Journal of Advanced Research 42 (2022) 17–28

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original Article

Identification of *Pseudo-R* genes in *Vitis vinifera* and characterization of their role as immunomodulators in host-pathogen interactions

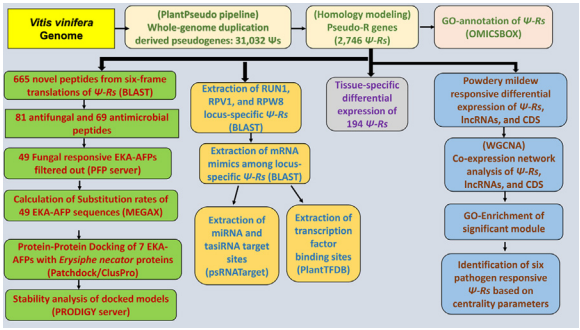
Naina Garewal [a], Shivalika Pathania [a], Garima Bhatia [a,b], Kashmir Singh [a,*]

[a] Department of Biotechnology, Panjab University, Chandigarh, India
[b] Department of Biology, University of Pennsylvania, Philadelphia, USA[1]

HIGHLIGHTS

- Resistance genes associated pseudogenes (*pseudo-R* genes) derived from whole-genome duplications were identified in the genome of *Vitis vinifera* and were annotated for their roles in plant defence responses.
- Novel antifungal (EKA-AFPs) and antimicrobial peptides (*Ψ*-AMPs) identified and characterized can serve as promising anti-microbial candidates.
- The co-expression network analysis between pseudogenes-lncRNAs-genes revealed six pathogen-responsive *pseudo-R* genes as significant during pathogen invasion.
- *Pseudo-R* genes were also exhibiting tissue-specific expression patterns.
- *Pseudo-R* genes play by regulating the gene expression either directly by acting as mRNA mimics for miRNA/tasiRNA targeting or indirectly by lncRNA mediated regulation of miRNA/tasiRNAs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 25 April 2022
Revised 13 July 2022
Accepted 29 July 2022
Available online 3 August 2022

Keywords:
Pseudogene
lncRNAs
Integrated-network analysis

ABSTRACT

Introduction: Duplication events are fundamental to co-evolution in host-pathogen interactions. Pseudogenes (*Ψs*) are dysfunctional paralogs of functional genes and resistance genes (*Rs*) in plants are the key to disarming pathogenic invasions. Thus, deciphering the roles of *pseudo-R* genes in plant defense is momentous.

Objectives: This study aimed to functionally characterize diverse roles of the resistance *Ψs* as novel gene footprints and as significant gene regulators in the grapevine genome.

Methods: PlantPseudo pipeline and HMM-profiling identified whole-genome duplication-derived (WGD) *Ψs* associated with resistance genes (*Ψ-Rs*). Further, novel antifungal and antimicrobial peptides were characterized for fungal associations using protein–protein docking with *Erysiphe necator* proteins. miRNA and tasiRNA target sites and transcription factor (TF) binding sites were predicted in *Ψ-Rs*.

Peer review under responsibility of Cairo University.

* Corresponding author.
E-mail address: kashmirbio@pu.ac.in (K. Singh).
[1] Current address.

https://doi.org/10.1016/j.jare.2022.07.014

Resistance genes
Powdery mildew
Antimicrobial

Finally, differential co-expression patterns in *Ψ-Rs*-lncRNAs-coding genes were identified using the UPGMA method.

Results: 2,746 *Ψ-Rs* were identified from 31,032 WGD *Ψs* in the genome of grapevine. 69-antimicrobial and 81-antifungal novel peptides were generated from *Ψ-Rs*. The putative genic potential was predicted for five novel antifungal peptides which were further characterized by docking against *E. necator* proteins. 395 out of 527 resistance loci-specific *Ψ-Rs* were acting as parental gene mimics. Further, to explore the diverse roles of *Ψ-Rs* in plant-defense, we identified 37,026 TF-binding sites, 208 miRNA, and 99 tasiRNA targeting sites on these Ψ-Rs. 194 *Ψ-Rs* were exhibiting tissue-specific expression patterns. The co-expression network analysis between *Ψs*-lncRNA-genes revealed six out of 79 pathogen-responsive *Ψ-Rs* as significant during pathogen invasion.

Conclusions: Our study provides pathogen responsive *Ψ-Rs* integral for pathogen invasion, which will offer a useful resource for future experimental validations. In addition, our findings on novel peptide generations from *Ψ-Rs* offer valuable insights which can serve as a useful resource for predicting novel genes with the futuristic potential of being investigated for their bioactivities in the plant system.

Introduction

Pseudogenes (*Ψs*) are paralogous genic fragments incapable of forming fully functional proteins due to disarming mutations such as indels and frame-shift mutations. *Ψs* are of two main types: processed and non-processed (also known as duplicated *Ψs*). Duplicated *Ψs* will have typically lost their function with time due to lack of selection pressure on their genic structure. This often leads to the generation of novel genes, possibly through *Ψs* gaining new functions by escaping detrimental mutations and becoming functional again [1]. Although inert, *Ψs* are involved in the formation of long non-coding RNAs (lncRNAs), micro RNAs (miRNAs), *trans*-acting siRNAs (tasiRNAs), etc. and modulate their parental genes by acting as anti-sense RNA molecules, siRNA molecules, competitive endogenous miRNA targets via target mimicry, regulation of translational machinery, and through interactions with RNA binding proteins, [2–4]. The close co-relation of lncRNA and *Ψs* as being biologically active has led to speculations *Ψs* could form new lncRNAs to modulate their parent genes via target mimicry. [5–7].

Powdery mildew (PM) caused by a biotrophic fungus *E. necator* is a colossal terror for extreme yield losses of grapevine crops worldwide. Resistance (*R*) genes comprising primarily of nucleotide binding site-leucine rich repeat genes (NLRs) are involved in conferring post-penetration resistance against PM pathogens. The NLRs arise after reiterated tandem duplications in the genome and random shuffling through evolution leading to formation of novel receptors in conjunction with pathogen effectors. The occurrence of novelty in R-genes is largely preserved especially in the wild genotypes of the vegetatively propagated crop plants [8].

Interestingly, considerable research has been directed towards *Ψs* such as those involved in cancer after the discovery of PTEN-*Ψ* being functionally active [5]. Similarly in plants, the majority of research circumscribes their genome-wide identification, phylogenetic analysis, plant-host studies, etc. [9]. In addition, *Ψs* have reportedly been transcriptionally active in some plants [10,11]. The emerging significance of *Ψs* in innate immune response in various cancers such as gastric cancer, breast cancer, pancreatic cancer, hepatic cancer etc. laid the foundations of our study [12]. In 2015, *Plasmodium falciparum* resistance-associated locus pfcrt aided in the identification of novel *Ψs* in malaria patients [13]. In addition, certain *Ψs* are able to transcribe mRNA fragments into peptides which serve as target antigens and induce immune responses in humans [14]. Plant antimicrobial peptides are a primary component of the barrier defense system against various phytopathogenic infections [15]. Studies involving pseudogene derived antimicrobial peptides were previously conducted in *Drosophila melanogaster* and *Escherichia coli* [16,17]. Although, the studies involving the artificial expression of pseudogene-based peptides are still in their infancy, but exhibit great potentials of pseudogenes being expressed as novel and functional proteins [18]. To the best of our knowledge, this is the first ever report predicting novel antifungal peptides from pseudogenes in plants and investigating *in-silico* their roles as potential antifungal agents against the *E. necator*.

Materials and methods

Identification of whole-genome duplication derived Ψs

Genome-wide identification of whole-genome duplication (WGD)-derived *Ψs* was conducted by using PlantPseudo pipeline [19]. The *V. vinifera* genome (PN40024-12X.v2) was downloaded from Ensembl Plants. The inputs of pipeline were unmasked genome, repeat-masked genomic sequences, non-redundant set of protein sequences in the genome, and GFF3 (General Feature Format) files for lncRNAs, genome, and repeat-masked genomic sequences. All FASTA and genome.gff3 files were obtained from Ensembl Plants, except the repeat masked genomic sequences gff3 file, which was obtained from its FASTA file using pblat sequence alignment tool [20]. *V. vinifera* lncRNA annotation was obtained from previous studies [21,22].

Identification and characterization of Pseudo-R genes (Ψ-Rs)

Further, to identify *R*-genes associated *Ψs*, a Hidden Markov Model (HMM) profile was generated for the reference *R*-genes downloaded from the Plant Resistance Genes Database (PRGdb-3.0) [23]. The multiple sequence alignment (MSA) for reference *R*-genes was performed using Clustal Omega from the EBI web server and their FASTA sequences were used as an input to build the HMM profile using module 'hmmbuild' in HMMER v3.3 [24]. Further, the HMM profile obtained was searched against the *V. vinifera* protein sequences using module 'hmmsearch' to shortlist sequences with a significant e-value of 1e-03. These were further examined using the 'hmmscan' module in HMMER to confirm the presence of NLR domains. The identified *Ψ-Rs* were then plotted using Circos software along with lncRNA and WGD-*Ψs*. The *Ψ-Rs* were functionally annotated using OmicsBox 2.0 (formerly known as Blast2GO) with default parameters [25]. Gene Ontology (GO) terms were assigned to the annotated sequences based on three categories such as biological process (BP), molecular function (MF), and cellular component (CC) along with pathway enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

Table 1
Molecular function GO terms and substitution rates of novel EKA-peptides of fungal defense response.

EKA-AFPs	PFP-Molecular function GO-prediction	dN/dS	Selection type
EKA-15	FUNGUS DEFENSE RESPONSE, ncRNA Processing	0.82186235	Negative selection
EKA-22	FUNGUS DEFENSE RESPONSE, Defense response	1.44003056	Positive selection
EKA-49	FUNGUS DEFENSE RESPONSE, Defense response	1.03092784	Positive selection
EKA-74	FUNGUS DEFENSE RESPONSE, ncRNA Processing, Defense response, pathogenesis, response to stress	0.96357013	Negative selection
EKA-75	FUNGUS DEFENSE RESPONSE, ncRNA Processing, Defense response, innate immune response, response to stress	1.52645721	Positive selection
EKA-51	Defense response, suppresses host, negative regulator of MAPK pathway, negative regulator of innate immune response, pathogenesis, response to stress, response to symbiont	1.03711519	Positive selection
EKA-53	Defense response, suppresses host, negative regulator of MAPK pathway, negative regulator of innate immune response, pathogenesis, response to stress, entry in host, response to symbiont	0.82724719	Negative selection

Identification and characterization of novel peptides from Ψ-Rs

Ψ-Rs were six-frame translated and their corresponding open reading frames (ORFs) were extracted using Emboss-sixpack and Emboss-getorf, respectively, from The European Molecular Biology Open Software Suite EMBOSS-v 6.6.0 suite. A length filter of a minimum of 100 bp (33 aa) was applied to extract putative peptides. These peptides were further used for the identification of novel pseudo-peptide sequences. BLASTp search was conducted against the local non-redundant (nr-2019) proteins database downloaded from NCBI using default parameters. The unique no-hit sequences were extracted and cross-verified for their novelty by searching against the nr-2021 database. The unique peptides were next examined for their anti-microbial and anti-fungal potentials. The filtered novel peptides with potent antimicrobial potential were thus renamed as Ψ-AMPS's, while those with antifungal potential were renamed as "EKA" peptides following the nomenclature provided in the previous literature [26].

The anti-microbial potential of peptides was predicted using the database of anti-microbial activity and structure of peptides (DBAASPv3.0) [27]. The filtered peptides were further characterized by BLAST search in the Plant Peptide Database (PlantPepDB) [28]. The Ψ-AMPs were characterized based on matched peptide families mainly "defensins", "thionins", "thaumatin", "lipid-transfer", "cyclotides", and "miscellaneous". The novel peptides were also screened for their anti-fungal potential using the Anti-fungal peptides webserver (Antifp server) and renamed as EKA-AFPs [29]. The physicochemical characterizations were done using ProtParam and Antimicrobial Peptides Database-v3 as 'Peptide length' 'APD hydrophobic ratio', 'Total net charge', 'pI', 'grand-average of hydropathicity index (GRAVY)', 'molecular weight', 'Boman index', 'aliphatic index' and 'instability index' for Ψ-AMPs and EKA-AFPs [30,31].

The EKA peptides were further screened based on their GO-MF predictions in the Protein function prediction (PFP) server into ten defense-related classes i.e., "fungal defense", "ncRNA processing", "defense response", "suppression of host immune-response", "stress-related MAPK pathway activation", "innate immune-response", "pathogenesis", "response to stress", "entry into host" and "response to symbiont" [32]. The EKA sequences directly and even distantly involved in "fungal defense" irrespective of their low prediction scores were screened out for further analysis and their substitution rates (dN/dS) were calculated using MEGAX. The $dN/dS < 1$ indicates negative selection, $dN/dS = 1$ indicates neutral selection and $dN/dS > 1$ indicates positive selection in pseudogenes with respect to their parent genes.

Stability prediction and structural analysis of novel EKA-AFPs

Secondary structures of EKA-AFPs were predicted using JPred (v4) webserver [33]. In addition, the free-energy of EKA-AFPs was examined using the UNAFold webserver [34]. Further, tertiary structures of peptides were devised using I-TASSER webserver [35] and the best 3-D structure models were selected based on their Confidence scores (C-score) and Root mean square deviation (RMSD) values.

Target protein prediction of EKA-AFPs

Targets were predicted for seven selective EKA-AFPs using 3-D structures of 66 *E. necator* target protein sequences [36] downloaded from Protein Model Database (PMDB) [37]. Protein-protein docking between EKA-AFPs and targets predicted their most compatible associations based on their maximum dock-score and surface area using the PatchDock algorithm [38]. The refined docked structures for EKA-AFPs and their corresponding targets were regenerated using ClusPro 2.0 [39]. The best-docked models were then selected based on their lowest energy and weighted mean scores. The downloaded models were further visualized using PyMOL 2.5.0 [40]. These interactions were further validated using a Protein binding energy prediction (PRODIGY) webserver [41] by calculating their binding affinities and dissociation constants (K_d).

Extraction of Ψ-Rs as potent mRNA mimics of R-gene mediated resistance

The QTL mapped and experimentally validated R-loci sequences of RUN1, RPV1 and RPW8 from *V. vinifera* were downloaded from the UniProt [42] to setup a local protein database. Blastx with default parameters identified *Ψ-Rs* aligning to *R*-loci and were subsequently differentiated into coding and non-coding transcripts using Coding Potential Calculator (CPC 2.0) [43]. Protein coding sequences (CDS) of *V. vinifera* were downloaded from Ensembl Plants and their local nucleotide BLAST database was setup. BLASTn of all the CDS sequences was done with default parameters in order to filter out *Ψ-Rs* associated with various mature mRNA sequences. The filtered *Ψ-Rs* with coding and non-coding templates were further used for studying passive interactions (with non-coding regulators such as miRNA, and tasiRNA's) as well as active interactions with coding regulators such as TF's.

Potential interactors of Ψ-mRNA mimics of R-gene mediated resistance

Plant small RNA target analysis server (psRNATarget) [44,45] was used for predicting miRNA *& tasiRNA* target sequences from the mRNA associated *Ψ-Rs* (coding and non-coding). *V. vinifera* 186-miRNA sequences from miRBase and 37-tasiRNA sequences from Plant non-coding RNA's Database (PNRD) were used for this analysis. [46]. Similarly, TF-binding sites were assessed in *Ψ-R*

regions by using Plant transcription factor database (PlantTFDB-v5.0) [47].

Differential expression (DE) analysis of Ψ-Rs

To analyze the expression footprints of *Ψ-Rs*, high-throughput RNA sequencing data from NCBI Sequence Read Archive (SRA) database [48–51] for 10 developmental stages of three tissues i.e., leaf (young, medium, large-sized and mature), seedless berries (veraison, intermediate and mature) and inflorescence (3, 5 and 7-days after 100 % cap-fall) were used to study tissue specific DE patterns of *Ψ-Rs* (Table S1). In addition, the RNA-Seq data for PM-infection in grapevine from bio-project no. PRJNA395634 [52] for two treatment conditions i.e., control and 36 h post infection (hpi) was used for studying expression profiles of *Ψ-Rs* in response to infection (Table S1). The expression analysis was done using Trinity-v2.03 package. The FPKM analysis using RSEM method and edgeR (with default parameters) was used to identify DE of all *Ψ-Rs* in developmental stages (p-value:0.001, c-value:2) and PM-response (p-value:0.05, c-value:1) [53]. Genes with sample outliers were excluded at this step to reduce noise, leaving behind most informative genes, followed by data normalization. Heat maps of DE *Ψ-Rs* were plotted using pheatmap library in Rstudio-v4.1.2. The PM-responsive DE patterns were also predicted for 29,927 CDS sequences and 61,078 lncRNA sequences to be subsequently used for co-expression analysis [21,22].

Co-expression network construction and integration of Ψ-Rs, lncRNAs & CDS

Integrative network analysis of Ψ-lncRNAs-CDS network was performed to determine the functionality of putative *Ψ-Rs* at systems-level against PM in *V. vinifera* based on similar expression patterns across coding and non-coding interacting partners. Three separate weighted gene co-expression networks (WGCNs) (*Ψs*-CDS, *Ψs*-lncRNAs, and lncRNAs-CDS) were constructed using DE values for control and treatment conditions. All WGCNs were constructed using Pearson correlation coefficient (PCC) as weight which was calculated using the PERL script [54].

PCC was enumerated using the following equation:

$$PCC = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{(x-\bar{x})^2}\sqrt{(y-\bar{y})^2}}$$

Where x and $\bar{x}$ as well as y and $\bar{y}$ represent expression data and the corresponding mean of interacting genes. Mean of expression profile data is determined by sample size which is different for both the sets of genes.

A biologically significant PCC threshold was obtained independently for all three (*Ψs*-CDS, *Ψs*-lncRNAs, and lncRNAs-CDS) networks, with both positive and negative set of PCCs, by analyzing their topological parameter like network density (ND) at different PCC cutoffs [55].

ND was computed as follows:

$$ND = \frac{2m}{n(n-1)}$$

Where m and $\frac{n(n-1)}{2}$ specify the number of observed edges and possible links of nodes, respectively.

The interaction pairs in each network were shortlisted at selected PCC threshold, and then subjected to implement 't-test' for the selection of statistically significant interactions (with p value ≤ 0.05) to reduce the rate of false positives in each network [56]. Subsequently, false discovery rate (FDR) estimation [57] was calculated, using "stats" library of R-statistical package-v3.4.4, to further reduce the false positives through multiple testing correction. Finally, the resulting interaction pairs were used to construct three independent WGCNs, and Cytoscape-v3.7.0 [58] was used to visualize these networks. The PCC values for all interacting pairs were applied as edge weights to construct individual networks, where weights constitute the strength of correlation (co-expression). Additionally, to elucidate the scale free nature of all WGCNs, random networks with same number of nodes and edges, corresponding to each weighted network, were constructed using 'igraph' library in R package, so as to compare for their network topology to interpret their biological nature. The complexity of system-level analysis for three (*Ψs*-CDS, *Ψs*-lncRNAS, and lncRNAs-CDS) co-expression networks was established by their clustering through Markov cluster (MCL) algorithm [59], of plugin clusterMaker [60] from cytoscape, and was implemented to achieve immensely interconnected but comparatively smaller and functional gene modules at default inflation value of 2.5. Topological analysis of integrated disease network was also obtained using 'NetworkAnalyzer' tool in cytoscape. To perform GO enrichment analysis for significant modules, DE *Ψ-Rs* were first assigned unique identifier by comparing against grapevine proteome, followed by implementation of 'topGO' library in R package. Significantly enriched terms were determined by comparing them against background reference, obtained from the OMICSBOX (*Ψ-Rs*, lncRNAs) and *Vitis* ontology (CDS). Statistically significant terms were obtained using p-values from Fisher's Exact Tests with the classic model. To perform pathway enrichment analysis of significant modules, DE *Ψ-Rs* were first assigned unique TAIR identifier by comparing their protein sequences against *Arabidopsis* proteome using BLASTp program [61], followed by identification of significant pathways using DAVID-v6.7 [62].

Results and discussion

Pseudogenes of NLRs acting in plant immune response pathways

Several genome-wide *Ψ* identification studies have been conducted in plants such as *Arabidopsis thaliana*, *Oryza sativa*, Soyabean, Barley, *Populus trichocarpa* etc. [11,63–65]. In our study, PlantPseudo pipeline identified 31,032 WGD *Ψs* based on duplication blocks for 9,246 parent genes from the genome of *V. vinifera* (Table S2). Chromosomal (Chr) distribution of the predicted *Ψs* demonstrated Chr18 possessing maximum and Chr11 having least number (Fig. 1a). The pipeline also predicted 7,411 lncRNA's associated with WGD-derived *Ψs* (Table S3) with maximum number of lncRNA's occurring on Chr13 and minimum on Chr10 (Fig. 1a). Further, the *Ψ* associations with lncRNAs were characterized into four types as Body associated, Co-promoter associated, Promoter associated and Tail to Tail associated with majority *Ψs* existing as co-promoter associated with lncRNAs (Fig. 1b).

V. vinifera is susceptible to a plethora of pathogenic microorganisms. *E. necator* causing PM infection is the most notorious pathogen and was used in our analysis. *R*-genes, activators of an innate immune response in plants are regarded as core components for delivering plant immune responses via effector-triggered immunity (ETI). Also, *NLRs* (major class of *R*-genes) are encoded in the genome as clustered tandem repeats and thus can act as paralogous R-gene reserves which protest against foreign attacks by the pathogenic microorganisms. Since Ψs have been reportedly known to induce a functional immune response in human, chicken, rabbit, and other vertebrates [66]. Role of *Ψs* in plant immune response especially for ETI have been scarcely analyzed in rice and wheat plants [67,68]. Thus, in our study 2,746 *Ψs* (after removing uncharacterized Chr sequences) specifically involved in *R*-gene mediated immune response were filtered out using homology modelling

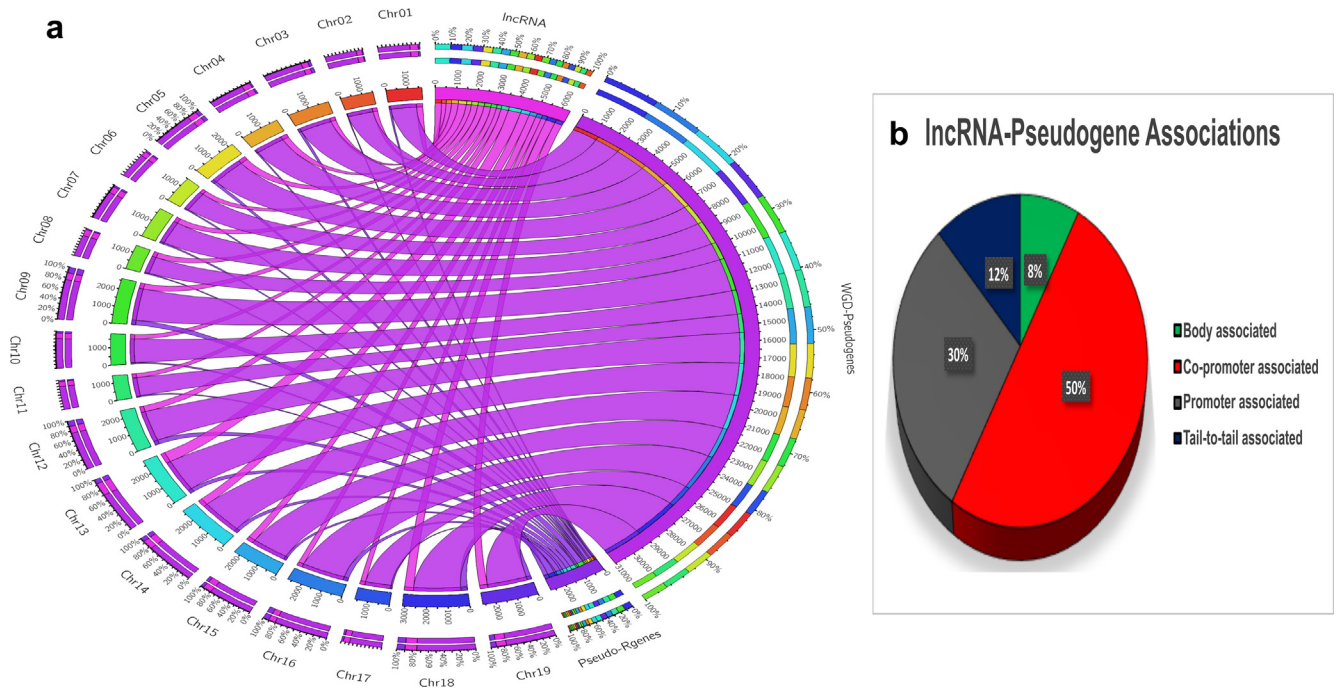


Fig. 1. Distribution and characterization statistics of noncoding genes. (a) Circos plot depicting chromosomal distribution of WGD-derived pseudogenes, lncRNAs and pseudo-R genes in the genome of *V. vinifera* **(b)** Categorical distribution of putative pseudogenes based on distance of their proximal upstream regions containing lncRNAs.

from WGD-derived *Ψs*. Their chromosomal distribution depicted chr09, chr12, and chr18 possessing majority of 289, 254 and 328 *Ψ-Rs* respectively (Fig. 1a).

Subsequently, these Ψ-Rs were functionally characterized *in-silico* to identify their putative roles in plants. Their functional characterizations unveiled their annotations based on GO terms for BP, CC, and MF. GO-based MF predicted several *Ψ-R* sequences involved in ADP binding, ATP binding, protein serine/threonine kinase activity, protein kinase activity etc which have been demonstrated to be crucial for NLRs in disease resistance pathways. CC-based GO distribution affiliated the majority of *Ψ-Rs* being membrane-associated, followed by cytoplasm and nucleus, which was in conjunction with activity of NLRs acting as primary receptors for pathogen recognition. *Ψ-Rs* involved in BP were primarily annotated to be involved in defence-response, protein-phosphorylation, signal-transduction, phosphorylation, DNA-integration etc. Predicted functions of *Ψ-Rs* were from defence-responses to bacterium (23*Ψ-Rs*), oomycetes (24*Ψ-Rs*), fungus (20*Ψ-Rs*) also reaffirms their significance in NLR-mediated immune response [69]. Also, *Ψ-Rs* were involved in several developmental processes such as lateral-root formation, development of root-cap, stomatal-complex, pollen, flower, and phloem etc. in addition to hormone signalling pathways of auxin, gibberellic acid, abscisic acid, brassinosteriod etc. Further, pathway enrichment of *Ψs* using KEGG inferred their roles in 112 pathways including important pathways of MAPK-signalling pathway in plants (ko:04016), Plant-Pathogen interaction pathway (ko:04626), Plant hormone signalling pathways (ko:04075) etc which are integral pathways of plant defense response.

Ψ-derived peptides as novel antifungal agents against PM

Duplications in the whole genome are considered key players for appearance of new genes via neofunctionalization event. Duplicated *Ψs* arise from block duplication events and undergo a multi-step process for their formation. The existence and characterization of novel *Ψs* arising because of segmental duplications in the human genome established strong correlations between segmental duplications and existence of *Ψs* [70]. Further, the clustered tandem duplication pattern of evolution existing in NLRs catered to the idea behind identification novel *Ψ*-R peptides in our study. Six-frame translation and length cut-off (33 aa) filter on 2,746 *Ψ-Rs* resulted in 1,923 peptide sequences. BLAST alignment of these peptides against nr proteins reported 665 novel peptide sequences.

The novel peptides were screened for their anti-fungal potential of Ψ-peptides using "Antifp" database reported 81 peptides possessing antifungal potential (EKA-AFPs). They were characterized further based on their physio-chemical properties, stability and ligand binding efficiencies (Table S4) which pointed to most being good antifungal candidates (details in SI). These EKA-AFP sequences were then functionally annotated for MF-based GO prediction using PFP server. The PFP server predicts functional matrixes based on information of weakest hits in PSI-BLAST and determines common functional domains even from distant orthologs. The GO MF-based screening was done for different host-pathogen interaction categories (Fig. 2b) to filter out total 48 out of 81 protein sequences to be involved in plant immune responses and were further analyzed for their substitution rates (Table S5). Seven peptides annotated to be involved in fungal defense responses were filtered out i.e., EKA-15, EKA-22, EKA-49, EKA-74 and EKA-75 as fungal defense peptides and EKA-51 and EKA-53 as promoters of fungal infection in plants based on PFP MF-GO terms (Table 1). Further their substitution rate analysis determined EKA-15, EKA-74 and EKA-53 being under negative selection while EKA-22, EKA-49, EKA-75 and EKA-51 being under positive selection (Table 1), which is significant for neofunctionalization of genes. Sequence based physio-chemical characterization reported EKA-AFPs as cationic and hydrophilic (except EKA-22) in nature. EKA-51 and EKA-53 were predicted to be multi-functional based on their boman indexes (Table 2). Finally, the secondary structure analysis of the EKA-AFPs predicted EKA-51, EKA-53 and EKA-75 to be forming most stable secondary structures (Table 2). The tertiary structures of seven out of 81 EKA-AFPs were primarily used for docking with *E. necator* target proteins.

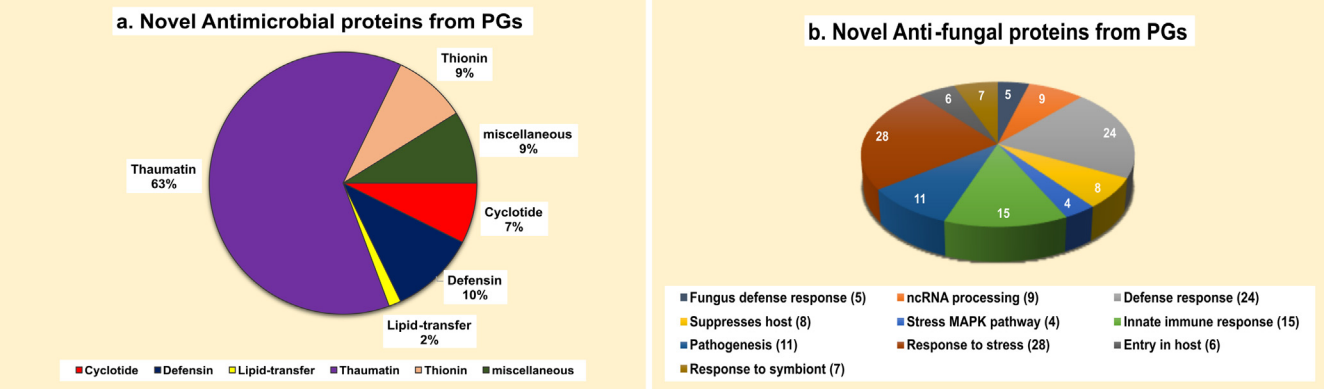


Fig. 2. Characterization of novel peptides from pseudo-R genes. (a) Distribution of novel peptides (Ψ-AMPs) obtained from pseudo-R genes into Plant antimicrobial families. **(b)** Functional characterization of novel antifungal EKA-peptides based on plant-host interaction responses.

Table 2
Properties of EKA-peptides of fungal defense response based on their sequence stability.

EKA peptides	mFOLD Free-energy change (ΔG) (KJ/mol)	I-TASSER 3-D model scores	Net charge	APD-Hydrophobicity ratio	Boman index (kcal/mol)	Aliphatic index	GRAVY index
EKA-15	−31	c-score: −2.19 rmsd: 6.2 ± 3.8 Å	0	33 %	1.97	76.67	−0.5788
EKA-22	-23.4	c-score: −1.49 rmsd: 4.8 ± 3.1 Å	2	47 %	0.71	114.71	0.241
EKA-49	−14.5	c-score: −2.5 rmsd: 7.1 ± 4.1 Å	6.5	35 %	1.54	100	−0.403
EKA-51	−56.9	c-score: −1.79 rmsd: 6.3 ± 3.9 Å	9	38 %	2.86	73.58	−0.651
EKA-53	−48.3	c-score: −1.51 rmsd: 5.1 + -3.3	7	41 %	3.05	67.44	−0.526
EKA-74	−31.5	c-score: −1.51 rmsd: 4.8 ± 3.1 Å	4	38 %	0.9	117.35	−0.197
EKA-75	−45.2	c-score: −3.04 rmsd: 9.2 ± 4.6 Å	7	22 %	3.01	52.41	−1.341

Table 3
Protein-protein docking parameters of fungal defense responsive EKA-peptides and their corresponding targets in *E. necator*.

EKA-peptides	Target protein	Target Protein PMDB ID	Patch Score	Weighted Score of lowest energy ClusPro2.0 model	Binding affinity of the complex (kcal/mol)	Dissociation constant of the complex at 25 ℃
EKA-15	Adenylosuccinatesynthetase	PM0082752	16,034	−960.1	−15.8	2.40E-12
EKA-22	Putative myosin class v myosin	PM0082816	15,736	−1032.7	−13.7	9.20E-11
EKA-49	Adenylosuccinatesynthetase	PM0082752	19,400	−1354.9	−17.1	3.00E-13
EKA-51	mRNA-capping enzyme subunit alpha	PM0082805	16,818	−1315.7	−16.3	1.00E-12
EKA-53	Adenylosuccinatesynthetase	PM0082752	17,422	−937.6	−12	1.50E-09
EKA-74	Serine/threonine-protein kinase Tel1	PM0082778	16,118	−771.7	−8.8	3.40E-07
EKA-75	Non-specific serine/threonine 23 protein kinase	PM0082756	18,840	−1090.6	−12.9	3.20E-10

Docking scores were obtained from individually docking all the seven EKA-AFPs with 66 protein sequences of *E. necator*. *E. necator* was used as a template pathogen for studying role of AFPs as potential antifungal agents due to easy availability of target genes from the pathogen [36] using molecular docking. The identified target sequences of EKA-AFPs in *E. necator* are shown in Table 2. Out of five target proteins of *E. necator* 'adenylosuccinate synthetase' (ubiquitous enzyme playing significant role in purine biosynthesis) is a target for three peptides i.e., EKA-15, EKA-49 (positive defense-regulator) and EKA-53 (negative defense-regulator). The best docked models were generated in ClusPro 2.0 (Table 3) and visualized in PyMOL-2.5 (Fig. 3). The binding affinity (ΔG) and dissociation constant (K_d) values of the predicted protein–protein complexes using the PRODIGY webserver (Table 3) identified EKA-49 and EKA-51 as the best docked models overall while EKA-53 and EKA-74 were the weakest models in terms of protein–protein interaction between EKA-AFPs and their corresponding targets.

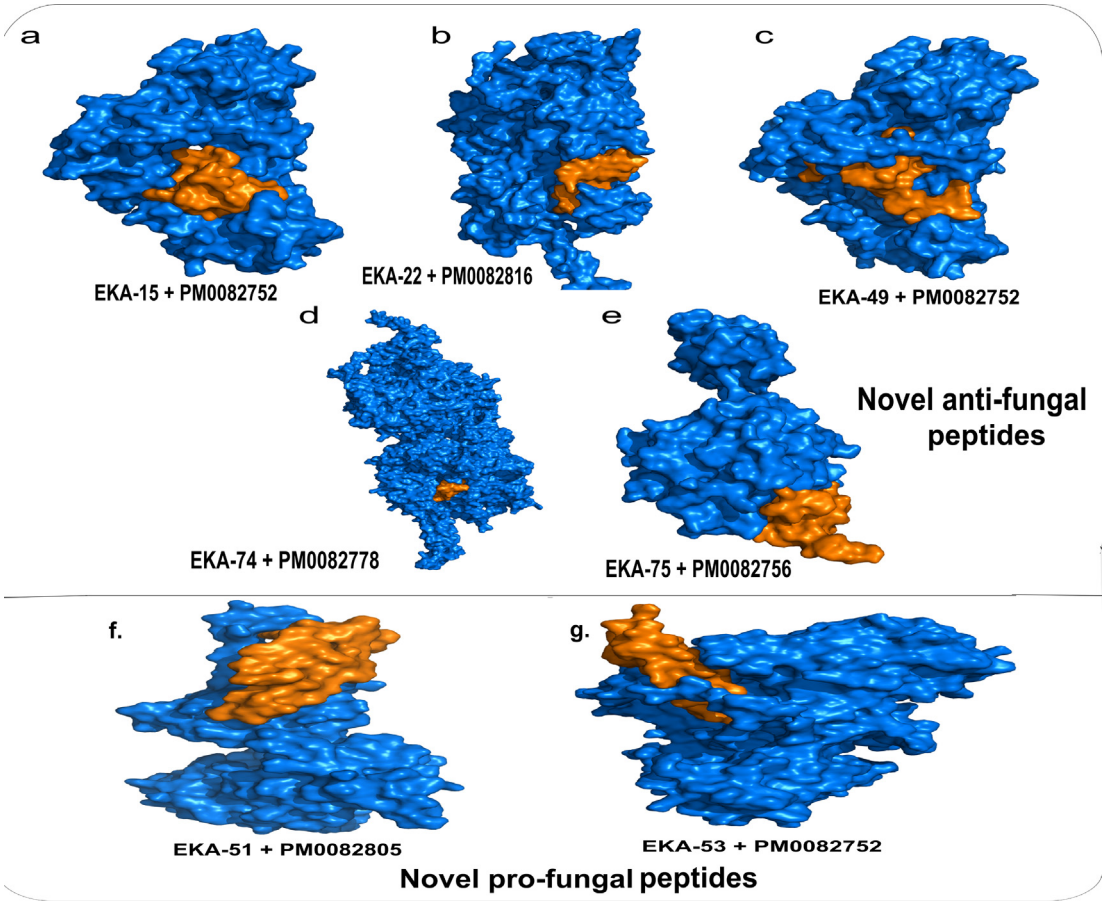


Fig. 3. Protein-protein docking of EKA-peptides and their corresponding *Erysiphe necator* target proteins: Protein-protein docking models obtained from ClusPro 2.0 of anti-fungal peptides i.e., positive regulators of defense response **(a)** EKA-15 with Adenylosuccinatesynthetase of *E. necator* (PM0082752). **(b)** EKA-22 with Putative myosin class v myosin of *E. necator* (PM0082816). **(c)** EKA-49 with Adenylosuccinatesynthetase of *E. necator* (PM0082752) **(d)** EKA-74 with Serine/threonine-protein kinase Tel1 of *E. necator* (PM0082778) **(e)** EKA-75 with Non-specific serine/threonine 23 protein kinase of *E. necator* (PM0082756); and pro-fungal peptides i.e., negative regulators of defense response **(f)** EKA-51 with mRNA-capping enzyme subunit alpha of *E. necator* (PM0082805) **(g)** EKA-53 with Adenylosuccinatesynthetase of *E. necator* (PM0082752).

Additionally, anti-microbial potential of 665 novel peptides identified 69 anti-microbial pseudogenic-peptides (Ψ-AMPs). Antimicrobial peptides have been widely characterized in amphibians, insects, non-vertebrates as well as humans [71]. These novel Ψ-AMPs belonged to five antimicrobial families of (Fig. 2a) Cyclotides (5-Ψ-AMPs), Defensins (7-Ψ-AMPs), Lipid-transfer peptides (1-Ψ-AMPs), Thaumatins (42-Ψ-AMPs), and Thionins (6-Ψ-AMPs) which play integral roles in plant defenses. Their physiochemical characterizations (Table S6) advocated to Ψ-AMPs as ideal anti-microbial candidates with properties like majority of natural antimicrobial peptides (details in SI). Most naturally occurring antimicrobials have peptide length of 10–50 aa, molecular weight 2–9 kDa, are positively charged, contain high hydrophobic amino acids, and display helical structures [72]. AMPs work by pore formation on the membranes of the pathogens causing metabolite leakage and ultimately cell death. This mode of action is facilitated by their amphipathic nature and positive charge which allows for peptides to interact with membrane lipids leading to their accumulation at the membrane surface [15]. All the 69 novel AMPs predicted in our study were cationic in nature and were medium sized peptides which makes them promising anti-microbial candidates (Table S6). Novel anti-microbial peptides derived from Ψs have been functionally annotated in *Drosophila* [16] till now. Eight out of 69 antimicrobial peptides had antifungal potential in them and were analyzed for their respective targets and binding efficiencies (Table S6). Thus, supporting our hypothesis that Ψs might be playing roles in generation of new antifungal genes and act as major driving force for modulation of coevolution between host-pathogen interactions [73].

Ψ-Rs as modulators of gene expression machinery

Resistance Locus specific Ψ-*R* mapping identified 102 Ψ-*Rs* from coding and 425 Ψ-*Rs* from non-coding transcripts aligning to the resistance hotspots. Majority of Ψ-*Rs* were observed on Chr09 and Chr18 for coding and non-coding transcripts, respectively. Further, Ψ-*Rs* acting as potential mRNA mimics were screened out to identify 90 coding and 305 non-coding Ψ-*Rs* which might be regulating their parent genes in the *Vitis* genome. These numbers strongly advocate for potential of interaction mimicry in Ψ-*Rs* for CDS owing to their sequence homology.

Since endogenous RNAs (lncRNAs, miRNA and tasiRNAs) undergo activation or silencing in response to pathogenic invasion and regulate expression of disease resistance genes through transcription or post-transcriptional gene silencing. Thus, Ψ-*Rs* identified as mature mRNA mimics were further screened for prediction of miRNA and tasiRNA target sequences on them. This was done to understand the putative pre-existing target sites in Ψs which might be playing key roles in regulation by sequestration of miRNA and tasiRNAs thereby modulating post-translational gene silencing through Ψs. A considerable number of targets for both miRNA (138 Ψs out of 305) and tasiRNA (63 Ψs out of 305) were predicted in the non-coding Ψ-*Rs* and more targets for miRNA's (70 out of 90 Ψs) and tasiRNA's (36 out of 90 Ψs) were present in coding tran-

scripts (Fig. 4a). Out of 90 coding *Ψ-Rs*, 32 sequences were common targets for both miRNA and tasiRNA's. Similarly, out of 305 non-coding *Ψ-Rs*, 41 sequences were common targets for miRNA and tasiRNA's. Thus, indicating *Ψ-Rs* might be modulating their parent genes via RNA mediated gene silencing and vice versa. In addition to target prediction, *Ψ-Rs* of CDS were also aligned with the predicted lncRNA's through BLASTn and consequently 6 *Ψ-Rs* for coding and 22 for non-coding transcripts were aligned (Fig. 4e). In addition, 3 *Ψs* from coding and 6 from non-coding transcripts were common among lncRNAs and *Ψ-Rs* (mRNA-mimics) of miRNA and tasiRNA targets (Fig. 4e), strongly pointing towards important regulatory roles of *Ψ-Rs* towards CDS in co-ordination with other non-coding RNA's. Since lncRNAs act by sequestering miRNA/tasiRNAs and *Ψs* can lead to lncRNA generation, this could point to the *Ψ-Rs* being the master regulators of gene regulation either directly (by acting as mRNA mimics for miRNA/tasiRNA targeting) or indirectly (by lncRNA mediated regulation of miRNA/tasiRNAs). Interestingly, a lot of the miRNA and tasiRNA target sites were also found on pseudogenes forming novel AFPs (Fig. 4c) and Ψ-AMPs (Fig. 4d), thus pointing to the importance of *Ψ-Rs* in regulation of novel genes through post-transcriptional gene silencing events.

According to the previous literature, the regulation by *Ψs* include siRNA, miRNA and lncRNA formation and associations with TFs and RNA polymerase II [19]. Our results are in conjunction with the hypothesis, as our analysis also found 37,026 TF binding sites in pseudogenic regions for 283 TFs which were common with *R*-loci of *V. vinifera*. The TF-binding sites identified were 15,011 binding sites (for coding transcripts) and 22,015 binding sites (for non-coding transcripts) for total 283 TFs. 9 major TF-families were identified involved in stress response (Fig. 4b) and MYB (89), NAC (87) and bZIP (85) TF families were present in majority in coding and MYB (251), NAC (197) and bHLH (193) in the non-coding transcripts. These TF-binding sites in *Ψ* regions might function as promoters or enhancers and regulate their parental gene expression levels by serving as gatekeepers of expression in general [19].

Synergistic effects of Ψ-Rs, lncRNA's and genes in modulation of immune response

Ψ-Rs were analyzed for their DE patterns in 10 different developmental stages i.e., young, medium, large, mature leaf tissues, 3-day, 5-day, 7-day inflorescence tissues and veraison, intermediate, mature berry tissues. A total 194 sequences were DE *Ψ-Rs* with a fourfold-change (c-value: 2) and p-value of 0.001 in different tissue stages in comparison to the other two (Fig. 5c) i.e., inflorescence (180 *Ψ-Rs*), berry (111 *Ψ-Rs*) and leaf tissue (185 *Ψ-Rs*). Further, the comparative DE between two tissue stages were also determined (Fig. 5d) i.e., inflorescence vs berry (110 Ψ-Rs), berry vs leaf (108 *Ψ-Rs*) and leaf vs inflorescence (172 *Ψ-Rs*). Heat Map depicting the tissue-specific DE-patterns is shown in Fig. 5a. Most DE *Ψ-Rs* were in leaves followed by inflorescence and least *Ψs* were seen in berry tissues. Previously, in plants tissue specific expression patterns of *Ψs* were predicted in rye and soyabean [9].

To the best of our knowledge, this is first ever study to inspect role of *Ψs* in disease resistance of *V. vinifera* against pathogen *E. necator* infection. In this study, *Ψ-Rs* were also screened for PM infection based DE patterns. *V. vinifera* cv. Thompson Seedless leaves infected with PM disease after 36 h post infection RNA-seq data which was used for the screening process. As a result, 79 DE

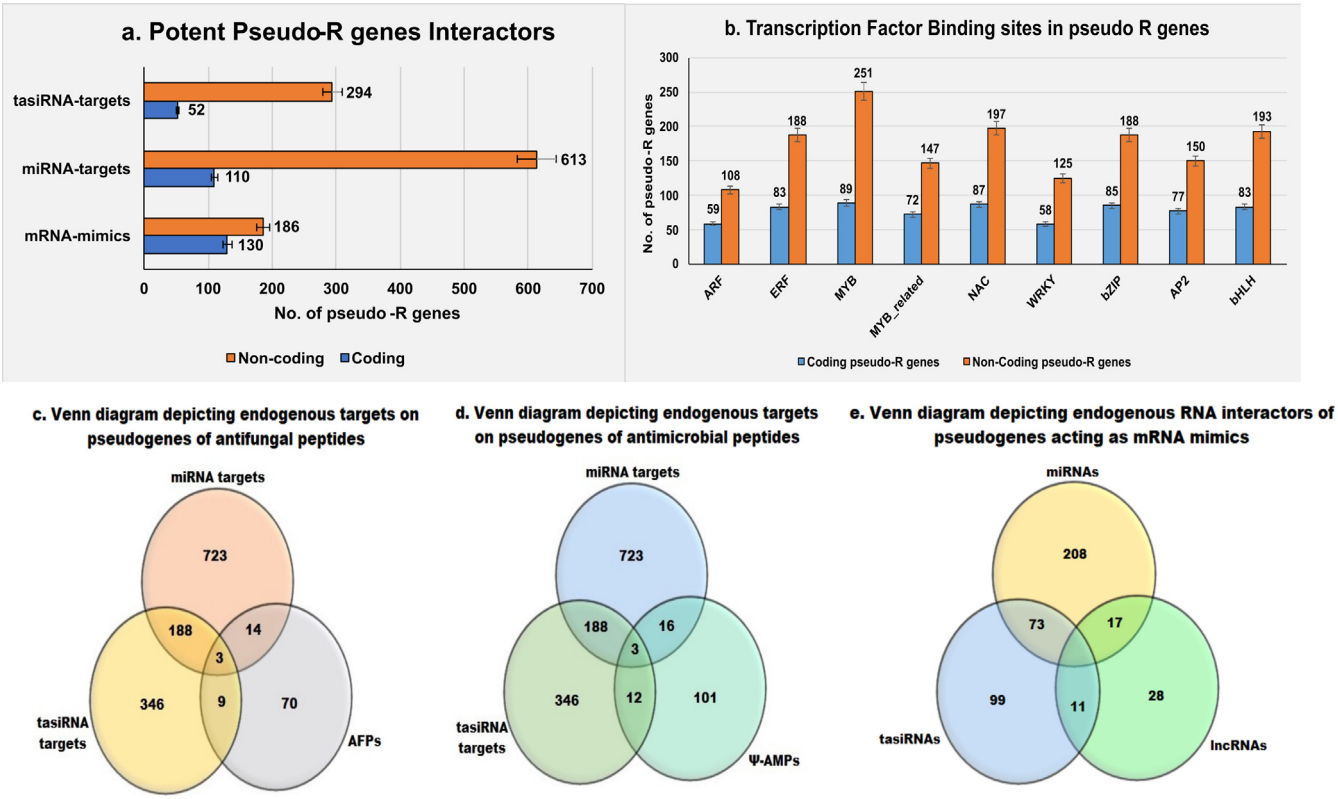


Fig. 4. Potential interactions of pseudo-R genes with mRNAs (resistance locus specific), miRNAs, tasiRNAs, lncRNAs and transcription factors: (a) Pseudo-R genes potentially involved in mRNA mimicry and in-turn acting as miRNA and tasiRNA targets from both coding and non-coding transcripts **(b)** Abundance estimation of stress-responsive transcription factor families in pseudogenic counterparts of resistance genes **(c)** Venn diagram depicting endogenous targets on pseudogenes of antifungal peptides **(d)** Venn diagram depicting endogenous targets on pseudogenes of antimicrobial peptides **(e)** Venn diagram depicting endogenous RNA interactors of pseudogenes acting as mRNA mimics.

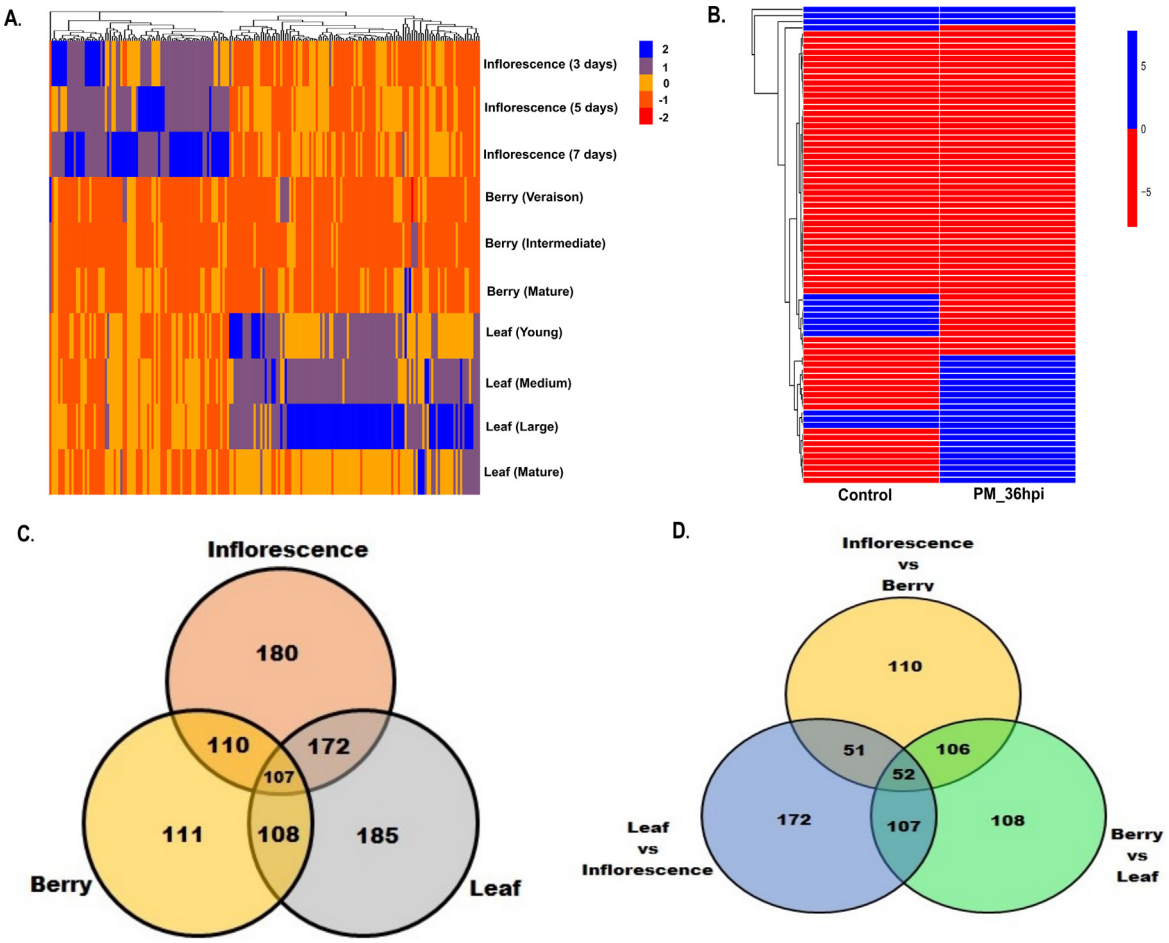


Fig. 5. Differential expression pattern analysis of pseudo-R genes: (a) Tissue-specific differential expression analysis of pseudo-R genes (194 pseudo-R genes at p-value: 0.001; 4-fold change) in ten developmental stages from three tissues i.e., 180 pseudo-R genes for Inflorescence (3-day, 5-day, and 7-day), 111 pseudo-R genes for Seedless Berry (Veraison, Intermediate and Mature) and 185 pseudo-R genes for Leaf (young, medium, large, and mature). Blue-purple colors demonstrate up-regulation while orange-red colors depict down-regulated pseudo-R genes **(b)** Differential expression analysis of pseudo-R genes in response to powdery mildew (PM) biotic stress response using samples of control and 36 hpi leaves of *Vitis vinifera* (79 pseudo-R genes at p-value-0.05; 2-fold change). Blue colour corresponds to upregulated pseudo-R genes in response to PM stress while red colored pseudo-R genes were experiencing downregulation. **(c)** Venn diagram depicting DE in individual tissues of pseudo-R genes in comparison with the other two. **(d)** Venn diagram depicting comparative DE in two different types of tissues for pseudo-R genes.

Ψ-Rs were found with a twofold change with a p-value of 0.05 as depicted in heatmap (Fig. 5b). Consequently, 50 *Ψ-Rs* were seen to be up-regulated in 36 h post-infected leaves of *V. vinifera* against PM infection, whereas 13 *Ψ-Rs* were down-regulated in control leaves of *V. vinifera* in contrast to infected leaves. Similarly, expression of lncRNAs [21,22] and CDS of *V. vinifera* led to 306 lncRNAs and 4,202 CDS differentially expressing in response to PM stress and were used for analyzing their co-expression analysis.

Subsequently, the co-expression analysis between DE *Ψ-Rs*, lncRNAs, and CDS was conducted for PM stress response in *V. vinifera* [74]. Previously, non-coding RNAs and mRNA co-expression networks with lncRNA-mRNA-miRNA based ce-analysis had been conducted in humans for pre-eclampsia indicating gained momentum in Ψ-based network interaction studies [75]. In addition, *Ψ*-based network associations has been widely studied in human cancers to affirm their prominent roles in immune system either as prognostic markers or as general mediators of cell survival [76]. In our analysis, possible interactions between *Ψ*-lncRNA-CDS for modulation of plant immune response have been analyzed in-lieu of their importance in cancer biology in humans.

Consequently, three independent WGCNs (*Ψs*-CDS, *Ψs*-lncRNAs, and lncRNAs-CDS) were generated and integrated to interpret system-level roles of putative *Ψ-Rs* in disease resistance pathways. The congruity of *Ψ-Rs* with CDS and lncRNAs and vice-versa, from DE data was used to calculate PCC. Weighted networks persistently utilize calculated values of PCC as attributes that computes robustness of interactions among nodes to ascertain their biologically pertinent modules. False positives reduce and connections of interacting partners strengthen by defining a specific PCC threshold. Further, ND identifies biologically significant topological modules by determining statistically relevant PCC threshold. Therefore, variations in ND as a function of various PCC cutoffs was investigated, and PCC with minimal ND was considered as a threshold (shortlisted based on statistical significance). For all these weighted co-expressions, same criterion of threshold selection was implemented, and a PCC threshold of 0.95 and −0.90 was selected for positive and negative set of correlation values, respectively, in *Ψs*-CDS network (Table S7). Similarly, PCC threshold values for −0.80 (*Ψs*-lncRNAs) (Table S8) and −0.90 (lncRNAs-CDS) (Table S9) were selected as negative set of correlation values while PCC threshold of 0.95 was selected for positive set correlation values of both networks. A total of 46,214, 4,692, and 2.61,706 interactions were obtained for *Ψs*-CDS, *Ψs*-lncRNAs, and lncRNAs-CDS networks, respectively, at their respective thresholds for both positive and negative correlation values. All

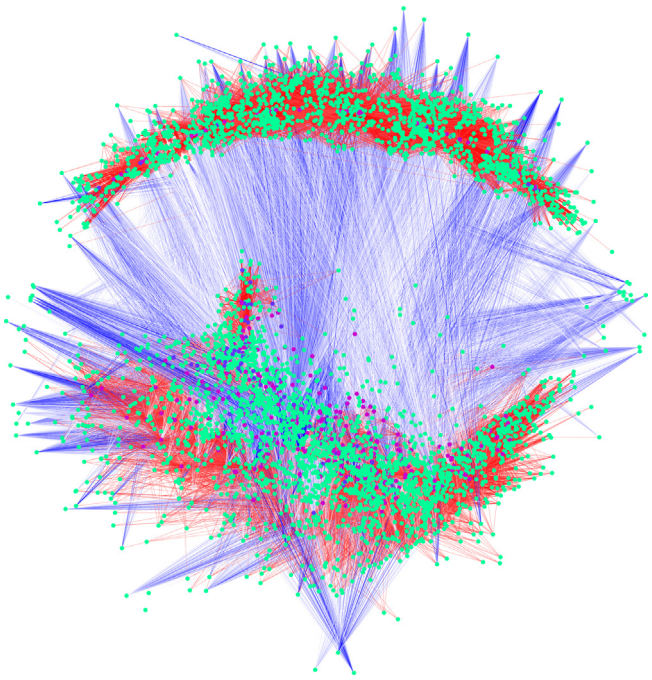


Fig. 6. Complete integrated PGs-lncRNAs-CDS co-expression network. This integrated co-expression network was comprised of pseudogenes (PGs), long noncoding RNAs (lncRNAs), and coding sequences (CDS), where PGs, lncRNAs, and CDS were depicted as circular nodes with blue, magenta, and green color, respectively. In this network, strong positive and negative set of interactions were also represented as red and blue edges, respectively. Cytoscape is used for the visualization of the network.

these interactions were found to be statistically significant at this threshold with p-value of ≤ 0.05.

These three independent networks were integrated to generate *Ψs*-lncRNAs-CDS network with 4,587 nodes (*Ψs*, 78; lncRNAs, 306; CDS, 4,203) and 3,12,612 edges, where nodes and edges exemplified as *Ψs*/lncRNAs/CDS and connections between them, respectively. The strength of interactions among nodes was further quantified using PCC values as edge attributes and differentiated with red and blue color based on high positive and negative PCC, respectively (Fig. 6), during network clustering. Scale-free networks are usually verified by comparing the biological networks against random network based on specific topological properties [77]. Hence, a weighted-random network was constructed with equivalent number of nodes (4,587) by generating random connections among edges (3,12,612) for 10,00,000 iterations. As compared to bell-shaped organization of random network, the degree distribution (DD) of integrated co-expression network was highly skewed that also fitted the power law which illustrates the presence of large number of nodes with very few connections oppressed by some immensely connected ones satisfies scale-free behavior of integrated [77]. Negative value of assortativity (-0.679) also augmented the scale-free nature of integrated co-expression network when compared to random network (0.00154). Since integrated co-expression network have to follow scale-free behavior, such networks are most likely to hit a node with only few neighbors and robust against random perturbations to disrupt only a small portion [78].

Module-based analysis was performed to evaluate their integrated response in disease-resistance pathways. Additionally, recognition of gene clusters in a network determines coordination between interacting partners of each module comprising of a significant set of interacting *Ψ-Rs* [79]. Network partitioning was implemented to determine modules through the active implementation of MCL algorithm which divided the large network into 5 sub-modules [80]. These 5 modules (1–5) were large in size with 1944 (*Ψs*, 63; lncRNAs, 207; CDS, 1674), 978 (*Ψs*, 6; lncRNAs, 22; CDS, 950), 669 (*Ψs*, 8; lncRNAs, 12; CDS, 649), 185 (*Ψs*, 0; lncRNAs, 54; CDS, 131), and 151 (*Ψs*, 1; lncRNAs, 9; CDS,141) nodes, and were typically referred by their number hereafter (Table 4). Since the nodes from these five modules counts for the ~ 85 % of nodes, rest ~ 14 % lacking interactions and were considered independent nodes. GO-enrichment of module 1 identified their role in defense responses (details in SI).

Further, significant key *Ψs* were present in module 1 (M1) had maximum number of *Ψs* (63; ~80.77 %) as compared to other modules 2–5 which is 6, 8, 0, and 1, respectively (Table 4). The topological analysis of integrated disease network (*Ψs*-lncRNAs-CDS) identified most central genes acting as possible candidate genes to regulate host-pathogen interactions in PM (Table S10). A total of six *Ψs* shared top10 positions when sorted by the centrality parameters (Table 5). Interestingly, all six *Ψs* were having comparatively high values for all these parameters and therefore considered as key genes to make physical interactions with pathogen targets and modulate plant immunity. These *Ψs* were considered as significant candidates for future validations to extend their role in disease-associated mechanisms in grapevine against PM.

Limitations and future prospects of the study

Our analysis although thorough is preliminary and based on computational validations only. In future, extensive wet-lab experimentations can help establish potential functional roles for these *Ψs*.

Conclusions

In our study, we identified *Ψ-Rs* of *R*-gene mediated resistance from WGD-derived *Ψs* occupy diverse roles in plant defense modulations. In addition, translation of *Ψ-R* sequences identified novel promising anti-microbial candidates with the supporting hypothesis that *Ψs* might be acting as genomic footprints for the generation of new antifungal genes. Thus, they might be driving the coevolution of host-pathogen interactions. Also, the network of *Ψs* with different players of gene expression machinery in conjunction with endogenous RNA regulators modulate the expression of *Ψ*-parent genes thereby directing the plant towards resistance or susceptibility.

Table 4
Distribution of Ψs, lncRNAs, and CDS in the most significant modules obtained from clustering of integrative co-expression (Ψs-lncRNAs-CDS) network.

S. No.	Type of Gene	Module 1 (M1)	Module 2 (M2)	Module 3 (M3)	Module 4 (M4)	Module 5 (M5)
1	Pseudogenes (Ψs)	63	6	8	0	1
2	Long noncoding (lncRNAs)	207	22	12	54	9
3	Coding sequences (CDS)	1674	950	649	131	141
4	Total	1944	978	669	185	151

Table 5
List of common shortlisted Ψs based on three network parameters such as betweenness (BC), degree (DC), and stress (SC) centralities for module 1 from integrated co-expression network with their GO terms.

S. No.	Pseudogenes (Ψ-Rs)	DC	BC	SC	GO term
1	Chr04_240041-240296_VIT_04s0008g00330	806	0.03396	8.7E + 07	GO:0006464, GO:0016301, GO:0043167, GO:0005575
2	Chr04_793408-793803_VIT_04s0008g00370	798	0.03911	1.4E + 08	GO:0006464, GO:0016301, GO:0043167, GO:0005575
3	Chr01_3558981-3560585_VIT_01s0011g03960	784	0.03564	1E + 08	GO:0006464, GO:0016301, GO:0016791 GO:0043167, GO:0005575,
4	Chr04_321505-321961_VIT_04s0008g00370	713	0.0247	1E + 08	GO:0006464, GO:0016301, GO:0043167, GO:0005575
5	Chr04_240649-240892_VIT_04s0008g00440	710	0.03403	2E + 08	GO:0006464, GO:0016301, GO:0043167, GO:0005575
6	Chr16_11036454-11038464_VIT_16s0022g00080	605	0.02274	1.4E + 08	GO:0016301

CRediT authorship contribution statement

Naina Garewal: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Shivalika Pathania:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Garima Bhatia:** Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Kashmir Singh:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Authors would like to thank SERB, India for funding this project (Grant number CRG/2020/000762). Shivalika Pathania thanks UGC, India for providing DS Kothari Post-Doctoral Fellowship.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2022.07.014.

References

[1] Mascagni F, Usai G, Cavallini A, Porceddu A. Structural characterization and duplication modes of pseudogenes in plants. Sci Reports 2021;2021(111):11. doi: https://doi.org/10.1038/s41598-021-84778-6.
[2] Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional regulation by pseudogenes. Biochimie 2011;93:1916–21. doi: https://doi.org/10.1016/J.BIOCHI.2011.07.024.
[3] Grandér D, Johnsson P. Pseudogene-expressed RNAs: emerging roles in gene regulation and disease. Curr Top Microbiol Immunol 2015;394:111–26. doi: https://doi.org/10.1007/82_2015_442.
[4] Jayarathna DK, Rentería ME, Sauret E, Batra J, Gandhi NS. Identifying complex lncrna/pseudogene–mirna–mrna crosstalk in hormone-dependent cancers. Biology (Basel) 2021;10:1014. doi: https://doi.org/10.3390/biology10101014.
[5] Lai Y, Li J, Zhong L, He X, Si X, Sun Y, et al. The pseudogene PTENP1 regulates smooth muscle cells as a competing endogenous RNA. Clin Sci 2019;133:1439–55. doi: https://doi.org/10.1042/CS20190156.
[6] Oliveira-Mateos C, Sánchez-Castillo A, Soler M, Obiols-Guardia A, Piñeyro D, Boque-Sastre R, et al. The transcribed pseudogene RPSAP52 enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent let-7 inhibition. Nat Commun 2019;10:1–18. doi: https://doi.org/10.1038/s41467-019-11910-6.
[7] Lou W, Ding B, Fu P. Pseudogene-Derived lncRNAs and their miRNA sponging mechanism in human cancer. Front Cell Dev Biol 2020;8:85. doi: https://doi.org/10.3389/FCELL.2020.00085/BIBTEX.
[8] Coleman C, Copetti D, Cipriani G, Hoffmann S, Kozma P, Kovács L, et al. The powdery mildew resistance gene REN1 co-segregates with an NBS-LRR gene cluster in two Central Asian grapevines. BMC Genet 2009;10(1). doi: https://doi.org/10.1186/1471-2156-10-89.
[9] Garewal N, Goyal N, Pathania S, Kaur J, Singh K. Gauging the trends of pseudogenes in plants. Crit Rev Biotechnol 2021;41:1–16. doi: https://doi.org/10.1080/07388551.2021.1901648.
[10] Yamada K, Lim J, Dale JH, Chen H, Shinn P, Palm CJ, et al. Empirical analysis of transcriptional activity in the arabidopsis genome. Science (80-) 2003;302:842–6. doi: https://doi.org/10.1126/SCIENCE.1088305/SUPPL_FILE/YAMADA.PDF.
[11] Thibaud-Nissen F, Ouyang S, Buell CR. Identification and characterization of pseudogenes in the rice gene complement. BMC Genomics 2009;10:1–13. doi: https://doi.org/10.1186/1471-2164-10-317.
[12] Nwadiugwu MC. Expression, interaction, and role of pseudogene Adh6-ps1 in cancer phenotypes 11779322211040592. Bioinform Biol Insights 2021;15. doi: https://doi.org/10.1177/11779322211040591.
[13] Gadalla NB, Malmberg M, Adam I, Oguike MC, Beshir K, Elzaki S-E, et al. Alternatively spliced transcripts and novel pseudogenes of the Plasmodium falciparum resistance-associated locus pfcrt detected in East African malaria patients. J Antimicrob Chemother 2015;70(1):116–23.
[14] Chen Xu, Wan L, Wang W, Xi W-J, Yang A-G, Wang T. Re-recognition of pseudogenes: from molecular to clinical applications. Theranostics 2020;10(4):1479–99.
[15] Nawrot R, Barylski J, Nowicki G, Broniarczyk J, Buchwald W, Goździcka-Józefiak A. Plant antimicrobial peptides. Folia Microbiol (Praha) 2014;59:181–96. doi: https://doi.org/10.1007/S12223-013-0280-4/TABLES/7.
[16] Varughese D, Nair VVI, Thomas V, Raj N, Ramakrishnan SP, Khan J, et al. Function annotation of peptides generated from the non-coding regions of D. melanogaster genome. Bioinformation 2016;12:202–8.
[17] Thomas V, Raj N, Varughese D, Kumar N, Sehrawat S, Grover A, et al. Predicting stable functional peptides from the intergenic space of E. coli. Syst Synth Biol 2015;9(4):135–40.
[18] Shidhi PR, Suravajhala P, Nayeema A, Nair AS, Singh S, Dhar PK. Making novel proteins from pseudogenes. Bioinformatics 2015;31:33–9. doi: https://doi.org/10.1093/BIOINFORMATICS/BTU615.
[19] Xie J, Li Y, Liu X, Zhao Y, Li B, Ingvarsson PK, et al. Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. Plant Cell 2019;31(3):563–78.
[20] Wang M, Kong L. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinf 2019;20:1–4. doi: https://doi.org/10.1186/S12859-019-2597-8.
[21] Bhatia G, Sharma S, Upadhyay SK, Singh K. Long non-coding RNAs coordinate developmental transitions and other key biological processes in grapevine. Sci Rep 2019;9:1–14. doi: https://doi.org/10.1038/S41598-019-38989-7.
[22] Bhatia G, Upadhyay SK, Upadhyay A, Singh K. Investigation of long non-coding RNAs as regulatory players of grapevine response to powdery and downy mildew infection. BMC Plant Biol 2021;21:1–16. doi: https://doi.org/10.1186/S12870-021-03059-6.
[23] Osuna-Cruz CM, Paytuvi-Gallart A, Di Donato A, Sundesha V, Andolfo G, Cigliano RA, et al. PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. Nucleic Acids Res 2018;46:D1197–201. doi: https://doi.org/10.1093/NAR/GKX1119.
[24] Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 2013;41:e121. doi: https://doi.org/10.1093/NAR/GKT263.
[25] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–35. doi: https://doi.org/10.1093/NAR/GKN176.
[26] Dhar PK, Thwin C, Tun K, Tsumoto Y, Maurer-Stroh S, Eisenhaber F, et al. Synthesizing non-natural parts from natural genomic template. J Biol Eng 2009;3(1):2.
[27] Pirtskhalava M, Amstrong AA, Grigolava M, Chubinidze M, Alimbarashvili E, Vishnepolsky B, et al. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. Nucleic Acids Res 2021;49:D288–97. doi: https://doi.org/10.1093/nar/gkaa991.
[28] Das D, Jaiswal M, Khan FN, Ahamad S, Kumar S. PlantPepDB: a manually curated plant peptide database. Sci Rep 2020;10:1–8. doi: https://doi.org/10.1038/s41598-020-59165-2.
[29] Agrawal P, Bhalla S, Chaudhary K, Kumar R, Sharma M, Raghava GPS. In silico approach for prediction of antifungal peptides. Front Microbiol 2018;9:323. doi: https://doi.org/10.3389/fmicb.2018.00323.
[30] Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res 2016;44(D1):D1087–93.
[31] Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. Nucleic Acids Res 2021;49:W216–27. doi: https://doi.org/10.1093/NAR/GKAB225.

[32] Khan IK, Wei Q, Chitale M, Kihara D. PFP/ESG: automated protein function prediction servers enhanced with Gene Ontology visualization tool. Bioinformatics 2015;31:271–2. doi: https://doi.org/10.1093/BIOINFORMATICS/BTU646.
[33] Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res 2015;43:W389–94. doi: https://doi.org/10.1093/NAR/GKV332.
[34] Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol 2008;453:3–31. doi: https://doi.org/10.1007/978-1-60327-429-6_1.
[35] Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–38. doi: https://doi.org/10.1038/nprot.2010.5.
[36] Padmashree AP, Imran S, Prakash T, Ravi L. Construction of 3D model of protein drug targets for Erysiphe necator a fungal plant pathogen causing powdery mildew. Biomed Pharmacol J 2020;13:1505–11. doi: https://doi.org/10.13005/BPJ/2024.
[37] Castrignanò T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. The PMDB protein model database. Nucleic Acids Res 2006;34:D306–9. doi: https://doi.org/10.1093/NAR/GKJ105.
[38] Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2002;2452:185–200. doi: https://doi.org/10.1007/3-540-45784-4_14.
[39] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2017;12(2):255–78.
[40] DeLano WL. The PyMOL molecular graphics system. CCP4 Newsl Protein Crystallogr 2002;40:82–92.
[41] Vangone A, Bonvin AMJJ. Contacts-based prediction of binding affinity in protein–protein complexes. Elife 2015;4:. doi: https://doi.org/10.7554/ELIFE.07454e07454.
[42] Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 2004;32:D115–9. doi: https://doi.org/10.1093/NAR/GKH131.
[43] Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 2017;45:W12–6. doi: https://doi.org/10.1093/NAR/GKX428.
[44] Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 2011;39:W155–9. doi: https://doi.org/10.1093/NAR/GKR319.
[45] Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res 2018;46:W49–54. doi: https://doi.org/10.1093/NAR/GKY316.
[46] Yi X, Zhang Z, Ling Y, Xu W, Su Z. PNRD: a plant non-coding RNA database. Nucleic Acids Res 2015;43:D982–9. doi: https://doi.org/10.1093/NAR/GKU1162.
[47] Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2017;45(D1):D1040–5.
[48] Leinonen R, Sugawara H, Shumway M. Collaboration on behalf of the INSD. The Sequence Read Archive. Nucleic Acids Res 2011;39(Database):D19–21.
[49] Pervaiz T, Haifeng J, Haider MS, Cheng Z, Cui M, Wang M, et al. Transcriptomic analysis of grapevine (cv. Summer Black) Leaf, using the illumina platform. PLoS One 2016;11. doi: https://doi.org/10.1371/journal.pone.0147369.
[50] Wen YQ, Zhong GY, Gao Y, Bin LY, Duan CQ, Pan QH. Using the combined analysis of transcripts and metabolites to propose key genes for differential terpene accumulation across two regions. BMC Plant Biol 2015;15:1–22. doi: https://doi.org/10.1186/S12870-015-0631-1.
[51] Domingos S, Fino J, Cardoso V, Sánchez C, Ramalho JC, Larcher R, et al. Shared and divergent pathways for flower abscission are triggered by gibberellic acid and carbon starvation in seedless Vitis vinifera L. BMC Plant Biol 2016;16:1–26. doi: https://doi.org/10.1186/S12870-016-0722-7.
[52] Hu Y, Li Y, Hou F, Wan D, Cheng Y, Han Y, et al. Ectopic expression of Arabidopsis broad-spectrum resistance gene RPW8.2 improves the resistance to powdery mildew in grapevine (Vitis vinifera). Plant Sci 2018;267:20–31.
[53] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 2013;8(8):1494–512.
[54] Pathania S, Acharya V. Computational analysis of "-omics" data to identify transcription factors regulating secondary metabolism in Rauvolfia serpentina. Plant Mol Biol Rep 2015;341:283–302. doi: https://doi.org/10.1007/S11105-015-0919-1.
[55] Aoki K, Ogata Y, Shibata D. Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol 2007;48:381–90. doi: https://doi.org/10.1093/PCP/PCM013.
[56] Mccluskey A, Mb B, Frca C, Ghaaliq A, Mb L. Statistics II: central tendency and spread of data. Contin Educ Anaesth Crit Care Pain 2007;7:127–30. doi: https://doi.org/10.1093/BJACEACCP/MKM020.
[57] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57:289–300. doi: https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.
[58] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.
[59] Van Dongen S. Graph clustering via a discrete uncoupling process. SIAM J Matrix Anal Appl 2008;30:121–41. doi: https://doi.org/10.1137/040608635.
[60] Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, et al. ClusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinf 2011;12(1). doi: https://doi.org/10.1186/1471-2105-12-436.
[61] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinforma 2009;101:110–9. doi: https://doi.org/10.1186/1471-2105-10-421.
[62] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57. doi: https://doi.org/10.1038/NPROT.2008.211.
[63] Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH. Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol 2009;151:3–15. doi: https://doi.org/10.1104/PP.109.140632.
[64] Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, et al. An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. Plant J 2010:86–99. doi: https://doi.org/10.1111/J.1365-313X.2010.04222.X.
[65] Chen M, Herde M, Witte CP. Of the nine cytidine deaminase-like genes in arabidopsis, eight are pseudogenes and only one is required to maintain pyrimidine homeostasis in vivo. Plant Physiol 2016;171:799–809. doi: https://doi.org/10.1104/PP.15.02031.
[66] Balakirev ES, Ayala FJ. Pseudogenes: are they "Junk" or functional DNA? Annu Rev Genet 2003;37:123–51. doi: https://doi.org/10.1146/annurev.genet.37.040103.103949.
[67] Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, et al. Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site–leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. Genetics 2009;182:1303–11. doi: https://doi.org/10.1534/GENETICS.109.102871.
[68] Tian H, MacKenzie CI, Rodriguez-Moreno L, van den Berg GCM, Chen H, Rudd JJ, et al. Three LysM effectors of Zymoseptoria tritici collectively disarm chitin-triggered plant immunity. Mol Plant Pathol 2021;22:683–93. doi: https://doi.org/10.1111/MPP.13055/FORMAT/PDF.
[69] Nguyen QM, Iswanto ABB, Son GH, Kim SH. Recent advances in effector-triggered immunity in plants: new pieces in the puzzle create a different paradigm. Int J Mol Sci 2021;22:4709. doi: https://doi.org/10.3390/IJMS22094709.
[70] Khurana E, Lam HYK, Cheng C, Carriero N, Cayting P, Gerstein MB. Segmental duplications in the human genome reveal details of pseudogene formation. Nucleic Acids Res 2010;38:6997–7007. doi: https://doi.org/10.1093/nar/gkq587.
[71] Wiesner J, Vilcinskas A. Antimicrobial peptides: the ancient arm of the human immune system. Virulence 2010;1:440–64. doi: https://doi.org/10.4161/VIRU.1.5.12983.
[72] Broekaert WF, Cammue BPA, De Bolle MFC, Thevissen K, De Samblanx GW, Osborn RW, et al. Antimicrobial peptides from plants. CRC Crit Rev Plant Sci 1997;16(3):297–323.
[73] Bolouri Moghaddam MR, Vilcinskas A, Rahnamaeian M. Cooperative interaction of antimicrobial peptides with the interrelated immune pathways in plants. Mol Plant Pathol 2016;17:464–71. doi: https://doi.org/10.1111/MPP.12299.
[74] Li W, Li Y. The pseudogene URAHP promotes proliferation and regulates the pathogenesis of preeclampsia. Am J Transl Res 2020;12:4715.
[75] He Y, Chen Y, Tong Y, Long W, Liu Q. Identification of a circRNA-miRNA-mRNA regulatory network for exploring novel therapeutic options for glioma. Res Sq 2021;9. doi: https://doi.org/10.7717/PEERJ.11894.
[76] Xing L, Zhang X, Guo M, Zhang X, Liu F. Application of machine learning in developing a novelty five-pseudogene signature to predict prognosis of head and neck squamous cell carcinoma: a new aspect of "Junk Genes" in biomedical practice. DNA Cell Biol 2020;39:709–23. doi: https://doi.org/10.1089/DNA.2019.5272.
[77] Alon U. An introduction to systems biology : Design principles of biological circuits. 1st ed. New York: Chapman and Hall/CRC; 2006. https://doi.org/10.1201/9781420011432.
[78] Emmert-Streib F, Dehmer M. Robustness in scale-free networks: comparing directed and undirected networks. Int J Mod Phys C 2008;19:717–26. doi: https://doi.org/10.1142/S0129183108012510.
[79] Wang Z, Maity A, Hsiao CK, Voora D, Kaddurah-Daouk R, Tzeng JY. Module-based association analysis for omics data with network structure. PLoS One 2015;10. doi: https://doi.org/10.1371/journal.pone.0122309.
[80] Feechan A, Anderson C, Torregrosa L, Jermakow A, Mestre P, Wiedemann-Merdinoglu S, et al. Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species Muscadinia rotundifolia identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. Plant J 2013;76(4):661–74.