

Article

Functional Annotation of Hypothetical Proteins Derived from Suppressive Subtraction Hybridization (SSH) Analysis Shows NPR1 (Non-Pathogenesis Related)-Like Activity

Murugesan Chandrasekaran ¹, Chandrasekar Raman ², Kandasamy Karthikeyan ³ and Manivannan Paramasivan ^{4,*}

¹ Department of Food Science and Biotechnology, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea; chandrubdubio@gmail.com

² Department of Biochemistry and Molecular Biophysics, Kansas State University, 238 Burt Hall, Manhattan, KS 66506, USA; biotechchandru@gmail.com

³ Department of Food Science and Nutrition, Periyar University, Salem 636011, India; karthibiochem@gmail.com

⁴ Department of Microbiology, Bharathidasan University, Tiruchirappalli, Tamilnadu 620024, India

* Correspondence: manivannan3065@gmail.com; Tel.: +91-431-240-7082

Received: 1 November 2018; Accepted: 9 January 2019; Published: 28 January 2019



Abstract: *Fusarium* wilt is considered the most devastating banana disease incited by *Fusarium oxysporum* f. sp. *cubense* (FOC). The present study addresses suppressive subtraction hybridization (SSH) analysis for differential gene expression in banana plant, mediated through FOC and its interaction with biocontrol agent *Trichoderma asperellum* (pr2). SSH analysis yielded a total of 300 clones. The resultant clones were sequenced and processed to obtain 22 contigs and 87 singleton sequences. BLAST2GO (Basic Local Alignment Search Tool 2 Gene Ontology) analysis was performed to assign known protein function. Initial functional annotation showed that contig 21 possesses p38-like endoribonuclease activity and duality in subcellular localization. To gain insights into its additional roles and precise functions, a sequential docking protocol was done to affirm its role in the defense pathway. Atomic contact energies revealed binding affinities in the order of miRNA > phytoalexins > polyubiquitin, emphasizing their role in the *Musa* defense pathway. Contig 21 and polyubiquitin showed an atomic contact energy value of -479.60 kJ/mol, and even higher atomic contact energies were observed for miRNA (-804.86 , -482.28 , -494.75 kJ/mol), demonstrating its high RNA-binding properties. Phytoalexin contig 21-interacting interfacial residues were identified as rigid (10)/non-rigid (2) based on Bi, N values, and B-factor per residue. Hence, based on these results, contig 21 was characterized as a NPR1 (non-pathogenesis-related protein) homolog that is involved in plant defense and systemic induced resistance.

Keywords: *Fusarium oxysporum* f. sp. *cubense*; functional annotation; *Musa* defense pathway; suppressive subtraction hybridization; *Trichoderma asperellum*

1. Introduction

Fusarium wilt, also called Panama disease, is caused by the soilborne fungus *Fusarium oxysporum* f. sp. *cubense* (FOC). It hinders banana production worldwide [1]. The role of FOC in the spread of the disease can be corroborated by the production of chlamydospores; when the spores contact susceptible lateral or feeder banana roots, the spores germinate, and infection is established. After infection and enumeration, FOC occupies the xylem vessels. The infected banana plants show symptoms

such as yellowing, wilting, and vascular discolorization. Fungicide control for combating FOC includes a corm injection protocol using carbendazim [2]. At this time, proteins termed hypothetical or uncharacterized are becoming increasingly abundant, due to difficulty in annotation protocols. Further deciphering the function of proteins, rather than their available structures, defines the impractical situation for researchers worldwide. Although chemical fungicides are increasingly being used in agriculture, the role of hypothetical proteins and their regulatory roles is still an underexploited arena. Nevertheless, it has been regarded as toxic and harmful not only for the plant, but also its surrounding environment [3]. Deleterious effects resulting from the usage of chemical fertilizers have now forced farmers and agricultural researchers to look for an exclusive and effective alternate to abate chemical use in controlling FOC. Hence, the use of biocontrol agents in combating plant diseases has become an increasingly important research topic, and has been given a prominent role in ecofriendly agricultural crop protection strategies [4,5]. *Trichoderma* spp. are regarded as an exemplary form of biocontrol agent against plant diseases in comparison with chemical fertilizers [6]. *Trichoderma* spp. are ubiquitous and have been used for biocontrol on a wide variety of crop plants [7,8]. Over the last decade, *Trichoderma*-based agro-products, such as biopesticides and biofertilizers, have been increasingly used as ecologically useful alternatives in crop protection, and their application has been observed to have the potential to alleviate soil fertility issues and result in rises in crop productivity and yield [6,7]. *Trichoderma* has the potential to suppress the activity of multiple plant pathogens, including FOC of banana [9,10].

The advent of the omics era has resulted in the escalating accumulation of gene and protein sequences in public databases. Now, the analysis and annotation of accurate function-related data for available sequences pose huge problems to biologists. From the proteome perspective, protein sequences corresponding to comprehensive structures establish the need for bioinformatics approaches in the elucidation of three-dimensional structures of proteins. In the present century, deciphering the function of a protein relies largely on computational tools abridging the laborious experimental protocols. A significant proportion of a genome is demarcated as 'hypothetical' and 'conserved hypothetical'. The former represents the proteins, which lack experimental proof for translational machinery but are originated from an open reading frame [11]. By contrast, conserved hypothetical proteins refer to proteins with phylogenetic lineages with no known definitive function [12]. The functional annotation of proteins in any genome, whether prokaryotic or eukaryotic, yields a considerable amount of proteins as hypothetical, which possess novel and uncharacterized functional properties [13]. Generally, almost half of the proteins in a genome are constituted by hypothetical proteins [14], stressing the need for manual annotation through computational tools. In the recent past, it has been demonstrated that hypothetical proteins have a strong association with evolutionary significance [15,16].

Suppressive subtraction hybridization (SSH) has been a very useful technique for the identification of differentially expressed genes, and reveals the activity, process, and function of a characteristic gene in complex pathways [17,18]. SSH has been applied to decipher plant–pathogen interactions in the banana plant for *Fusarium* wilt [19,20], *Mycosphaerella eumusae* [21], and *Mycosphaerella fijiensis* [22]. A considerable group resulting from SSH are hypothetical proteins with unknown function, which thus necessitate manual annotation for further functional characterization. The present study is unique, and the computational model used could be rationalized by assigning roles to the plant defense mechanism of *Musa*. It has been extensively shown by many researchers that the plant–pathogen interaction pathway in *Musa* depicts miRNA production, phytoalexin accumulation, and defense-related gene induction. The abovementioned steps were taken into account for the elucidation and establishment of a functional role for hypothetical proteins derived from the SSH library of biocontrol–FOC interactions in banana. In the present study, an attempt is made to characterize functional roles for a hypothetical protein, contig 21, based on an in silico analysis. The assessment encompasses sequential docking in the interpretation of RNA–protein interaction, protein–protein interaction, and protein–ligand interaction. Interacting interfacial residues involved in protein–ligand interactions were categorized

into rigid/non-rigid residues based on solvent accessibility. SSH analysis of contig 21 resulted in hypothetical proteins being annotated for their function and involvement in the plant defense pathway. Post-translational modification was also performed, in an effort to gain insights into additional functional roles.

2. Materials and Methods

2.1. Plant Materials and RNA Isolation

Banana cv. Grand Naine (AAA), grown in mud pots, was inoculated with *Fusarium* wilt pathogen (*Fusarium oxysporum* f. sp. *cubense*) and challenged with *Trichoderma asperellum*. Petri plates were antagonistically streaked to assess antifungal activity. Total RNA was isolated for differential analysis of treatment with FOC alone and *T. asperellum* alone.

2.2. SSH Library Construction

Suppression subtractive hybridization was performed using PCR-Select cDNA subtraction kit (Clontech, Mountain View, CA, USA). Forward and reverse subtraction libraries were constructed using cDNA samples of control versus treatment. The subtracted cDNAs were subjected to two rounds of PCR to normalize and enrich cDNA populations. The PCR products were subcloned into a pGEM-T easy cloning vector and transformed into *Escherichia coli* JM109 competent cells. Luria-Bertani medium with isopropyl β -D-1-thiogalactopyranoside (IPTG) and X-gal were used for screening recombinants. Single white colonies were picked and grown overnight at 37 °C. Glycerol stocks were prepared, and all the clones were stored at 80 °C until further use.

2.3. DNA Sequencing and Initial Processing

Positive clones were enumerated for plasmid isolation and validated by colony PCR using nested primers and, finally, plasmids from positive clones were isolated and sequenced using M13 primers. The resultant sequences were trimmed and end-clipped using CodonCode Aligner initially. Later, vector sequences still present were removed by VecScreen (<http://www.ncbi.nlm.nih.gov/VecScreen>) and poly A tails and adapter sequences were eliminated by trimest (<http://emboss.bioinformatics.nl/cgi-bin/emboss/trimest>). Sequences were assembled by CAP3 (<http://doua.prabi.fr/software/cap3>) and annotated by BLAST2GO (<http://www.blast2go.com>).

2.4. Functional Annotation

2.4.1. Homology Search

Contigs and singletons obtained were subjected to BLAST-P, PSI-BLAST [23,24], and HHPRED (<http://toolkit.tuebingen.mpg.de/>) for analyzing similarity and identity. The hits obtained were demarcated into relevance to *Musa* genomes, and PRofile ALIgNmEnt (PRALINE—<http://www.ibi.vu.nl/programs/pralinewww/>) and ClustalX (2.0) (<http://www.clustal.org/clustal2/>) were used to align the functional residues of the protein.

2.4.2. Physicochemical Characterization

Theoretical physicochemical parameters, such as molecular weight, isoelectric point, aliphatic index, instability index, and grand average of hydropathicity (GRAVY) were calculated by using ExPASy's ProtParam server (<http://web.expasy.org/protparam/>) to predict protein stability. ProtScale (<http://web.expasy.org/protscale/>) was used for hydrophobicity plotting based on the Kyte and Doolittle method for analyzing permissive sites in the profile.

2.4.3. Subcellular Localization

PSORT (<https://www.genscript.com/psort.html>) tool was used for predicting subcellular localization of proteins. Further, CELLO (<http://cello.life.nctu.edu.tw/>), a support vector machine-based classification system, was used for precise characterization of subcellular localization. TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) and HMMTOP (<http://www.enzim.hu/hmmtop/index.php>) [25] were utilized for the assessment of transmembrane domains in the proteins based on hidden Markov models. SignalP 4.1 [26] and Signal 3L [27] were used to predict signal peptide sequences and their cleavage sites.

2.4.4. Protein Classification and Function Prediction

PANTHER (<http://www.pantherdb.org/>) [28] was used to categorize protein classification based on protein families, subfamilies, and trees, based on gene ontology. MEME Suite [29] was utilized to identify motif patterns in the proteins. Additionally, InterProScan (<https://www.ebi.ac.uk/interpro/>) was used to discover motifs and assign function [30]. PFP-FunDSeqE server (<http://www.csbio.sjtu.edu.cn/bioinf/PFP-FunDSeqE/>) was used to assign fold patterns based on functional domain and evolutionary significance.

2.4.5. Homology Modeling of miRNA and Target Proteins

miRNAs in functionally annotated proteins were predicted from miRBase and conformations in the secondary structure were assessed using GeneBee server which, in turn, was used for 3D modeling (http://www.genebee.msu.su/services/rna2_reduced.html). Suitable templates were chosen using R3D BLAST and aligned using ClustalX. The template and alignment are fed on to MODERNA server (<http://iimcb.genesilico.pl/modernaserver/submit/model/>), for 3D modeling of the miRNAs. Model building involves the addition of modification, exchange of residues, copying of residues, the addition of loops, and removal of modifications.

For homology modeling of proteins, templates were selected using the homology detection and structure prediction by HMM–HMM comparison employing PSI-BLAST (<http://toolkit.tuebingen.mpg.de/>). The availability of authentic structures in the Protein Data Bank (PDB) was comparatively checked in NCBI Entrez, SWISS PROT, and SCOP databases. The templates chosen had an E-value of <3, and similarity of more than 85%. The proteins were modeled by employing MODELLER 9.14 [31]. Secondary structure was predicted by PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) [32].

2.5. Model Validation

The protein models generated were assessed for stability in the SAVES server (<http://services.mbi.ucla.edu/SAVES/>) and PROSA2 (<https://prosa.services.came.sbg.ac.at/prosa.php>). The stereochemical analysis was performed by RAMPAGE server (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). Putative protein domains and their boundaries were analyzed by DOMPRED server [33]. RMSD values and stability of the proteins modeled were evaluated by QMEAN server (<https://swissmodel.expasy.org/qmean/>), and the models were viewed using CHIMERA (<https://www.cgl.ucsf.edu/chimera/>). Furthermore, both the modeled proteins were superposed for calculating individual RMSD of alpha carbons, backbone, and heavy chains, employing the SuperPose server (<http://wishart.biology.ualberta.ca/SuperPose/>) [34] and DALI server (<http://ekhidna2.biocenter.helsinki.fi/dali/>).

2.6. Molecular Dynamics Simulation

The 3D molecule was minimized locally in vacuo by constraining the backbone of the helices in order to give a first optimization of the rough geometry derived from homology modeling. Using GROMOS96 of SWISS-PDB VIEWER [35]. GROMOS96 is a molecular dynamics computer simulation package for the study of biomolecular structures. Hydrogens were initially added and amino acids with

a more than 30% solvent accessible surface were chosen for further simulations. Energy minimization parameters include 20 steps of deepest descent, and other default conditions in the GROMOS96 package (<http://www.gromos.net/>).

2.7. Analysis of Pockets and Clefts

Pocket detection and occupancy in the modeled proteins were found using CASTp (<http://sts.bioe.uic.edu/castp/index.html?lycs>). The solvent accessible surface area (SASA) was found by GETAREA (<http://curie.utmb.edu/getarea.html>) [36]. The atomic SASA covered by each cleft was calculated by utilizing the radius of water probe 1.4 Å and the area/energy/residue was calculated. The dielectric constant was set to 80.0, and Poisson–Boltzmann method of computation for 20 cycles was used for calculating the electrostatic potential in SWISS-PDB, USA viewer (<https://spdbv.vital-it.ch/>).

2.8. Preparation of Phytoalexins

ACD/ChemSketch (<https://www.acdlabs.com/resources/freeware/chemsketch/>) was used to create two-dimensional Chemical Markup Language (CML) files of phytoalexin structures. OPENBABEL GUI v2.4.1 (http://openbabel.org/wiki/Main_Page) was used to convert cml files to Protein Data Bank (PDB) files. The root-mean-square deviation (RMSD)-based energy minimization was performed in vacuo to give a first optimization of the rough structure. Hydrogens were initially added to a receptor molecule, and AMBER and Gasteiger charges were added to fix unusual bonds in the 3D structure, which utilizes CHARMM force field parameters. The 2D model was optimized, and energy use was minimized using a clean geometry option in ArgusLab 4.0. [37].

2.9. Docking and Analysis of the Interaction

PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/>) was employed for RNA–protein docking and protein–protein docking. The atomic contact energy-occupying area of the docked complex was analyzed. Molecular Docking server was employed for protein–ligand docking (<http://www.dockingserver.com/web>). The geometry of interactions and HB plot of interacting residues was calculated.

2.10. Prediction of Rigid/Non-Rigid Nature of Interacting Residues

Normalized B-factor for ligand-binding residues was predicted using BFPred server, and individual B-factors were statistically depicted for summarizing the rigid/non-rigid nature of the ligand-binding residues. The normalized B-factor per residue (B_i, N) was calculated based on

$$B_{i,N} = B_i - \langle B_i \rangle + \sigma B_i \quad (1)$$

where B_i —B-factor of residue, $\langle B_i \rangle$ —mean B-factor, σ —standard deviation, $B_i, N \leq 0.04$ was considered as rigid whereas $B_i, N \geq 0.04$ were regarded non-rigid.

3. Results

3.1. SSH Analysis

A total of 31 hypothetical proteins were chosen for functional annotation. Amplification of 28S rRNA between subtracted and unsubtracted cDNAs after 28 PCR cycles in subtracted sample confirms successive subtraction.

3.2. Sequence Analysis

The obtained contigs and singleton sequences were separately clustered on the basis of unknown/uncharacterized or hypothetical proteins. The Basic Local Alignment Search Tool (BLASTX) and homology detection by HHPRED, which is a free protein function and protein structure prediction

server, resulted in Musa-specific hypothetical proteins. Initial sequence analysis provided the basic insight that contig 21 could have potential functional attributes. BLASTX results for the hypothetical proteins obtained from differentially expressed genes in biocontrol– FOC interaction reveal three contigs of relevance to the *Musa* genome (Table 1).

Table 1. Basic Local Alignment Search Tool (BLASTX) results for suppressive subtraction hybridization (SSH)-derived clones.

Name of the Protein	BLASTX Results	E Value	Identity(%)
CONTIG5	PREDICTED: uncharacterized protein LOC103975166 (<i>Musa acuminata</i> subsp. <i>Malaccensis</i>). ACCESSION XP_009388351	2×10^{-36}	97%
CONTIG13	hypothetical protein AUEXF2481DRAFT_37807 (<i>Aureobasidium subglaciale</i> EXF-2481). ACCESSION KEQ97279	9.5	36%
Contig 21	PREDICTED: uncharacterized protein LOC10399244 (<i>Musa acuminata</i> subsp. <i>Malaccensis</i>). ACCESSION XP_009419208	7×10^{-58}	100%
12F	Unknown protein (<i>Arabidopsis thaliana</i>). ACCESSION AAK68824	2×10^{-30}	66%
49R	PREDICTED: uncharacterized protein LOC103975166 (<i>Musa acuminata</i> subsp. <i>Malaccensis</i>). ACCESSION XP_009388351	1×10^{-35}	97%
51R	PREDICTED: phytochrome E isoform X2 (<i>Prunus mume</i>). ACCESSION XP_008225375	8×10^{-29}	71%
68F	hypothetical protein HMPREF1063_05178, partial (<i>Bacteroides dorei</i> CL02T00C15). ACCESSION EIY16078	6×10^{-14}	78%
94F	PREDICTED: uncharacterized protein LOC10399244 (<i>Musa acuminata</i> subsp. <i>Malaccensis</i>). ACCESSION XP_009419208	4×10^{-38}	100%
964F	alpha-hemolysin (<i>Burkholderia cenocepacia</i>). ACCESSION ACJ54934	5×10^{-6}	64%
335R	PREDICTED: uncharacterized protein LOC10399244 (<i>Musa acuminata</i> subsp. <i>Malaccensis</i>). ACCESSION XP_009419208	7×10^{-41}	100%
542F	hypothetical protein ECP02989421_5339 (<i>Escherichia coli</i> P0298942.1). ACCESSION ENA26139	6×10^{-11}	43%
934F	hypothetical protein HMPREF1602_03442 (<i>Escherichia coli</i> 907889). ACCESSION ESD39518	2×10^{-7}	35%
959F	hypothetical protein, partial (<i>Plasmodium chabaudi</i>). ACCESSION XP_731877	6×10^{-5}	53%

3.3. Homology Modeling and Sequential Docking

The sequences identified from SSH analysis yielded seven proteins that have functions in plant defense. The protein sequences comprise of bZIP transcription factor, polyubiquitin, calmodulin-binding protein, endochitinase, isoflavone reductase, and mannose glucose-binding lectin. The above proteins were modeled for comprehensive three-dimensional structures (Figure 1) along with the hypothetical protein contig 21, which was functionally annotated as NPR1 (non-pathogenesis-related protein) homolog. Stereochemical analysis of the protein significantly confers with 85.4% residues in the favored region and the least number of residues in the outlier region (Figure 2). With this view, nucleotide-binding properties and protein interaction of contig 21 were assessed after miRNA prediction.

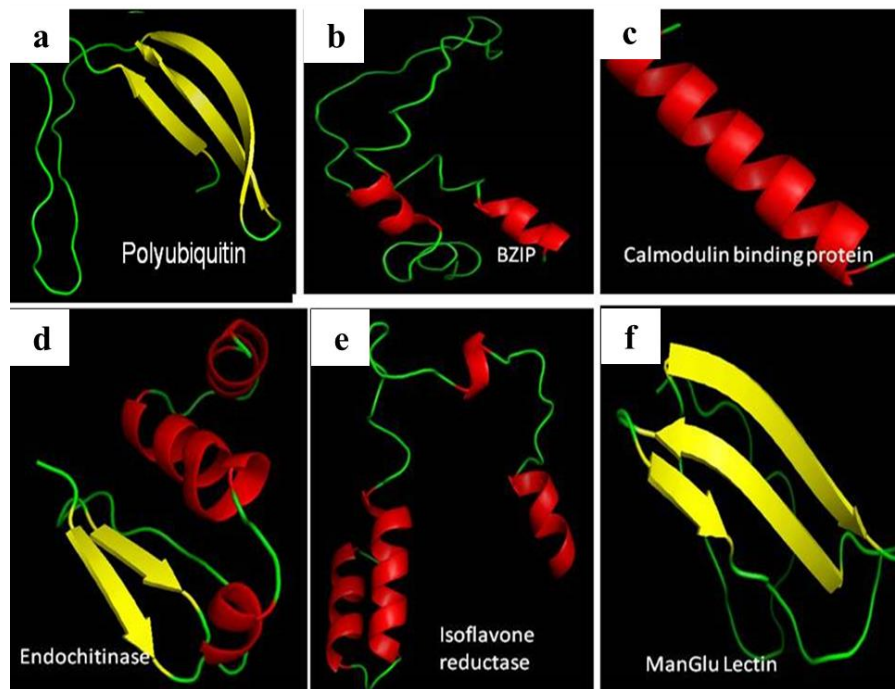


Figure 1. Homology-modeled structures of proteins involved in plant defense indicating (a) polyubiquitin, (b) bZIP transcription factor, (c) calmodulin-binding protein, (d) endochitinase, (e) isoflavone reductase, and (f) mannanase.

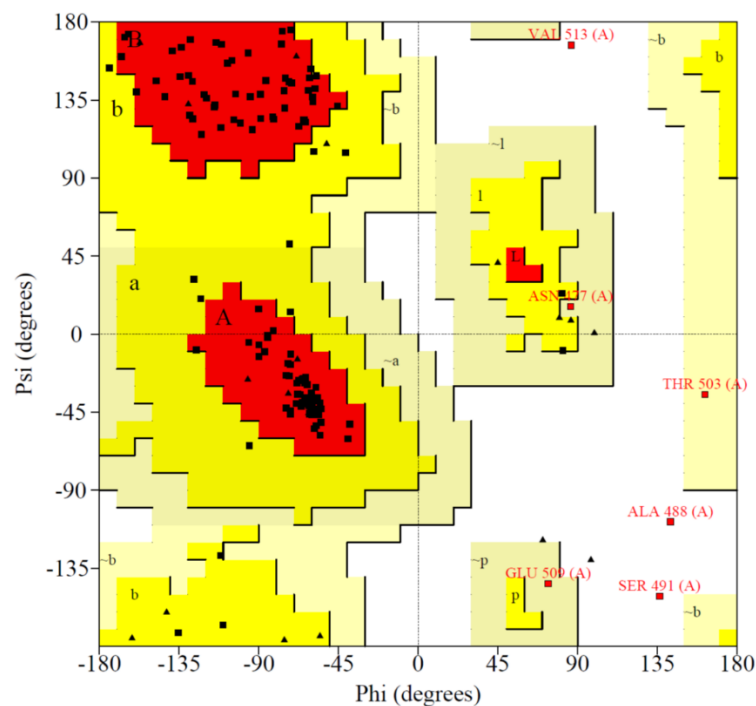


Figure 2. Stereochemical nature of contig 21 showing least outlier residues with good stability (85.4%). Number of residues in favored region (~98.0% expected): 111 (85.4%). Number of residues in allowed region (~2.0% expected): 11 (8.5%). Number of residues in outlier region: 8 (6.2%). Ramachandran plot depicting dihedral angles Psi and Phi of amino acid residues, indicating most favoured regions are in red curves (ABL) and additional allowed regions are (a,b,l,p) in dark yellow curves. VAL—Valine, ASN—Asparagine, THR—Threonine, ALA—Alanine, SER—Serine, GLU—Glutamic acid, Psi—Pound per Square Inch.

3.4. miRNA, Phytoalexin, and Polyubiquitin-Binding Capacity of Contig 21

In contig 21, three putative miRNAs were identified by miRbase, namely >+36–57 AUGGUGAAAUUUGCAAACACUC, >+298–319 GAAGGGAACUCGAUCUAUCUGA, >–489–510 UUUGACGUUGGAGUCCAGUUC. miRNA has the greatest binding energy, followed by phytoalexins and finally, polyubiquitin (Table 2).

Table 2. Binding energies of the docked complexes.

Name of the Docked Complex	Atomic Contact Energy
contig 21–calmodulin-binding protein	288.75
contig 21–manglu lectin	302.46
contig 21–bzip	71.57
contig 21–isoflavone reductase	183.13
contig 21–endochitinase	227.07
contig 21–polyubiquitin	479.60
contig 21–miRNA 1	804.86
contig 21–miRNA 2	482.28
contig 21–miRNA 3	494.75

The highest binding atomic contact energies (ACE) were observed for miRNA followed by phytoalexins and polyubiquitin, emphasizing contig 21's role in the *Musa* defense pathway. The three-dimensional structures of miRNAs predicted with unusual bond angles show that they are single-stranded in nature (Figure 3). With the emphasis of RNA-binding properties, phytoalexin anigorufone, *cis*-2,3-dihydro-2,3-dihydroxy-9-phenylphenalenone, 4-phenylphenalenone trenbolone, and 4'-ethoxyirenonolone were docked onto contig 21.

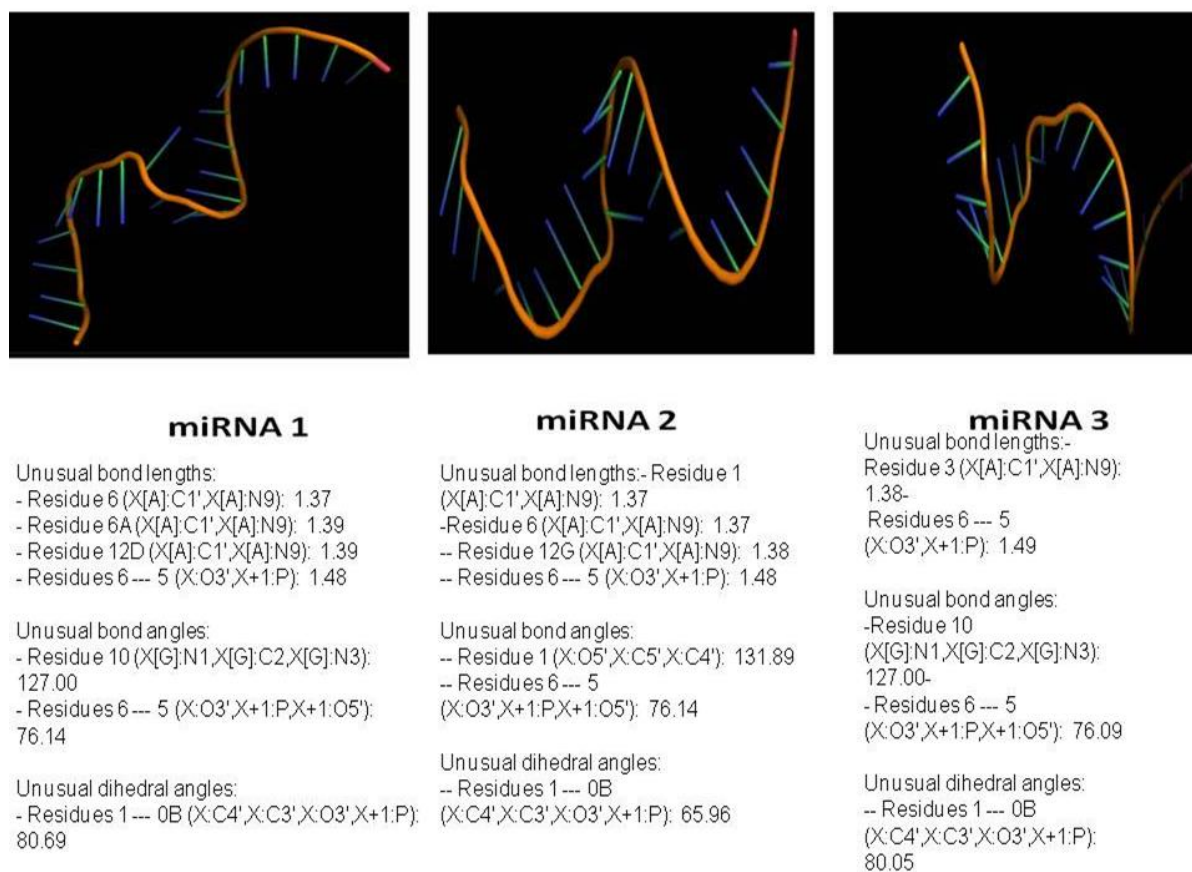


Figure 3. Three dimensional structures of modeled miRNAs with unusual bond angles.

The modeled hytoalexins have comprehensive structures and, upon binding, show distinct geometry of interactions (Table 3). Docked poses of the four phytoalexins show similar binding, and this can be correlated to same active site-binding pattern for all the four phytoalexins docked onto contig 21 (Figure 4).

Table 3. The geometry of interactions between contig 21 and phytoalexins.

Name of the Protein	Binding Free Energy (kcal/mol)	Estimated Inhibition Constant, K_i (μM)	Electrostatic Energy (kcal/mol)	Total Intermolecular Energy (kcal/mol)	Frequency (%)	Interaction Surface
Contig 21-1	-6.95	8.04	-0.05	-7.84	50	785.434
Contig 21-2	-7.58	2.77	-0.04	-8.14	50	679.576
Contig 21-3	-7.17	5.58	-0.11	-8.10	50	702.381
Contig 21-4	-7.29	4.53	+0.04	-7.49	50	664.13

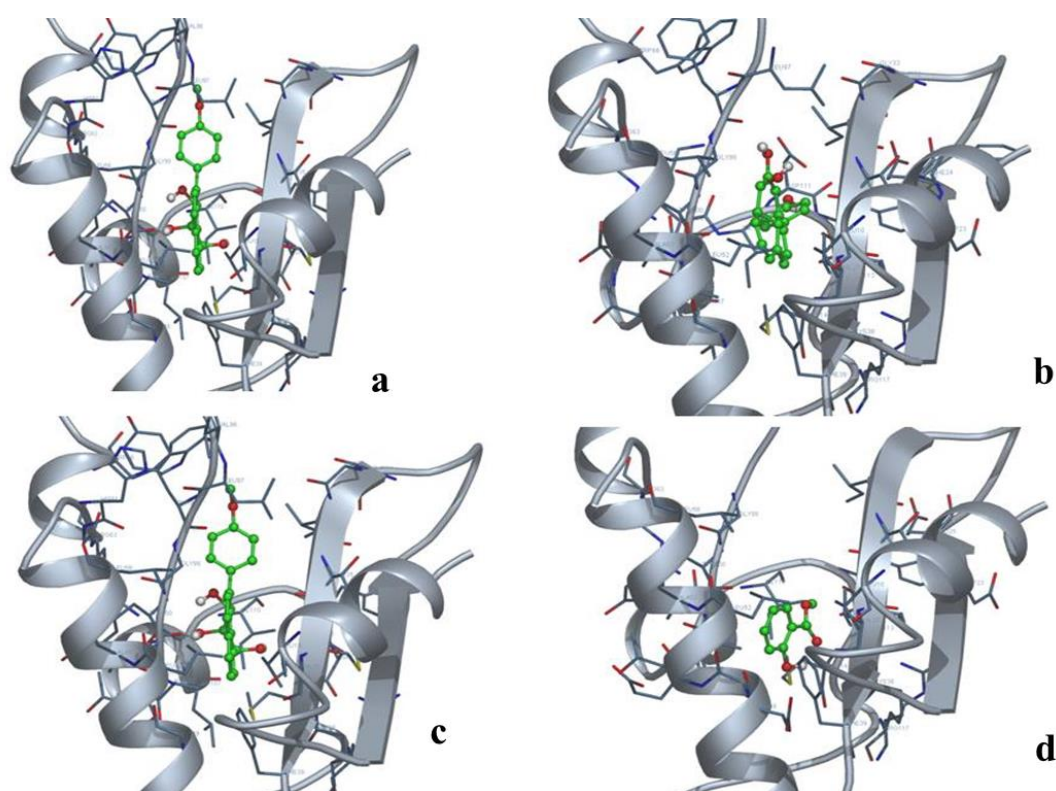


Figure 4. Docked poses for phytoalexins and contig 21 interaction: (a) anigorufone–contig 21, (b) i 2,3-dihydro-2,3-dihydroxy-9-phenylphenalenone–contig 21, (c) 4-phenylphenalenone trenbolone–contig 21; (d) 4'-methoxyirenonone–contig 21.

Active site-interacting residues are Phe39, Asp113, Leu101, Met114, Trp66, Asp111, Leu52, Met36, Ala102, Arg63, Val37, and Ile34. Furthermore, the interacting residues were characterized as rigid/non-rigid based on their B-factor profiles. B-factor values pictorially showing B_i , N values of individual residues clearly indicate that increased binding affinity is due to the contribution of the rigidity of interacting residues (Figure 5 and Table 4). Ten rigid residues and 2 non-rigid residues show the efficiency of interaction.

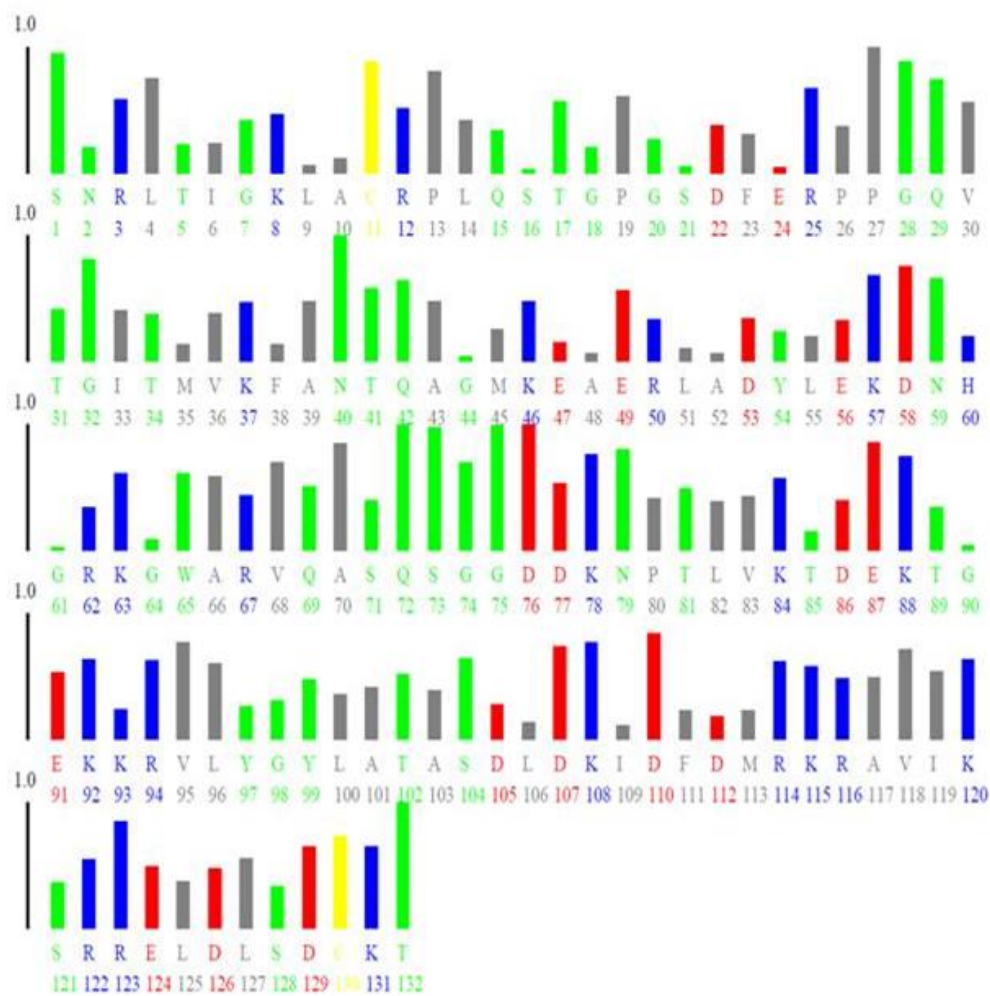


Figure 5. B-factor analysis of interacting residues between phytoalexins and contig 21. Ile 34—polar uncharged, Met 36—hydrophobic, Phe 39—hydrophobic, Leu 52—hydrophobic, Arg 63—positively charged, Trp 66—positively charged, Leu 101—positively charged, Ala 102—polar uncharged, Asp 111—hydrophobic, Asp 113—hydrophobic, Met 114—positively charged; blue—positively charged, red—negatively charged, green—polar uncharged, gray—hydrophobic residues.

Table 4. B-factor values for interacting residues and prediction of rigid/non-rigid nature of interacting residues.

Residue Number	Name of the Residue	Individual B-Factor	B_i, N^*
34	E	0.227	0.0023
36	D	0.661	0.0354
37	N	0.808	0.0466
39	G	0.668	0.0359
52	G	0.58	0.0292
63	T	0.659	0.0352
66	K	0.737	0.0411
101	R	0.193	−2.6675
102	E	0.084	−0.0085
111	S	0.294	0.0074
113	R	−0.076	−0.0207
114	F	−0.347	−0.414

* normalized B-factor per residue.

3.5. Duality in Localization in Deciphering Contig 21 as NPR1 Homolog

Subcellular localization analysis shows contig 21 to be both nuclear and cytoplasmic. Further, there were no signal peptides and transmembrane domains present. NPR1 occurs as an inactive oligomer in the cytoplasm. Based on gene ontology, a function for contig 21 was deciphered as endoribonuclease p38-like activity, which is involved in RNA interference and possesses transferase activity. Hence, from the present results, it is evident that contig 21 has dual localization, like NPR1, and functions as a regulator of transcription.

3.6. Ubiquitination Activities of Contig 21

Diversified ubiquitination patterns showed contig 21 as a NPR1 homolog (Table 5). Post-translational modification mapping of contig 21 yielded significant differential patterns for ubiquitination alone, rather than glycosylation, sumoylation, acetylation, and other modifications.

Table 5. Prediction of polyubiquitination patterns in contig 21.

Peptide	Position	Score
ANTQAGMKEAERLAD	24	2.33
DKNPTLVKTDEKTGE	62	1.51
TLVKTDEKTGEKKKRV	66	1.01
TDEKTGEKKRVLYGY	70	1.02
DEKTGEKKRVLYGYL	71	0.96
KIDFDMRKRAVIKSR	93	1.09
MRKRAVIKSRRELDL	98	1.94
ELDLSDCCKTSFRFCH	109	0.45
TAGVLHVKGTCVSVL	128	1.04
GRTRVTIKERGLQRQ	160	0.64
LQRQRGGKPSYSGDG	171	2.23
FSGPSACKALNSEPL	195	2.74

4. Discussion

Fusarium wilt management involving chemical fungicides is rigorous and not economically feasible. Alternative strategies and environmentally friendly protocols for sustainable agriculture of banana and other crops have been in dire need worldwide. Transcriptome refers to the comprehensive listing of transcripts along with their splice variants. SSH analysis has been adopted as an economical mode for the identification of genes with ease and looks out for novel genes in a genome [38]. Bioinformatic analysis, coupled with SSH, has been utilized for the identification of differentially expressed immune response in insects [39]. However, studies pertaining to SSH analysis in banana corroborated with functional annotation of a hypothetical protein renders the present study significant. PSI-BLAST analysis showed a biological role of contig 21 in banana. The recently published *Musa* genome of 523 Mb [40] has resulted in a considerable amount of ESTs (expressed sequence tags).

NPR1 oligomer–monomer transition involves cysteine residues at 82, 150, 155, 160, and 216 positions [41]. However, at the 82nd position alanine, instead of cysteine, is found in many homologs. Duality in subcellular localization viz., nucleus and cytoplasm, indicate that contig 21 could have potential implications for the similar monomer–oligomer transition. In contig 21, the 82nd residue was occupied by Ala, characteristic of NPR1. Leucine-rich repeats in the motif pattern of contig 21 are also revealing the function corresponding to NPR1. NPR1 natively possess protein–protein interaction motifs, a zinc finger, and ankyrin repeat domains with DNA-binding capability [42]. Higher atomic contact energies were observed for miRNAs (−804.86, −482.28, −494.75 kJ/mol) followed subsequently by polyubiquitin (−479.60), suggesting a role for contig 21 in ubiquitination. miRNAs in regulating genes important for plant defense have been reported [43]. The miRNA-binding capabilities of contig 21 indicate that it has regulatory significance in resistance to *F. oxysporum*. Differentially expressed miRNAs in tomato after *F. oxysporum* infection are involved in plant

defense [44]. Nevertheless, in the case with banana, this has not been previously covered. Hotspots in the interface of protein complexes identify binding sites unique from the rest of the surface [45]. Hence, assessment of rigid/non-rigid nature decides the interaction sites and characterization of hotspot residues. Redox changes lead to the conversion of the oligomer to a monomer, that translocates to the nucleus and interacts with TGA transcription factors [42,46].

The results postulate that the hypothetical protein, contig 21, is involved in cross-communication between the salicylic acid and jasmonic acid-dependent defense signaling pathway, like that of NPR1 [47]. As the SSH analysis resulted in proteins that belong to plant defense, the presence of NPR1 characteristically determines whether *Fusarium* resistance is constitutively expressed in banana. Resistance to *F. oxysporum*, apart from ethylene, jasmonic acid, and the salicylic acid pathway, requires the *NPR1* gene [48]. Thus, the functional annotation of contig 21 provides us a clue that it functions as an NPR1 homolog with regulatory attributes. NPR1 has its role in plant defense through ubiquitination-mediated activity [49]. Overall, contig 21 has been deciphered as an NPR1 homolog based on SSH analysis and in silico perspectives.

5. Conclusions

Functional annotation of hypothetical proteins derived from the SSH library of banana after FOC–biocontrol interaction reveals contig 21 as an NPR1 homolog involved in systemic resistance. miRNA-binding properties, phytoalexin interaction, and ubiquitination affirm the role of contig 21 as an NPR1 homolog. However, additional in vitro and in vivo testing is necessitated to determine its definitive function.

Author Contributions: M.C. and M.P. planned and designed the research; M.C. and C.R. performed the experiments; K.K., M.C., and C.R. analyzed the data; K.K. and C.R. editing; M.P. wrote the manuscript with assistance from M.C.

Funding: This research received no external funding.

Conflicts of Interest: Authors declare there is no conflict of interest.

References

1. Lin, Y.H.; Chang, J.Y.; Liu, E.T.; Chao, C.P.; Huang, J.W.; Chang, P.F.L. Development of a molecular marker for specific detection of *Fusarium oxysporum* f. sp. *cubense* race 4. *Eur. J. Plant Pathol.* **2008**, *123*, 353–365. [[CrossRef](#)]
2. Pegg, K.G.; Moore, N.Y.; Bentley, S. *Fusarium* wilt of banana in Australia: A review. *Aust. J. Agric. Res.* **1996**, *47*, 637–650. [[CrossRef](#)]
3. Getha, K.; Vikineswary, S. Antagonistic effects of *Streptomyces violaceusniger* strain G10 on *Fusarium oxysporum* f.sp. *cubense* race 4: Indirect evidence for the role of antibiosis in the antagonistic process. *J. Ind. Microbiol. Biotechnol.* **2002**, *28*, 303–310. [[PubMed](#)]
4. Monte, E. Understanding Trichoderma: Between biotechnology and microbial ecology. *Int. Microbiol.* **2001**, *4*, 1–4. [[PubMed](#)]
5. Chandrasekaran, M.; Chun, S.C. Expression of PR-protein genes and induction of defense-related enzymes by *Bacillus subtilis* CBR05 in tomato (*Solanum lycopersicum*) plants challenged with *Erwinia carotovora* subsp. *carotovora*. *Biosci. Biotechnol. Biochem.* **2016**, *80*, 2277–2283. [[CrossRef](#)] [[PubMed](#)]
6. Harman, G.E.; Howell, C.R.; Viterbo, A. *Trichoderma* species—Opportunistic, virulent plant symbionts. *Nat. Rev. Microbiol.* **2004**, *2*, 43–56. [[CrossRef](#)] [[PubMed](#)]
7. Waghunde, R.R.; Shelake, R.M.; Sabalpara, A.N. *Trichoderma*: A significant fungus for agriculture and environment. *Afr. J. Agric. Res.* **2016**, *11*, 1952–1965.
8. Mukherjee, P.K.; Buensanteai, N.; Moran-Diez, M.E.; Druzhinina, I.S.; Kenerley, C.M. Functional analysis of non-ribosomal peptide synthetases (NRPSs) in *Trichoderma virens* reveals a polyketide synthase (PKS)/NRPS hybrid enzyme involved in induced systemic resistance response in maize. *Microbiology* **2012**, *158*, 155–165. [[CrossRef](#)]

9. Cotxarrera, L.; Trillas-Gay, M.I.; Steinberg, C.; Alabouvette, C. Use of sewage sludge compost and *Trichoderma asperellum* isolates to suppress Fusarium wilt of tomato. *Soil Biol. Biochem.* **2002**, *34*, 467–476. [[CrossRef](#)]
10. Segarra, G.; Casanova, E.; Avile's, M.; Trillas, M.I. *Trichoderma asperellum* strain T34 controls Fusarium wilt disease in tomato plants in soilless culture through competition for iron. *Microb. Ecol.* **2010**, *59*, 141–149. [[CrossRef](#)]
11. Koonin, E.V.; Galperin, M.Y. *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics*; Kluwer Academic Publishers: Philip Drive Norwell, MA, USA, 2003.
12. Stephens, R.S.; Kalman, S.; Lammel, C.; Fan, J.; Marathe, R. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **1998**, *282*, 754–759. [[CrossRef](#)]
13. Desler, C.; Suravajhala, P.; Sanderhoff, M.; Rasmussen, M.; Rasmussen, L.J. In silico screening for functional candidates amongst hypothetical proteins. *BMC Bioinform.* **2009**, *10*, 289. [[CrossRef](#)]
14. Loewenstein, Y.; Raimondo, D.; Redfern, O.C.; Watson, J.; Frishman, D.; Linial, M.; Orengo, C.; Thornton, J.; Tramontano, A. Protein function annotation by homology-based inference. *Genome Biol.* **2009**, *10*, 207. [[CrossRef](#)]
15. Mazandu, G.K.; Mulder, N.J. Function prediction and analysis of *Mycobacterium tuberculosis* hypothetical proteins. *Int. J. Mol. Sci.* **2012**, *13*, 7283–7302. [[CrossRef](#)]
16. Shahbaaz, M.; Hassa, M.I.; Ahmad, F. Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS ONE.* **2013**, *8*, e84263. [[CrossRef](#)]
17. Diatchenko, L.; Lau, Y.F.; Campbell, A.P.; Chenchik, A.; Moqadam, F.; Huang, B.; Lukyanov, S.; Lukyanov, K.; Gurskaya, N.; Sverdlov, E.D. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6025–6030. [[CrossRef](#)]
18. Fernandez, L.; Torregrosa, L.; Terrier, N.; Sreekantan, L.; Grimplet, J.; Davies, C.; Thomas, M.R.; Romieu, C.; Ageorges, A. Identification of genes associated with flesh morphogenesis during grapevine fruit development. *Plant Mol. Biol.* **2007**, *63*, 307–323. [[CrossRef](#)]
19. Van Den Berg, N.; Berger, D.K.; Hein, I.; Birch, P.R.J.; Wingfield, M.J.; Viljoen, A. Tolerance in banana to *Fusarium* wilt is associated with early up-regulation of cell wall-strengthening genes in the roots. *Mol. Plant Pathol.* **2007**, *8*, 333–341. [[CrossRef](#)]
20. Raman, T.; Gopalakrishnan, V.; Perumal, G. Identification of differentially expressed genes from *Fusarium oxysporum* f. sp. *cubense* and *Trichoderma asperellum* (pr2) interaction in the susceptible banana cultivar Grand Naine. *Turk. J. Bot.* **2016**, *40*, 480–487. [[CrossRef](#)]
21. Uma, S.; Sudhakar, B.; Thangavelu, R.; Backiyarani, S.; Saraswathi, M.S.; Saravankumar, A.S. Differential gene expression analysis for Sigatoka (*Mycosphaerella eumusae*) in resistant *Musa acuminata* cv. *Manoranjitham* (AAA). In Proceedings of the Plant & Animal Genomes XVIII Conference, San Diego, CA, USA, 9–13 January 2010.
22. Portal, O.; Izquierdo, Y.; De Vleeschauwer, D.; Sanchez-Rodriguez, A.; Mendoza-Rodriguez, M. Acosta-Suarez. Analysis of expressed sequence tags derived from a compatible *Mycosphaerella fijiensis*-banana interaction. *Plant Cell Rep.* **2011**, *30*, 913–928. [[CrossRef](#)]
23. Altschul, S.F.; Koonin, E.V. Iterated profile searches with PSI-BLAST—A tool for discovery in protein databases. *Trends Biochem. Sci.* **1998**, *23*, 444–447. [[CrossRef](#)]
24. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
25. Tusnád, G.E.; Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **2001**, *17*, 849–850. [[CrossRef](#)]
26. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
27. Shen, H.B.; Chou, K.C. Signal-3L: A 3-layer approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* **2007**, *363*, 297–303. [[CrossRef](#)]
28. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **2003**, *13*, 2129–2141. [[CrossRef](#)]

29. Timothy, L.; Bailey, E.C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, Stanford, CA, USA, 14–17 August 1994; AAAI Press: Menlo Park, CA, USA, 1994; pp. 28–36.
30. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: Protein domains identifier. *Nucleic Acid. Res.* **2005**, *33*, W116–W120. [[CrossRef](#)]
31. Eswar, N.; Marti-Renom, M.A.; Webb, B.; Madhusudhan, M.S.; Eramian, D.; Shen, N.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling with MODELLER. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; Volume 15, pp. 5.6.1–5.6.30.
32. Bryson, K.; McGuffin, L.J.; Marsden, R.L.; Ward, J.J.; Sodhi, J.S.; Jones, D.T. Protein structure prediction servers at University College London. *Nucleic Acids Res.* **2005**, *33*, W36–W38. [[CrossRef](#)]
33. Marsden, R.L.; McGuffin, L.J.; Jones, D.T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.* **2002**, *11*, 2814–2824. [[CrossRef](#)]
34. Maiti, R.; Gary, H.; Domselaar, V.; Zhang, H.; Wishart, D.S. SuperPose: A simple server for sophisticated structural superposition. *Nucleic Acid. Res.* **2004**, *32*, W590–W594. [[CrossRef](#)]
35. Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723. [[CrossRef](#)]
36. Fraczekiewicz, R.; Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.* **1998**, *19*, 319–333. [[CrossRef](#)]
37. Thompson, M.A. *ArgusLab 4.0.1. Planaria Software*; LLC: Seattle, WA, USA, 2004.
38. Goswami, S.; Kumar, R.R.; Dubey, K.; Singh, J.P.; Tiwari, S.; Kumar, A.; Smita, S.; Mishra, D.C.; Kumar, S.; Grover, M.; et al. SSH Analysis of Endosperm Transcripts and Characterization of Heat Stress Regulated Expressed Sequence Tags in Bread Wheat. *Front. Plant Sci.* **2016**, *7*, 1230. [[CrossRef](#)]
39. Badapanda, C. Suppression subtractive hybridization (SSH) combined with bioinformatics method: An integrated functional annotation approach for analysis of differentially expressed immune-genes in insects. *Bioinformatics* **2013**, *9*, 216–221. [[CrossRef](#)]
40. D’Hont, A.; Denoeud, F.; Aury, J.M.; Baurens, F.C.; Carreel, F.; Garsmeur, O.; Noel, B.; Bocs, S.; Droc, G.; Rouard, M.; et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **2012**, *488*, 213–217. [[CrossRef](#)]
41. Sandhu, D.; Tasma, M.; Frasc, R.; Bhattacharyya, M.K. Systemic acquired resistance in soybean is regulated by two proteins, orthologous to *Arabidopsis* NPR1. *BMC Plant Biol.* **2009**, *9*, 105. [[CrossRef](#)]
42. Després, C.; Chubak, C.; Rochon, A.; Clark, R.; Bethune, T.; Desveaux, D.; Fobert, P.R. The *Arabidopsis* NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity to the basic domain/leucine zipper transcription factor TGA1. *Plant Cell* **2003**, *15*, 2181–2191. [[CrossRef](#)]
43. Pumphlin, N.; Voinnet, O. RNA silencing suppression by plant pathogens: Defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* **2013**, *11*, 745–760. [[CrossRef](#)]
44. Ouyang, S.; Park, G.; Atamian, H.S.; Han, C.S.; Stajich, J.E. MicroRNAs suppress NB domain genes in tomato that confer resistance to *Fusarium oxysporum*. *PLoS Pathog.* **2014**, *10*, e1004464. [[CrossRef](#)]
45. Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. Protein–protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5772–5777. [[CrossRef](#)]
46. Mou, Z.; Fan, W.; Dong, X. Inducers of plant systemic acquired resistance regulate *NPR1* function through redox changes. *Cell* **2003**, *113*, 935–944. [[CrossRef](#)]
47. Pieterse, C.M.; van Loon, L.C. *NPR1*: The spider in the web of induced resistance signaling pathways. *Curr. Opin. Plant Biol.* **2004**, *7*, 456–464. [[CrossRef](#)] [[PubMed](#)]
48. Berrocal-Lobo, M.; Molina, A. Ethylene response factor 1 mediates *Arabidopsis* resistance to the soilborne fungus *Fusarium oxysporum*. *Mol. Plant Microbe Interact.* **2004**, *17*, 763–770. [[CrossRef](#)] [[PubMed](#)]
49. Trujillo, M.; Shirasu, K. Ubiquitination in plant immunity. *Curr. Opin. Plant Biol.* **2010**, *13*, 402–408. [[CrossRef](#)] [[PubMed](#)]

