of each codon: positive (dN > dS); neutral (dN = dS); and negative (dN < dS). Analyses were conducted using the SLAC algorithm [36]). The Tajima's D test is to distinguish between genes evolving randomly ("neutrally") and those evolving under a non-random process, including directional selection or balancing selection and, gene family expansion or contraction using the MEGA6 software [37]. All the *Cor a TLP* -coding DNA sequences were aligned using ClustalW [31].

Single- and multi-targeting guide RNAs (gRNAs) design

Single- and multi-targeting guide RNAs (gRNAs) for the CRISPR-Cas9 system were designed using CRISPR-Local, a high-throughput tool [38]. Reference gRNAs database (RD)-build model was used to build the whole-genome gRNA database from reference genomes of *Corylus* spp. and corresponding annotation files. The RD-build program was executed with -U 15 -D 3 settings (Supplementary Tables 6–9). The screening of all possible on-target gRNAs was realized with the Rule Set 2 algorithm [39]. The seqMap program was used to predict the effects of each off-target site with the highest cutting frequency determination (CFD) score for each gRNA [40]. All target and off-target data determined across the entire genome were exported into RD. Database (DB)-search model was used to obtain sorted results from all annotated Cor a genes. Paralogs (PL)-search model was used to output gRNAs targeting Cor a gene homologs (Supplementary Table 10).

Validation of prediction results

All gRNAs predicted in *C. avellana* were used to perform molecular validation. DNA was extracted from leaf tissue of genotype *Corylus avellana* cv Mortarella using DNeasy Plant mini kits (Qiagen, Valencia, CA, USA). PCR was executed with 25 ng of genomic or complementary DNA, 10 pmol primers, 1 U of Taq DNA polymerase Kit (Invitrogen, Carlsbad, CA, USA), 10 pmol dNTPs, and 2 mM MgCl2 in 25 µl reaction volumes. Amplification was performed using the following cycling conditions: 1 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min 30 s at 60 °C and 2 min at 72 °C, with a final extension for 7 min at 72 °C. Amplicons were separated by electrophoresis on agarose gel (1.5%), and photographed by a GelDoc apparatus. Primers were designed with Primer3 (http://frodo.wi.mit.edu), with a length between 18 and 22 bp. The length of the amplified fragments and the Tm of the specific primers were reported in Supplementary Table 11. In particular, 13 out of 52 guides were used as forward primers in combination with two different reverse primers for on and off-targets in *C. avellana*, respectively (Supplementary Table 11).

Results

Allergenic genes annotation in four *Corylus* spp.

Thirteen *Corylus avellana* (Cor a) proteins were identified and reported as allergens triggering specific immunoglobulin E (IgE) reactivity (Supplementary Table 2). In order to identify their coding genes in the published genome assemblies of *Corylus* spp., we used the BLAST-based reciprocal best hits (RBHs) method. A total of 52 *Cor a* genes were discovered in four analyzed proteomes (Supplementary Tables 3, 4). Four *Corylus* species presented all the 13 *Cor a* orthologs located on nine

chromosomes (1, 2, 3, 4, 5, 6, 8, 9, and 11) suggesting that these allergens were preserved along the speciation *continuum* (Fig. 1). Moreover, the average amino acidic identity among *Corylus* orthologs was more than 90% (Supplementary Table 3). A lineage-specific micro-synteny profile for all allergens was observed in *C. avellana*, *C. americana* and *C. heterophylla*. Differently, *Cor a 8*, *Cor a 10*, *Cor a 13*, *Cor a 15*, *Cor a 16* and *Cor a TLP* in *C. mandshurica* were located on different chromosomes, as highlighted in Fig. 1. Two *Cor a 1* ortholog genes (CamerRush.01G109100.1 and CamerRush.01G109000.1), with confidence score = 1, were identified in *Corylus americana* genome. Despite this, CamerRush.01G109100.1 showed a higher global identity (97.5%) to *Cor a 1*. *Cor a 6* orthologs showed the highest identity (99.1%) among the four hazelnut species. While the wide diversity range was observed for *Cor a 1* (90.2%) and *Cor a 16* (90.4%) orthologs (Supplementary

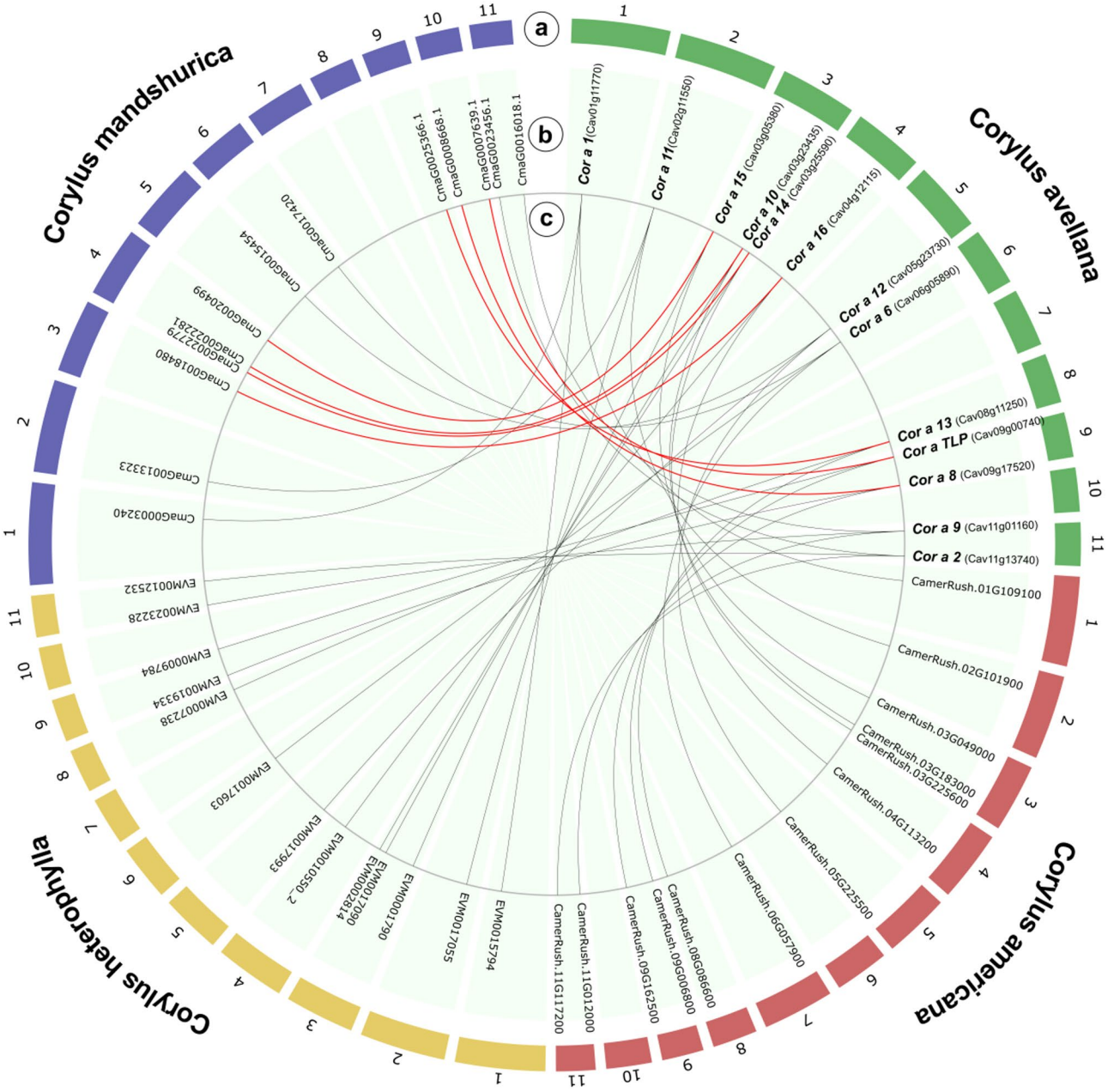

Fig. 1 Circos plot summarizes the genomic positions of hazelnut allergenic loci and their orthology relationships. Insert legend provides information for the data rings. (**a**) The 11 chromosomes of *C. avellana*, *C. americana*, *C. heterophylla*, and *C. mandshurica* genome assemblies. The length of each circle segment is proportionated to the size of chromosomes. (**b**) The gene IDs of allergens annotated in analyzed *Corylus* spp. (**c**) The orthology relationships between putative allergens, while red lines indicated no synthenic *Cor a* genes

Table 3). Unexpectedly, RBH analysis identified no *Cor a 16* locus among predicted *C. avellana* gene models [12]. Therefore, we executed an augustus-PPX prediction on *C. avellana* genome, and a new locus named Cav04g12115 with an amino acidic identity ~ 87% to *Cor a 16* gene (ID: ON060881) was predicted (Fig. 2A). Similarly, we have reannotated *Cor a 16* (ID: ON060881) and *Cor a 10* (ID: AJ295617) loci in *C. heterophylla* and *C. avellana* genomes, respectively. Initially, very low identity values (< 70%) were detected between homology gene pairs *Cor a 16*: ON060881-EVM0010550 and *Cor a 10*: AJ295617-Cav03g20430 (Fig. 2B and C). Our predicted *Cor a 16* (EVM0010550_2) in *C. mandshurica* was located on the same region of EVM0010550 gene model (Fig. 1B), while *Cor a 10* (Cav03g20435) locus resulted longer than ~ 1.2Kb compared to Cav03g20430 gene model (Fig. 2C). Hence, coding DNA sequences of re-annotated exon-intron structures showed higher amino acidic identities (> 90%) to *Cor a 10* and *Cor a 16* compared to protein sequences released by consortia (Fig. 1B).

Isoforms and isoallergens identification

Plant genomes frequently host gene groups that have evolved from a common ancestor following gene duplication events. To discover potential 'isoallergens' and 'isoforms' of Cor a genes, we have extensively analyzed the groups of closely related proteins identified by OrthoFinder [29]. Isoallergens were defined as proteins from the same species that show at least 67% amino acid identity, while isoforms are considered variants of the same allergen, typically with > 90% identity [41]. A total of 56 isoallergens (out of orthologs) found in four *Corylus* species showed average protein identities ≥ 80% to the reference Cor a allergens (Table 1). Isoallergens were found for *Cor a 1*, *Cor a 2*, *Cor a 9*, *Cor a 10*, *Cor a 13 and Cor a TLP*; while, *Cor a 6*, *Cor a 8*, *Cor a 11*, *Cor a 12*, *Cor a 14*, *Cor a 15*, and *Cor a 16* showed no paralogs. Among these, six *Cor a 9* and four *Cor a 13* proteins showed sequence identities higher than 93% and could be classified as isoforms (Table 1; Supplementary Table 5). Among the hazelnut isoallergens, major differences emerged for *Cor a 1* and *Cor a TLP*, which stood out with higher gene copy numbers in *C. avellana* compared to the other species. By contrast, the number of *Cor a 2* isoallergens was conserved among the four species.

Phylogenetic relationships between nut allergenic genes

Three *Cor a* loci (*Cor a 1*, *Cor a 2*, and Cor a TLP) were strongly duplicated in *Corylus* genomes. To retrace their evolutionary trajectories, we infer three phylogenetic analyses for ortholog and paralog loci. The evolutionary relationships were analyzed using a maximum likelihood method (Fig. 3). For comparative purposes, we included two well-characterized reference genes (*Cor*

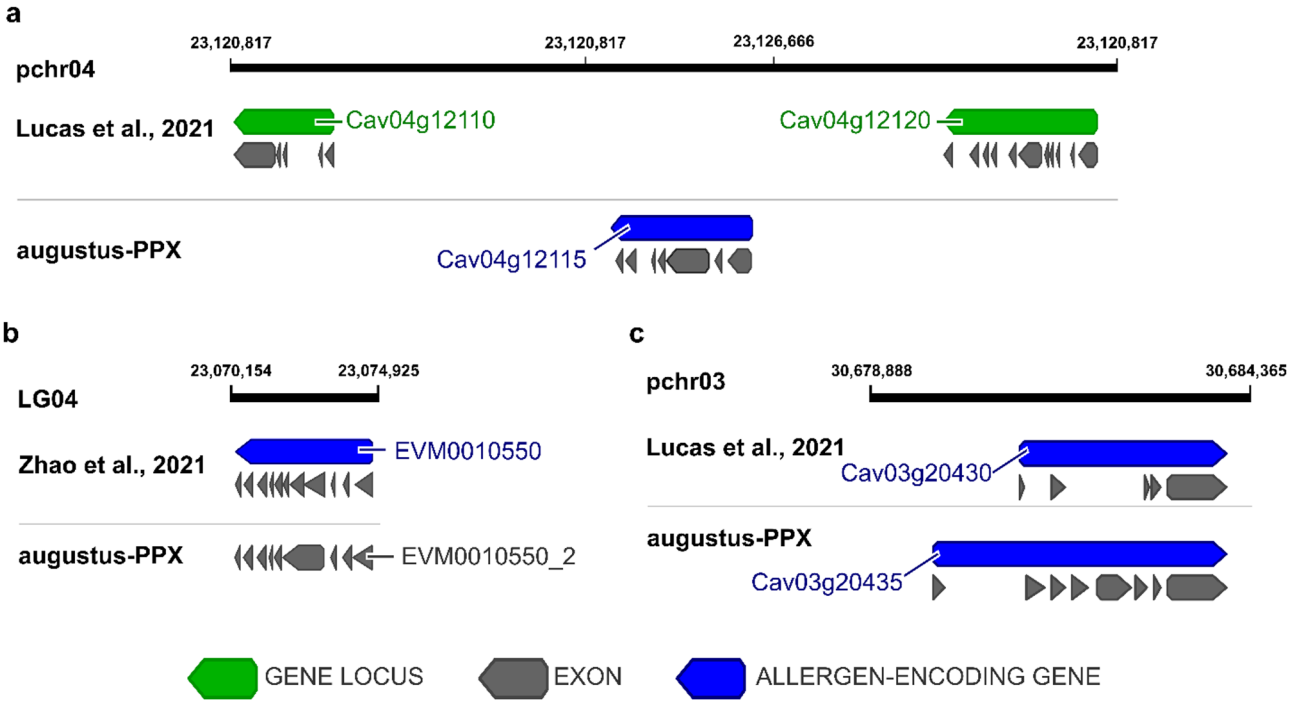

Fig. 2 Agustus-PPX *Cor a 16* and *Cor a 10* gene prediction in two hazelnut genome assemblies. (**a**) Genomic region including *Cor a 16* locus (Cav04g12115) on *Corylus avellana* chromosome 4. (**b**) Re-annotation of exon-intron structures erroneously predicted for *Cor a 16* (EVM0010550) on *Corylus mandshurica* linkage group (LG) 4. (**c**) Re-annotation of exon-intron structures erroneously predicted for *Cor a 10* (Cav03g20435) on *Corylus avellana* chromosome 3 (pchr = pseudo-chromosome; LG = linkage-group)

Table 1 Hazelnut isoallergenic genes annotated in four *Corylus* species

Short gene name	Number of isoallergenic genes[A]				Average identity (%)[B]
	C. avellana	C. americana	C. heterophylla	C. mandshurica	
Cor a 1	4 (81.8) [a]	4 (82.0)	2 (83.9)	3 (84.7)	84.7
Cor a 2	3 (83.6)	3 (77.6)	3 (82.8)	3 (81.7)	86.3
Cor a 6	1	1	1	1	99.1
Cor a 8	1	1	1	1	95.0
Cor a 9	2 (96.5)	2 (95.0)	1	2 (96.1)	96.4
Cor a 10	1	3 (82.2)	2 (81.7)	3 (85.1)	81.5
Cor a 11	1	1	1	1	95.3
Cor a 12	1	1	1	1	98.2
Cor a 13	1	1	2 (93.6)	2 (96.5)	96.8
Cor a 14	1	1	1	1	95.4
Cor a 15	1	1	1	1	97.9
Cor a 16	1	1	1	1	90.4
Cor a TLP	15 (73.2)	8 (86.8)	5 (80.9)	5 (82.0)	80.0

[a]Cor a homologs (paralogs and orthologs) predicted by OrthoFinder [29]. Paralogs identity (%) has been reported in brackets. [b]Amino acidic average identity of homologous gene groups

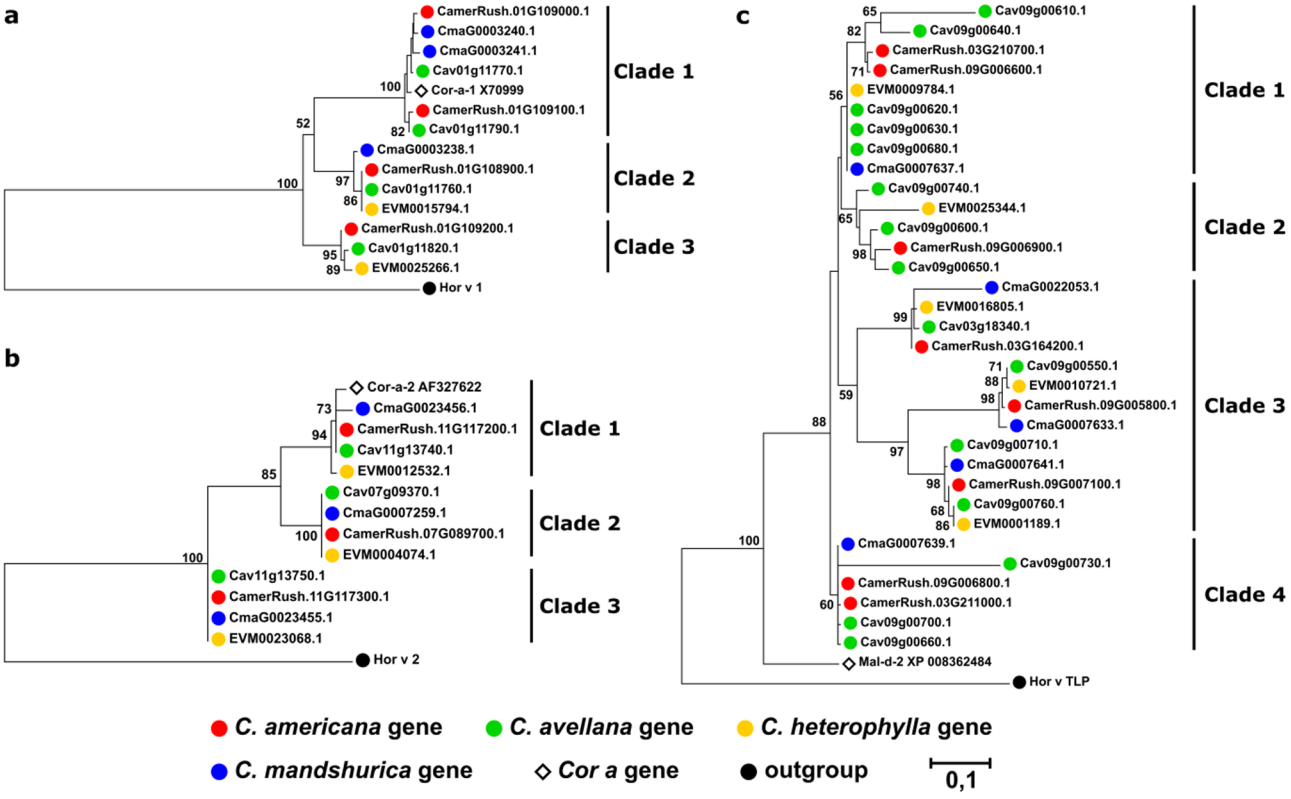

Fig. 3 Natural diversification of the Cor a 1 (**panel a**), Cor a 2 (**panel b**), and Cor a TLP (**panel c**) orthogroups identified in C. avellana, C. americana, C. heterophylla and C. mandshurica genomes. The evolutionary history of 58 Corylus genes was used together with three well-characterized allergenic genes to perform a maximum likelihood analysis. Labels showing the bootstrap values higher than 50 (out of 100) are indicated above the branches. The taxa to which the protein sequences belong are indicated by colored spots

a 1 and *Cor a 2*), the *Mal d 2* gene from *Malus domestica* (ID: XM008362484; Supplementary Table 2), and three out-group genes (Hor v 1, Hor v2 and Hor v TLP) identified in *Hordeum vulgare* (L.) (Fig. 3). A total of 13, 12, and 33 homologs to *Cor a 1*, *Cor a 2*, and *Cor a TLP* proteins, respectively, were aligned. The sequences were grouped into phylogenetic clades supported by bootstrap values ≥ 50%, allowing the definition of three, three, and four clades for *Cor a 1*, *Cor a 2* and *Cor a TLP* allergens, respectively (Fig. 3). Clade 1 of *Cor a 1* orthogroup showed two homologs in C. *avellana*, C. *mandshurica*, and C. *americana* (Fig. 3A). Interestingly, no

Table 2 Estimation of Non synonymous and synonymous substitutions mean dissimilarity (ω = dN /dS) for each Corylus TLP gene family and phylogentic clade

Aligned CDS sequences	*n*. TLP sequences	Π[A]	Tajima's Test (D)	SLAC method[B]		
				(ω)	PSCs (*n*.)	NSCs (*n*.)
TLP clade 1	9	0.070 (0.209)[C]	-1.035	0.411	0	3
TLP clade 2	5	0.054 (0.006)	-0.464	0.507	0	0
TLP clade 3	13	0.152 (0.033)	1.091	0.303	0	19
TLP clade 4	6	0.071 (0.024)	-1.226	0.552	0	0
C. americana TLP family	8	0.130 (0.018)	0.152	0.286	0	13
C. avellana TLP family	15	0.257 (0.034)	-0.599	0.388	0	24
C. heterophylla TLP family	5	0.167 (0.013)	0.201	0.296	0	6
C. mandshurica TLP family	5	0.187 (0.011)	0.372	0.289	0	10

In the table are reported: number of TLP sequences, the average evolutionary divergence (π) and the direction and magnitude of natural selection for each analyzed Cor a TLP gene family and phylogentic clade

[A]In brackets was reported the standard error estimate by a bootstrap procedure (100 replicates)

[B]Number of positive (PSCs: positively selected codon sites) and negative (NSCs: negatively selected codon sites) codons statistically significant ($P < 0.1$) in TLP protein-coding genes

[C]Poorly aligned sites were removed using GBlocks

C. heterophylla genes were included in clade 1, and the most similar to *Cor a 1* (EVM0015794.1) was collapsed in clade 2. Moreover, no *C. mandshurica* genes were clustered in clade 3 (Fig. 3A). The Panel B of the phylogenetic tree provides insights into the evolutionary relationships among *Cor a 2* homologs. A total of 12 genes were included in this analysis, and in each clade, a member of each species was grouped (Fig. 3B), suggesting gene duplication events probably occurred before *Corylus* speciation. Finally, panel C summarizes the evolutionary analysis of homologs to *Cor a TLP*. As expected, *Mal d* was significantly separated from hazelnut phylogenetic clades (Fig. 3C). Interestingly, a total of 18 *Mal d* homologs were found in *C. avellana*, of which 17 were distributed in a short genomic region of chromosome 9 (~ 112 kb) (Fig. 2C; Supplementary Fig. 1). Similarly, *C. americana* showed eight homologs distributed among all the clades. Differently, *C. mandshurica* and *C. heterophylla* homologs were included in three clades.

Selection pressure acting on *Cor* a *TLP* genes

To infer the direction and magnitude of natural selection acting on *Corylus TLP* genes, the non-synonymous substitution (dN) and synonymous substitution (dS) ratio was used. Neutrality tests, performed on the Cor a TLP gene families and TLP phylogentic clades, yielded average Tajima's D and ω (dN/dS) values ranging from 0.669 to – 1.226 and from 0.552 to 0.102, respectively (Table 2). Single codon analysis highlighted the presence of 6, 10, 13 and 24 negatively selected sites in the *C. heterophylla*, *C. mandshurica*, *C. americana* and *C. avellana Cor a TLP* families, respectively (Table 2). Interestingly, the negative Tajima's D value found for the TLP family of C. avellana confirms the low frequency of polymorphisms and a gene family size expansion (Table 2). Globally, a restricted evolutionary divergence was observed into the phylogentic clades, indeed the average nucleotide diversity (π) ranged from 0.019 to 0.152 (Table 2). Finally, Tajima's highest negative D values were found for C1 (–1.035) and C4 (–1.226) clades (Table 2), in which the highest number of paralogs were revealed. Hovewer, confirming low frequency of polymorphisms, indicating gene family size expansion.

Identification of single and multiple gRNAs for Cor a allergens

CRISPR-Cas9 genome engineering offers a real opportunity for developing hypoallergenic cultures, and gRNA design is a pivotal step. To provide high-throughput gRNAs considering the genetic variations among *Corylus* genomes with different genetic backgrounds, we designed gRNAs with the highest CFD score for *Corylus* sp. allergenic loci [38]. For instance, the gRNAs targeting *Cor a 1* allergens in *C. avellana*, *C. heterophylla*, and *C. mandshurica* showed high on-target scores (0.746, 0.758, 0.813, respectively), indicating a strong potential for targeting this allergen. In contrast, the gRNA sequence in *C. americana* had a slightly lower on-target score (0.686). For *Cor a 2*, all the analyzed species had similar on-target scores (0.705, 0.695, 0.678 and 0.629). However, the off-target scores for *C. americana* and *C. heterophylla* were barely higher (0.017 and 0.016). Focusing on the off-target scores, the lowest values were found for the Cor a 11 in *C. avellana* and *C. heterophylla*, *Cor a 6* in *C. heterophylla*, and *Cor a 13* in *C. americana* (off-target value: 0). However, we found gRNA sequences with relatively low off-target value for several genes (Supplementary Tables 6–9). Indeed, the highest off-target value was related to the gRNA identified on *Cor a 1* allergen in *C. avellana* (0.26), while all the other gRNA sequences showed lower off-target values. Moreover, the gRNA sequences for *Cor a 6*, *Cor a 8*, *Cor a 9*, *Cor a 10*, *Cor a 11*, *Cor a 12*, *Cor a*

13, *Cor a 14*, *Cor a 15*, and *Cor a TLP* showed relatively consistent on-target scores among the four species (all above of 0.62), indicating a comparable potential for targeting these allergens.

The gRNA specificity was molecularly validated for predicted *C. avellana* guides (Supplementary Table 7). Revealed amplicons on target genes confirmed gRNA-matching specificity (Supplementary Fig. 2). Finally, we identified multiple CRISPR-Cas9-gRNAs for silencing species-specific isoallergens (Supplementary Table 10). This allowed the selection of seven gRNAs targeting 15 Cor a paralogs. In particular, we identified common gRNAs for the simultaneous silencing of all *Cor a 13* and two *Cor a 1* homologs in *C. heterophylla* (Supplementary Table 10). Similarly, *Cor a 9* homologs identified in *C. avellana* and in *C. americana* could be simultaneously targeted using a common gRNA (Supplementary Table 10). Finally, three gRNAs were identified in *C. mandshurica* to simultaneously silence all *Cor a 1*, *Cor a 9* and *Cor a 13* homologs (Supplementary Table 10).

Discussion

Genomes of hazelnut species *Corylus avellana*, *Corylus americana*, *Corylus heterophylla*, and *Corylus mandshurica* were recently assembled and released, paving new ways for genetic research within this genus [12, 22–24]. These species are particularly appreciated for their nutritional values [42], and genetic studies have given particular emphasis to the identification and characterization of hazelnut Cor a proteins, which have been described as the primary allergenic proteins responsible for IgE-mediated allergy induction [18]. Recently, CRISPR-Cas9 technology revolutionized plant breeding programs by enabling the direct manipulation of specific genes, enhancing speed and precision in obtaining improved genotypes [43]. However, so far, no studies have been reported about the application of CRISPR-Cas9 in hazelnuts, most likely due to the absence of efficient regeneration protocols and available user-friendly software for gRNA design [17, 44]. This lack of knowledge and resources hampered progress in exploring hazelnut genetic modifications for developing hypoallergenic crops [18].

In light of the limited genome editing studies and challenges associated with available bioinformatics tools for gRNA design [45, 46], the investigation of evolutionary relationships among allergenic Cor a genes in four *Corylus* species and identification of gRNAs lists for targeting *Corylus* allergens individually and in combination represent a further step towards genetic manipulation and biotechnological-mediated breeding in hazelnut species [43]. The conservation of Cor a genes throughout the evolutionary divergence of the *Corylus* species, was initially confirmed with the identification of 52 Cor a genes across *C. avellana*, *C. americana*, *C. heterophylla*, and *C. mandshurica* genomes. Moreover, gene duplication of *Cor a 1* was identified in *C. americana*, suggesting a recent and potentially functional duplication event [47]. Notably, we identified the *Cor a 16* locus on chromosome four in *C. avellana* cv. 'Tumbol' named Cav04g12115, which was unpredicted by Lucas et al. [12]. Meanwhile, new exon-intron structures were predicted for *Cor a 16* in *C. mandshurica* and *Cor a 10* in *C. avellana*. These newly predicted genes and structures [48] expand our knowledge of allergen genetic positions and clarify the structural organization of allergenic genes, enhancing the understanding of hazelnuts' allergenic profile, which is a crucial step for accurately characterizing gene functional domains for future genome editing studies [49]. Moreover, the investigation of orthologous relationships and isoallergenic gene variations across different *Corylus* species highlighted high amino acidic sequence similarities among identified hazelnut allergens, while sequence variations could be attributed to the divergent speciation path [46, 50]. Understanding gene relationships and sequence variations enhances insights into the genetic basis of hazelnut allergenicity and potential cross-reactivity sites. Furthermore, the presence of hazelnut isoallergens with varying copy numbers across the genomes suggested genetic diversity in the allergenic composition of different *Corylus* species [51]. For instance, *Cor a 1* and *Cor a TLP* exhibited differences in the number of isoallergenic genes. *Cor a TLP* showed almost the double copy number in *C. avellana* compared to the other *Corylus* genomes. In accordance with Lucas et al. [12], we inferred that the *Cor a TLP* was encoded by Cav09g00740. These proteins are members of pathogenesis response proteins-5 (PR-5), which are known to be involved in plant hypersensitive response to infections and, due to their high sequence similarity, may all be considered as potential allergens in hazelnuts [15, 52, 53]. Indeed, almost all (14 out of 15) of the *Cor a TLP* homologs formed a gene cluster spanning a genomic region of ~112 kb of chromosome 9, originating by tandem gene duplications [54].

Hence, the genetic diversity and evolutionary history of *Cor a 1*, *Cor a 2*, and *Cor a TLP* allergens have been further elucidated across the different *Corylus* species by the phylogenetic analysis, which revealed evolutionary relationships among the four hazelnut genomes. By employing a phylogenetic reconstruction, we identified complex relationships between these genes and the established isoallergen nomenclature [55]. Three maximum likelihood analyses collapsed the 58 Cor a proteins into ten phylogenetic clades [12]. Our findings indicated that *C. avellana* and *C. mandshurica* showed multiple homologs closely related to *Cor a 1*, while in *C. heterophylla* orthologs to *Cor a 1* showed a more divergent evolutionary path [56]. By contrast, the analysis of *Cor a 2* homologs revealed species-specific patterns, with each species

having representative genes in different clades, indicating unique evolutionary trajectories [57]. We can speculate that gene duplication events of the *Cor a 2* homologs preceded the speciation events among the four species. Furthermore, the investigation of *Cor a TLP* orthologs demonstrated variations in gene numbers and distribution among the species, with tandem duplications observed on chromosome 9 in *C. avellana*, indicating the functional relevance of *Cor a TLP* in plant defense and potentially in allergenic induction [12, 58]. Therefore, despite the distinct evolution paths, all identified orthologs conserved high sequence similarities and should be considered as potential allergens. Indeed, the understanding of Cor a sequence similarities provides insights into the evolutionary history of hazelnut allergenic genes across different *Corylus* species, shedding light on the genetic relationships, divergence, and conservation of these allergenic molecules. The ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide substitutions among coding DNA sequences of phylogenetic clades was below 1. This indicates that negative selection has been acting against extreme polymorphic variants and in favor of purifying selection [59]. Moreover, Tajima's negative D values of phylogentic clades grouping highest number of Cor a TLP paralogs supported the hypothesis of recent gene duplication events [60].

Considering the limited research focusing on the specific application of CRISPR-Cas9 for targeting hazelnut allergens, we identified single and multi-targeting gRNAs for Cor a allergens across different *Corylus* species, contributing to addressing the underexplored field of CRISPR-Cas9 targeting hazelnut allergens. By employing CRISPR-Local, a high-throughput tool for gRNA design, we generated a comprehensive list of Cor a gRNAs for all analyzed genomes. Our analysis prioritized gRNAs based on their on-target and off-target scores. The screened gRNAs showed an on-target score > 0.66, with a putative cutting efficiency ranging from 50 to 75% [36, 55]. In general, identified gRNAs showed high potential for targeting Cor a allergens across the species [38]. The importance of utilizing gRNAs with low off-target effects is widely known [61–64]. Here, most of the identified gRNAs showed low off-target values, suggesting a low risk of off-target effects for the majority of gRNAs. Like most algorithms, the cut specificity score is not always a perfect predictor of off-target effects. As a general rule, however, the higher the CFD score, the more likely that a guide matches an off-target site [36]. Our work provides gRNAs without off-target sites or with low off-target scores, as also established in other studies [55, 60, 65]. In our work, we also provide multiple gRNAs designed to target paralog genes [66]. The analysis led to the selection of different gRNAs capable of targeting fifteen Cor a paralogs that could be used to edit multiple genes in *Corylus* species, as also evidenced in recent studies employing CRISPR-Cas9 multiplex genome editing to target simultaneously different allergenic genes and obtaining less allergenic plants [67]. Interestingly, in *Corylus mandshurica*, a gRNA was designed to target three different *Cor a 1* paralogs, demonstrating the feasibility of targeting multiple genes simultaneously [66, 68]. Most of the detected common sequences exhibited high on-target scores, indicating a potential high efficiency in targeting the selected genes. Despite the absence of efficient regeneration protocols, recent studies have successfully demonstrated the use of Agrobacterium-mediated transient transformations for expressing the CRISPR-Cas9 system in plants, leading to efficient genome editing. In particular, Hamada et al., [69] reported a protocol to perform genome-editing in plants, without requiring callus culture and regeneration through the biolistic delivery of CRISPR-Cas9 expression cassette in embryos. Moreover, other studies generated stable genome-edited plants using particle bombardment in embryos demonstrating an efficient method to obtain CRISPR-Cas9 edited plants in species reluctant to callus culture [70]. Furthermore, the recent application of CRISPR-Cas9 system in nut species such as walnut [71] indicates a growing trend toward overcoming challenges in nut-related species. Therefore, our findings highlight the importance of in-depth research for efficient gRNA design for the CRISPR-Cas9 system, paving the way for future research to develop hypoallergenic hazelnut varieties through targeted gene editing.

Conclusion

Our findings elucidate Cor a genes evolutionary divergence and similarities within the *Corylus* genus, and provide insights into the genomic basis of hazelnut allergome. This research contributed to discovering new potential allergens in *Corylus*, pointing out new Cor a orthologs and paralogs. Furthermore, gRNAs identified in this study could be considered ideal candidates for targeting single and multiple Cor a homologs. In conclusion, this research is a first step for developing potential hypoallergenic *Corylus* varieties.

Abbreviations

Cor a	Corylus avellana
FA	Food allergy
nsLTP	Non-specific lipid transfer protein
HSP	Heat shock protein
PR	Pathogen-related
TLP	Thaumatin like-protein
SSPs	Seed storage proteins
Mal d	Malus domestica
Pru p	Prunus persica
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR-associated protein 9
gRNA	Guide RNA
PPX	Protein Profile eXtension
BIC	Bayesian Information Criterion

CFD	Cutting frequency determination
RD	Reference gRNAs database
DB	Database
PL	Paralogs
IgE	Immunoglobulin E
RBHs	Reciprocal best hits
LG	Linkage group
Hor v	Hordeum vulgare

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06685-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements
Plant material of ***C. avellana*** cv. Mortarella was kindly provided by Dr. Antonio Napolitano of the Campania hazelnut consortium (CO.N.CI) (Nola, Italy).
We thank Dr. Barbara Auzino, of Plant Genetics and Biotechnology Unit of Department of Agricultural Science at University of Naples "Federico II", for excellent technical assistance.

Author contributions
C.G.A. inspired the study, analyzed and interpreted the results, wrote the article's draft, and revised the final version. G.A. conceived and designed the study, interpreted the results, performed the analyses, generated the figures, and revised the final version of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the PNRR-PRIN Fund, CUP: E53D23020840001P (CucuR-gene) to G.A.

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 April 2024 / Accepted: 6 May 2025
Published online: 19 May 2025

References

1. CI&DETS/ESAV, Polytechnic Institute of Viseu, Department of Food Industry, Viseu, Portugal, Guiné, Correia RPF. PMR. Hazelnut: A Valuable Resource. ETP Int J Food Eng. 2020;:67–72.
2. Aydoğan M, Demiryürek K, Abacı Nİ. World hazelnut trade networks. Acta Hortic. 2018;:429–36.
3. McWilliam V, Koplin J, Lodge C, Tang M, Dharmage S, Allen K. The prevalence of tree nut allergy: A systematic review. Curr Allergy Asthma Rep. 2015;15:54.
4. Kumar S, Verma AK, Das M, Dwivedi PD. Molecular mechanisms of IgE mediated food allergy. Int Immunopharmacol. 2012;13:432–9.
5. He S, Zhang H, Zeng X, Chen D, Yang P. Mast cells and basophils are essential for allergies: mechanisms of allergic inflammation and a proposed procedure for diagnosis. Acta Pharmacol Sin. 2013;34:1270–83.
6. Blanc F, Bernard H, Ah-Leung S, Przybylski-Nicaise L, Skov PS, Purohit A, et al. Further studies on the biological activity of hazelnut allergens. Clin Transl Allergy. 2015;5:26.
7. Masthoff LJ, Hoff R, Verhoeckx KCM, Van Os-Medendorp H, Michelsen-Huisman A, Baumert JL, et al. A systematic review of the effect of thermal processing on the allergenicity of tree nuts. Allergy. 2013;68:983–93.
8. Lauer I, Foetisch K, Kolarich D, Ballmer-Weber BK, Conti A, Altmann F, et al. Hazelnut (*Corylus avellana*) vicilin Cor a 11: molecular characterization of a glycoprotein and its allergenic activity. Biochem J. 2004;383:327–34.
9. Offermann LR, Bublin M, Perdue ML, Pfeifer S, Dubiela P, Borowski T, et al. Structural and functional characterization of the hazelnut allergen Cor a 8. J Agric Food Chem. 2015;63:9150–8.
10. Datema MR, Zuidmeer-Jongejan L, Asero R, Barreales L, Belohlavkova S, De Blay F, et al. Hazelnut allergy across Europe dissected molecularly: A EuroPrevall outpatient clinic survey. J Allergy Clin Immunol. 2015;136:382–91.
11. Valcour A, Lidholm J, Borres MP, Hamilton RG. Sensitization profiles to hazelnut allergens across the united States. Ann Allergy Asthma Immunol. 2019;122:111–e1161.
12. Lucas SJ, Kahraman K, Avşar B, Buggs RJA, Bilge I. A chromosome-scale genome assembly of European Hazel (*Corylus Avellana* L.) reveals targets for crop improvement. Plant J. 2021;105:1413–30.
13. Nebbia S, Lamberti C, Cirrincione S, Acquadro A, Abbà S, Ciuffo M, et al. Oleosin Cor a 15 is a novel allergen for Italian hazelnut allergic children. Pediatr Allergy Immunol. 2021;32:1743–55.
14. Mattsson L, Holmqvist M, Porsch H, Larsson H, Pontoppidan B, Valcour A, et al. A new vicilin-like allergen in hazelnut giving rise to a spectrum of IgE-binding low-molecular-weight N-terminal fragments. Clin Exp Allergy. 2022;52:1208–12.
15. Palacín A, Rivas LA, Gómez-Casado C, Aguirre J, Tordesillas L, Bartra J, et al. The involvement of Thaumatin-Like proteins in plant food Cross-Reactivity: A multicenter study using a specific protein microarray. PLoS ONE. 2012;7:e44088.
16. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun. 2018;9:3048.
17. Mehlenbacher SA, Molnar TJ. Hazelnut Breeding. In: Goldman I, editor. Plant Breeding Reviews. 1st edition. Wiley; 2021. pp. 9–141.
18. Evangelista AR, Amoroso CG, Nitride C, Andolfo G. Seed storage allergens tackled via next-generation research assistant. Front Food Sci Technol. 2024;4:1372770.
19. Rowley ER, VanBuren R, Bryant DW, Priest HD, Mehlenbacher SA, Mockler TC. A draft genome and High-Density genetic map of European hazelnut (*Corylus Avellana* L.). 2018.
20. Pavese V, Cavalet-Giorsa E, Barchi L, Acquadro A, Torello Marinoni D, Portis E et al. Whole-genome assembly of *Corylus avellana* Cv Tonda gentile Delle Langhe using linked-reads (10X Genomics). G3 GenesGenomesGenetics. 2021;11:jkab152.
21. Liu J, Wei H, Zhang X, He H, Cheng Y, Wang D. Chromosome-Level genome assembly and hazelomics database construction provides insights into unsaturated fatty acid synthesis and cold resistance in hazelnut (Corylus heterophylla). Front Plant Sci. 2021;12:766548.
22. Zhao X, Yi L, Ren Y, Li J, Ren W, Hou Z, et al. Chromosome-scale genome assembly of the yellow Nutsedge (*Cyperus esculentus*). Genome Biol Evol. 2023;15:evad027.
23. Li Y, Sun P, Lu Z, Chen J, Wang Z, Du X, et al. The *Corylus Mandshurica* genome provides insights into the evolution of Betulaceae genomes and hazelnut breeding. Hortic Res. 2021;8:54.
24. Brainard SH, Sanders DM, Bruna T, Shu S, Dawson JC. The first two chromosome-scale genome assemblies of American hazelnut enable comparative genomic analysis of the genus *Corylus*. Plant Biotechnol J. 2024;22:472–83.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
26. Remm M, Storm CEV, Sonnhammer ELL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol. 2001;314:1041–52.
27. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
28. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45:D200–3.
29. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.

30. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27:757–63.
31. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. Bioinformatics. 2007;23:2947–8.
32. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a Maximum-Likelihood approach. Mol Biol Evol. 2001;18:691–9.
33. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38:3022–7.
34. Felsenstein J, Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783–91.
35. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci. 2004;101:11030–5.
36. Pond SLK, Frost SDW. A genetic algorithm approach to detecting Lineage-Specific variation in selection pressure. Mol Biol Evol. 2005;22:478–85.
37. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
38. Sun J, Liu H, Liu J, Cheng S, Peng Y, Zhang Q, et al. CRISPR-Local: a local single-guide RNA (sgRNA) design tool for non-reference plant genomes. Bioinformatics. 2019;35:2501–3.
39. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized SgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34:184–91.
40. Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics. 2008;24:2395–6.
41. Chapman MD, Pomés A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. J Allergy Clin Immunol. 2007;119:414–20.
42. Mehlenbacher SA. HAZELNUTS (CORYLUS). Acta Hortic. 1991;:791–838.
43. Savadi S, Muralidhara BM, Preethi P. Advances in genomics of cashew tree: molecular tools and strategies for accelerated breeding. Tree Genet Genomes. 2020;16:61.
44. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
45. Ferrucci A, Lupo M, Turco S, Pavese V, Marinoni DT, Botta R, et al. A roadmap of tissue culture and biotechnology in European hazelnut (Corylus Avellana L). Plant Physiol Biochem. 2023;205:108167.
46. Naeem M, Alkhnbashi OS. Current bioinformatics tools to optimize CRISPR/Cas9 experiments to reduce Off-Target effects. Int J Mol Sci. 2023;24:6261.
47. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of gene duplication in plants. Plant Physiol. 2016;171:2294–316.
48. Andolfo G, Dohm JC, Himmelbauer H. Prediction of NB-LRR resistance genes based on full-length sequence homology. Plant J. 2022;110:1592–602.
49. Savadi S, Mangalassery S, Sandesh MS. Advances in genomics and genome editing for breeding next generation of fruit and nut crops. Genomics. 2021;113:3718–34.
50. Yang Z, Zhao T-T, Ma Q-H, Liang L-S, Wang G-X. Resolving the speciation patterns and evolutionary history of the intercontinental disjunct genus Corylus (Betulaceae) using Genome-Wide SNPs. Front Plant Sci. 2018;9:1386.
51. Fiore MC, Marchese A, Mauceri A, Digangi I, Scialabba A. Diversity assessment and DNA-Based fingerprinting of Sicilian hazelnut (Corylus Avellana L.) germplasm. Plants. 2022;11:631.
52. Hoffmann-Sommergruber K. Pathogenesis-related (PR)-proteins identified as allergens. Biochem Soc Trans. 2002;30:930–5.
53. Amoroso CG, D'Esposito D, Aiese Cigliano R, Ercolano MR. Comparison of tomato transcriptomic profiles reveals overlapping patterns in abiotic and biotic stress responses. Int J Mol Sci. 2023;24:4061.
54. Andolfo G, Sanseverino W, Rombauts S, Van De Peer Y, Bradeen JM, Carputo D, et al. Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important *Solanum* R locus dynamics. New Phytol. 2013;197:223–37.
55. Costa J, Mafra I, Carrapatoso I, Oliveira MBPP. Hazelnut allergens: molecular characterization, detection, and clinical relevance. Crit Rev Food Sci Nutr. 2016;56:2579–605.
56. Yang Z, Ma W, Wang L, Yang X, Zhao T, Liang L, et al. Population genomics reveals demographic history and selection signatures of hazelnut (*Corylus*). Hortic Res. 2023;10:uhad065.
57. Helmstetter AJ, Buggs RJA, Lucas SJ. Repeated long-distance dispersal and convergent evolution in Hazel. Sci Rep. 2019;9:16016.
58. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, et al. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 2019;20:38.
59. Iovieno P, Andolfo G, Schiavulli A, Catalano D, Ricciardi L, Frusciante L, et al. Structure, evolution and functional inference on the mildew locus O (MLO) gene family in three cultivated Cucurbitaceae spp. BMC Genomics. 2015;16:1112.
60. Andolfo G, Sánchez CS, Cañizares J, Pico MB, Ercolano MR. Large-scale gene gains and losses molded the NLR defense arsenal during the Cucurbita evolution. Planta. 2021;254:82.
61. Manghwar H, Li B, Ding X, Hussain A, Lindsey K, Zhang X, et al. CRISPR/Cas systems in genome editing: methodologies and tools for SgRNA design, Off-Target evaluation, and strategies to mitigate Off-Target effects. Adv Sci. 2020;7:1902312.
62. Amoroso CG, Panthee DR, Andolfo G, Ramìrez FP, Ercolano MR. Genomic tools for improving tomato to biotic stress resistance. In: Kole C, editor. Genomic designing for biotic stress resistant vegetable crops. Cham: Springer International Publishing; 2022. pp. 1–35.
63. Huang Y, Shang M, Liu T, Wang K. High-throughput methods for genome editing: the more the better. Plant Physiol. 2022;188:1731–45.
64. Veluchamy A, Teles K, Fischle W. CRISPR-broad: combined design of multi-targeting gRNAs and broad, multiplex target finding. Sci Rep. 2023;13:19717.
65. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016;17:148.
66. Bewg WP, Harding SA, Engle NL, Vaidya BN, Zhou R, Reeves J, et al. Multiplex knockout of trichome-regulating MYB duplicates in hybrid Poplar using a single gRNA. Plant Physiol. 2022;189:516–26.
67. Camerlengo F, Frittelli A, Sparks C, Doherty A, Martignago D, Larré C, et al. CRISPR-Cas9 multiplex editing of the α-Amylase/Trypsin inhibitor genes to reduce allergen proteins in durum wheat. Front Sustain Food Syst. 2020;4:104.
68. Hu Y, Patra P, Pisanty O, Shafir A, Belew ZM, Binenbaum J, et al. Multi-Knock—a multi-targeted genome-scale CRISPR toolbox to overcome functional redundancy in plants. Nat Plants. 2023;9:572–87.
69. Hamada H, Liu Y, Nagira Y, Miki R, Taoka N, Imai R. Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. Sci Rep. 2018;8:14422.
70. Liu Y, Luo W, Linghu Q, Abe F, Hisano H, Sato K, et al. In planta genome editing in commercial wheat varieties. Front Plant Sci. 2021;12:648841.
71. Chang Y, Song X, Zhang Q, Zhang P, Lei X, Pei D. Robust CRISPR/Cas9 mediated gene editing of JrWOX11 manipulated adventitious rooting and vegetative growth in a nut tree species of walnut. Sci Hortic. 2022;303:111199.

Publisher's note