International Journal of Environmental Sciences & Natural Resources
ISSN: 2572-1119

Juniper Publishers
key to the Researchers

Research Article
Volume 30 Issue 4 - June 2022
DOI: 10.19080/IJESNR.2022.30.556296

Int J Environ Sci Nat Res

Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids

François Anthony*, Gennady Ershov and Alexis Dereeper

Institut de Recherche pour le Dévelopement (IRD), UMR Interactions Plantes Microorganismes Environnement (IPME), France

Submission: June 20, 2022; **Published:** July 04, 2022

***Corresponding author:** François Anthony, Institut de Recherche pour le Dévelopement (IRD), UMR Interactions Plantes Microorganismes Environnement (IPME), 34394 Montpellier Cedex 5, France

Abstract

Most resistance (*R*) genes contain a nucleotide-binding site (NBS) domain characterised by several conserved motifs. Recent whole-genome sequencing data gave us the opportunity to explore the evolutionary history of NBS genes in euasterid clades, including tomato, potato, and for the first time coffee and monkey-flower. Two eurosid species (*Arabidopsis, grapevine*) were used as outgroups. A workflow based on hidden Markov model searches was designed to identify genes with a complete NBS domain. The coffee genome has the highest number of NBS genes reported in plants. Eight conserved motifs were easily identified in the NBS domain of euasterids, including the P-loop, RNBS-A, kinase-2, RNBS-B, RNBS-C, GLPL, RNBS-D and MHDV. Differences were detected between the composition, clustering and origin of the NBS genes in euasterid species and those in eurosid species. The study of complex clusters with at least ten NBS genes revealed several patterns of tandem duplication with transfer to a contiguous site or to a more distant one. Tandem duplication appeared to be a continuous mechanism over time since eight gene pairs had zero diversity. The study of orthologous relationships revealed that most NBS genes arose from duplication of paralogues in a few orthologous groups. Evolution of NBS genes was inferred from an analysis of synonymous and non-synonymous substitutions in the orthologous groups. Traces of 11 major large-scale duplication events were observed and dated in the euasterid genomes. Specific ancestral signatures of large-scale duplication events were identified in the genomes.

Keywords: Euasterid; Evolution; Nucleotide-binding site domain; Orthology; Resistance gene analogues

Introduction

Pathogens have been a major threat in agriculture and breeding programmes for decades [1,2]. Changes in pathogen distribution and disease severity could be a response to certain aspects of climate change, and may increase crop yield losses [3,4]. In the face of pathogen diversity (bacteria, fungi, insects, nematodes, oomycetes, virus), plants have set up a sophisticated defence system to detect attacks and activate innate immune responses [5,6]. Plant responses are governed by nonspecific transmembrane pattern recognition receptors (PRRs) and cytoplasmic immune receptors encoded by resistance (*R*) genes. The products of *R* genes play a critical role in recognizing proteins (effectors), which are introduced into plant cells by pathogens, and in triggering various defence responses including localized cell death [7,8]. Despite the diversity of pathogen attacks, R proteins share a high degree of homology and present a number of conserved motifs and domains among plant species [9]. Since the discovery of the first *R* gene in a plant in 1992, the *Hm1* gene in maize [10], more than 100 *R* genes have been cloned from different plant species (http://prgdb.crg.eu/wiki/). The majority encode a nucleotide-binding site (NBS) domain and leucine-rich repeat (LRR) domains. The NBS domain forms part of a larger domain known as NB-ARC, which is present in the human apoptotic protease-activating factor-1 (APAF-1), the *Caenorhabditis elegans* death-4 protein (CED-4) and plant R proteins [11,12]. This domain contains the three-layered α-β fold and subsequent short α-helical region characteristic of the AAA+ ATPase domain superfamily [13]. In NBS-LRR-encoding genes, the LRR domain interacts with the product of pathogen *AVR* genes directly or indirectly and is thus involved in recognising R protein specificity [5,14,15]. The deduced NBS-LRR proteins can be divided into two subfamilies based on their N-terminal features [16]. (i) TIR-NBS-LRR (TNL) proteins contain an N-terminal domain which is similar to both the intracellular signalling domains of *Drosophila* Toll and the mammalian Interleukin-1 receptor (TIR). (ii) non-TIR-NBS-LRR

proteins often present a predicted N-terminal coiled-coil (CC) structure and are collectively named non-TIR proteins. In addition to architectural differences, NBS-LRR-encoding genes in these subfamilies differ considerably in their phyletic distribution and downstream signalling pathways, suggesting possible divergence in their functions [6,17].

Several conserved motifs have been identified throughout the NBS domain of non-TIR and TIR proteins, including P-loop, RNBS-A, kinase-2, RNBS-B, RNBS-C, GLPL, RNBS-D and MHD [16]. The functional importance of these motifs is documented by the effect of mutations of motif residues which lead to either loss-of-function or auto-activation (i.e., a hypersensitive response in the absence of a pathogen or AVR protein) of the NBS-LRR protein [9]. The RNBS-A, kinase-2 and RNBS-D motifs display different features in non-TIR and TIR proteins, and these can be used as specific signatures to separate the two subfamilies of NBS-encoding (NBS) genes. Based on these highly conserved motifs, *R* gene analogues (RGAs) have been discovered using different genome-wide approaches with degenerate primers [18,19], BLAST [20,21] or HMMER searches [22,23]. Hundreds of NBS-LRR genes are generally detected in plant genomes, underlining their duplication dynamics and the key role played by these genes.

In recent years, genomic organisation, phylogenetic reconstruction and evolutionary patterns of NBS-LRR-encoding genes have been investigated extensively in plant genomes [24-28]. The *Arabidopsis* genome has become a reference for many genomic studies, including *R* gene analyses, because it was the first plant genome to be sequenced [29]. The main findings pointed to a major role of clustering in *R* gene expansion and to some basic evolutionary mechanisms, such as interlocus gene conversions within clustered *R* genes, and tandem and segmental duplication [30,31]. Although *R* genes in a cluster often display evidence for intergenic exchange, paralogues can also diverge considerably from one another [32]. Therefore, the evolution of *R* genes in plant genomes appears to be a complex process.

Recent advances in whole-genome sequencing gave us the opportunity to explore RGAs in the euasterid clade. Whole-genome data from coffee [33] and monkey-flower [34] were used for the first time in a comparative study of *R* genes, along with those from potato [35] and tomato [36]. Potato and tomato are members of the Solanaceae family, which contains several major food crops. Coffee belongs to the Rubiaceae family, which is also a family of economic importance. Monkey-flower belongs to an intermediate lineage (Figure 1). Data from two genomes belonging to the eurosid clade, *Arabidopsis thaliana* [29] and grapevine [37], were included in the analysis as outgroups. We designed a workflow to identify *R* genes presenting a complete NBS domain and used the same procedure for all genomes. We then conducted a comparative analysis to characterise composition, clustering and selection pressures. Our findings enabled us to retrace the evolutionary history of NBS genes in euasterids and provided insights into genome evolution.

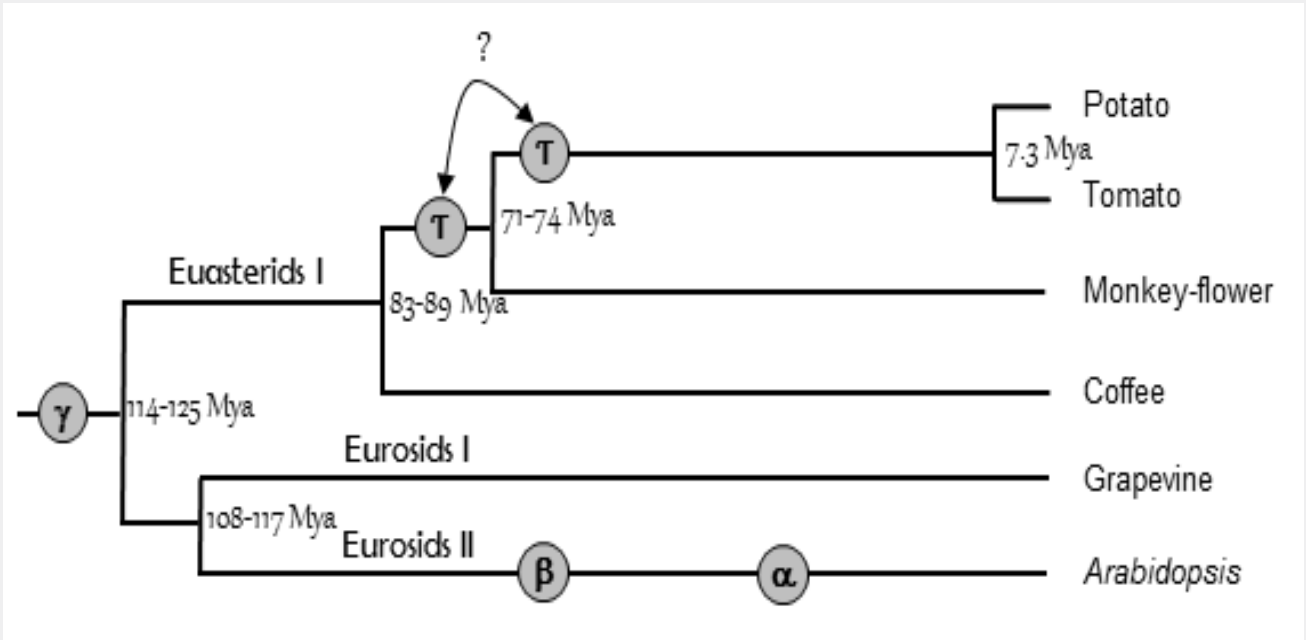


Figure 1: Dendogram representing the phylogenetic relationships between euasterids and eurosids. The time-scale of divergence was published in Wikström et al. (2001). Polyploidization events are indicated according to Tang et al. (2008) and The Tomato Genome Consortium (2012).

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Materials and Methods

Identification of NBS genes

Predicted gene sequences were downloaded from http://coffee-genome.org/ for *Coffea canephora*, http://www.phytozome.net/ for *Mimulus guttatus*, http://solanaceae.plantbiology.msu.edu/ for *Solanum tuberosum*, http://solgenomics.net/ for *S. lycopersicum*, http://www.genoscope.cns.fr/spip/Vitis-vinifera-whole-genome.html for *Vitis vinifera*, and http://www.arabidopsis.org/ for *Arabidopsis thaliana*. Only a scaffold assembly was available for the monkey-flower genome. For each genome, predicted protein sequences matching the Pfam NBS family (NB-ARC domain PF00931) (http://pfam.sanger.ac.uk) were identified using HMMER search 3.0 (http://hmmer.janelia.org) with an E-value cut-off of 10-60 (Figure S1). These sequences were aligned using HMMER align and used to construct a NBS hidden Markov model (HMM) specific to each genome using HMMER build 3.0. New HMM searches (E-value cut-off of 0.01) led us to select specific sets of NBS candidate genes. The candidate genes were submitted to the National Center for Biotechnology Information's (NCBI) Conserved Domains tool [38] for validation of the presence of an NBS domain (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Only sequences with a complete NBS domain at both N- and C-termini were retained for subsequent analyses. The identified sequences are available at the GreenPhyl website (http://www.greenphyl.org/cgi-bin/index.cgi) [39].

Subsequently, the NCBI-CD tool was used to determine whether the corresponding NBS proteins presented TIR and LRR motifs. CC domains were specifically identified using COILS/PCOILS version 2.2 [40] ($P \geq 0.9$) and PAIRCOIL2 [41] ($P \leq 0.025$). The NBS genes were then classified according to detailed information on protein motifs and domains. Multiple alignments of amino acid sequences were performed for each genome using MAFFT [42]. The resulting alignments were manually cleaned to remove sequences with poor ends and incomplete motifs using MEGA (Molecular Evolutionary Genetics Analysis) version 5.2 [43].

Prediction of conserved motif structures

The structural diversity of the NBS proteins identified was studied through analysis of conserved motifs and domains. Non-TIR-NBS-LRR (CNL) and TIR-NBS-LRR (TNL) protein sequences were characterised using MEME (Multiple Expectation maximization for Motif Elucidation) [44] with specific conditions: (i) the optimum motif width was set at ≥ 6 and ≤ 50 and (ii) the maximum number of motifs to find was successively set at 15 and 30. The consensus sequences of each motif were aligned using MUSCLE [45]. Similarity and dissimilarity among sequences were further checked in the complete alignment regions. Core sequences containing amino acids conserved in all euasterid consensus sequences were determined for each motif.

Gene clusters

As clustering of *R* genes frequently occurs in plant genomes, physical clusters of NBS genes were investigated in the six plant genomes studied. We used the same parameters to define a cluster as Holub [46]: a chromosome region which contains more than three genes in a distance of less than 200kb. In monkey-flower, this approach was used on scaffolds since pseudomolecules were not available. Complex clusters, i.e., clusters containing at least 10 NBS genes, were further investigated. Their nucleotide sequences were aligned by MUSCLE and the sequences corresponding to the NBS domains were extracted using the previously built alignments of amino acid sequences. Both ends of the NBS domains were defined using the core sequences of the P-loop and MHD motifs previously determined in euasterids (see section headed "Prediction of conserved motif structures"), e.g., IVGxGGxGKTT and MHDxxxD for CNL proteins respectively. Nucleotide diversity (π) between pairs of NBS domain sequences belonging to the same cluster were calculated using EggLib (Evolutionary Genetics and Genomics Library) tools version 2.1.5 [47]. Gene families were defined based on an π criterion < 20% [48]. Recent duplications of NBS genes were then identified assuming that low levels of diversity (< 5%) between nucleotide sequences corresponded to the most recent duplications [49]. Levels of nucleotide diversity of less than 10% and 20% were also used to retrace the chronology of duplication events in multi-gene families.

Orthology

A clear distinction between paralogues and orthologues is critical for the construction of a robust evolutionary classification of genes [50]. The NBS genes from the six plant genomes were assigned to orthologous groups (orthogroups) using the OrthoMCL database 5 [51]. The sequences within the orthogroups were aligned using MAFFT. Their NBS domains were then cut following the method described above (see section headed "Gene clusters"). *R* genes (http://prgdb.crg.eu/wiki/) belonging to the same orthogroups were included in the datasets. Phylogenetic trees were constructed based on the neighbour-joining (NJ) method [52] with a Jones-Taylor-Thornton correction model implemented in MEGA. Branch lengths were assigned by pairwise calculation of the genetic distances. The confidence at each branching node was assessed by bootstrap analysis [53] with 500 replicates. Missing data were treated by pairwise deletion of the gaps. The distribution of the genes in the orthogroups was then used to classify the clusters as homogeneous when all sequences shared a common ancestor, or as heterogeneous in the case of more distantly related NBS genes [54].

Ks and Ka analyses

The timing of the divergence of homologous genes and the selective pressures on duplicated genes were estimated by calculating the synonymous (Ks) and nonsynonymous (Ka)

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

substitutions per site between NBS genes. The nucleotide sequences within orthogroups were aligned using MAFFT and the sequences corresponding to the NBS domains were cut. Pairs with a nucleotide diversity (π) < 0.05 were eliminated because their Ks (as denominator) were too small to obtain a reliable estimate [55]. Ks and Ka values from pairwise alignments of NBS domain were calculated using the Nei-Gojobori [56] method of model averaging using KaKs-calculator 1.2 software [57]. Assuming that synonymous changes approximate the neutral rate of molecular evolution [58], the relative age-distribution of gene duplicates within a genome can be inferred indirectly from the distribution of Ks [59]. A Ka:Ks ratio >1 generally indicates a positive or diversifying selection for amino acid substitution. Conversely, a Ka:Ks ratio <1 corresponds to a negative or purifying selection [60].

Results

Identification and classification of NBS genes

A total of 5,998 candidate genes encoding NBS domains were detected in the six plant genomes with our HMM workflow (Figure S1) but only 2,151 genes were validated using NCBI-CD (Table S1). Both the absolute number and relative proportion of NBS genes were significantly higher in coffee (670, 2.62%) and grapevine (459, 1.51%) than in the other genomes (113-373, 0.42-1.11%). Among the validated genes, around 20% presented an incomplete NBS domain according to NCBI-CD and were not retained for further analyses. Differences in the composition of conserved domains associated with the NBS domain were observed between euasterids and eurosids. In euasterids, both the number and proportion of NBS genes containing CC domains were much higher than those containing TIR domains. Around 52.9% of NBS genes presented CC domains in euasterids while only 4.8% presented TIR domains. No TIR domains were detected in monkey-flower and only a few (9) in coffee. The proportion of NBS genes with LRR regions ranged from 13.4 to 24.5% in euasterid genomes. This proportion was higher in grapevine (51.7%) and *Arabidopsis* (64.8%). Including other NCBI conserved domains, the number of conserved domains associated with NBS ones ranged from six to 14 in euasterid genomes while there were 18 in *Arabidopsis* and 22 in grapevine (Table S2). Out of a total of 39 conserved domains, only 11 were shared by euasterids and eurosids. Except the LRR_8 motif, the most frequently conserved domains associated with NBS domains were related to the ATPase family (AAA and AAA_16), which can act as DNA helicases and transcription factors.

Table 1: Subclasses of NBS-encoding genes in six plant genomes.

Predicted Protein Domains*	Letter Code	Coffee	Monkey -flower	Potato	Tomato	Grapevine	Arabidopsis
Total NBS-encoding genes		**559**	**268**	**265**	**151**	**404**	**108**
Subtotal NBS-LRR type genes		124	36	65	28	209	70
TIR-NBS-LRR	TNL	3		19	8	33	38
TIR-NBS-LRR-TIR	TNLT					3	
TIR-NBS-LRR-TIR-NBS-LRR	TNLTNL				1	1	
TIR-NBS-TIR-NBS-LRR	TNTNL						1
TIR-TIR-NBS-LRR	TTNL						1
LRR-TIR-NBS-LRR	LTNL					1	
TIR-D-TIR-NBS-LRR	TDTNL						1
TIR-NBS-NBS-LRR	TNNL			1			
W-TIR-NBS-LRR	WTNL						1
NBS-LRR-TIR	NLT					1	1
CC-NBS-LRR	CNL	65	26	19	8	80	21
CC-NBS-CC-NBS-LRR	CNCNL					1	
CC-NBS-CC-NBS-LRR-NBS-LRR	CNCNLNL					1	
NBS-LRR	NL	55	10	25	10	86	6
NBS-LRR-NBS-LRR	NLNL					2	
X-NBS-LRR	XNL	1		1	1		
Subtotal NBS type genes		435	232	200	123	195	38
TIR-NBS	TN	1		14	4	23	7
TIR-NBS-TIR	TNT				1	2	
TIR-TIR-NBS	TTN						1

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

TIR-LRR-TIR-NBS	TLTN					1	
CC-NBS	CN	259	112	104	70	82	22
CC-NBS-CC-NBS	CNCN	1		1			
CC-NBS-NBS-NBS	CNNN					1	
CC-LRR-NBS	CLN					2	
CC-LRR-CC-NBS	CLCN					1	
LRR-CC-NBS	LCN					1	
NBS-CC-NBS	NCN	1					
CC-D-NBS	CDN			2			
D-CC-NBS	DCN			1			
NBS	N	170	120	73	42	78	7
NBS-NBS	NN					2	
D-NBS	DN	3		5	6		
X-NBS	XN					2	1
No subclasses	33	10	5	12	10	21	13

* CC coiled-coil domain, D DUF3542 domain, LRR leucine-rich repeat domain, NBS nucleotide binding site, TIR Toll/Interleukin-1-receptor, W WRKY domain, X RPW8 gene of resistance.

Table 2: Major motif sequences of NBS domains in CNL and TNL proteins of six plant genomes. Core motifs (i.e., amino acids conserved in all euasterids consensus sequences) are indicated in bold characters.

CNL Proteins			
Motif	**Genome**	**Consensus Sequence**	**Sequence Expression**
P-loop	Coffee	VIS**IVG**M**GG**L**GKTTLA**QKV**YN**	V[IV][SP][IV]VGMGG[LI]GKTTLA[QK][KL][VI][YF]N
	Monkey-flower	LP**IVG**V**GG**L**GKTTLA**QLV**YN**D	[LV][PS]I[VI]G[VM]GGLGKTTLA[QK]L[VA]YND
	Potato	IP**IVG**M**GG**L**GKTTLA**KAV**YN**D	[IV][PG][IV]VGMGG[LV]GKTTL[AV]K[AR][VIL]YND
	Tomato	VCIIG**IVG**A**GG**I**GKTTLA**QNI**YN**E	[VP][CS]I[IV]G[IV][VW]GAGG[IV]GKTTL[AV]Q[NL][IL][YN]N[ED]
	Grapevine	VIG**IVG**M**GG**V**GKTTLA**QLI**YN**	[VI]I[GSP][IL][VY]GMGG[VL]GKTTL[AL][QK][LK][IV][YN]N
	Arabidopsis	EDGVGIMGLY**G**M**GG**V**GKTTL**LTQI**N**	ED[GE][VI]G[IT][ML]GL[YH]GMGGVGKTTLLT[QK][IL][NH]
RNBS-A-non TIR	Coffee	DDRVKNH**F**DxKA**W**V**C**VSDxFD	[DN]DR[VI][KR][NK]HF[DE]x[KR]AWVCVS[DEQ]xFD
	Monkey-flower	RVVEH**F**DKRI**W**V**C**VSDNFDEKEIAKAIIE	RVV[EKN]HF[DE][KL]R[IA]W[VI]CVSD[NP]F[DN]E[KV][ET][IL][AL]K[AG][IM][IV]E
	Potato 1	VKKH**F**GLKA**W**F**C**VSEAYDAFRITKGLLQEIGSF-DLKVDDNLNQLQVKLKE	V[KQ][KN]HF[GV]LKAWFCVSEAYDAFRITKGLLQEIGS[FT]DLK[-VA]DDNLNQLQVKLKE
	Potato 2	IEKH**F**EKRV**W**L**C**LPEMSETKSFLELILESLTER-KVEVQSRDIIVKKLQDE	I[EK][KQ][HQT]FEKRVWLCLPEMSETKSFL[EQ][LQ]IL[EQ]SLT[EK]RK[VL][EK]VQ[SRT]RD[IL]IV[KM][KT]L[QR]DE
	Tomato	GF**F**DIKI**W**I**C**VSNDFD	[GH][FH]FD[IK][KR]IW[IV][CY]VS[NR][DP]FD
	Grapevine	SKD**F**DVVI**W**V**C**VSDEFDLEKIQKAILNKL	[SK]K[DH]FD[VL][VR][IA]WV[CV]VS[DK][EP]F[DN][LVI]E[KR][IV][QT][KE][AV]IL[NE][KAS][LIV]
	Arabidopsis	NKFSELGGG**F**DIVI**W**VVVS	NKF[SV][EK]LG[GS][GE]FD[IV]VIW[VI]VVS
Kinase-2	Coffee	GKRYLLV**LDD**V**W**NED	GK[RK][YF]L[LI]VLDDVWNE[DE]
	Monkey-flower	GKKFLLV**LDD**V**W**NEDQTKWSEL	[GN]K[KR][FY]LLVLDDVW[NT]ED[QD][TE]KW[SE][EP]L
	Potato 1	V**LDD**V**W**NDNYNEWDDL	VLDDVWN[DE][ND]Y[NS][EAK]WDDL
	Potato 2	KKYLLV**LDD**L**W**RVDSTLWDEF	[KR][KMR]YLLVLDD[LFM]WRVD[SL][TIP][LSV]W[DHNV]EF
	Tomato	FLLI**LDD**V**W**EEDD	FL[LI][IV]LDDVW[ENS]E[DI]D
	Grapevine	KDLNEKQxKIxEVLKGKKFLLV**LDD**V**W**EE	[KN][DST]L[NDE][EQ][KL][QA]x[KE][IL]x[EKR]VLK[GT]K[KR]F[LV]L[VL]LDD[VI]W[EN][ER]
	Arabidopsis	ASDIYNVLKKKRFVLL**LDD**I**W**EKVDLEAI	A[SV]DI[YH][NR]VL[KR][KR]K[RK]FVLLLDD[IL]W[ES][KE]VDL[ET][AKE]I

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

RNBS-B	Coffee	xGAK**G**SK**I**LV**TTR**SED**V**ATxM	xGAKGS[KRW]I[LI][VL]TTR[SN][EK][DR]VATx[MV]
	Monkey-flower	KNALACGST**G**SS**I**IV**TTR**LKK**V**ADIMGT	[KR]N[AV]L[AK]CG[SG]TGS[SK]I[IL]VTTR[LN][KE][KR]VA[DI][IM]MGT
	Potato 1	RNLFVQGDI**G**SK**I**IV**TTR**KES**V**ALIMGNEQ-ISMD	RN[LV][FL]VQGDIGS[KR]IIVTTR[KS]ESVA[LS][IM]MG[NS]E[QA]IS[MV][DGL]
	Potato 2	VDTLRGINTSR**G**NC**I**LV**TTR**MEQ**V**ASTVA	[VM]D[TS]LRG[IV]NTSRGN[CF]IL[VM]TTRM[EK]QVAS[TI]VA
	Tomato	**G**SK**I**I**T**S**R**SLD**V**CKIMG	[GR][SG][KR][IVM][IL][IV]T[ST]R[SF][LET][DEGKR]V[CAV][KGRS][IKMQR][MLV][GK]
	Grapevine	PDGAN**G**SK**I**IV**TTR**SED**V**ASxMRA	PDG[AQ]N[GK]SK[IV][IV][VF]TTR[SN]E[DN]V[AC]SxM[RG]A
	Arabidopsis	PYPSREN**G**CKVAF**TTR**SKE**V**C	P[YP]P[ST]RE[NK][GK][CS]K[VI][AV]FTTRS[KLR][ED]VC
RNBS-C	Coffee	xTSxPHxLGx**L**SDDDC**W**SLFEKKAFGGGE	xTSx[PT][HY]xLGxLS[DE][DES]D[CS]W[LS][LI][FL][QE][KR]KAF[GA]G[GKR][ES]
	Monkey-flower	IHYLKG**L**SDEHC**W**MLLRERAF	[IPA]H[YH]L[KG]GLSDE[HE][CS]W[ML]L[LFM]R[ER][RIK]A[FL]
	Potato	**L**SEDHS**W**SLFKRHAFENMDPP	LSE[DE]H[SC]WS[LI]FK[RQ][HR]A[FG][EDV][ND]M[DE][V][PE]
	Tomato	HKVxT**L**DEDES**W**ALFMEKAGD	[HM]K[VL]xTL[DS]E[DN][ED][SC]WALFM[EK][KN]A[GF][DS]
	Grapevine	HLKC**L**SWEDC**W**SLFxKKAFEN	[HK][LV][KE]CLSW[ED][DE][CAS]WSLFxK[KH][AV][FG]E[ND]
	Arabidopsis	GRMGVDKPMEVQC**L**EPDDA**W**	GRMG[VAD][DHE][KD]P[MI][EK]V[QS]CL[ES][PE][DNE][DE]AW
GLPL	Coffee	LEEIGKK**I**AKK**C**G**GLPLA**AKVI**G**G**LLR**F**K**	L[EK]E[IL]GK[KE]I[ALV]K[KR]CGGLPLA[AI][KS][VTA][IL]GGLLRFK
	Monkey-flower	EAIGKQ**I**ANK**C**K**GLPLA**AKTL**G**G**LL-RF**K**RTEEEWNYVLESEIWELPEEET	E[AN]IGK[QE]I[AV][NK]KC[KA]G[LV]PLAAK[TA]LG[GS]LL-RFK[RN]T[EL][EKN]EW[NE][YNS]V[LK][EN]SEIWEL[PE][EQ][EV]E[TD]
	Potato	LEEMGKQ**I**VEK**C**K**GLPLA**LKTL**G**G**LLRS**K**	[LI]E[ES][MVI]GK[QR]I[VA][EK][KE]C[KQ]GLPLA[LA][KS][TV][LI][GA][GS][LM]LR[SG]K
	Tomato	EIEDIAKK**I**ARE**C**D**GLPLA**IITI**G**S**LLR**G**K**ND-VEEWEDVL	[EN][IL][EQ][DP][IL][AG]K[KR][IV]AR[EK]CDGLPLA[IA][IK][TV][ILM][GA][ST][LS][LM]R[GF]KNDVEEWED[VA][LQ]
	Grapevine	K**I**VKK**C**K**GLPLA**AKTL**G**R**LLRS**K**KTPEEW	K[IV][VA][KE][KE]CKGLPLA[AL][KI]T[LI]G[RG][LA][LM][RA]SKKT[PE]EEW
	Arabidopsis	SHPDIPELARIVAAK**C**C**GLPLA**LNVI**G**ETMAC**K-**RTVQEWRHAIDVLTSYA	S[HD]PDIPELA[RK][IK]VA[AKQ]KC[CR]GLPLALNVIGE[TA]M[AS][CS]K[RE][TM][VI]QEWRHAI[DN]VL[TN][SR][YS]A
RNBS-D	Coffee	LPPxLKK**C**FA**Y**CSIF**PKD**FEI	[LP]PPxLK[KR]CF[AL]Y[CL][SA][IV][FY]P[KE]D[FSY]E[IM]
	Monkey-flower	ILPALLLSYHHLPPALKQ**C**FA**Y**CAVF**PKD**TxIR-KEELIFMWMAHGYISSK	[ILV][LF]P[AH]L[LR]LSY[HN][HE]L[PS][PL][AS]L[KR][QR]CF[AS]YC[AG][VI]FP[KE]D[TS]xI[RD][KV][ED]ELIFMWMA[HM]G[YF][IL][SG][SP][KN]
	Potato	LKR**C**FA**Y**FAIF**PKD**YEFEKDQLIHLWIAN-GLLPP	LK[RK]CF[AS]Y[FC]A[IM][FY]PKD[YF][EP]F[EG]K[DE]Q[LV]I[HQ]LW[IM]A[NE]G[LF][LV]P[PQ]
	Tomato	PPDLQR**C**F**L**YCSLY**PKD**IEID	PP[DH][LI][QK]RCFLYCSL[YF]P[KA]D[IY][EP][IT][DP]
	Grapevine	**C**FA**Y**CSIF**PKD**YEIEKEELILLWMAEGFLQES	CF[AL]YC[SA][IL]FP[KE]DYE[IF]EK[EK]ELILLW[MI][AG]EG[FL][LI][QD][EQS]S
	Arabidopsis	**C**F**L**YCALF**P**E**D**YEIEKEKLIEYWICEGFIDE	CFLYC[AS]LFPEDY[EK]I[EK]KE[KD]L[IV][ED]YWI[CG]EG[FI]I[DN][EGP]
MHD	Coffee	SCK**MHD**LVH**D**LAQSVSKEEC	SC[KR]MHDLV[HR]D[LFM]A[QL][SF][VIK][ST]KE[EK][CN]
	Monkey-flower	**MHD**LVH**D**LAQSIMENKGPGMK	MHD[LI][VI]HD[LF]AQ[SF][IL][MR][EK]N[KV][GI][PS][GR][MT][KQ]
	Potato	FK**MHD**LVN**D**LA	[FCV][KL]MHD[LV]V[NHR]D[LV]A
	Tomato	CWWAEDCLGEHDTYEEAYNRGITMIEELKDA-CLLEIEAHDCVK**MHD**VIR**D**	CWWAE[DG][CF]LGEHDTYE[EN][AV]YN[RT]GIT[MT]IE[ET]LKD[AV]CLLE[IK][EH][AN][HL]D[CS]VK[ML]HDV[IV]RD
	Grapevine	FK**MHD**LIR**D**LAQWVA	[FV][KV]MHD[LV][IV][RHN]D[LM]A[QL]W[VIL][AS]
	Arabidopsis	VK**MHD**VVREMALWIASDLGKQKENFI	VKMHDVVREMALWIASD[LF]GKQKEN[FCI][IV]
CNL Proteins			
Motif	**Genome**	**Consensus Sequence**	**Sequence Expression**

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

P-loop	Coffee	QVGIDSRVHKVNALLNLGSDEVHFI**GIWG**MS-**GIGKTT**I**AKA**VFNRISIHF	[QP]VGI[DS]SR[VL][HD][KI][VL][NL][AR][LK]L[NE][LV][GK][SG][DGN][EGK]V[HQR][FI]IG[IF][WH]GM[SG]GIGKTT[IL]A[KRW]A[VL]F[ND][RK][IL][SV][IST]HF
	Potato	RIV**GIWG**IG**GIGKTT**I**AKA**IFD	RIVGIWG[IM]GG[IV]GKTTIA[KR]AI[FY]D
	Tomato	VRMI**GIWG**MG**GIGKTT**L**AKA**VYNQLFxQF	VR[MI][IV]GI[WY]GMGG[IV]GK[TS]T[LI]A[KR]A[VI][YF][ND]Q[LI]FxQF
	Grapevine	VRMV**GI**Y**G**IG**GIGKTT**I**AKA**IYNEISHQFEGSSF	VR[MV][VI]GI[YWC]G[IM]GGIGKTT[IL]A[KR]A[IV]YN[EK]IS[HY]QF[ED]G[SAC]SF
	Arabidopsis	VRMV**GI**W**G**PA**GIGKTT**I**ARA**LYNQLSSxF	V[RK][MI][VI]GIWG[PM][ASP]GIGKTTIARAL[YFH][NS][QR]LSSxF
RNBS-A-TIR	Coffee	EGAI**FL**HE**V**R**E**QSKSLEILQ	EGAIFLH[ED]VR[EKQ]QS[KE]SL[EGK][INL]LQ
	Potato	QFKASC**FL**AD**V**K**E**NA	[QK]F[KE][AG][SA]CFLA[DN][VI]KEN[AKS]
	Tomato	EGSC**FL**AD**VRE**	[EDK]GS[CS]FL[ADS][DN]V[RK]E
	Grapevine	GLLQLQQQL**L**HDI**LE**	[GD]L[LI]QLQ[QK][QEK]LL[HS][DGQ][IL]L[EK]
	Arabidopsis	GLDEYGLKLHLQEQL**L**SKILNQKDIKIxHLGV	G[LS]D[ED]Y[GS][LA]KL[HR]LQ[EK]Q[LF]LS[KE]I[LI]N[QH]K[DG][IM][KR][IV]xHLG[VA]
Kinase-2	Coffee	CD**KKVLIVLDDVD**HL**DQL**DA**LAG**	[CH][DRY]K[KP]VL[ILV]VLDD[VI]D[HD][LA][DNS]QL[DEN]ALAG
	Potato	**KKVLIVLDDI**D**H**R**D**H**L**EY**LAG**	[KM]KVLIVLDD[IV]DHRD[HQ]L[ED]YLAG
	Tomato	YEGKRLIKERLGSM**KVLIVLDDV**D**D**RD**QL**E-A**LAG**	[YAD][EK]G[KV][RS][LI]I[KQ]ERLGS[MR][KR]VLIVLDDVD[DH]R[DS]Q[LI]E[AY]LAG
	Grapevine	K**KVLIVLDD**VD**D**LKQ**L**EY**LAG**	K[KR]VL[IV][VI][LF]D[DN]V[DN]D[LP][KT][QI]L[EK]YL[AV][GE]
	Arabidopsis	IEERLKDQ**KVLI**I**LDDVD**	[IV][EK]ERLKD[QK]KVLI[IV]LDDVD
RNBS-B	Coffee	GHD**WFGA****GSRII**I**TT**KNKHLLVTHEVD	[GKM][HR][DEH]WF[GY][ADE]GSRIIITT[KR][NT]K[HE][LV]L[VP][TQ][HN][EGI]VD
	Potato	DVG**WFG**N**GSRII**V**TT**RNKHL	D[VL][GD]WFGNGSR[IV][IV][VA]TTR[ND][KR][HQ]L
	Tomato	ERD**WFG**S**GSRII**I**TT**R	[EG]R[DES]WFGSGSRII[IV]TTR
	Grapevine	NHD**WFG**P**GSRII**I**TT**RDKHL	[NE][HQ]DWFGP[GK]S[RT]IIIT[TS]R[DN]K[HQR][LV]
	Arabidopsis	**FG**P**GSRII**V**TT**EDKQLLKAHGINNIY	FGPGSRII[VI]TT[EQ]DK[QE][LI]LK[AQ]HG[IV]NN[IV]Y
RNBS-C	Coffee	Q**LF**SS**HAFK**EDY**PA**ED	[QC][LY][FI][SP][SWY][HN]AF[KG][EKR][DE][YK]P[AN][EKM][DG]
	Potato	YEVSTLPDHEAMK**LF**NQ**HAFK**KEV**P**D**E**C	Y[EK]V[SP]TL[PL]D[HD][ED]A[MI][KQ]LF[NS]Q[HY]AF[KR][KE]EVPDE[CR]
	Tomato	YEVKLLTDDEAIQLFSxHAFKKExPPED	YEVK[LP]LTD[DN]E[AS]IQLF[SN]xHAF[KG]KExP[PD][EK]D
	Grapevine	HGVDALYEVKKLNYKEAIELFSLYAFKQN	[HY][GE]VDALYE[VA][KE][KE][LF]NY[KDE]EAI[EQ]LFSL[YHW]A[FL]K[QH]N
	Arabidopsis	VDFPSxEEALQIFCQSAFGQSSPP	V[DG]FPSx[EK]EAL[QE]I[FL]C[QL][SY]AF[GKR][QK][SN]S[PA]P
GLPL	Coffee	Y**E**ELSE**E**I**V**HY**A**GC**LPLALKV**LGSF	[YF][EKM][EK][LI]S[EIV][EKQ]IV[HN][YL][AT]G[CG]LPLAL[KE]V[LI]GS[FS]
	Potato	F**E**KFSL**EV****V**NH**A**KG**LPLALKV**W	F[EKM][KE][FL][ST]LEVV[NR]HAKGLPLALKV[WL]
	Tomato	F**E**ELAL**EV****V**QY**A**GG**LPLALKV**L	FE[ED]LA[LN][EQ]V[VI]QY[AS][GK]GLPLALK[VI]L
	Grapevine	YxELSRRV**V**DY**A**KGL**PLALKV**LGSFLFGK	[YF]x[END]LSRR[VI][VI][DG]YA[KQ]GLPLAL[KE]VLGSFLFG[KM]
	Arabidopsis	DGF**E**ELAx**E**VTKL**A**GN**LPLGLRV**LGS	DGFE[EK]LAx[EK]V[TA]KL[AC]GNLPL[GA]L[RS]V[LM]GS
RNBS-D-TIR	Coffee	YGRDMAEWRSEVERLKRIPEDEIMEKLEVSF-NGLDEVEKE**IFLDIACFF**	Y[GK]R[DER][MK][AT]EW[RI][SD][EV][VL]E[RK]LK[RQ]I[PR][EP][DN][EH][IL][MQ][ED][KV]L[EKQ][VI]SF[NK]GL[DKN][ED][VQ]EK[ER][IV]FLD[IL]ACFF
	Potato	QE**IFLDIACFF**RGKE	Q[EK]IFLDIACFFRGKE
	Tomato	QK**IFLDIACFF**RGK	[QK]KIFLDIACFFR[GE]K
	Grapevine	EWESALDKLKKIPNMEIQNVLKISFDGLDDTE-KE**IFLDIACFF**K	EW[EK]S[AE]LDKL[KE][KR]IPNM[EK]IQNVL[KR]IS[FY]DGLD-D[TK][EQ]K[EDN]IFLD[IV]ACFFK
	Arabidopsis	**FL**H**IACFF**NGENVDYVKxLLA	FLHIAC[FL]FN[GY]EN[VI]DYVKx[LM]LA

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

MHD	Coffee	**MH**C**LIQ**EMGWH**IV**RQKAPDEPGKHSRLW-VAEEICDVLARDKATENIVGMM	MH[CD][LQ]I[QR][ED]MG[WR][HQ]IV[RQ][QR][KE][AS][PY][DA][ED][PA]GK[HR]SRLW[VS][AQ][EG]EI[CM][DM]VL[AK][RN][DR]K[AGV]T[ER][NS][IV][VE]G[MI][MTW]
	Potato	VMQILESCDFGAEYGLDVLIDKSLVFISEY-DRIE**MH**D**LIQ**DMGKY**IV**KMQ	VMQILESCD[FS]GAEYGLDVL[IV]DKSLVFIS[EK][YD][DN][RTK]I[EQ]MHDLI[QE]DMG[KR]Y[IV]VKMQ
	Tomato	SLLFISDxDTLE**MH**D**LIQ**DMGRE**IV**RLE	[SR][LC]LFISDx[DN]T[LI][EW]MHDL[IV][QR][DE]M[GA][RW][EY]IVR[LQ][EG]
	Grapevine	**MH**D**LIQ**QMGWE**IV**RQECPKEP	MHDL[IL]Q[QE]MG[WR]EI[VI]R[QE][EQ][CS][PL][KE][ED]P
	Arabidopsis	IV**MH**N**LLQ**QLGRE**IV**	[IV][VE]MHNLL[QE]Q[LM]G[RK]EI[VI]

Based on their N-terminal and C-terminal NCBI-CD domains, the NBS genes formed 15 subclasses in euasterids and 26 in eurosids (Table 1). The majority of NBS genes were classified in four subclasses: CNL, CN, NL, and N. The NBS genes without TIR or CC N-terminal domain were then aligned and divided into TIR-NBS and non-TIR-NBS genes according to their NBS signature [61].

Analysis of conserved motif structures

Structural divergence and conserved motifs shared among NBS domains were examined by analysing the predicted CNL and TNL proteins using MEME software. Eight conserved motifs were easily identified in the NBS domain, including the P-loop, RNBS-A, kinase-2, RNBS-B, RNBS-C, GLPL, RNBS-D and MHDV. Divergence differed within and between CNL and TNL proteins (Table 2). A high level of similarity was found between CNL and TNL proteins in the P-loop, kinase-2, RNBS-B, GLPL, and MHDV motifs. The RNBS-A, RNBS-C, and RNBS-D motifs were dissimilar. The core motifs of euasterid consensus sequences were remarkably conserved in several motifs: 14 residues were conserved in the kinase-2 motif of TNL proteins (KVLIVLDDxDxxDxLxxLAG), 13 in the P-loop motif of CNL (IVGxGGxGKTTLAxxxYN) and TNL (GIWGxxGIGKTTxAKA) proteins, 12 in the GLPL motif of CNL (IxxxCxGLPLAxxxxGxLLRxK) proteins, and 11 in the GLPL motif of TNL (ExxxxExVxxAxxLPLALKV) proteins. Two consensus sequences of RNBS-A, kinase-2 and RNBS-B motifs were identified in the CNL proteins of potato. Some substitutions around the core motif residues were specific to euasterid (most CNL and TIR motifs) and Solanaceae (RNBS-A, Kinase-2, RNBS-C, RNBS-D and MHD of TNL proteins and RNBS-C of CNL proteins) sequences (Table 2).

Genomic organisation of NBS genes

Clustering analysis of NBS genes revealed different genomic organisations in euasterids. One hundred and nineteen clusters according to Holub's [46] definition were detected in the six plant genomes (Table S3). The number of clusters varied from eight in tomato to 40 in coffee (Table 3). The proportion of clustered NBS genes was 25.2% in tomato, 38.8% in coffee, 40.4% in potato, and 57.8% in monkey-flower. The average number of NBS genes per cluster was higher in monkey-flower (9.1) and grapevine (8.1) than in other genomes (4.8-5.6). In addition to clusters, NBS genes were also grouped in triplets and doublets. Finally, the ratio of NBS genes in a cluster or in a tandem array was 58.9% in tomato, 66.4% in potato, 70.1% in coffee, and 79.9% in monkey-flower. The maximum number of NBS genes in a cluster was 33 genes (monkey-flower CL16).

Table 3: Genomic organization of NBS-encoding genes in six plant species. A gene cluster is defined as a region of 200 kb or less that contains at least four NBS-encoding genes (Holub, 2001).

	Coffee	Monkey-flower	Potato	Tomato	Grapevine	Arabidopsis
Singleton no.	167	54	89	62	80	30
% NBS genes	29.90%	20.10%	33.60%	41.10%	19.80%	27.80%
Doublet no.	47	14	24	12	29	10
% NBS genes	16.80%	10.40%	18.10%	15.90%	14.40%	18.50%
Triplet no.	27	10	7	9	13	7
% NBS genes	14.50%	11.20%	7.90%	17.90%	9.70%	19.40%
Cluster no.	40	17	19	8	28	7
Clustered genes	217	155	107	38	227	37
% NBS genes	38.80%	57.80%	40.40%	25.20%	56.20%	34.30%
Maximal members of a cluster	13	33	10	8	20	9
Average members per cluster	5.4	9.1	5.6	4.8	8.1	5.3

The physical organisation and sequence diversity of complex clusters of at least ten NBS genes were then characterised. Ten complex clusters were identified in euasterids and four in grapevine, but none in *Arabidopsis* (Figure 2). All complex clusters in euasterids were composed of non-TIR sequences while in grapevine there were two clusters (CL09, CL16) with

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

TIR sequences. The grapevine CL09 contained four TIR members followed by ten non-TIR members, suggesting the presence of two distinct clusters. The largest complex clusters covered more than 909 kb (monkey-flower CL09 and grapevine CL16), whereas the shortest cluster spanned only 151kb (potato CL04) (Table 4). The number of predicted genes in a complex cluster varied from ten (potato CL04) to 115 (monkey-flower CL16). The majority of clustered NBS genes (97/165) occupied neighbouring positions in the complex clusters. Examples are monkey-flower CL16 which had 33 members, of which 23 were consecutive, and potato CL04, which is a chain of 10 NBS genes with no non-NBS genes. Average length of clustered NBS genes was lower in euasterids (2,723-3,804nt) than in grapevine (4,991-11,352nt). Similar differences were also observed in the non-NBS genes located in the complex clusters (1,812-3,086nt in euasterids *vs.* 3,785-8,924nt in grapevine). The average distance between clustered NBS genes ranged from 12,424nt (potato CL04) to 46,741nt (coffee CL23), reflecting variable gene density. Nucleotide diversity (π) between pairs of clustered NBS genes revealed different diversification patterns among complex clusters. The minimum values were low in all clusters (< 0.06) whereas the maximum values were either low, i.e., 0.153 in potato CL04, 0.157 in monkey-flower CL14 and 0.183 in coffee CL23, or high, i.e. 0.454 in monkey-flower CL09, 0.464 in CL13 and 0.469 in CL16 (Table 4). Computed nucleotide diversity (π) between each pair of NBS domain sequences was then used to analyse relationships among the genes clustered in complex clusters (Figure 3). In monkey-flower CL14 (Figure 3c), coffee CL23 (Figure 3e) and potato CL04 (Figure 3f), all gene pairs presented low diversity and belonged to different gene families, suggesting a recent origin for these clusters. By contrast, the other complex clusters (monkey-flower CL09, CL13 and CL16) showed more diversified sequences, suggesting an ancient origin (Figure 3a, 3b & 3d). The most recent duplications (π < 5%) were observed in five out of six complex clusters of euasterids (Figure 3a-3e) and in the four complex clusters of grapevine (Figure 3g-3j). They involved two (e.g., monkey-flower CL09 and CL13) or three (e.g. monkey-flower CL14) genes and were often accompanied by a translocation or an inversion.

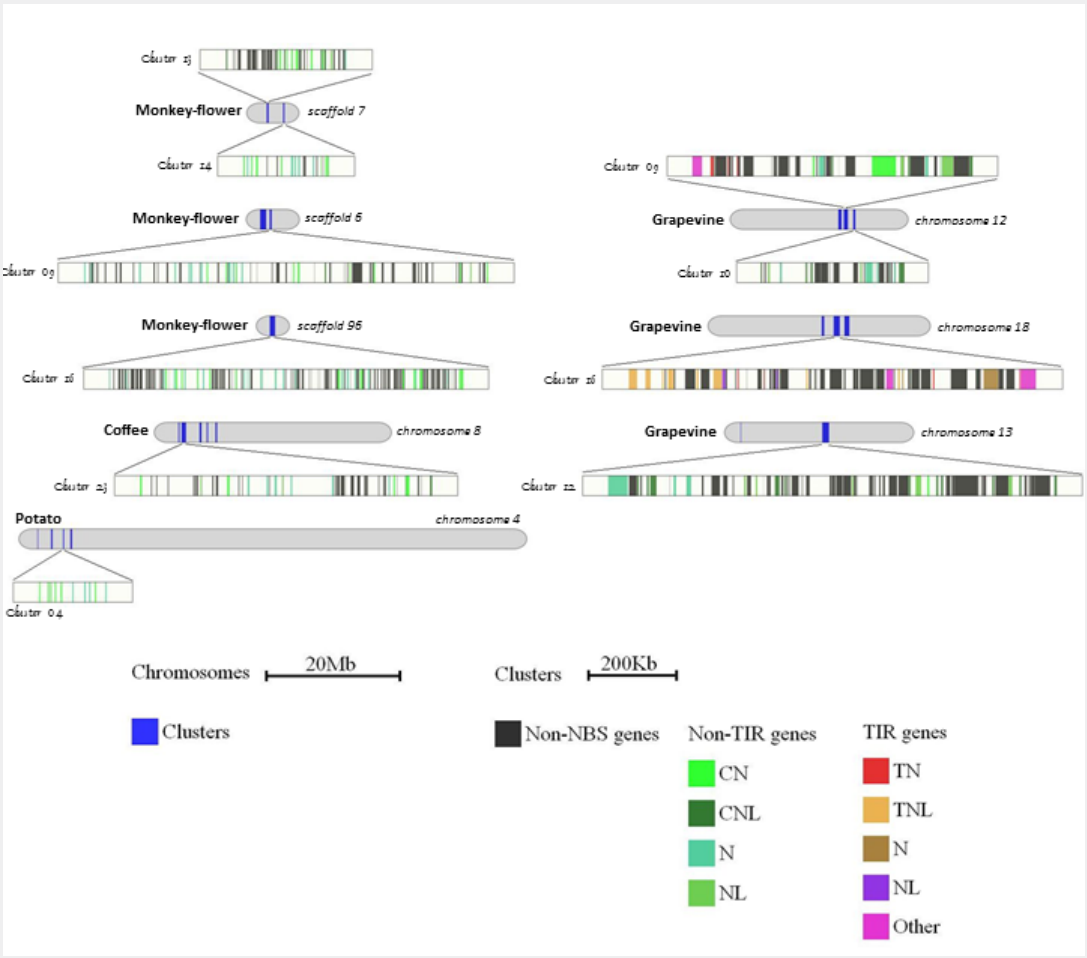


Figure 2: Representation of complex clusters of NBS genes (≥10) showing their position on chromosomes or scaffolds, according to their subclass

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

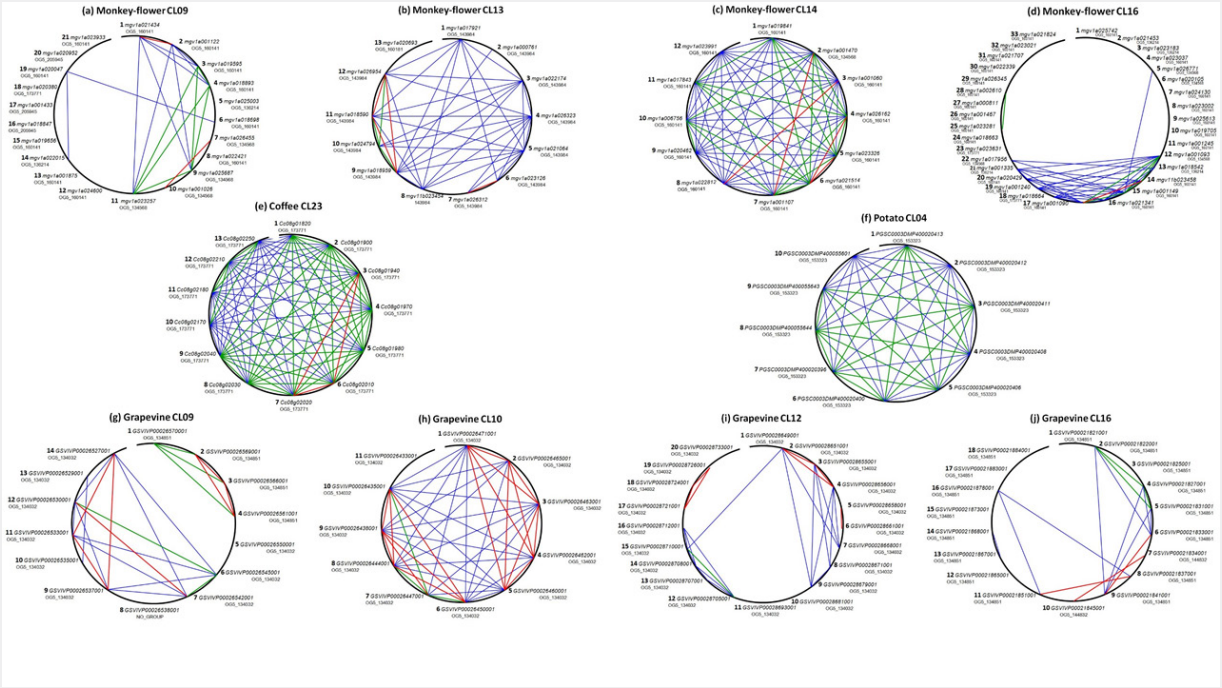


Figure 3: Relationships among NBS genes in complex clusters based on nucleotide diversity (π). Diversity between gene pairs is as follows: < 5% in red, < 10% in green, and < 20% in blue. The orthogroups are indicated below the gene names

Orthologous relationships of NBS genes

The 1,775 NBS genes identified in this study were assigned to orthologous groups (orthogroups) using OrthoMCL. In euasterids, 1,122 genes (90.3%) were distributed in 23 orthogroups (Table 5). Eight additional orthogroups were identified in eurosids. The orthogroups were composed of either CNL or TNL sequences. The proportion of genes assigned to an orthogroup was lower in coffee (82.6%) than in the other genomes (> 92%). Three orthogroups showed a large expansion in all euasterid genomes: OG5_160141 with 513 genes (45.7% of the genes assigned to OrthoMCL groups), OG5_134032 with 177 genes (15.8%) and OG5_173771 with 116 genes (10.3%). Together these orthogroups represented 64.8% of euasterid NBS genes. Eight orthogroups were composed of sequences from all euasterid genomes and were retained for phylogenetic analysis of NBS domains. They contained about three quarters of the NBS genes of euasterids. Two hundred and thirty grapevine genes but only five *Arabidopsis* genes classified in these orthogroups.

Table 4: Characterization of complex clusters of NBS-encoding genes (10 genes) sorted by NBS-encoding gene number.

Genome	Cluster Name	Cluster Length (nt)	Identified NBS Genes	Consecutive NBS Genes	Predicted Genes	Average Length of NBS Genes (nt)	Average Length of non-NBS Genes (nt)	Average Distance between NBS Genes (nt)	Nucleotide Diversity (π)
Monkey-flower	CL16	797 194	33	23	115	2 966	2 535	21 192	0.028-0.469
Monkey-flower	CL09	909 376	21	10	64	2 828	3 086	40 475	0.012-0.454
Grapevine	CL12	851 750	20	11	66	7 918	8 182	42 588	0.012-0.425
Grapevine	CL16	916 923	18	9	55	10 230	8 924	40 711	0.018-0.381
Grapevine	CL09	630 846	14	6	39	11 352	8 685	33 708	0.019-0.427*
Coffee	CL23	657 092	13	8	44	3 804	2 506	46 741	0.019-0.183
Monkey-flower	CL13	270 789	13	8	40	3 243	2 649	17 587	0.030-0.464
Monkey-flower	CL14	192 046	12	10	19	3 056	1 812	12 948	0.000-0.157
Grapevine	CL10	323 940	11	2	34	4 991	3 785	24 459	0.009-0.269
Potato	CL04	151 464	10	10	10	2 723	-	12 424	0.057-0.153

* 0.019 – 0.084 between TIR sequences and 0.023 – 0.218 between non-TIR sequences.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Table 5: Distribution of NBS-encoding genes in the orthogroups defined by OrthoMCL database. The *R* genes were downloaded from the plant resistance gene database (http://prgdb.crg.eu/wiki).

Orthogroup	Class	Sub-total							Total	R Genes associated
		Coffee	Mimulus	Potato	Tomato	Euasterids	Grapevine	Arabidopsis		
OG5_129219	CNL			1		*1*			*1*	*LR10, MLA1, MLA10, MLA12, MLA13, MLA6, Pi36, Pi-ta, Tm-2, Tm-2a*
OG5_129307	TNL							33	*33*	*RAC1, RPP1, RPP4, RPP5, SSI4*
OG5_133425	TNL							1	*1*	
OG5_133794	CNL	1				*1*	12	20	*33*	*RPS5*
OG5_134032	CNL	95	23	45	14	*177*	142	1	*320*	*I-2, Pl8, R3a, Rdg2a, Rpi-blb1, Rps1-k-1, Rps1-k-2*
OG5_134568	CNL	4	17	4	1	*26*	2	1	*29*	*RPM1*
OG5_134851	TNL			20	4	*24*	67	7	*98*	*P2, RPS4*
OG5_136214	CNL	2	13		2	*17*		14	*31*	*HRT, RCY1, RPP8*
OG5_137195	CNL	16		19	9	*44*	23		*67*	*Dm3(RGC2B), VAT*
OG5_141458	TNL							6	*6*	*RRS1*
OG5_143984	CNL	21	22	1	1	*45*	18		*63*	
OG5_144832	TNL	8		16	9	*33*	6		*39*	*Bs4, Gro1-4, KR1, L6, M, N, RY-1*
OG5_144910	TNL							2	*2*	
OG5_147224	TNL							1	*1*	*Pm3*
OG5_153323	CNL	2		22	15	*39*		3	*42*	*Pi2, Pi9, Piz-t, RPP13*
OG5_158177	CNL			2	1	*3*	14		*17*	
OG5_160010	CNL	1		1	1	*3*	1	4	*8*	
OG5_160079	TNL	1		1	2	*4*	2	2	*8*	
OG5_160115	TNL							3	*3*	
OG5_160141	CNL	196	140	104	73	*513*	31	1	*545*	*Bs2, Gpa2, Hero, Mi1-2, R1, Rpi-blb2, Rx, Rx2*
OG5_170181	CNL	5				*5*		1	*6*	
OG5_170771	CNL	1			1	*2*	10	2	*14*	
OG5_173771	CNL	75	24	10	7	*116*	11		*127*	*Prf, Sw-5*
OG5_177968	CNL	3	2			*5*	3	1	*9*	*RPS2*
OG5_184596	CNL	1	1	2	1	*5*	14		*19*	
OG5_190299	CNL	4		3	3	*10*	1	1	*12*	
OG5_205945	CNL	23	4	3	3	*33*	3		*36*	
OG5_212041	CNL							2	*2*	
OG5_212667	CNL	1	1	8	3	*13*	9	2	*24*	

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

OG5_242951	CNL	2	1			*3*			*3*	
OG5_245007	CNL						32		*32*	
Total in orthogroups		462	248	262	150	*1122*	401	108	*1631*	
% NBS genes		82.60%	92.50%	98.90%	99.30%	*90.30%*	99.30%	100%	*92.90%*	
Without orthogroup		97	20	3	1	*121*	3	0	*124*	
% NBS genes		17.40%	7.50%	1.10%	0.70%	*9.70%*	0.70%	0%	*7.10%*	

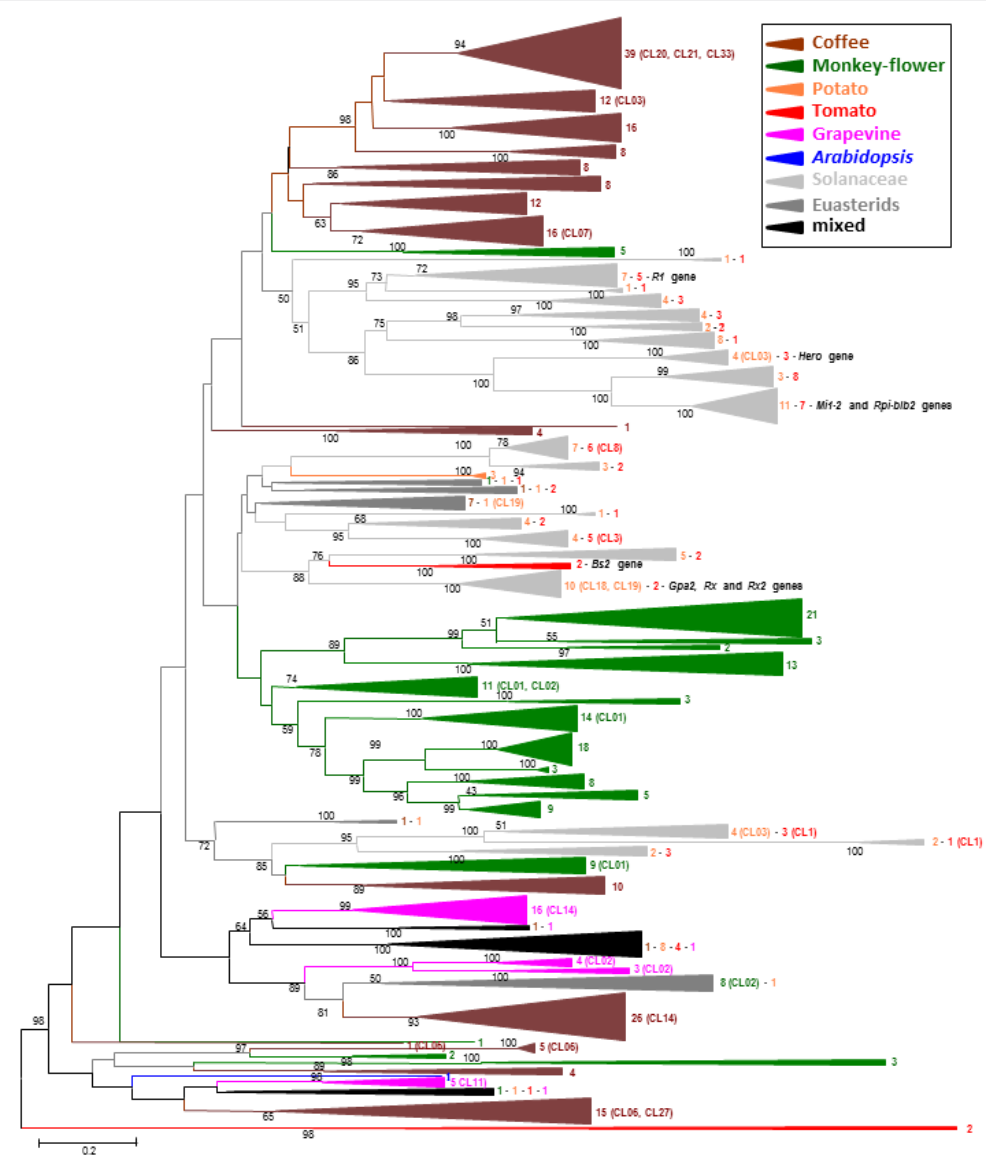


Figure 4: Neighbour-joining tree of NBS domains from R gene analogues and cloned R genes (in bold), belonging to the orthogroup OG5_160141. The complete tree was based on 553 sequences. Clades containing genes from individual plants were collapsed into single branches and the number of genes in each branch is indicated on the right. Different taxa are shown in different colours; clades with R gene analogues from several plants are in light grey (potato and tomato), dark grey (euasterids) and black (euasterids and eurosids). Homogeneous gene clusters (i.e., fully represented in the orthogroup) are shown in brackets after the number of NBS genes. Bootstrap percentages ≥ 50% are indicated at branches

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

For each orthogroup, sequences and *R* genes (http://prgdb.crg.eu/wiki) were aligned using MAFFT. Regions corresponding to the NBS domain were then selected and the evolutionary history of the sequences was inferred using the neighbour-joining (NJ) method [52]. Allelic diversities appeared to be structured in branches specific to each genome or to the Solanaceae family (Figure 4, S2 & S3). Indeed, most genes in potato and tomato grouped together and formed Solanaceae branches. Eighteen *R* genes were classified in euasterid orthogroups, 13 of which belonged to the Solanaceae family. Except for the *Rpi-blb1* gene, the other *R* genes in tomato, potato and their wild relatives were classified in accordance with their origin, i.e., in branches which only contained Solanaceae NBS genes (Figure 4, S2 & S3a). Similarly, the *RPM1* gene grouped with an *Arabidopsis* NBS gene (Figure S3d). Different mutation rates were associated with NBS gene expansion in the orthogroups and were three-fold higher in OG5_160141 (Figure 4) and OG5_134032 (Figure S2), than in OG5_143984 (Figure S3b).

Based on their orthogroup, the clusters of NBS genes were separated into "homogeneous" clusters when all clustered genes derived from a single ancestor, or in "heterogeneous" clusters when the genes originated from more than one ancestor. Thirty-four homogeneous and 45 heterogeneous clusters were identified in euasterids (Table 6). The proportions of homogeneous and heterogeneous clusters were similar in all genomes except for monkey-flower which had fewer homogeneous clusters (3/16). By contrast, the grapevine genome had a majority of homogeneous clusters (20/28). Most homogeneous clusters were classified in distinct clades, the others being mixed with NBS genes from the same genome (Figure 4, S2 & S3). Two complex clusters were homogeneous (coffee CL23 to OG5_173771 and potato CL04 to OG5_153323), while the others were heterogeneous. Cross analysis of gene positions and origins in the heterogeneous complex clusters revealed discontinuous distribution of orthogroups inthe clusters, suggesting translocation events that concentrated locally NBS genes from different orthologues (Figure 3).

The NBS genes not assigned to an orthogroup represented specific sequences with no known orthologues. These genes were distributed unevenly in the genomes: 97 genes from coffee, 20 from monkey-flower but only three from potato and grapevine, and one from tomato (Table 5). They formed four clusters in coffee and one cluster in monkey-flower (Table 6). In coffee, chromosomes 1 and 3 carried respectively 22 and 17 NBS genes without orthologues.

Table 6: Homogeneous and heterogeneous clusters of NBS-encoding genes based on OrthoMCL database 5 (Chen *et al*., 2007).

	Coffee	Monkey-flower	Potato	Tomato	Grapevine	Arabidopsis
Clusters without orthogroup	4	1	0	0	0	0
No clustered NBS genes without orthogroup	18	6				
No homogeneous clusters	18	3	9	4	20	4
No NBS genes in homogeneous clusters	103	18	51	20	150	24
No heterogeneous clusters	18	13	10	4	8	3
No NBS genes in heterogeneous clusters	96	131	56	18	77	13

Synonymous vs. non-synonymous substitutions

The evolution of the NBS domains was investigated by analysing synonymous and non-synonymous substitutions in orthogroups containing NBS genes from the four euasterid genomes. Gene pairs with a nucleotide diversity (π) < 0.05 or with N/A (not applicable) value for Ks or Ka were excluded. Eight gene pairs had zero diversity, suggesting recent duplications: two in monkey-flower (*mgv1a000811-mgv11b017348*, *mgv1a019980-mgv1a022166*), two in potato (*PGSC0003DMP400013227-PGSC0003DMP400011616*, *PGSC0003DMP400047655-PGSC0003DMP400047657*), one in tomato (*Solyc00g102400.2.1-Solyc06g008800.1.1*) and three in grapevine (*GSVIVP00004211001-GSVIVP00000786001*, *GSVIVP00005246001-GSVIVP00026462001*, *GSVIVP00030986001-GSVIVP00030997001*). Kruskal-Wallis tests showed differences in π, Ka, Ks and Ka:Ks between orthogroups and genomes (Table S4). The average ratio Ka:Ks was higher in OG5_160141 (0.789-0.811) and OG5_134032 (0.718-0.952) than in OG5_173771 (0.426-0.643). Differences between homogeneous and heterogeneous clusters were detected only for Ka using the Kolmogorov-Smirnov test (Table S5). All clusters were under purifying selection (Ka:Ks < 1), except one (monkey-flower CL04), under diversifying selection (Ka:Ks > 1). Its low diversity ($0.07 < \pi < 0.11$) showed that this cluster was recent.

The relative frequency of Ks revealed variable distribution patterns in the orthogroups. In OG5_160141, all genomes presented a Gaussian-like distribution with maximum Ks frequencies of around 0.6-0.7 for coffee and monkey-flower and 0.7-0.8 for potato and tomato (Figure 5a). A minor peak (0.1-0.2) was also present in potato. Such distribution patterns clearly showed that NBS genes of OG5_160141 arose from a major large-scale duplication event. In contrast, complex Ks distributions were observed in the other orthogroups. In OG5_134032, the distribution pattern was bimodal in all genomes, with the maximum around 0.6-0.7 and

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

0.1-0.2 (Figure 5b). In OG5_173771, the distribution was complex with several peaks in all genomes (Figure 5c). Relative frequencies of Ka showed similar distribution patterns to Ks in the species (Figure S4). Both Ks and Ka distributions were consequently shaped by the same evolutionary events.

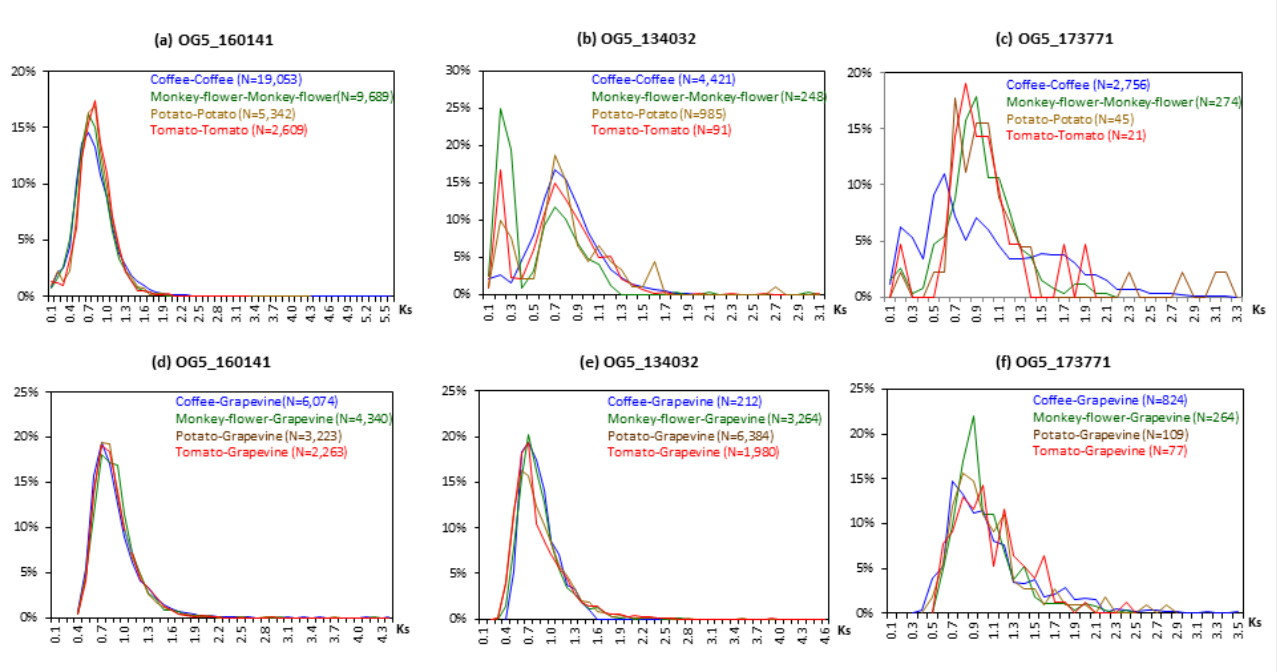


Figure 5: Distribution of Ks frequency in the NBS domain of homologues from the major euasterid orthogroups. The relative frequency of sequence pairs is shown on the vertical axes and the Ks values at 0.1 intervals on the horizontal axes. The number of gene pairs is given in brackets

The timing of duplication was estimated by building an evolutionary clock using grapevine as reference and based on fossil deposits for time calibration [62]. The Ks peaks of the three major euasterid orthogroups (Figure 5) were dated, and then grouped in 11 major events according to periods of time (Table S6). Three periods of time were observed following the divergence of Eurosids (Figure 6). The oldest period (83-116Mya) lasted until the divergence of coffee and showed differential contributions of orthogroups to today's genomes. Five events affected the ancestral sequence of euasterids but ancestral NBS genes were conserved in the four euasterid genomes only in the third event and in a single orthogroup (OG5_134032). The other events left traces of duplication in one (second and fifth events) or two (first and forth) genomes. Interestingly the fourth event appeared to be at the origin of NBS gene expansion in OG5_160141 of both coffee and monkey-flower genomes, but without leaving any traces in the Solanaceae genomes. The following period of time (27-83Mya) occurred between these intensive duplication activities in the ancestral sequence and the recent period. This intermediate period was the longest, but only two minor events happened in a single orthogroup (OG5_173771). The most recent period (16-27Mya) included four minor duplication events, which all occurred during a relatively short period. One orthogroup (OG5_134032) was particularly concerned by these events.

An analysis of the NBS genes involved in duplication events revealed differential impacts on genome organisation. Duplication events involved several chromosomes in all genomes. For example, coffee chromosomes 3, 8 and 11 were the most affected by the events, with 53, 99 and 59 genes involved, respectively (Table S7). In potato, events 1 and 10 were the only ones to involve NBS genes of chromosomes 2, 7 and 12, as a specific signature left in OG5_160141. The coffee and Solanaceae genomes still had traces of important ancestral events (4 and 1, respectively) since all the chromosomes except one of potato were affected, indicating large-scale duplication events.

Discussion

Identification of NBS genes

Plant resistance to a range of pathogenic organisms (bacteria, fungi, insects, nematodes, oomycetes and viruses) is conferred by a diverse group of disease resistance proteins [63]. Classification of these proteins is primarily based on predicted domains and motifs. One of the largest families of resistance proteins encodes NBS domains. The recent availability of the complete genomic sequences of *C. canephora* and *M. guttatus* enabled us to perform comparative analyses of NBS domains in the euasterid clade. This clade encompasses many economically important crops such

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

as coffee, potato and tomato, which were included in this study, others like cinchona, jasmine, sesame, and other Solanaceae crops (eggplant, pepper, sweet potato, tobacco, etc.). The cultivated varieties are susceptible to many pests and diseases, and production depends on huge amounts of pesticides. We designed a workflow to perform an automatic search for NBS genes using the predicted gene sequences of a given genome. The procedure was applied to six plant genomes with the aim of constructing datasets of genes with a full NBS domain rather than making a complete inventory. Such predicted genes are stored on the GreenPhyl website [39].

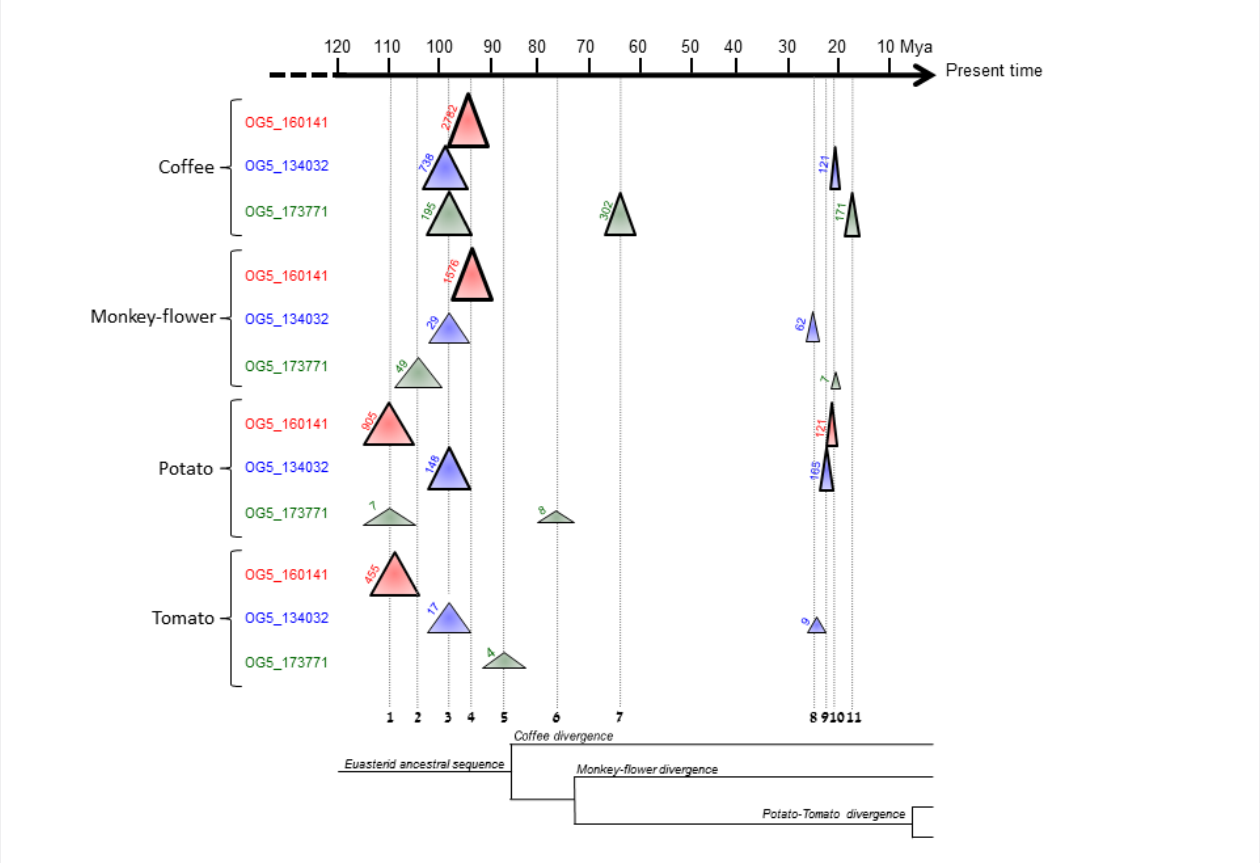


Figure 6: Large-scale duplication events in the major euasterid orthogroups based on Ks frequency (Table S6). The events are symbolized by triangles whose base corresponds to the duration of the event. Their size is proportional to the number of gene pairs indicated on the left, using the following scale:

Comparison with previous studies [35,36,61,64] revealed small differences in the number of NBS genes detected. However these variations can be explained by the different methods used for the search (HMM, Basic Local Alignment Search Tool) and validation (NCBI, PFAM, InterProScan, InterPro) (data not shown). Instead of inventorying the NBS genes, our study focused on sequences having a complete NBS domain, excluding those with partial domains. This ensured that at least 80% of both N- and C-termini sequences were included in the alignment found by Reverse Position Specific-BLAST [38,65]. The NBS domain of euasterids is composed of eight conserved motifs in both TIR and non-TIR sequences. The core sequences of each motif resemble each other and are similar to eurosid motifs such as *Populus trichocarpa* [66], *Theobroma cacao* [67], *Zea mays* [68] and *Cucumis sativus* [69]. In addition, they resemble the motifs of CNL sequences in monocots like *Oryza sativa* [49] and other more distant genera [70] confirming the remarkable conservation of the NBS domain in the plant kingdom.

In addition to conservation of the NBS domain, euasterid genomes displayed specific features compared with those of the eurosids. The conserved domains with which the NBS domain of euasterids is associated differ from those in eurosids. Out

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

of a total of 39 conserved domains recognized by NCBI, seven were specific to euasterids and 21 to eurosids. This points to different genomic composition in the two clades. Similarly, only nine subclasses out of a total of 33 were common to euasterids and eurosids. The differences also included the TIR and LRR subclasses, which are underrepresented in euasterids. Expansion of TIR genes in eurosids but absence in monkey-flower has already been reported by Kim et al. [71]. The underrepresentation of LRR domains in euasterids suggests that the mechanisms of interaction with pathogen-derived molecules differ in euasterids and eurosids. Other differences were found in the orthogroups. Few eurosid sequences were assigned to the three main euasterid orthogroups. This suggests the phylogenetic origins of NBS genes differ in the two clades.

Insights into genome organisation and evolution

Our study of euasterid NBS genes provided insights into genome organisation and evolution. Examination of orthologous relationships may inform chromosomal rearrangements and guide the assembly of non-anchored sequences. Our results confirm that the coffee genome presents no sign of whole-genome polyploidisation since the γ triplication was seen to be at the origin of the core eudicots [33]. However, our study revealed ancient traces of large-scale duplications in the orthogroups whose impacts varied with the genome. The Solanaceae genomes have conserved genes of an ancestral duplication event around 110Mya, but not of the event around 94Mya which was at the origin of NBS gene expansion in the coffee and monkey-flower genomes. In return, no coffee and monkey-flower gene was found to date from the ancestral duplication displayed in the Solanaceae. This demonstrates that NBS genes evolved leaving specific signatures in genomes and that their evolution included significant elimination phases. This corresponds to the birth-and-death model of evolution proposed for the vertebrate major histocompatibility complex genes and the immunoglobulin genes [72,73]. New genes are created by repeated gene duplication and some duplicate genes are maintained in the genomes for a long time while others are deleted or become nonfunctional. Four duplication events occurred relatively recently (between 16 and 27Mya), mainly in one orthogroup (OG5_134032) (Figure 6 & Table S6).

One of these events, the 10th, was previously observed in tomato and potato, and dated at 18.3-23.3Mya using substitution rates of cereals and *Arabidopsis* [74]. Analysis of new sequence data in related lineages will help identify the species which have undergone these duplication events and estimate their relative importance.

The NBS genes in the eurosid genomes included in our study provided information on their evolutionary history. The NBS genes in grapevine showed a high level of conservation of amino acid motifs similar to those in euasterids, especially in coffee. This is in accordance with previous results on microsynteny and collinearity observed in short regions of the grapevine and coffee genomes [75-77]. By contrast, the NBS genes in *Arabidopsis* differed in composition and organisation. The α and β whole-genome duplications, which occurred in the *Arabidopsis* genome following the divergence between eurosids I and II (Figure 1), likely split genomic rearrangements and relocated fragments of duplicated chromosomes around the genome, making comparison difficult.

Expansion and diversification of NBS genes

The proportion of euasterid genes that were predicted to encode a NBS domain is similar to estimates for other plant species, and ranges between 0.64 and 1.11%, except for coffee whose genome reveals an extreme expansion of NBS genes, up to 2.62% of the total number of predicted genes. This exceeds the highest proportion (2.05%) of NBS genes found in plants [78]. Expansion in the coffee genome is also reflected in the total number of NBS genes detected by HMM searches, which was higher in coffee than in the other genomes. The number of detected NBS genes in monkey-flower is probably underestimated since they were searched for in 2,216 scaffolds rather than in 14 pseudomolecules [34]. However a mega-cluster of 33 NBS genes (CL16) was identified in a scaffold, which is a record in plants.

Most NBS genes (90.3%) arose from the duplication of paralogues in only a few orthogroups. This expansion is clearly visible in orthogroup OG5_160141, which contains nearly half (45.7%) of the euasterid sequences identified in our study and 35.1% of the coffee sequences. Evidence for diversification in the orthogroups was revealed by examination of the evolutionary distances between branches in our phylogenetic analyses. The speed of evolution varied within the orthogroups, revealing independent diversification dynamics. Expansion in orthogroups was associated with dispersion in the genomes, sometimes involving blocks of NBS genes or single NBS genes. At chromosome level, the genomic dispersion of NBS genes could benefit from chromosomic rearrangements and segmental duplications. When homozygous, these insertions may contribute to the divergence of intergenic regions, since they tend to decrease the chance of misalignment and therefore of unequal crossing-over [79]. They thus open new evolutionary contexts for duplicated gene diversification and possibility for escape from homogenization within clusters [80]. Tandem and segmental duplications, but not whole-genome duplication, hence play a major role in NBS gene expansion in euasterids and in plants in general [48]. Besides expansion and diversification of paralogues, the NBS genes with no known orthologues are a valuable source of diversity despite their small numbers.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

The physical distribution of NBS genes revealed that most euasterid sequences are organised in doublets (15.6%), triplets (12.8%) or clusters (41.6%). The coffee genome has at least twice as many clusters (40) as the other euasterid genomes (8-19) in relation with gene expansion. The study of complex clusters with at least ten NBS genes revealed several patterns of tandem duplication with transfer to a contiguous site or to a more distant one. Tandem duplication is undoubtedly an important mechanism to stimulate gene expansion in clusters. As reported in *Arabidopsis*, tandemly clustered *R* genes may be a reservoir of genetic variation from which new disease resistant specificities can evolve [79]. Eight recent duplications ($\pi = 0$) were detected in our study and three complex clusters comprised a single gene family with low diversity between all gene pairs ($\pi < 0.2$). Tandem duplication is thus a continuous mechanism over time. Moreover, it appears to be a universal mechanism since it has also been described in monocots and eurosids [81].

Among the clusters identified in euasterids, the majority were heterogeneous, comprising sequences derived from several orthologues. They probably originated from random associations among NBS genes from different orthologues rather than from diversification within homogeneous clusters as a result of diversifying selection. Heterogeneous clusters can derive from ectopic rearrangement events [61]. However, they present similar diversity and the same mode of evolution as the homogeneous ones, confirming strong homogenization of neighbouring sequences. Such homogenization is likely the result of frequent intergenic exchanges among NBS genes which are closely related and physically linked [82].

Evolutionary history of NBS genes

NBS genes have an ancient and complex origin in plant genomes. Their abundance allowed us to reconstruct their expansion and evolution in the euasterid clade. Our study based on orthology analysis revealed 11 large-scale duplication events. All but the 10th event have not yet been observed in plant genomes, which reinforces the interest of our orthology approach to access ancient genomic rearrangements. The oldest events pinpointed by our study occurred in the euasterid ancestral sequence and left specific signatures in present-day genomes. By contrast, the latest events happened in separate genomic environments after divergence occurred in coffee and monkey-flower. The synchronicity of these events in separate genomes may be related to massive and simultaneous pathogen attacks. The oldest duplication events in the ancestral sequence may represent an adaptive response to ancient pathogen diversification and spread.

Our approach based on orthologous relationships enabled us to retrace the history of NBS genes in the euasterid clade. The orthogroups have their own evolutionary dynamics with variable speeds of diversification and expansion in the genomes. Most orthogroups have been shaped by two large-scale duplication events, showing thus a possible reactivation in time. The synchronicity of large-scale duplication events in different genomes demonstrates the impact of living conditions on NBS gene expansion and diversification in the orthogroups. That is why the NBS genes represent a valuable source of information to understand genome evolution.

Acknowledgment

G. Ershov was funded by the “Chaires croisées” programme of IRD (Institut de Recherche pour le Développement) (*PlanteResiste* contract). Thanks to M. Noirot for helpful comments on diversity and evolution.

References

1. McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol 40: 349-379.
2. McDonnell AV, Woffenden BJ (2003) Plant disease resistance genes: recent insights and potential applications. Trends Biotechnol 21(4): 178-183.
3. Gregory PJ, Johnson SN, Newton AC, Ingram JSI (2009) Integrating pests and pathogens into the climate change/food security debate. J Exp Bot 60(10): 2827-2838.
4. Shaw MW, Osborne TM (2011) Geographic distribution of plant pathogens in response to climate change. Plant Pathol 60(1): 31-43.
5. Dangl JL, Jones JD (2001) Plant pathogens and integrated defense responses to infection. Nature 411(6839): 826-833.
6. Collier SM, Moffett P (2009) NB-LRRs work a “bait and switch” on pathogens. Trends Plant Sci 14(10): 521-529.
7. Dangl JL, Dietrich RA, Richberg MH (1996) Death don’t have no mercy: cell death programs in plant-microbe interactions. Plant Cell 8(10): 1793-1807.
8. Ma Y, Berkowitz GA (2011) Danger at your door: pathogen signals and programmed cell death in plants. New Phytol 192(1): 1469-8137.
9. van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, et al. (2008) Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. J Exp Bot 59(6): 1383-1397.
10. Johal GS, Briggs SP (1992) Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258(5084): 985-987.
11. van der Biezen EA, Jones JD (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol 8(7): R226-R227.
12. Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50(3): 203-213.
13. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the α- and β-subunits of ATP synthase myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1(8): 945-951.
14. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J 19(15): 4004-4014.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

15. McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7(4): 212.

16. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BWS, et al. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20(3): 317-332.

17. Tarr DEK, Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2: 197.

18. Miller RNG, Bertioli DJ, Baurens FC, Santos CMR, Alves PC, et al. (2008) Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: isolation, RFLP marker development, and physical mapping. BMC Plant Biol 8: 15.

19. Bakker E, Borm T, Prins P, van der Vossen E, Uenk G, et al. (2011) A genome-wide genetic map of NB-LRR disease resistance loci in potato. Theor Appl Genet 123(3): 493-508.

20. Mun JH, Yu HJ, Park S, Park BS (2009) Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. Mol Genet Genomics 282(6): 617-631.

21. Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, et al. (2011) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. New Phytol 192(1): 1469-8137.

22. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, et al. (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Mol Genet Genomics 281(6): 609-626.

23. Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G (2012) Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group Phureja. PLoS ONE 7(4): e34775.

24. Shao ZQ, Zhang YM, Hang YY, Xue JY, Zhou GC, et al. (2014) Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the Legume family. Plant Physiol 166(1): 217-234.

25. Zhong Y, Yin H, Sargent DJ, Malnoy M, Cheng Z-M (2015) Species-specific duplications driving the recent expansion of NBS-LRR genes in five Rosaceae species. BMC Genomics 16(1): 77.

26. Gu L, Si W, Zhao L, Yang S, Zhang X (2015) Dynamic evolution of NBS-LRRR genes in bread wheat and its progenitors. Mol Genet Genomics 290(2): 727-738.

27. Wei C, Chen J, Kuang H (2016) Dramatic number variation of *R* genes in Solanaceae species accounted for by a few *R* gene subfamilies. PLoS One 11(2): e0148708.

28. Zhang YM, Shao ZQ, Wang Q, Hang YY, Xue JY, et al. (2016) Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae. J Integr Plant Biol 58(2): 165-177.

29. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

30. Richly E, Kurth J, Leister D (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol Biol Evol 19(1): 76-84.

31. McDowell JM, Simon SA (2006) Recent insights into *R* gene evolution. Mol Plant Pathol 7(5): 437-448.

32. Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant *R*-genes. Science 292(5525): 2281-2285.

33. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, et al. (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345(6201): 1181-1184.

34. Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, et al. (2013) Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. Proc Natl Acad Sci USA 110(48): 19478-19482.

35. The Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B, et al. (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355): 189-195.

36. The Tomato Genome Consortium (2012) The tomato sequence provides insights into fleshy fruit evolution. Nature 485(7400): 635-641.

37. The French-Italian Public Consortium for Grapevine Genome Characterization, Jaillon O, Aury JM, Noel B, Policriti A, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161): 463-468.

38. Marchler-Bauer A, Bryant SH (2004) CD-search: protein domain annotations on the fly. Nucleic Acids Res 32: W327-W331.

39. Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, et al. (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. Nucleic Acids Res 39: D1095-D1102.

40. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252(5009): 1162-1164.

41. McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: Improved prediction of coiled coils from sequence. Bioinformatics 22(3): 356-358.

42. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33(2): 511-518.

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10): 2731-2739.

44. Timothy LB, Bodén M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202-W208.

45. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5): 1792-1797.

46. Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2: 516-527.

47. De Mita S, Siol M (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet 13: 27.

48. Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics 280(3): 187-198.

49. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, et al. (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271(4): 402-415.

50. Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet 39: 309-338.

51. Chen F, Mackey AJ, Vermunt JK, Roos DS (2007) Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS ONE 2(4): e383.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

52. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4): 406-425.

53. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4): 783-791.

54. Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet 20(3): 116-122.

55. Yang S, Feng Z, Zhang X, Jiang K, Jin X, et al. (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol Biol 62(1-2): 181-193.

56. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol Biol Evol 3(5): 418-426.

57. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, et al. (2006) KaKs Calculator: calculating Ka and Ks through model selection and model averaging. Geno Prot Bioinfo 4(4): 259-263.

58. Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267(5608): 275-276.

59. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17(10): 1483-1498.

60. Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18(9): 486-487.

61. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-Wide Analysis of NBS-LRR–Encoding Genes in Arabidopsis. Plant Cell 15(4): 809-834.

62. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proc R Soc Lond B 268(1482): 2211-2220.

63. Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54: 23-61.

64. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2(12): e1326.

65. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, et al. (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43(D1): D222-D226.

66. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, et al. (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Mol Biol 66(6): 619-636.

67. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, et al. (2011) The genome of *Theobroma cacao*. Nat Genet 43(2): 101-108.

68. Cheng Y, Li X, Jiang H, Ma W, Miao W, et al. (2012) Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize. FEBS J 279(13): 2431-2443.

69. Wan H, Yuan W, Bo K, Shen J, Pang X, et al. (2013) Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genomics 14: 109.

70. Yue JX, Meyers BC, Chen JQ, Tian D, Yang S (2011) Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol 193(4): 1049-1063.

71. Kim J, Lim CJ, Lee BW, Choi JP, Oh SK, et al. (2012) A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. Mol Cells 33(4): 385-392.

72. Nei M, Gu X, Sitnikova T (1997) Evolution by birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94(15): 7799-7806.

73. Ota T, Nei M (1994) Divergent evolution and evolution by birth-and-death process in the immunoglobulin VH gene family. Mol Biol Evol 11(3): 469-482.

74. Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16(7): 1667-1678.

75. Guyot R, de la Mare M, Viader V, Hamon P, Coriton O, et al. (2009) Microcollinearity in an ethylene receptor coding gene region of the *Coffea canephora* genome is extensively conserved in *Vitis vinifera* and other distant dicotyledonous sequenced genomes. BMC Plant Biol 9: 22.

76. Guyot R, Lefbvre-Pautigny F, Tranchant-Dubreuil C, Rigoreau M, Hamon P, et al. (2012) Ancestral synteny shared between distantly-related plant species from the asteroid (*Coffea canephora* and *Solanum* sp.) and rosid (*Vitis vinifera*) clades. BMC Genomics 13: 103.

77. Cenci A, Combes MC, Lashermes P (2010) Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. Mol Genet Genomics 283(5): 493-501.

78. Jia YX, Yuan Y, Zhang Y, Yang S, Zhang X (2015) Extreme expansion of NBS-encoding genes in *Rosaceae*. BMC Genet 16: 48.

79. Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8(11): 1113-1130.

80. Walsh JB (1995) How often do duplicated genes evolve new functions? Genetics 139(1): 421-428.

81. Zhang R, Murat F, Pont C, Langin T, Salse J (2014) Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genomics 15: 187.

82. Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8(2): 129-134.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

supplementary

Table S1: Results of HMM searches for NBS, LRR and TIR domains and of COILS/PCOILS (Lupas et al. 1991) and PARCOIL (McDonnell et al. 2006) identification for coiled-coil structures in six plant genomes.

	Coffee	Monkey -flower	Potato	Tomato	Subtotal Euasterids	Grapevine	Arabidopsis	Total
No predicted genes	25,574	28,282	39,031	34,727	99,332	30,434	27,169	156,935
No putative NBS genes of high quality	399	188	184	108	879	286	46	1,211
No HMM candidate genes with NBS domains	1,496	853	1,168	758	4,275	1,177	546	5,998
No NBS genes validated by NCBI-CD	670	314	373	222	1,579	459	113	2,151
% of predicted genes	2.62%	1.11%	0.96%	0.64%	1.59%	1.51%	0.42%	1.37%
No validated genes with complete NBS domain	559	268	265	151	1,243	404	108	1,755
% of NCBI-CD validated genes	83.40%	85.40%	71.00%	68.00%	78.70%	88.00%	95.60%	81.60%
No genes with TIR domain	4	0	34	14	52	64	50	166
% of NBS genes	0.70%	0.00%	12.80%	9.30%	4.18%	15.80%	46.30%	9.46%
No genes with CC domain	325	128	127	78	658	169	43	870
% of NBS genes	58.10%	51.50%	47.90%	51.70%	52.90%	41.80%	39.80%	49.60%
No NBS genes with LRR regions	124	36	65	28	253	209	70	532
% of NBS genes	22.20%	13.40%	24.50%	18.50%	20.40%	51.70%	64.80%	30.30%
No TIR-NBS genes	9	0	36	15	60	75	52	187
% of NBS genes	1.60%	0.00%	13.60%	10.00%	4.80%	18.60%	48.10%	10.70%
No non-TIR-NBS genes	550	268	229	136	1,183	329	56	1,568
% of NBS genes	98.40%	100.00%	86.40%	90.00%	95.20%	81.40%	51.90%	89.30%

Table S2: NCBI conserved domains associated with the NBS domain in six plant genomes

Name	Coffee	Monkey-flower	Potato	Tomato	Grapevine	Arabidopsis
AAA superfamily	237	106	78	41	105	29
AAA_14 superfamily	4	2	7	8	2	
AAA_16	226	101	70	37	86	25
AAA_17 superfamily	2					
ABC ATPase superfamily					1	
AhaH superfamily			1			
AMN1 superfamily			4	2	7	6
BRX_N superfamily						1
C2 superfamily		1				
Cation efflux superfamily					1	
COG2263			1			
COG4886	1		1		2	
Cupin_2 superfamily				1		
Dimer_Tnp_hAT superfamily					1	
DUF640 superfamily						1
DUF3542 superfamily	45		8	10		
DUF4220 superfamily					1	
FYVE superfamily						1
GRAS superfamily					1	

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

HMA superfamily	1					
Hyccin superfamily					1	
LEA_2 superfamily						1
LRR_3 superfamily			1		20	39
LRR_4	18	2	1		24	4
LRR_5						1
LRR_8	137	39	77	32	225	41
LRR_RI superfamily	1		2		1	
LRR_TYP superfamily					1	
MMT1					1	
Mob1_phocein superfamily						1
PAH superfamily						1
Pepsin_retropepsin_like superfamily					1	
PH-like superfamily						1
Ras_like_GTPase superfamily					2	
RCC1 superfamily						6
Retropepsin_like					1	
RPW8 superfamily	1		1	2	6	2
TIR_2 superfamily	4		34	16	74	55
WRKY superfamily						1
Total	667	251	286	149	564	216
No domains / NBS gene	1.21	0.94	1.08	0.99	1.4	2

Table S3: Clusters of four NBS genes or more within 200 kb or less (Holub 2001) in six plant genomes.

Genome	Cluster	NBS Genes	Localization*	Start	End	Length
Coffee	CL01	9	chr1	419,818	968,243	548,425
Coffee	CL02	4	chr1	19,737,175	19,948,638	211,463
Coffee	CL03	7	chr1	26,195,001	26,513,689	318,688
Coffee	CL04	5	chr2	27,273,667	27,718,278	444,611
Coffee	CL05	4	chr2	33,529,134	33,764,772	235,638
Coffee	CL06	8	chr3	4,878,515	5,338,582	460,067
Coffee	CL07	4	chr3	7,986,357	8,149,709	163,352
Coffee	CL08	5	chr3	17,026,230	17,456,448	430,218
Coffee	CL09	5	chr3	21,935,775	22,226,095	290,320
Coffee	CL10	6	chr3	22,628,644	22,875,374	246,730
Coffee	CL11	5	chr3	28,961,770	29,110,403	148,633
Coffee	CL12	6	chr3	29,329,643	29,603,977	274,334
Coffee	CL13	6	chr4	12,697,375	13,095,583	398,208
Coffee	CL14	4	chr4	19,760,582	20,023,477	262,895
Coffee	CL15	5	chr5	1,674,987	2,036,590	361,603
Coffee	CL16	7	chr5	18,142,560	18,774,487	631,927
Coffee	CL17	6	chr5	21,250,780	21,401,768	150,988
Coffee	CL18	5	chr6	14,195,803	14,241,543	45,740
Coffee	CL19	5	chr6	21,431,123	21,779,230	348,107
Coffee	CL20	5	chr7	14,215,921	14,397,410	181,489

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Coffee	CL21	4	chr7	15,036,928	15,090,663	53,735
Coffee	CL22	7	chr8	1,629,285	1,781,383	152,098
Coffee	CL23	13	chr8	2,098,669	2,755,761	657,092
Coffee	CL24	6	chr8	4,786,562	5,072,536	285,974
Coffee	CL25	6	chr8	5,880,525	6,050,935	170,410
Coffee	CL26	4	chr8	7,201,292	7,424,466	223,174
Coffee	CL27	5	chr11	6,014,491	6,297,912	283,421
Coffee	CL28	4	chr11	15,691,634	15,916,405	224,771
Coffee	CL29	5	chr11	16,250,139	16,688,695	438,556
Coffee	CL30	4	chr11	26,171,147	26,362,348	191,201
Coffee	CL31	8	chr0	13,071,309	13,590,717	519,408
Coffee	CL32	4	chr0	21,183,074	21,427,717	244,643
Coffee	CL33	4	chr0	42,029,100	42,307,369	278,269
Coffee	CL34	4	chr0	68,715,522	68,855,591	140,069
Coffee	CL35	4	chr0	81,144,342	81,291,308	146,966
Coffee	CL36	4	chr0	93,670,814	93,731,494	60,680
Coffee	CL37	5	chr0	173,429,049	173,949,938	520,889
Coffee	CL38	5	chr0	180,732,504	181,022,493	289,989
Coffee	CL39	6	chr0	184,248,807	184,577,497	328,690
Coffee	CL40	4	chr0	187,532,750	187,756,575	223,825
Monkey-flower	CL01	4	scaffold_150	242,917	394,243	151,326
Monkey-flower	CL02	8	scaffold_16	245,035	515,115	270,080
Monkey-flower	CL03	4	scaffold_238	90,702	216,667	125,965
Monkey-flower	CL04	5	scaffold_240	92,791	280,045	187,254
Monkey-flower	CL05	4	scaffold_281	63,730	95,442	31,712
Monkey-flower	CL06	5	scaffold_55	424,981	510,306	85,325
Monkey-flower	CL07	6	scaffold_55	870,156	1,031,278	161,122
Monkey-flower	CL08	14	scaffold_57	698,547	1,421,743	723,196
Monkey-flower	CL09	21	scaffold_6	80,356	989,732	909,376
Monkey-flower	CL10	5	scaffold_6	1,513,541	1,812,047	298,506
Monkey-flower	CL11	6	scaffold_60	35,901	182,540	146,639
Monkey-flower	CL12	4	scaffold_69	994,278	1,030,999	36,721
Monkey-flower	CL13	13	scaffold_7	1,007,732	1,278,521	270,789
Monkey-flower	CL14	12	scaffold_7	3,473,731	3,665,777	192,046
Monkey-flower	CL15	5	scaffold_70	1,079,850	1,230,003	150,153
Monkey-flower	CL16	33	scaffold_96	26,653	823,847	797,194
Monkey-flower	CL17	6	scaffold_97	232,736	429,739	197,003
Potato	CL01	5	chr01	79,482,156	79,670,255	188,099
Potato	CL02	5	chr04	778,134	871,272	93,138
Potato	CL03	5	chr04	2,813,381	3,039,439	226,058
Potato	CL04	10	chr04	4,613,730	4,765,194	151,464
Potato	CL05	5	chr04	5,708,305	5,968,326	260,021
Potato	CL06	4	chr05	4,589,149	4,843,837	254,688

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Potato	CL07	7	chr06	851,351	1,167,783	316,432
Potato	CL08	5	chr06	47,328,471	47,537,661	209,190
Potato	CL09	7	chr08	702,201	862,136	159,935
Potato	CL10	4	chr08	3,733,192	3,827,992	94,800
Potato	CL11	6	chr08	48,381,495	48,602,017	220,522
Potato	CL12	9	chr09	59,398,259	59,742,918	344,659
Potato	CL13	6	chr09	60,622,644	60,967,756	345,112
Potato	CL14	4	chr10	52,682,205	52,874,129	191,924
Potato	CL15	5	chr11	1,198,844	1,589,104	390,260
Potato	CL16	7	chr11	5,842,698	6,071,020	228,322
Potato	CL17	4	chr11	43,970,064	44,219,141	249,077
Potato	CL18	4	chr12	1,684,447	1,758,197	73,750
Potato	CL19	5	chr12	6,342,492	6,418,862	76,370
Tomato	CL1	4	chr04	742,844	796,086	53,242
Tomato	CL2	8	chr04	2,637,079	2,738,484	101,405
Tomato	CL3	4	chr04	4,319,129	4,352,103	32,974
Tomato	CL4	5	chr05	1,779,356	2,153,957	374,601
Tomato	CL5	4	chr06	2,318,596	2,377,556	58,960
Tomato	CL6	5	chr06	36,704,723	36,756,163	51,440
Tomato	CL7	4	chr09	66,762,154	66,864,834	102,680
Tomato	CL8	4	chr11	1,181,636	1,245,275	63,639
Grapevine	CL01	7	chr1	792,601	953,756	161,155
Grapevine	CL02	7	chr3	1,039,812	1,109,586	69,774
Grapevine	CL03	8	chr5	21,002,768	21,410,746	407,978
Grapevine	CL04	6	chr6	15,876,112	16,304,165	428,053
Grapevine	CL05	4	chr9	3,431,024	3,647,634	216,610
Grapevine	CL06	9	chr9	4,441,751	4,842,503	400,752
Grapevine	CL07	6	chr9	5,892,087	6,121,512	229,425
Grapevine	CL08	5	chr12	14,228,796	14,725,224	496,428
Grapevine	CL09	14	chr12	15,043,216	15,674,062	630,846
Grapevine	CL10	11	chr12	16,460,669	16,784,609	323,940
Grapevine	CL11	5	chr13	393,371	461,935	68,564
Grapevine	CL12	20	chr13	12,662,945	13,673,050	1,010,105
Grapevine	CL13	8	chr13R	1,098,226	1,362,994	264,768
Grapevine	CL14	6	chr15R	2,382,003	2,489,269	107,266
Grapevine	CL15	5	chr18	15,067,355	15,429,784	362,429
Grapevine	CL16	18	chr18	16,860,211	17,776,163	915,952
Grapevine	CL17	8	chr18	18,482,769	19,211,374	728,605
Grapevine	CL18	5	chr18R	3,982,122	4,399,751	417,629
Grapevine	CL19	6	chr18R	4,782,708	5,260,032	477,324
Grapevine	CL20	5	chr19	5,356,328	5,392,638	36,310
Grapevine	CL21	5	chr19	11,752,849	12,047,436	294,587
Grapevine	CL22	4	chr1R	17,534,534	17,758,939	224,405

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Grapevine	CL23	12	chr1R	24,834,306	25,285,767	451,461
Grapevine	CL24	14	chr1R	117,814,740	118,316,557	501,817
Grapevine	CL25	5	chr1R	123,397,929	124,035,964	638,035
Grapevine	CL26	5	chr1R	129,273,569	129,572,299	298,730
Grapevine	CL27	15	chr1R	133,154,727	133,984,095	829,368
Grapevine	CL28	4	chr1R	137,578,402	137,742,004	163,602
Arabidopsis	CL1	4	chr1	4,140,948	4,181,247	40,299
Arabidopsis	CL2	9	chr1	21,690,962	21,997,691	306,729
Arabidopsis	CL3	4	chr1	22,551,486	22,615,943	64,457
Arabidopsis	CL4	5	chr1	23,494,935	23,783,449	288,514
Arabidopsis	CL5	7	chr4	9,488,584	9,551,007	62,423
Arabidopsis	CL6	4	chr4	10,440,102	10,647,070	206,968
Arabidopsis	CL7	4	chr5	18,182,038	18,332,229	150,191

Table S4: Means of nucleotide diversity (π), non-synonymous substitutions (Ka), synonymous substitutions (Ks), and ratio of non-synonymous substitutions to synonymous ones (Ka:Ks) in the NBS domain of three major euasterid orthogroups.

Orthogroup	Genome	No Gene Pairs	π	Ka	Ks	Ka:Ks
OG5_160141	Coffee	19,053	0.387	0.527	0.759	0.797
	Monkey-flower	9,689	0.388	0.528	0.732	0.811
	Potato	5,342	0.395	0.541	0.765	0.789
	Tomato	2,609	0.398	0.55	0.771	0.791
OG5_134032	Coffee	4,421	0.377	0.506	0.731	0.778
	Monkey-flower	248	0.303	0.407	0.496	0.952
	Potato	985	0.338	0.442	0.667	0.757
	Tomato	91	0.348	0.452	0.733	0.718
OG5_173771	Coffee	2,756	0.306	0.346	0.945	0.426
	Monkey-flower	274	0.393	0.528	0.881	0.643
	Potato	45	0.404	0.542	1.04	0.616
	Tomato	21	0.396	0.532	0.905	0.642

Table S5: Means of nucleotide diversity (π), non-synonymous substitutions (Ka), synonymous substitutions (Ks), and ratio of non-synonymous substitutions to synonymous ones (Ka:Ks) in the NBS domain of euasterid homogeneous and heterogeneous clusters.

Type	Genome	Cluster	No Gene Pairs	π	No Gene Pairs with π < 5%	Ka	Ks	Ka:Ks	No Gene Pairs with Ka>Ks
Heterogeneous	Coffee	CL15	10	0.106	0	0.089	0.197	0.473	0
Heterogeneous	Coffee	CL16	21	0.16	0	0.369	0.569	0.537	0
Heterogeneous	Coffee	CL17	15	0.196	0	0.256	0.397	0.573	1
Heterogeneous	Coffee	CL35	6	0.321	0	0.43	0.575	0.709	1
Heterogeneous	Monkey-flower	CL03	6	0.241	3	0.664	0.836	0.805	0
Heterogeneous	Monkey-flower	CL04	10	0.092	0	0.098	0.103	1.086	4
Heterogeneous	Monkey-flower	CL07	15	0.323	0	0.399	0.691	0.566	1
Heterogeneous	Monkey-flower	CL08	91	0.208	5	0.267	0.59	0.445	3
Heterogeneous	Monkey-flower	CL09	210	0.36	2	0.482	0.754	0.668	16
Heterogeneous	Monkey-flower	CL13	78	0.222	4	0.254	0.481	0.539	2
Heterogeneous	Monkey-flower	CL14	66	0.109	4	0.088	0.272	0.335	0

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Heterogeneous	Monkey-flower	CL15	10	0.261	0	0.291	0.733	0.409	0
Heterogeneous	Monkey-flower	CL16	528	0.343	1	0.434	0.726	0.654	41
Heterogeneous	Potato	CL07	21	0.103	1	0.095	0.173	0.559	0
Heterogeneous	Potato	CL08	10	0.13	0	0.111	0.286	0.396	0
Heterogeneous	Potato	CL10	6	0.354	1	0.599	0.873	0.714	0
Heterogeneous	Potato	CL12	36	0.202	3	0.297	0.479	0.525	1
Heterogeneous	Potato	CL13	15	0.182	1	0.191	0.545	0.37	0
Heterogeneous	Tomato	CL4	10	0.299	0	0.357	0.619	0.703	2
Heterogeneous	Tomato	CL5	6	0.064	3	0.083	0.151	0.551	0
Heterogeneous	Tomato	CL6	10	0.1	1	0.094	0.209	0.454	0
Homogeneous	Coffee	CL03	21	0.244	0	0.226	0.68	0.353	1
Homogeneous	Coffee	CL04	10	0.148	6	0.275	2.203	0.125	0
Homogeneous	Coffee	CL05	6	0.173	1	0.181	0.499	0.384	0
Homogeneous	Coffee	CL06	28	0.335	1	0.433	0.784	0.587	1
Homogeneous	Coffee	CL07	6	0.143	0	0.113	0.38	0.316	0
Homogeneous	Coffee	CL14	6	0.101	0	0.078	0.223	0.356	0
Homogeneous	Coffee	CL19	10	0.174	0	0.151	0.413	0.421	0
Homogeneous	Coffee	CL20	10	0.208	0	0.198	0.479	0.421	0
Homogeneous	Coffee	CL21	6	0.186	0	0.193	0.308	0.649	0
Homogeneous	Coffee	CL22	21	0.156	0	0.143	0.347	0.374	0
Homogeneous	Coffee	CL23	78	0.116	3	0.091	0.284	0.399	0
Homogeneous	Coffee	CL24	15	0.076	0	0.061	0.156	0.436	0
Homogeneous	Coffee	CL25	15	0.107	0	0.103	0.162	0.649	0
Homogeneous	Coffee	CL26	6	0.059	1	0.052	0.11	0.49	0
Homogeneous	Coffee	CL27	10	0.087	0	0.067	0.194	0.358	0
Homogeneous	Coffee	CL31	28	0.189	0	0.184	0.361	0.583	1
Homogeneous	Coffee	CL33	6	0.107	0	0.09	0.228	0.372	0
Homogeneous	Monkey-flower	CL01	6	0.378	0	0.457	1.088	0.464	0
Homogeneous	Monkey-flower	CL02	28	0.342	0	0.393	0.991	0.422	0
Homogeneous	Monkey-flower	CL11	15	0.167	0	0.131	0.451	0.291	0
Homogeneous	Potato	CL04	45	0.108	0	0.095	0.208	0.494	0
Homogeneous	Potato	CL05	10	0.065	3	0.066	0.119	0.602	1
Homogeneous	Potato	CL09	21	0.139	0	0.137	0.215	0.678	2
Homogeneous	Potato	CL11	15	0.116	0	0.089	0.288	0.336	0
Homogeneous	Potato	CL17	6	0.14	1	0.15	0.301	0.691	2
Homogeneous	Potato	CL18	6	0.18	2	0.149	0.414	0.379	0
Homogeneous	Potato	CL19	10	0.253	0	0.283	0.49	0.531	1
Homogeneous	Potato (TIR)	CL01	10	0.369	0	0.443	0.953	0.581	1
Homogeneous	Tomato	CL1	6	0.29	0	0.348	0.482	0.765	1
Homogeneous	Tomato	CL2	28	0.1	0	0.081	0.211	0.405	0
Homogeneous	Tomato	CL3	6	0.099	1	0.085	0.193	0.438	0
Homogeneous	Tomato	CL8	6	0.101	0	0.072	0.261	0.278	0

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

Table S6: Dating of large-scale duplication events (Mya) in the major euasterid orthogroups using grapevine as reference and based on fossil deposits for time calibration (Wikström et al. 2001). Eleven major events were highlighted according to time intervals.

Orthogroup	Genome	No Gene Pairs	Mean Ks	Minimum	Maximum	Event n°
OG5_173771	Potato-Potato	7	0.9353	105.7	115.9	1
OG5_160141	Potato-Potato	905	0.7499	105.2	115.4	1
OG5_160141	Tomato-Tomato	455	0.7476	104.1	114.2	1
OG5_173771	Monkey-flower-Monkey-flower	49	0.8474	99.6	109.3	2
OG5_134032	Coffee-Coffee	738	0.6531	95.1	104.3	3
OG5_134032	Potato-Potato	148	0.651	94.2	103.3	3
OG5_173771	Coffee-Coffee	195	0.8489	94.2	103.3	3
OG5_134032	Tomato-Tomato	17	0.6493	94	103.1	3
OG5_134032	Monkey-flower-Monkey-flower	29	0.6488	93.9	103	3
OG5_160141	Coffee-Coffee	2,782	0.6501	90.7	99.4	4
OG5_160141	Monkey-flower-Monkey-flower	1,576	0.6511	89.5	98.2	4
OG5_173771	Tomato-Tomato	4	0.7575	83.5	91.6	5
OG5_173771	Potato-Potato	8	0.6471	73.2	80.2	6
OG5_173771	Coffee-Coffee	302	0.5534	61.4	67.4	7
OG5_134032	Monkey-flower-Monkey-flower	62	0.1676	24.3	26.6	8
OG5_134032	Tomato-Tomato	9	0.1617	23.4	25.7	8
OG5_134032	Potato-Potato	165	0.1504	21.8	23.9	9
OG5_160141	Potato-Potato	121	0.1471	20.6	22.6	10
OG5_173771	Monkey-flower-Monkey-flower	7	0.172	20.2	22.2	10
OG5_134032	Coffee-Coffee	121	0.1374	20	21.9	10
OG5_173771	Coffee-Coffee	171	0.1487	16.5	18.1	11

Table S7: Number of NBS genes involved in the large-scale duplication events per chromosome and orthogroup.

Genome	Event	Orthogroup	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr0	Total
	1	OG5_160141	7	4		16	16	14	4	3	14	3	7	13	3	104
	1	OG5_173771	2			2	1	1			1		1		1	9
	3	OG5_134032	3		2		1			12	2	11	13		1	45
	6	OG5_173771	2			2	1	1			1		1		1	9
	9	OG5_134032					1			11	1	11	11		1	36
	10	OG5_160141	5	2		9	6	14	2		10	1	6	8	1	64
Potato		Total	19	6	2	29	26	30	6	26	29	26	39	21	8	267
	1	OG5_160141	2	2	3	13	20	8	3	2	5	3	6	5	1	73
	3	OG5_134032	1	2	2					4	1		4			14
	5	OG5_173771	1			1	3				1		1			7
	8	OG5_134032								4	1		3			8
Tomato		Total	4	4	5	14	23	8	3	10	8	3	14	5	1	102

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

	3	OG5_134032	5	18	9	2	11		1	7				N/A*	42	95
	3	OG5_173771		1	2	2	1	2		28		1	7	N/A	29	73
	4	OG5_160141	9	3	35	16	11	8	19	1	1	5	43	N/A	45	196
	7	OG5_173771		1	1	2	1	2		28		1	6	N/A	27	69
	10	OG5_134032	5	7	4	2	10		1	7				N/A	32	68
	11	OG5_173771			2	2		1		28			3	N/A	24	60
Coffee		Total	19	30	53	26	34	13	21	99	1	7	59	N/A	199	561

* N/A not applicable.

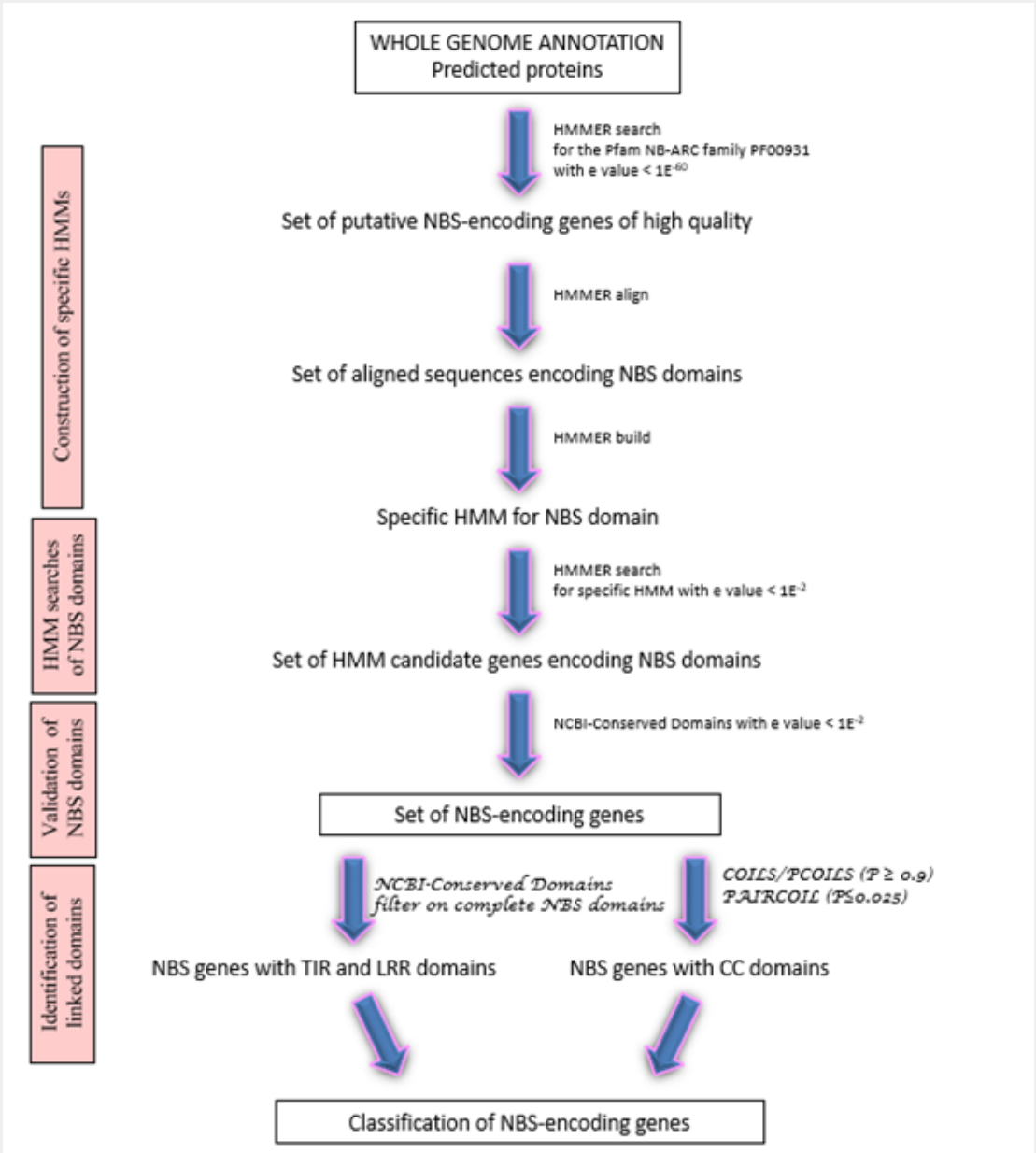


Figure S1: Workflow used for the identification of NBS genes.

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

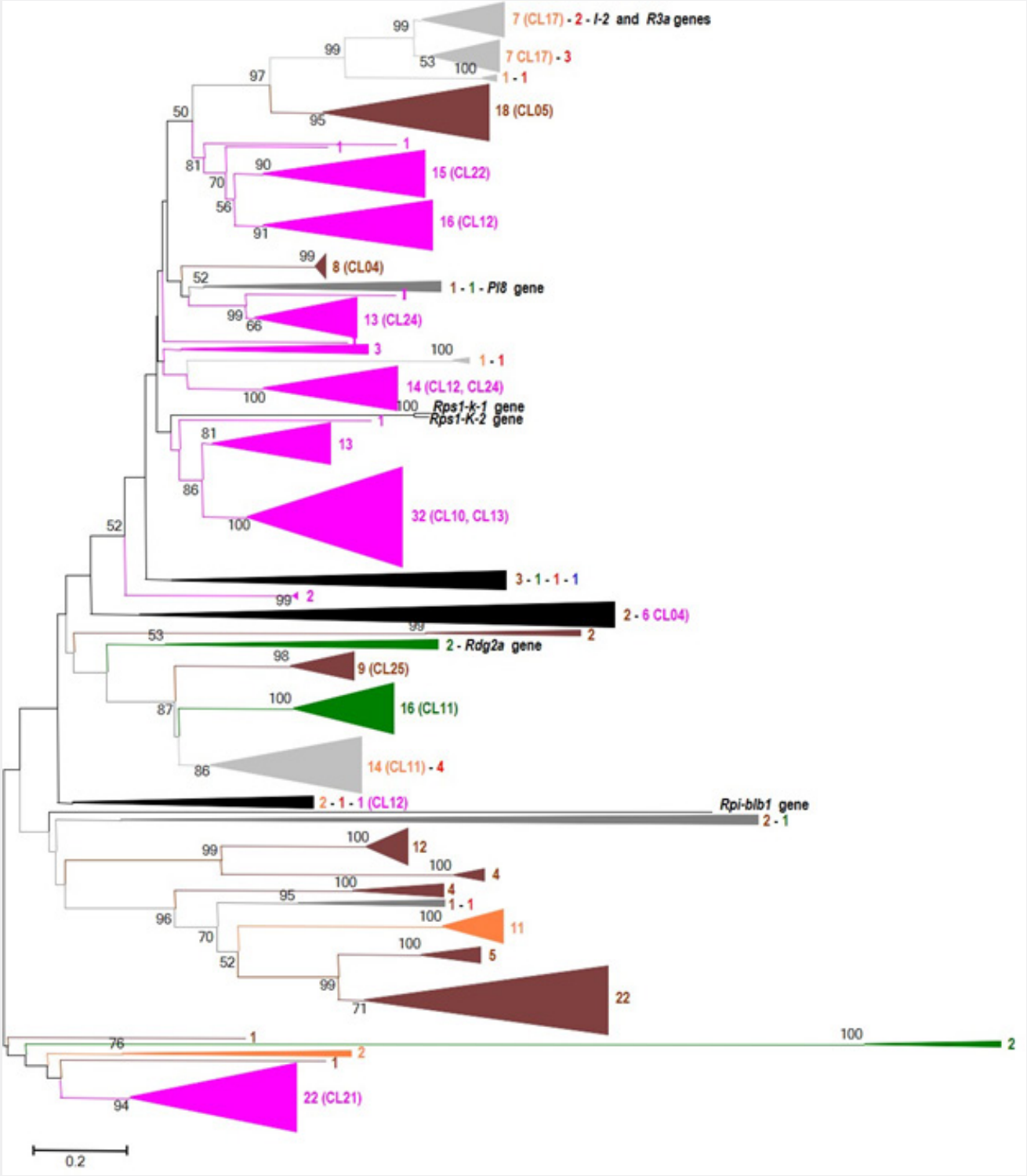
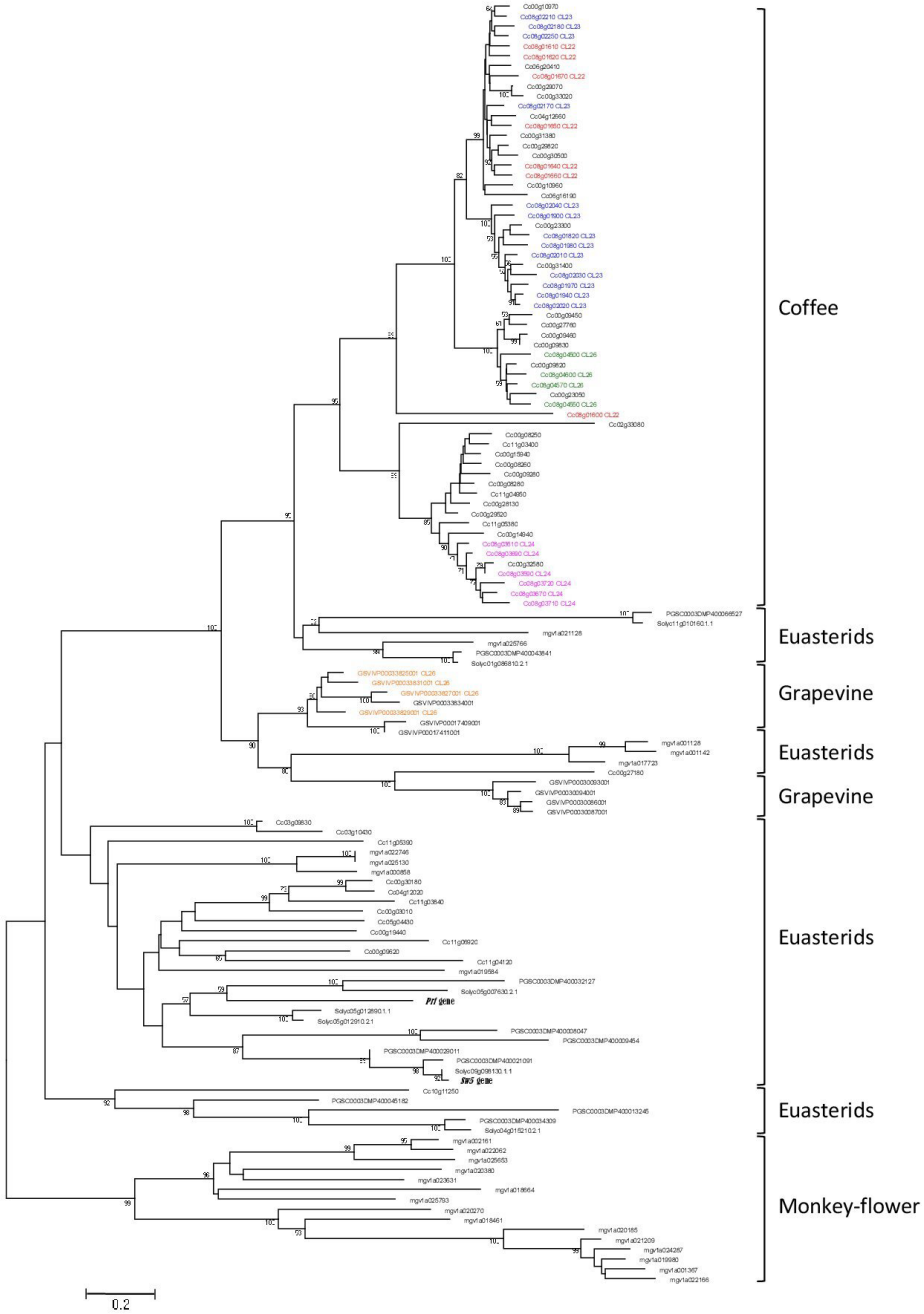


Figure S2: Neighbour-joining tree of NBS domains from R gene analogues and cloned R genes (in bold), belonging to orthogroup OG5_134032. The complete tree was based on 327 sequences. Clades containing genes from individual plants were collapsed into single branches and the number of sequences in each branch is indicated on the right. Different taxa are shown in different colours; clades with R gene analogues from several plants are in light grey (potato and tomato), dark grey (euasterids) and black (euasterids and eurosids). Homogeneous gene clusters (ie. fully represented in the orthogroup) are in brackets after the number of sequences. Bootstrap percentages □ 50% are indicated at branches.

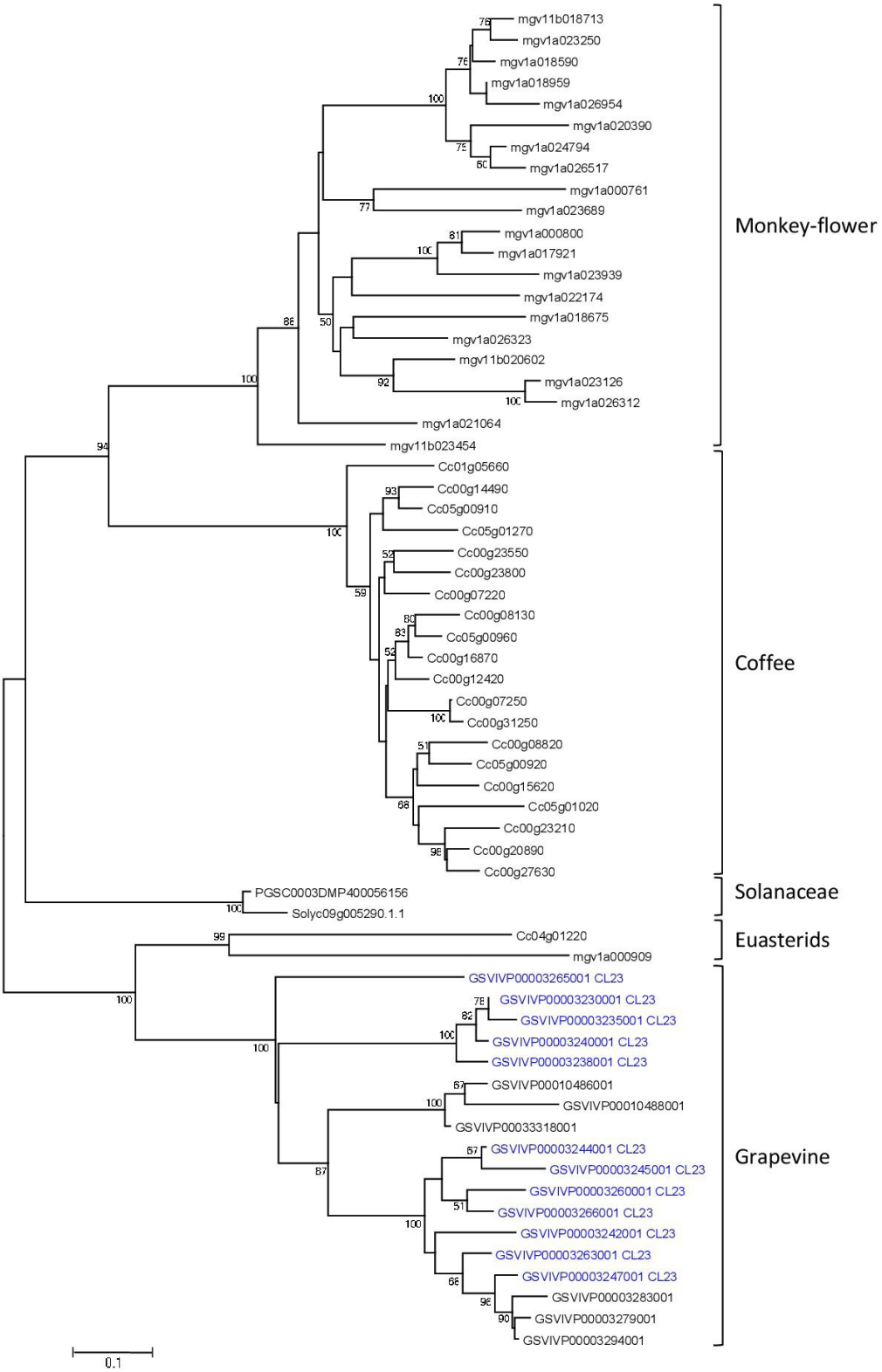
How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

(a) Orthogroup OG5_173771



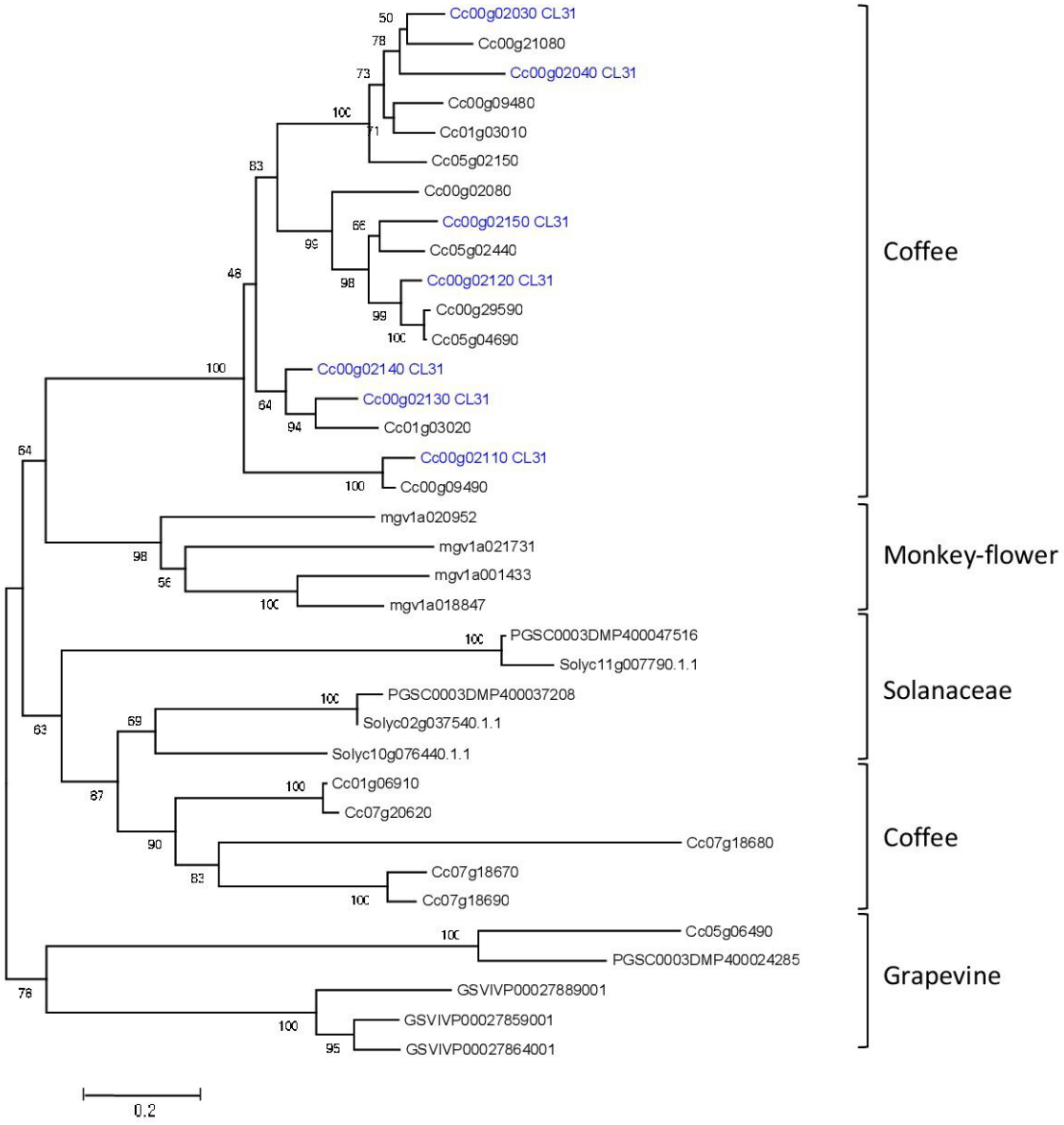
How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

(b) Orthogroup OG5_143984



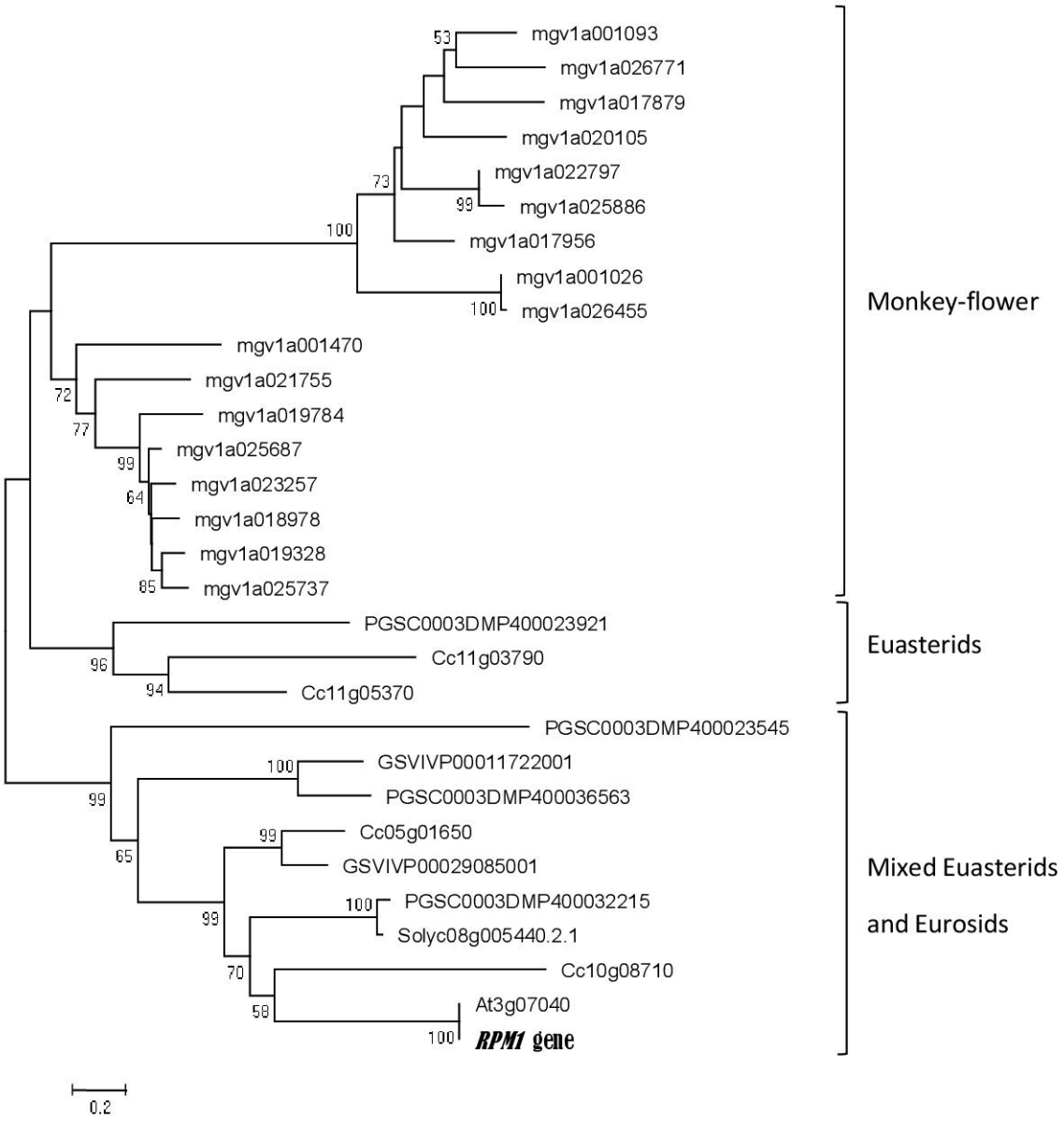
How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

(c) Orthogroup OG5_205945



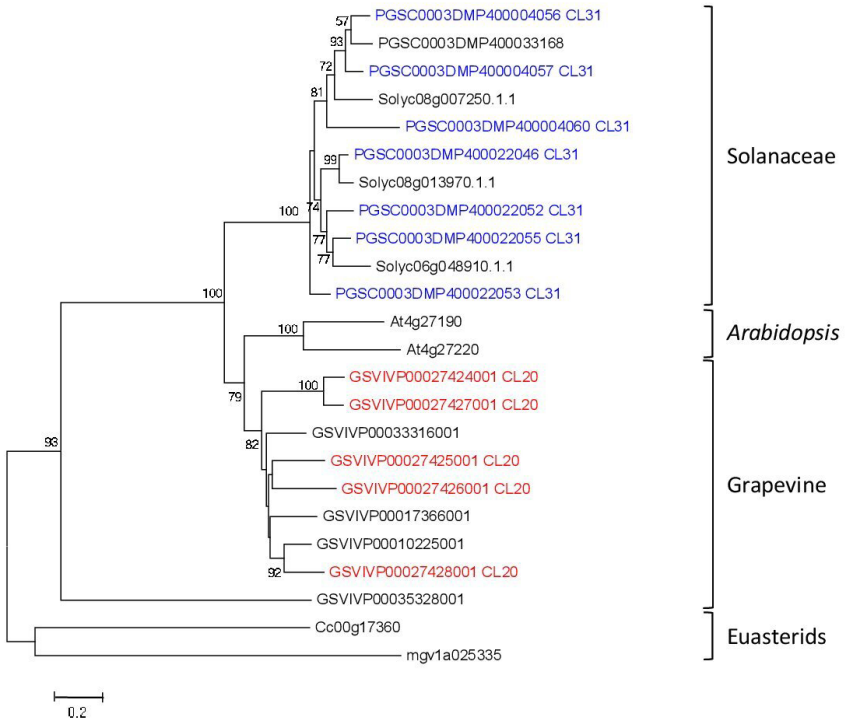
How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

(d) Orthogroup OG5_134568



How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

(e) Orthogroup OG5_212667



(f) Orthogroup OG5_184596

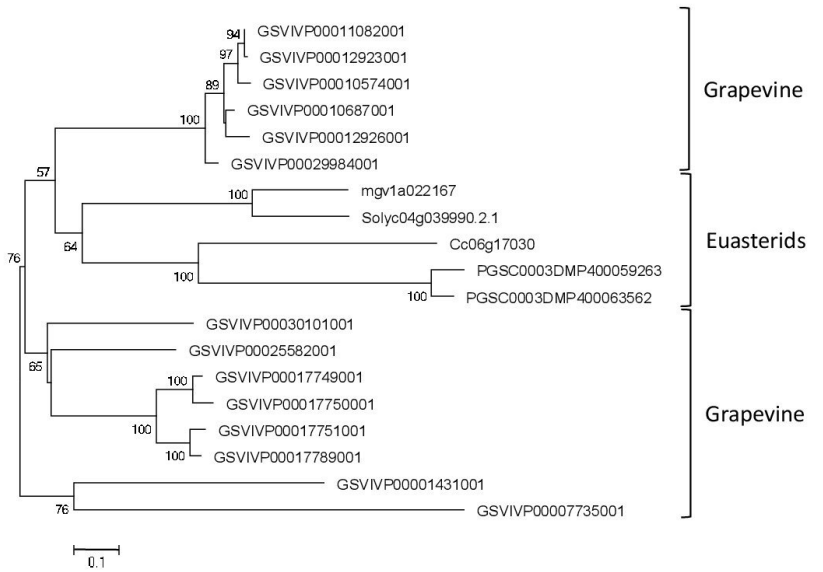


Figure S3: Neighbour-joining tree of NBS domains from R gene analogues and cloned R genes (in bold), belonging to orthogroups present in the four euasterid genomes. Homogeneous gene clusters (ie. fully represented in the orthogroup) are shown in different colors. Bootstrap percentages > 50% are indicated at branches

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296

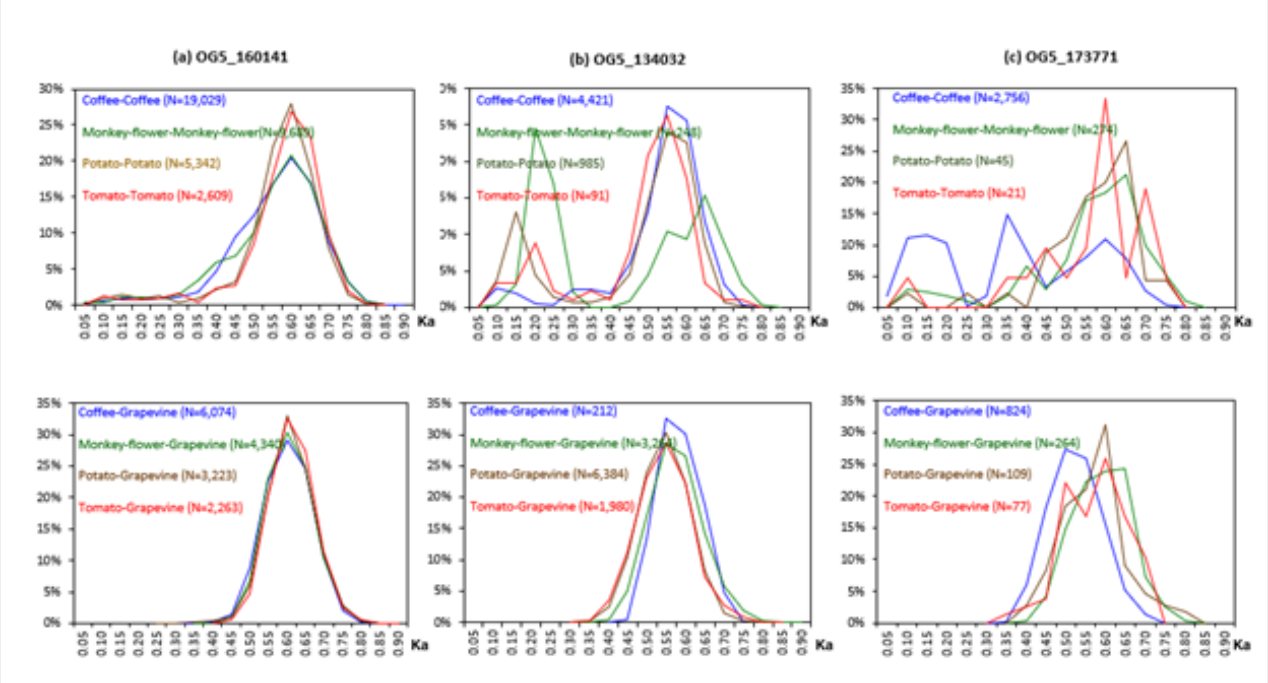


Figure S4: Distribution of Ka frequency in the NBS domain of homologues from the major euasterid orthogroups. The relative frequency of sequence pairs is shown on the vertical axes and Ka values at 0.1 intervals on the horizontal axes. The number of gene pairs is given in brackets.

DOI: 10.19080/IJESNR.2022.30.556296

How to cite this article: François A, Gennady E, Alexis D. Reconstructing the Evolutionary History of Nucleotide-Binding Site (NBS) Genes in Euasterids. Int J Environ Sci Nat Res. 2022; 30(4): 556296. DOI: 10.19080/IJESNR.2022.30.556296