Philippine Journal of Science
149 (3-a): 915-934, October 2020
ISSN 0031 - 7683
Date Received: 17 Apr 2020

Resistance Gene Analogs of Mango: Insights on Molecular Defenses and Evolutionary Dynamics

Darlon V. Lantican*, **Cris Q. Cortaga**, **Anand Noel C. Manohar**, **Fe M. dela Cueva**, and **Maria Luz J. Sison**

Institute of Plant Breeding, College of Agriculture and Food Science,
University of the Philippines Los Baños, College, Laguna, Philippines 4031

Mango is an economically important fruit crop largely cultivated in the tropics and, thus, is constantly challenged by a myriad of insect pests and diseases. However, no detailed analysis of its resistance gene analogs (RGAs) has been performed, which is a vital resource for plant breeding. Here, we analyzed the RGAs of mango *via de novo* assembly of transcriptomic sequences and mining of the recently published whole genome sequence (WGS). From the transcriptomic assembly, a core mango RGA database with 747 protein models was established. Meanwhile, 1,775 RGAs were identified in the mango WGS and classified based on conserved domains and motifs: 54 nucleotide binding site proteins (NBS), 107 NBS – leucine rich repeat proteins (NBS-LRR), 242 coiled-coil NBS-LRR (CNL), 79 toll/interleukin-1 receptor NBS-LRR (TNL), 78 coiled-coil NBS (CN), 30 toll/interleukin-1 receptor NBS (TN), 45 toll/interleukin-1 receptor with unknown domain (TX), 133 receptor-like proteins (RLP), 917 receptor-like kinases (RLK), 83 transmembrane coiled-coil domain protein (TM-CC), and seven NBS-encoding proteins with other domains. The transcriptome- and genome-wide RGAs have been functionally well-annotated through gene ontology (GO) analysis, and their expression profiles across different mango varieties were also examined. Phylogenetic analyses of expressed and genome-wide RGAs suggest highly divergent functions of the RGAs, which were broadly clustered into 6 and 8 major clades, respectively, based on their domain classification. From the mango RGA transcripts, 134 unique EST-SSR (expressed sequence tags – simple sequence repeat) loci were identified and primers were designed targeting these potential markers. Moreover, comparative analysis of mango with other plant species revealed 65 species-specific RGA families (396 orthologous genes) and detected 1,005 RGA gene duplication events. To date, this is the most comprehensive analysis of mango RGAs, which also provide insights into the dynamic mango-pest co-evolutionary arms race and offer a trove of markers for utilization in resistance breeding.

Keywords: mango, resistance gene analogs (RGAs), bioinformatics, plant defenses, evolution

INTRODUCTION

Mango (*Mangifera indica* L.) is a globally popular fruit crop with economical and agricultural significance, especially in the tropics where it is largely cultivated. One of the constraints in mango production worldwide include a myriad of insect pests (*e.g.* oriental fruit fly, mango hopper, cecid fly, *etc.*) and diseases (*e.g.* anthracnose, stem-end rot, scab, *etc.*), which can affect mango at different life stages thus significantly reducing fruit quality and yield (Bally 2006). These are major barriers to the export market and can impede international trade as many pests and diseases in mango are important quarantine considerations.

*Corresponding Author: dvlantican@up.edu.ph

Along the course of evolution, plants have generally developed mechanisms to recognize pathogen and insect attacks and to activate defense response against them (Dangl and Jones 2001; Acevedo *et al.* 2015). The molecules derived from pathogens, insects, and plant cell damage (*e.g.* due to pathogen and insect attack) that trigger defense responses are generally referred to as microbe-/pathogen-/herbivore-/damage-associated molecular patterns (MAMPs/ PAMPs/ HAMPs/ DAMPs) (Acevedo *et al.* 2015; Choi and Klessig 2016). These molecular patterns are recognized by the pattern recognition receptors (PRRs) in the plant cell and trigger defense responses as these are perceived by the plant as "non-self" and indicative of pest attack (Pitzschke 2013; Zipfel *et al.* 2006; Acevedo *et al.* 2015; Choi and Klessig 2016). Hence, these molecules are generally considered as elicitors of plant defense responses and initiate the first line of defense, known as PAMP or DAMP-triggered immunity (Jones and Dangl 2006; Hu *et al.* 2018). The first layer of defense, however, is often overcome by plant pests through effector proteins, leading to effector-triggered susceptibility. To counteract this, plants can have a second line of defense known as effector-triggered immunity (Boller and He 2009; Pieterse *et al.* 2012). In ETI, due to host-pest co-evolutionary "arms race," plants acquire a family of polymorphic and diverse resistance genes (*R* genes), which encode R proteins that recognize the attacker-specific effectors (Cui *et al.* 2015; Jones and Dangl 2006) to correspond for a gene-for-gene interaction (Thompson and Burdon 1992). This recognition results in intracellular signaling and switching "on" of plant defense genes leading to plant resistance against pest attack.

Collectively, the PRRs and *R* genes are called RGAs, which share conserved domains and motifs (Li *et al.* 2016). PRRs are predominantly categorized as RLKs and RLPs. RLKs have an extracellular sensing domain [either leucine-rich repeat (LRR) type or lysin motif (LysM) type], a transmembrane (TM) domain, and an intracellular kinase domain; while RLPs possess similar domain structure except that it lacks an intracellular kinase domain (Tör *et al.* 2009). The R proteins are intracellular immune receptors (effector-recognition receptors) and the majority of which belong to NBS-LRR class (Gururani *et al.* 2012; Neupane *et al.* 2018). The R proteins contain certain domains or motifs such as serine/threonine kinases, LRRs, NBS, TMs, leucine-zipper, coiled-coil, and toll/ interleukin-1 receptor (TIR) (Li *et al.* 2016). Depending on the domain/motif architecture combinations, the subgroups of NBS-encoding proteins are designated as NBS, CNL, TNL, CN, TN, NL, TX, and other NBS-protein that may have chimeric domain/motif architecture. Since RGAs share conserved structural features, it is possible to predict them extensively to obtain deeper insight into the underlying molecular defenses of the plant. RGAs have been widely studied and are useful for the breeding of resistant crops [see review by Sekhwal *et al.* (2015)]. One way to predict RGAs in plants is through bioinformatics analyses of next-generation sequencing (NGS) genomic and transcriptomic data (Lantican *et al.* 2019; Neupane *et al.* 2018; Zhang *et al.* 2016; Karthika *et al.* 2019). With the advent of this technology, it is possible to unravel gene networks and develop molecular markers tagging economically important traits, reveal other molecular information such as intron-exon boundaries and the presence of transposable elements, and discover novel biological processes (Goodwin *et al.* 2016).

The RGAs were first studied on a genome-wide scale using the model plant *Arabidopsis thaliana* and since then, thousands of RGAs from numerous plant genomes have been identified (Meyers *et al.* 2003; Sekhwal *et al.* 2015). Recently, the WGS of mango (Alphonso variety) has been published (Wang *et al.* 2020). Mango is believed to be an allotetraploid (4n = 40) (Chagné 2015) with a genome size of around 360 MB (Alphonso variety) according to the recent WGS effort (Wang *et al.* 2020). Despite the economical and agricultural importance of mango, especially in the tropics, no comprehensive analysis of mango RGAs has been performed. Thus, with the availability of valuable genomic and transcriptomic resources, this study aimed to systematically identify and characterize the RGAs of mango through the mining of the recent WGS data (from Alphonso variety) and *de novo* assembly of transcriptomes (from different varieties) available in biological repositories. The evolutionary dynamics of mango RGAs were also analyzed and compared to other well-known plant species. To our knowledge, this paper provides the most comprehensive identification, characterization, and evolutionary analysis of mango RGAs to date, which offer a trove of markers and evolutionary insights for utilization in resistance breeding against insect pests and pathogens.

MATERIALS AND METHODS

De Novo Transcriptome Assembly and Gene Models from Mango Whole Genome

To obtain the expressed RGAs in mango, raw transcriptome reads (n = 24) were accessed and retrieved from the public repository (NCBI; Table 1) and one directly from Hong *et al.* (2016) (obtained from paired-end sequencing of "Zill" mango variety). The raw read sequences were pre-processed by removing the adapter sequences and low-quality base score nucleotide sequences using Trimmomatic v0.36 (Bolger *et al.* 2014) with the following parameters: SLIDINGWINDOW: 5:30; LEADING:5; TRAILING:5; MINLEN:85. Each

Table 1. List of source transcriptome sequences and their predicted number of RGAs.

Source SRA files	Mango variety	Tissue	NBS coding								RLP	RLK	TM-CC	Total
			NBS	CNL	TNL	CN	TN	NL	TX	Other				
Hong *et al.* (2016)	"Zill"	Fr	78	9	1	15	3	17	11	0	45	150	14	343
SRR1448302	"Keitt"	Fr	33	9	2	17	6	22	10	1	43	136	6	285
SRR1561197	"Kent"	Fr	20	0	0	3	1	1	4	0	15	68	12	124
SRR1562177	"Kent"	Fr	15	0	0	3	0	1	9	0	10	41	8	87
SRR1956775	"Amrapali"	L	40	1	0	9	1	2	11	0	10	21	6	101
SRR2159471	"Tommy Atkins"	Fl, L, Fr, S	74	7	1	15	3	18	12	2	78	223	21	454
SRR2162836	"Amin Abrahimpur"	Fl, L, Fr, S	71	10	1	11	4	23	18	1	76	216	22	453
SRR2162878	"Burma"	Fl, L, Fr, S	24	0	0	3	0	2	11	0	37	112	11	200
SRR2162889	"*M. casturi* Purple"	Fl, L, Fr, S	85	6	1	7	2	14	19	1	67	177	20	399
SRR2162908	"Neelum"	Fl, L, Fr, S	66	7	1	14	5	18	4	1	72	214	20	422
SRR2162919	"Thai Everbearing"	Fl, L, Fr, S	80	10	2	19	1	13	15	2	71	210	24	447
SRR2162929	"Turpentine"	Fl, L, Fr, S	67	6	0	10	5	15	14	2	71	186	23	399
SRR2162953	"Turpentine," "Thai Everbearing," "Neelum," "*M. casturi* Purple", "Burma," "Amin Abrahimpur" (pooled)	Fl	64	11	1	17	1	24	20	4	66	225	17	450
SRR2165756	"Tommy Atkins"	S	8	0	0	0	0	1	7	0	8	28	4	56
SRR2736811	"Ataulfo"	Fr	36	4	0	11	0	5	8	0	19	69	18	170
SRR3192873	"Amrapali"	L	89	7	1	18	4	15	15	4	49	193	16	411
SRR3288569	"Keitt"	B	75	4	1	17	2	13	17	0	61	157	21	368
SRR3319054	"Chausa"	L	68	7	1	19	6	15	12	2	60	187	17	394
SRR3359450	"Dashehari"	Fr	38	3	1	7	0	10	7	1	19	130	14	230
SRR5966284	"Amrapali"	Fl	32	6	0	16	5	11	14	1	45	212	21	363
SRR5966306	"Amrapali"	Fl	45	1	0	16	3	11	18	1	52	246	25	418
SRR8449851	"Chokanan"	Fr	53	1	0	5	2	8	10	0	20	88	12	199
SRR8449858	"Golden Phoenix" and "Water Lily" (pooled)	Fr	103	10	2	19	6	23	15	2	56	168	12	416
SRR8926025	Mango cv. 1243	R; L; St, B, Fl; Fr	29	0	0	8	2	3	7	1	31	113	8	202
Total			1293	119	16	279	62	285	288	26	1081	3570	372	**7,391**
De-duplicated RGAs			53	17	2	29	4	27	17	6	158	362	72	**747**

Fr – fruit; Fl – flower, L – leaf, S – seed; B – buds; St – stem; R – roots

trimmed transcriptome FASTQ file was independently assembled *de novo* using Trinity pipeline (Grabherr *et al.* 2011) at default parameters. Upon assembly of the contigs, SuperTranscripts were then generated to provide an accurate representation of each expressed genes in the transcriptome assembly (Davidson *et al.* 2017). TransDecoder (Haas *et al.* 2013) was subsequently used to construct a polypeptide sequence database from the coding region of the SuperTranscripts based on nucleotide composition and open reading frame length with default parameters. Also, the collection of predicted gene models from the mango WGS was accessed directly from Wang *et al.* (2020) for further analysis.

RGA Identification and Classification

Mango transcriptome- and genome-wide RGA candidates belonging to NBS and TM-CC containing proteins, and membrane associated RLK and RLP families were identified in the generated gene models from genome annotation using RGAugury (Li *et al.* 2016), an automated RGA prediction pipeline. The input protein sequences were initially filtered using BLASTp search against RGAdb database integrated into the pipeline using an *e*-value cut-off of $1e^{-5}$. The domain and motif of the initial set of candidate RGAs were detected using nCoils (Lupas *et al.* 1991), phobius (Käll *et al.* 2004), pfam_scan (Finn *et al.* 2010), and InterProScan 5 (Zdobnov and Apweiler 2001) implemented within the RGAugury pipeline. A core set of RGAs from the transcriptome assembly was generated by concatenation and clustering (at 90% identity) of the identified RGAs from each polypeptide database using CD-HIT (Fu *et al.* 2012) and de-duplication based on BLAST description.

RGA Characterization and Annotation

GO annotations from the three domains of molecular function (MF), biological processes (BP), and cellular component (CC) were assigned to each protein represented in the identified expressed and genome-wide RGAs using BLAST2GO package (Conesa *et al.* 2005). The homology of the predicted polypeptide sequences of each RGAs to existing entries in UniProtKB/SwissProt protein database were determined using BLASTp (with *e*-value of $1e^{-5}$). The mapped BLAST hits were then merged to InterProScan (Zdobnov and Apweiler 2001) search output to produce the GO annotations.

Profiling of Mango RGA Expression

RNA-seq by expectation maximization (RSEM) (Li and Dewey 2011) was used to provide insights on the expression profile of each RGA obtained from different tissues and diverse mango varieties. The pre-processed RNA-seq reads were independently mapped to the constructed core mango RGA reference transcriptome using Bowtie2 (Langmead and Salzberg 2012). Transcripts per million (TPM) count values were then generated by RSEM from each paired-end reads. A csv (comma-separated value) file containing the RGAs and TPM values for each transcriptome data set was prepared and uploaded to Heatmapper (Babicki *et al.* 2016) for heat-map visualization of the RGA expression profiles.

Evolutionary Analysis

The FASTA amino acid sequences of the expressed and genome-wide mango RGAs were used to construct the phylogenetic trees. Multiple sequence alignment of the RGA sequences was performed using the CLUSTALW program (Thompson *et al.* 1994) with the following parameters: Gap Opening Penalty: 10; Gap Extension Penalty: 0.2. The phylogeny of these aligned sequences was reconstructed using the maximum likelihood statistical method using IQ-TREE (Nguyen *et al.* 2015) with best-fit substitution model selected based on Bayesian information criterion (BIC) through ModelFinder (Kalyaanamoorthy *et al.* 2017). The phylogenetic tree for expressed RGAs was generated using general "variable time" (VT) matrix (Müller and Vingron 2001) with empirical amino acid frequencies (+F) and FreeRate (+R5) rate heterogeneity across sites (Yang 1995; Soubrier *et al.* 2012). For the genome-wide RGAs, the tree was generated using general matrix (JTT) (Jones *et al.* 1992) with empirical amino acid frequencies (+F) and discrete Gamma (+G4) rate heterogeneity across sites (Yang 1994). The resulting phylogenetic trees were validated with 1,000 replicates of ultrafast bootstrapping (Hoang *et al.* 2018) and visualized using FigTree (v1.4.4) (Rambaut 2018).

Identification and Design of EST-SSRs

The FASTA file corresponding to the expressed RGA transcripts for analysis was uploaded into GMATA (Genome-wide Microsatellite Analyzing Toward Application) Software Package (Wang and Wang 2016). SSR (simple sequence repeat) loci within the data set were identified using default parameters [Min-length (nt) = 2; Max-length (nt) = 6; Min. repeat-times = 5]. Consequently, oligonucleotide primer pairs were designed at regions flanking the identified SSRs using the Primer3 (Untergasser *et al.* 2012; Koressaar and Remm 2007) integrated within the package. Default parameters were also used in primer design [Min. amplicon size (bp) = 120; Max. amplicon size (bp) = 400; Optimal annealing T_m (°C) = 60; Flanking sequence length = 400; Max. template length = 2000].

Orthologous RGA Gene Analysis

The genome-wide mango RGA protein sequences were compared against the RGA sequences from other sequenced plant species to provide further insights on the shared orthologs and gene duplication events of these RGAs. Gene models from the sequenced genomes of *Carica papaya* (ASGPBv0.4), *Musa acuminata* (v1), *Solanum lycopersicum* (iTAG2.40), *Theobroma cacao* (v1.1), *Zea mays* (Ensemble18), *Oryza sativa* (v7 JGI), *Arabidopsis thaliana* (TAIR10), *Citrus sinensis* (v1.1), and *Prunus persica* (V2.1) were downloaded from Phytozome (https://phytozome.jgi.doe.gov/). The RGAs from these plant species were identified using RGAugury pipeline (Li *et al.* 2016) at BLASTp *e*-value $1e^{-5}$. All of the identified RGAs from the ten organisms (including mango) were used as an input to OrthoFinder2 (Emms and Kelly 2019) at default settings. Rooted species tree for the analyzed plant species was constructed using Species Tree Inference from All Genes (STAG; Emms and Kelly 2018) and Species Tree Root Inference from Gene Duplication Events (STRIDE; Emms and Kelly 2017) algorithms while the gene duplication events were identified using the duplication-loss-coalescent model (Wu *et al.* 2014). The UpsetR plot was then generated using Intervene (Khan and Mathelier 2017) while the rooted species tree was constructed using FigTree (v1.4.4) (Rambaut 2018).

RESULTS AND DISCUSSION

Identification of Mango RGAs

Raw transcriptome sequences from diverse mango cultivars (Table 1) and gene models from the WGS of mango (Alphonso variety) (Wang *et al.* 2020) were retrieved and the RGAs were identified using RGAugury, an integrative NGS bioinformatics pipeline that efficiently predicts RGAs in plants (Li *et al.* 2016). In general, this pipeline works by identifying first the domains and motifs related to RGAs such as NBS, LRR, serine/threonine and tyrosine kinase (STTK), CC, LysM, TM, and TIR. Based on the presence of combinations of these domains and motifs, the RGA candidates are then identified and categorized into four major families: NBS-encoding, TM-CC, and membrane associated RLP and RLK.

For the transcriptome-wide, expressed RGAs, the number of proteins predicted from the sequence reads archive (SRA) files ranged from 56–454 (Table 1). In total, 7,391 RGAs were identified from the assembled transcriptome sequences (Table 1; Supplemental File 1). These RGAs were further clustered at 90% homology creating 2,892 clusters, followed by the automated selection of a representative protein sequence per cluster. The representative protein sequences were described using BLAST and BLAST2GO, and proteins with duplicated BLAST descriptions were removed retaining only unique/non-redundant protein type per RGA domain to generate 747 core RGAs with unique functions (Table 1; Supplemental Table 1; Supplemental File 2a). The expressed core RGA proteins were classified as follows: 53 NBS, 27 NBS-LRR, 17 CNL, two TNL, 29 CN, four TN, 17 TX, 158 RLP, 362 RLK, 72 TM-CC, and six NBS-encoding proteins with other domains (Table 1). For the genome-wide RGAs, a total of 1,775 proteins were predicted and were classified as follows: 54 NBS, 107 NBS-LRR, 242 CNL, 79 TNL, 78 CN, 30 TN, 45 TX, 133 RLP, 917 RLK, 83 TM-CC, and seven NBS-encoding proteins with other domains (Table 2; Supplemental Table 1; Supplemental File 2b).

Upon comparison of genome-wide RGAs of mango to other plant species, it appears that mango harbors more RGAs (1,775) than rice (1,537), cacao (1,171), *Arabidopsis* (979), corn (935), tomato (922), banana (769), and papaya (402) but has lower compared to orange (1,806) and peach (2,005) (Table 2). Eventually, the RGA content and characteristics of a plant have been frequently associated with plant resistance (Sekhwal *et al.* 2015). Like most crops, RLKs constitute the largest group of RGAs in mango comprising about 52% of the RGAs characterized, followed by NBS-encoding proteins, then RLPs (Table 2, Figure 1). In dicots such as mango, all NBS-encoding proteins are present (*i.e.* NBS, CNL, TNL, CN, TN, NL, TX, and other NBS-encoding proteins) (Table 2) but in monocots such as corn, rice, and banana, the TNL protein is usually absent (Table 2) (Tarr and Alexander 2009). It is hypothesized that after the divergence of dicots and monocots, the TNL genes might have been lost from the monocot lineage (Zhang *et al.* 2016). There were six expressed and seven genome-wide putative RGAs that showed chimeric domain/motif architecture and were classified as "other" NBS-encoding proteins (Tables 1 and 2). These other RGAs are described to have an unexpected domain combination of both TIR and CC domains (Li *et al.* 2016).

Mango Resistance (R)/ Defense Proteins and Expression Profiles

BLAST analysis of the expressed and genome-wide RGAs of mango revealed homology to a number of well-known R/defense proteins against pathogens and insects (Supplemental Table 1). Among these R proteins include RRS proteins (putative WRKY transcription factors), which confer resistance against *Colletotrichum higginsianum* and *Ralstonia solanacearum* (Narusaka *et al.* 2009; Saucet *et al.* 2015); RPP proteins (including At4g19530 and At4g19520 proteins), which confer resistance against downy mildew (*Peronospora parasitica*) (Parker *et al.* 1997; Sinapidou *et al.* 2004;

Table 2. Genome-wide distribution of RGAs in mango and its comparison to other sequenced plant genomes.

RGA codes		Mango (*Mangifera indica*)	Cacao (*Theobroma cacao*)	Papaya (*Carica papaya*)	Orange (*Citrus sinensis*)	Peach (*Prunus persica*)	Tomato (*Solanum lycopersicum*)	Arabidopsis (*Arabidopsis thaliana*)	Corn (*Zea mays*)	Rice (*Oryza sativa*)	Banana (*Musa acuminata*)
NBS	NBS	54	18	10	60	45	58	9	16	45	8
CC-NBS-LRR	CNL	242	187	10	158	185	74	48	69	223	50
TIR-NBS-LRR	TNL	79	15	4	31	157	23	98	0	0	0
CC-NBS	CN	78	19	5	65	5	5	1	12	37	7
TIR-NBS	TN	30	3	1	31	19	8	16	3	3	2
NBS-LRR	NL	107	67	22	189	183	93	29	52	197	43
TIR with unknown domain	TX	45	7	5	49	44	9	40	3	3	3
Other NBS-encoding	Other	7	0	3	7	30	1	17	0	0	0
RLP	RLP	133	159	25	215	176	77	65	35	102	95
RLK	RLK	917	634	282	895	1,040	507	591	649	856	489
TM-CC	TM-CC	83	62	35	106	121	67	65	96	71	72
Total RGAs		**1,775**	**1,171**	**402**	**1,806**	**2,005**	**922**	**979**	**935**	**1,537**	**769**

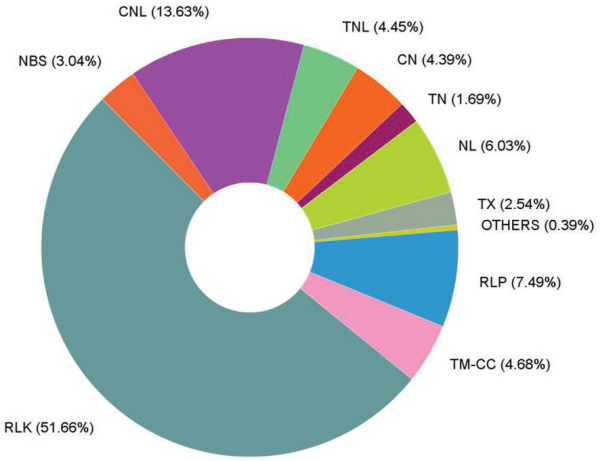

Figure 1. Distribution of the classifications based on conserved domains and motifs of genome-wide RGAs of mango.

Bittner-Eddy *et al.* 2000; Wan *et al.* 2019); the LRK10/Lr10 proteins (all of which are of RLK domains), which provide resistance against leaf rust (*Puccinia triticina*) (Shiu and Bleecker 2003; Feuillet *et al.* 1998); the RGA/RGA-blb and R1A-10 proteins, which provide resistance against the devastating late blight disease (*Phytophtora infestans*) (Lokossou *et al.* 2010; Kuang *et al.* 2005); the RLM proteins, which confer resistance against the fungal pathogen *Leptosphaeria maculans* (Staal *et al.* 2006, 2008); the Pik-2 and SasRGA5 (RGA5) proteins, which confer resistance against blast disease (*Magnaporthe oryzae*) (Zhai *et al.* 2011; Okuyama *et al.* 2011); and many homologs of TMV (tobacco mosaic virus) resistance protein N, which are responsible for TMV resistance (Dinesh-Kumar *et al.* 2000). The mango RGAs also contained numerous sequences with homology to various RPS proteins, including RPM1 and TAO1 proteins, which are important R proteins against the bacterium *Pseudomonas syringae*, a widely studied biotrophic pathogen (Eitas *et al.* 2008; Mackey *et al.* 2002; Warren *et al.* 1998; Kim *et al.* 2009).

RGA gene models showed homology to ERECTA protein, which confers quantitative resistance against the necrotrophic fungus *Plectosphaerella*, and bacterial wilt caused by *Ralstonia solanacearum* (Sánchez-Rodríguez *et al.* 2009). This R protein is also responsible for the regulation of efficient transpiration in plants (Masle *et al.* 2005). Homology to the R protein EDR1 (ENHANCED DISEASE RESISTANCE 1), a MAPKKK serine/threonine-protein kinase, was also found which confer *non*-host resistance against *Colletotrichum gloeosporioides* in *Arabidopsis thaliana* through the induction of plant defensins expression (Hiruma *et al.* 2011). Mango RGA showed homology to the R protein ADR2 (ACTIVATED DISEASE RESISTANCE 2), an NBS-encoding protein

that confers broad-spectrum resistance against many pathogens such as strains of *Pseudomonas syringae* and *Hyaloperonospora arabidopsis*, and full immunity to several races of the pathogen *Albugo candida*, the causal agent of white rust disease (Borhan *et al.* 2008; Aboul-Soud *et al.* 2009). This R protein is also reported to play a role in the response to UV stress (Piofczyk *et al.* 2015). Predicted mango RGAs homologous to the R-like proteins DSC 1 and 2 (DOMINANT SUPPRESSOR OF CAMTA3) were also found, which act as a guard of CAMTA3, a negative regulator of immunity, during infection of pathogens (Lolle *et al.* 2017). NBS-encoding RGAs in mango showed homology to SUMM2 (SUPPRESSOR OF *mkk1 mkk2* 2), which acts as an R protein that becomes activated when the MEKK1-MKK1/MKK2-MPK4 cascade in the basal defense response is disrupted by the pathogen effector HopAI1 (Zhang *et al.* 2012; Kong *et al.* 2012).

Other putative resistance proteins abundantly found in mango showed homology to various DRL-coded (Uniprot/SwissProt ID) potential R proteins from *Arabidopsis thaliana* which have not been fully studied (Supplemental Table 1). Homologs to CHS and CHL proteins were also found in mango RGAs, which are important R proteins against chilling/low temperature and also confer resistance against bacterial infection such as *Pseudomonas syringae* (Zbierzak *et al.* 2013). Aside from resistance to pathogens, R/defense proteins against insects were also identified by filtering the GO results of expressed RGAs with GO term "defense response to insect" (GO:0002213) (Supplemental Table 1). The RGAs found include the At4g11170 putative resistance protein, RPP5, DSC1 and 2, TRANSPORT INHIBITOR RESPONSE 1 (TIR1), and CORONATINE INSENSITIVE-1 (COI-1). TIR1 is an F-box protein and a major receptor of the plant hormone auxin and mediate in the auxin/indole acetic acid regulated signaling pathways involved in plant growth and development, as well as in the defense response against pathogens and insect pests (Dharmasiri *et al.* 2005). COI-1 is a jasmonic acid – isoleucine (JA-Ile) coreceptor and also an F-box protein involved in the JA signaling pathway, which regulates plant defense against insects and pathogens, wound healing, and other vegetative and reproductive developmental functions (Xie *et al.* 1998). The plant hormone JA is conjugated to isoleucine as a consequence of insect/pathogen attack, and JA-Ile binding with COI-1 stimulates degradation of JAZ coreceptor repressor proteins, thereby promoting the expression of JA-responsive genes (Pieterse *et al.* 2012). Further filtering of expressed RGAs containing the GO term "response to insect" (GO:0009625) revealed 17 RGAs that putatively play vital roles in the defense pathways against insects, all of which were characterized to possess RLK/RLP domains. The complete and detailed BLASTp analysis and description of the RGAs are provided in Supplemental Table 1.

Of all the mango varieties analyzed in this study, the three varieties that showed the lowest expression of RGAs were SRR8449851 (Chokanan), SRR3359450 (Dashehari), and SRR1562177 (Kent) (Figure 2; Supplemental Table 2). On the other hand, the following showed the highest expressions of the RGAs characterized: SRR2162953 (pooled RNA from different varieties), SRR2162919 (Thai Everbearing), SRR2162878 (Burma), SRR2162889 (*M. casturi* "purple"), SRR2159471 (Tommy Atkins), SRR2162929 (Turpentine), SRR2162836 (Amin Abrahimpur), and SRR2162908 (Neelum) (Figure 2; Supplemental Table 2). The relatively high expression of RGAs in these RNA-seq reads could be explained by the fact that they are pooled RNA samples from different varieties or from different plant tissues.

GO and Functional Annotation of Mango RGAs

GO analysis and functional annotation of the expressed and genome-wide mango RGAs provided a broader knowledge of the major roles of these proteins and their cellular localizations. The MFs of the RGAs are primarily associated with protein and nucleotide binding, and kinase activity (Figure 3a) as they are known extra- and intracellular binding receptors and relays defense signaling in the cells through a cascade of kinase activities (Rodriguez *et al.* 2010). The absence of an intracellular kinase in RLPs suggests that it relies on/ interacts with other domains such as RLK-type receptors to relay the signaling from the extracellular matrix to the intracellular component (Tör *et al.* 2009). Meanwhile, the BPs of the RGAs – especially in the expressed RGAs – are highly diverse, distributing the RGA sequences to a wide array of GO IDs/terms (Supplemental Table 3). The RGAs are mainly involved in protein autophosphorylation during signal transductions in the cell, defense/immune responses to different biotic (insect pests and diseases) and abiotic (water deprivation, ozone, UV stress, *etc.*) stresses, and in various developmental processes (*i.e.* from embryonic to floral/pollen development) (Supplemental Table 3; Figure 3b). Numerous RGAs were also described to be involved in hormone-mediated signaling pathways and systemic acquired resistance which include the plant hormones brassinosteroids, abscisic acid (ABA), salicylic acid (SA), auxin, jasmonic acid, ethylene, and gibberellic acid (GA) (Supplemental Table 3). These plant hormones create crosstalk to modulate defense signaling pathways and activate systemic resistance against a broad spectrum of pathogens and insect pests. As an example of this crosstalk, SA antagonizes JA to activate the immune response against biotrophic pathogens (Williams 2011). Likewise, JA antagonizes SA to activate two branches of the immune response, which also antagonizes each other: the ERF branch against necrotrophic pathogens

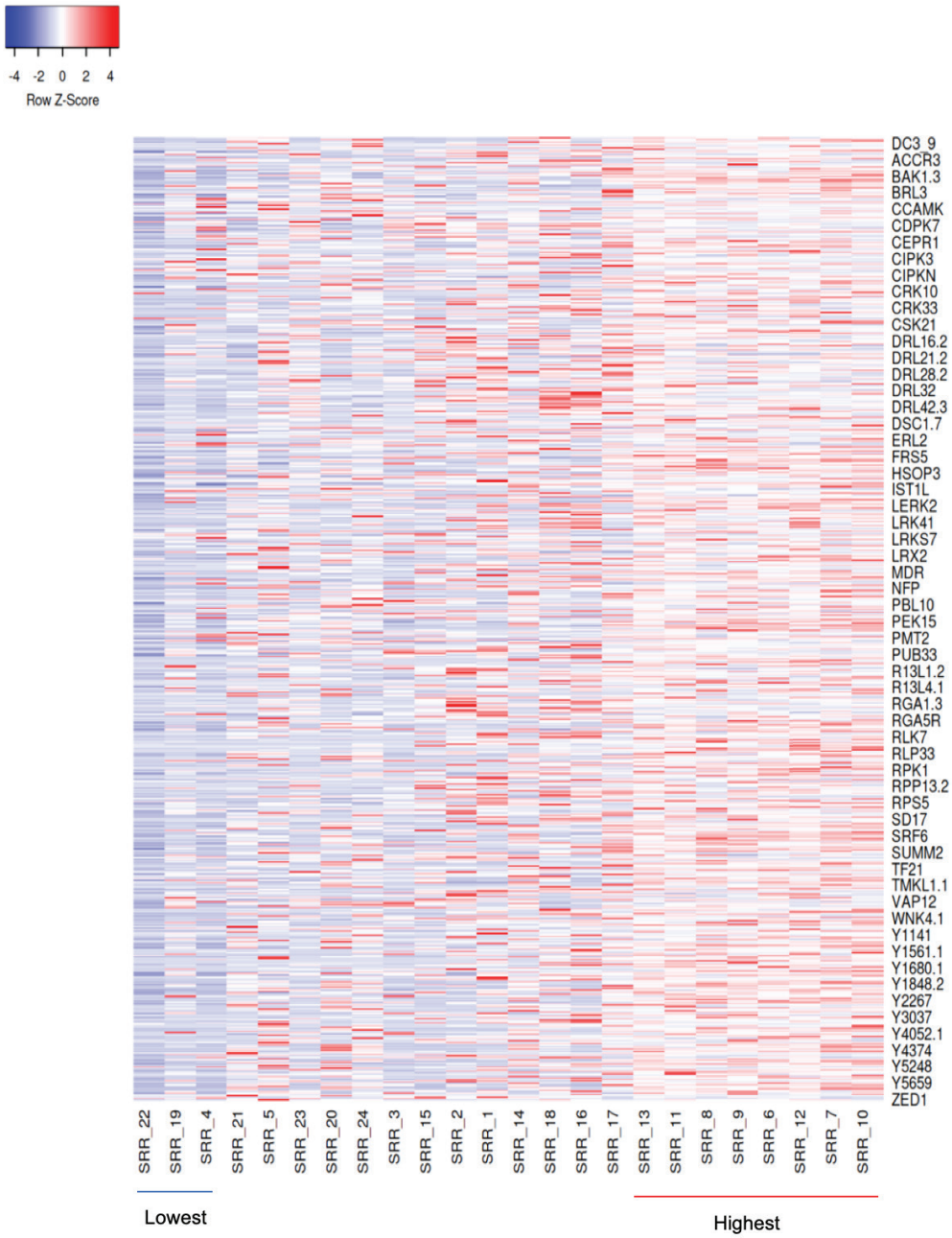

Figure 2. RGA expression profile from mango transcriptomes visualized as a heatmap. Count values were normalized within samples as TPM using RSEM. Y-axis corresponds to RGA gene IDs. X-axis corresponds to RNA-seq raw read samples: SRR_1. Hong *et al.* (2016) (ZILL); 2. SRR1448302 (KEITT); 3. SRR1561197 (KENT); 4. SRR1562177 (KENT); 5. SRR1956775 (AMRAPALI); 6. SRR2159471 (TOMMY ATKINS); 7. SRR2162836 (AMIN ABRAHIMPUR); 8. SRR2162878 (BURMA); 9. SRR2162889 (M. CASTURI "PURPLE"); 10. SRR2162908 (NEELUM); 11. SRR2162919 (THAI EVERBEARING); 12. SRR2162929 (TURPENTINE); 13. SRR2162953 (Pooled – TURPENTINE, THAI EVERBEARING, NEELUM, M. CASTURI "PURPLE", BURMA, AMIN ABRAHIMPUR); 14. SRR2165756 (TOMMY ATKINS); 15. SRR2736811 (ATAULFO); 16. SRR3192873 (AMRAPALI); 17. SRR3288569 (KEITT); 18. SRR3319054 (CHAUSA); 19. SRR3359450 (DASHEHARI); 20. SRR5966284 (AMRAPALI); 21. SRR5966306 (AMRAPALI); 22. SRR8449851 (CHOKANAN); 23. SRR8449858 (GOLDEN PHOENIX and WATER LILY); 24. SRR8926025 (MANGO cv. 1243)

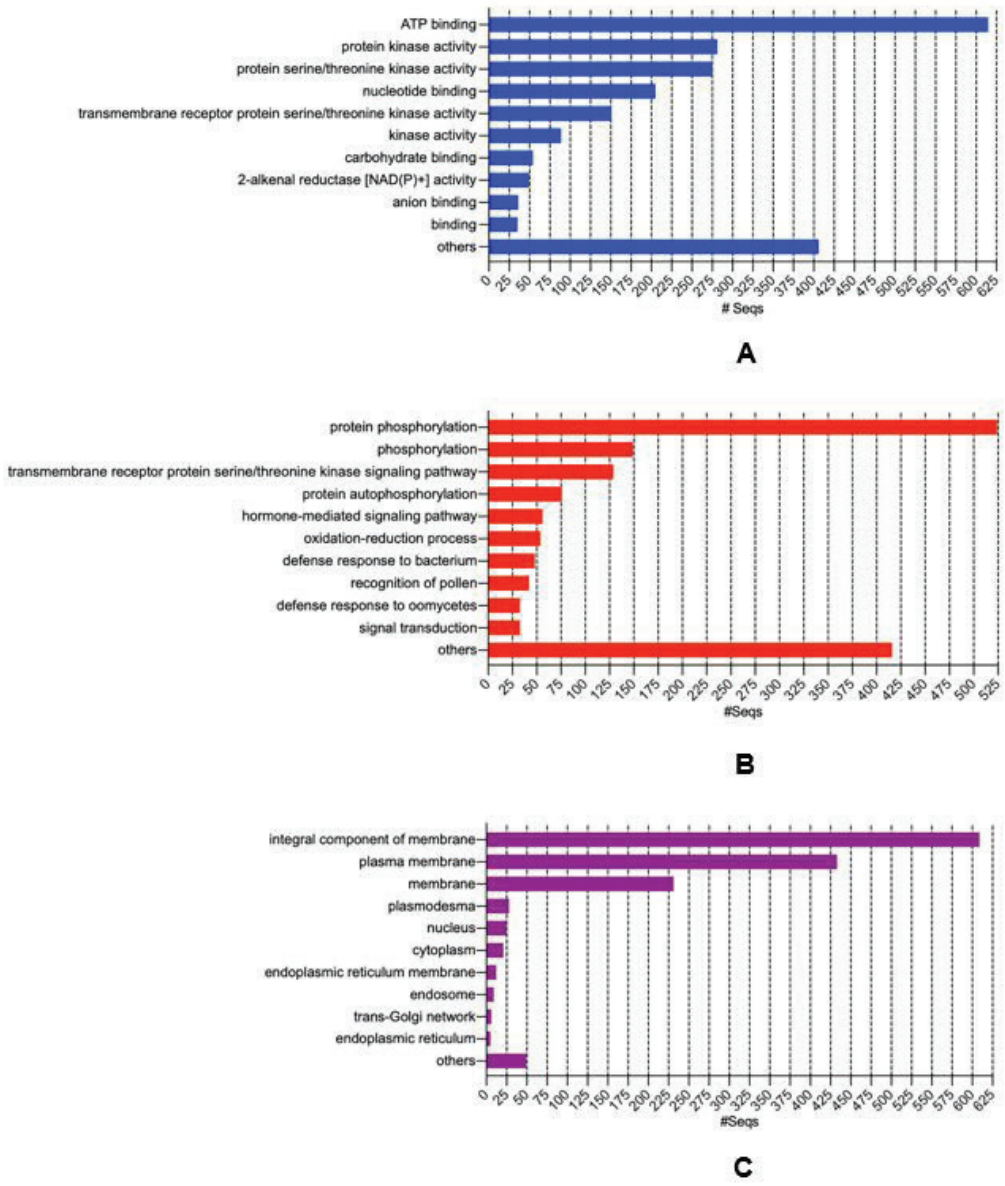

Figure 3. GO distribution of genome-wide RGAs of mango based on MF (A), BPs (B), and CC (C).

(synergistic interaction with ethylene) and the MYC branch against insects (synergistic interaction with ABA) (Pieterse *et al.* 2012). In the cell, the majority of the mango RGAs are localized in the cell/ plasma membrane, plasmodesma, cytosol/ cytoplasm, and nucleus (Figure 3c) which are primary sites for recognition of pathogen/insect invasion and effector proteins. In these CCs, the RGAs functions to convert extracellular stimuli into intracellular responses to activate defense cascades and counteract the attack. In general, the results of GO analysis and functional annotation show that plant immune response is a consequence of the activity of a wide range of plant hormones, resistance/defense-related proteins, biochemical and developmental processes, *etc.* that aim to suppress pest attack, thereby improving plant defense. The complete and detailed GO analysis and annotation of each RGAs (expressed and genome-wide RGAs) are provided in Supplemental Tables 1 and 3.

Evolutionary Relationships Among Mango RGAs

Using the evolutionary best-fit model selected according to BIC (Supplemental Files 3a and 3b), maximum likelihood phylogenetic trees were constructed using the 747 (Figure 4a; Supplemental File 4a) and 1,775 (Figure 4b; Supplemental File 4b) mango RGA proteins from transcriptomic and genomic data, respectively, to investigate their evolutionary relationships and diversity.

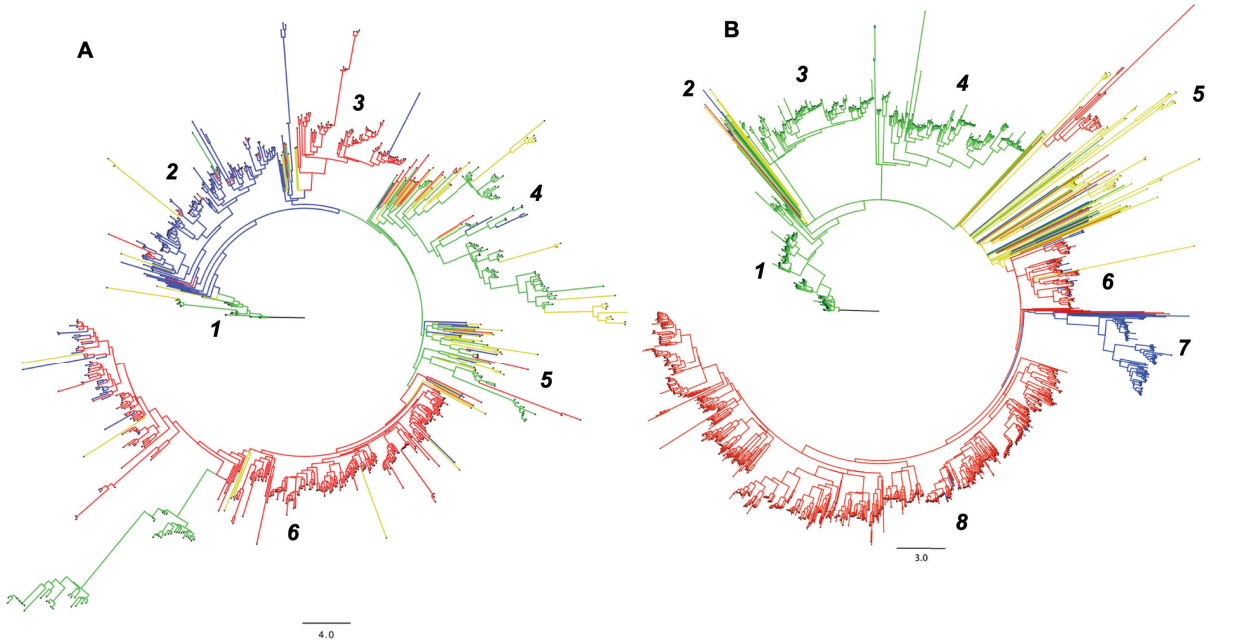

Figure 4. Maximum likelihood phylogenetic trees constructed using IQ-TREE from the sequence alignment of RGAs generated from *de novo* transcriptome assemblies (A) and WGS (B) of mango. Best-fit model was selected according to BIC using ModelFinder (Kalyaanamoorthy *et al.* 2017). Tree A was constructed using General VT matrix (Muüeller and Vingron 2004) with empirical amino acid frequencies (+F) and FreeRate (+R5) rate heterogeneity across sites (Yang 1995; Soubrier *et al.* 2012). Tree B was constructed using General Matrix (JTT) (Jones *et al.* 1992) with empirical amino acid frequencies (+F) and discrete Gamma (+G4) rate heterogeneity across sites (Yang 1994). The trees were tested with 1,000 replicates of ultrafast bootstrapping (Hoang *et al.* 2018). The numbers correspond to the clades.

It showed that the clustering of RGAs based on domain classification is very apparent but mingled with subclades from other RGA domains. As shown in the trees, the mango RGAs have become highly diversified and are broadly clustered into six (for expressed RGAs) and eight (for genome-wide RGAs) major clades, both with many subclades (Figures 4a and 4b, respectively).

Expressed RGAs having TM-CC domain didn't emerge as a separate cluster but formed subclades all across the tree suggesting that it shares similarity to all other RGA classes (Figure 4a). The clade 1 is composed mainly of NBS-encoding proteins and other domains such as TM-CC and RLP. Notably, the clade 1 appeared to be the most distantly related RGA group among other clades. The clade 2 is predominantly composed of RLPs with subclades from all other RGA domains. The clade 3 contains several RLPs and, most especially, numerous RLK proteins that formed a major nested subclade, indicating their close relationship with RLPs. The clades 4 and 5 are composed mainly of NBS-encoding proteins with subclades from all other RGA domains and, most especially, from TM-CC domain. The clade 6 is composed predominantly of RLKs and shared close relationships with some other RGAs. In this clade, NBS-encoding proteins that shared close similarity related to RLKs were observed to form a major nested subclade.

As expected, genome-wide RGAs exhibited a more structured grouping of domains (Figure 4b). Unlike in the expressed RGAs, genome-wide TM-CC proteins formed two major clades (clades 2 and 5) with subclades from all other domains. TM-CC proteins in clade 2 appeared to share close relationships with NBS-encoding proteins. Similar to expressed RGAs, the genome-wide NBS-encoding proteins also formed 3 major clades found in clades 1, 3, and 4. The clade 2 (predominantly composed of TM-CC proteins) and clades 3 and 4 (predominantly NBS-encoding proteins) appeared to be sister clades. The RLK proteins are primarily grouped in clades 6 and 8, with a major nested subclade in clade 5 that shared a close relationship mainly with TM-CC proteins. In these clades, RLPs also formed subclades; however, these proteins are predominantly found in clade 7, which includes some RLK proteins that share close relationships with RLPs. This clade appeared to be a sister clade of clade 8, which is composed predominantly of RLK proteins.

As observed also by Rody *et al.* (2019) in sugarcane, the TM-CC proteins can form widespread subclades across the phylogenetic tree and may show close relationships with NBS-encoding proteins. The distribution of RGAs in the phylogenetic trees may indicate that their functions

are highly divergent and form clusters that may be related to functions but not necessarily in the protein sequence (Chang *et al.* 2002; Zhang *et al.* 2016). This could be the case in mango RGAs, especially the NBS-encoding proteins and RLKs, which are broadly distributed into many major clades and major nested subclades. The NBS and RLKs are among the largest and diversified groups of RGAs in plants (Hulbert *et al.* 2001; Meyers *et al.* 2003; Tör *et al.* 2009; Li *et al.* 2016), which also holds true in mango based on this study.

Many factors have been attributed to the evolutionary pattern and diversity of RGAs in plants. One of which is the well-known co-evolutionary "arms race" model between the host plant and associated pests and diseases, which drives the selective forces to overcome each other (Anderson *et al.* 2010; Edger *et al.* 2015; Boller and He 2009). Other selection pressures such as climatic conditions (*e.g.* rainfall, temperatures, humidity, *etc.*) that favors pest and disease development have been correlated also to the diversity of RGAs (Sela *et al.* 2009). Whole-genome duplications (WGDs) and genomic reorganizations that occurred during ancient times, especially in angiosperms, have been associated with the expansion of RGA families and in creating new gene functions (Perazzolli *et al.* 2014; Chagné 2015; Michelmore and Meyers 1998). It is likely that in the evolutionary history of mango, it had undergone at least one round of polyploidization, *i.e.* allopolyploidy, which is a type of WGD *via* hybridization followed by genome doubling (Glover *et al.* 2016; Soltis and Soltis 2009). These are some factors that could have influenced the complex phylogenetic structure of RGAs of mango, given also that the plant's tropical growing condition provides a thriving environment to a plethora of insect pests and diseases.

RGA-derived EST-SSRs of Mango

Using the RGA transcripts (Supplemental File 5), a total of 151 EST-SSR loci were identified. Among these, 134 loci yielded unique potential markers. Mostly di-repeat motifs (2-mer) were observed, accounting for 64.24% of the markers, followed by tri-repeats (3-mer) at 30.46%. The remaining 5.3% is composed of tetra-, hexa-, and penta-repeat motifs. AT repeat is the most represented motif among the di-repeats at 12.5%, followed by TC and TA at 11.26% each (Figure 5a). Meanwhile, AG/CT is the most abundant paired repeat motif among the SSR loci analyzed at 16.56% (Figure 5b). Data on the distribution of RGA-linked SSRs (Figure 5c) revealed that RLK has the greatest number of EST-SSRs designed with 67 markers, followed by RLP and TM-CC motifs at 34 and 18 markers, respectively. This distribution corresponds directly to the number of RGAs per category identified in the whole mango RGA transcriptome.

Similar *in silico* approaches in EST-SSR marker development were performed by other studies focusing on other plants, such as bamboo (Cai *et al.* 2019), mint

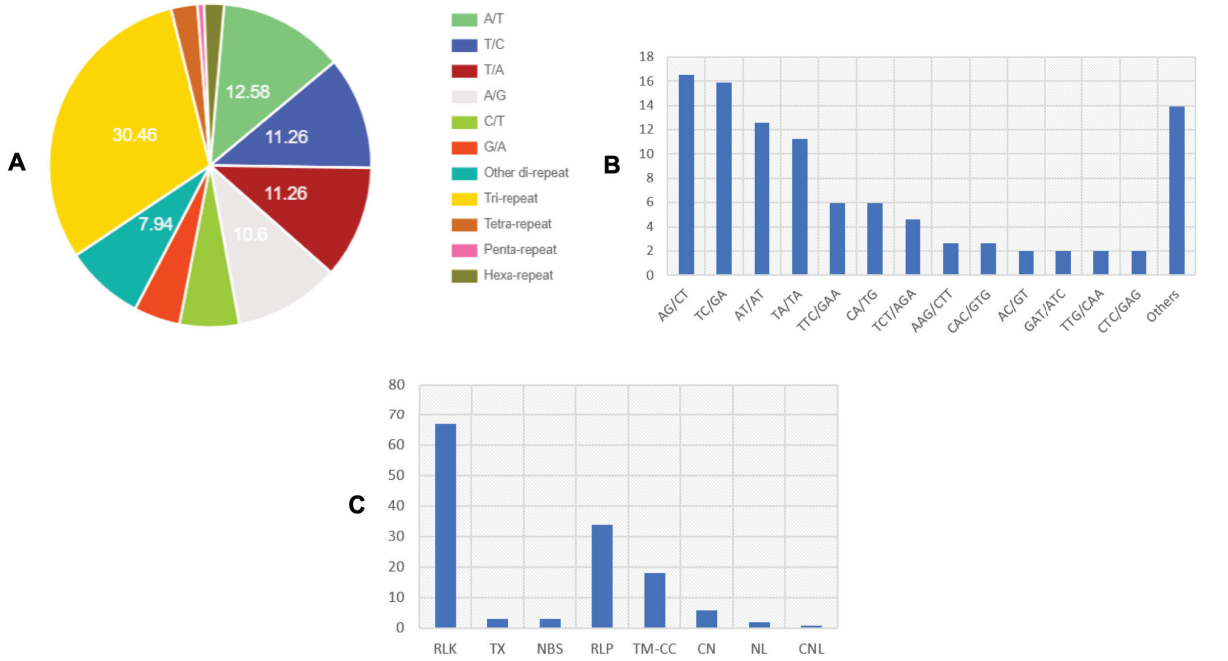

Figure 5. Distribution of the different repeat motifs that were detected in the mango RGA transcriptome (A). Paired-repeat SSR motif distribution of the mango RGA transcriptome (B). Distribution of EST-SSR markers designed across the different RGA domains identified in *de novo* mango transcriptomes (C).

(Kumar *et al.* 2015), castor (Thatikunta *et al.* 2016), Ethiopian potato (Gadissa *et al.* 2018), and *Elymus* species (Zhang *et al.* 2019). The EST-SSRs possess high-transferability (Liu *et al.* 2016) and, thus, may be used to different mango cultivars and other related species. Recently, SSRs or microsatellites have been increasingly found and characterized within protein-coding genes (*i.e.* in the transcriptome) and thus, with their repeat motif variations, can potentially influence on gene regulation, transcription, translation, protein function, and broadly, the organism's evolutionary pattern (Kashi and King 2006; Li *et al.* 2004; Hancock and Simon 2005; Liu *et al.* 2016). The transcriptomic SSR profile has been associated also to the adaptive evolution of the plant. As an example, the observed significant enrichment of SSRs in the transcriptome of *Helianthus annuus* (sunflower) was associated with biotic and abiotic stress responses (Pramod *et al.* 2014). Thus, the EST-derived SSR markers generated in this study can be used for functional and genetic analyses of *R* genes against specific biotic stress, once verified by wet-lab experiments. Moreover, these markers may also be used in genotyping, genetic diversity, linkage mapping, gene-based association studies, and marker-assisted selection in mango breeding. The sequences of the designed forward and reverse primers derived from expressed mango RGAs are found in Supplemental Table 4.

Comparative RGA and Gene Duplication Analyses

The comparative analysis of genome-wide RGAs of mango and other plant species (papaya, banana, tomato, rice, corn, cacao, orange, peach, and *Arabidopsis*) identified 968 RGA orthogroups consisting of 11,336 genes (92.2%), 456 species-specific orthogroups (2,232 genes in species-specific orthogroups; 18.1%) (Table 3). Fifty percent (50%) of all genes were in orthogroups with 23 or more genes (G50 = 23) and were contained in the largest 110 orthogroups (O50 = 110) (Table 3). There were 118 most conserved RGA families, in which all species have at least one representative gene in the orthogroups and one of these consisted entirely of single copy genes (Table 3). The single-copy ortholog is characterized as RLK, namely, probable LRR receptor-like serine/threonine-protein kinase At2g24230 (Supplemental File 6). Moreover, 10 commonly shared orthogroups (OG0000024, OG0000074, OG0000221, OG0000357, OG0000405, OG0000460, OG0000524, OG0000603, OG0000604, and OG0000605) were specifically identified in monocot species included in the analysis (rice, corn, banana) whereas 14 orthogroups (OG OG0000007, OG0000031, OG0000068, OG0000148, OG0000151, OG0000188, OG0000196, OG0000212, OG0000257, OG0000284, OG0000293, OG0000316, OG0000317, and OG0000344) were specific only to dicots (mango,

Table 3. Comparative genome-wide RGA analysis summary statistics of mango and other plant species.

Parameters	Arabidopsis	Papaya	Orange	Banana	Mango	Rice	Peach	Tomato	Cacao	Corn	Overall
Number of genes	979	402	1,806	769	1,775	1,537	2,005	922	1,171	935	12,301
Number of genes in orthogroups	922	392	1,619	709	1,664	1,352	1,792	888	1,122	876	11,336
Number of unassigned genes	57	10	187	60	111	185	213	34	49	59	965
Number of species-specific orthogroups	48	0	89	15	65	79	89	22	23	26	456
Number of genes in species-specific orthogroups	239	0	392	68	396	316	466	171	95	89	2,232
Percentage of genes in species-specific orthogroups	24.4	0	21.7	8.8	22.3	20.6	23.2	18.5	8.1	9.5	18.1
Number of orthogroups	–	–	–	–	–	–	–	–	–	–	968
Mean orthogroup size	–	–	–	–	–	–	–	–	–	–	11.7
Median orthogroup size	–	–	–	–	–	–	–	–	–	–	5
G50 (assigned genes in orthogroups)	–	–	–	–	–	–	–	–	–	–	26
G50 (all genes)	–	–	–	–	–	–	–	–	–	–	23
O50 (assigned genes in orthogroups)	–	–	–	–	–	–	–	–	–	–	90
O50 (all genes)	–	–	–	–	–	–	–	–	–	–	110
Number of orthogroups with all species present	–	–	–	–	–	–	–	–	–	–	118
Number of single-copy orthogroups	–	–	–	–	–	–	–	–	–	–	1

tomato, cacao, papaya, orange, peach, and *Arabidopsis*) (Figure 6a; Supplemental File 6). This suggests that the evolution of RGAs in these orthogroups happened after the divergence of dicot and monocot.

Among the conserved RGA gene families of the analyzed plant species, the four most frequent families are S-locus lectin protein kinase family protein (251 orthologous RGAs), family of putative disease resistance RPP13-like protein 1 (250 orthologs), family of disease resistance protein SUMM2-like (210 orthologs), and family of MDIS1-interacting receptor like kinase 2-like (167 orthologs) (Supplemental File 6). The expansion among the gene members of these top four most enriched orthogroups might have a direct implication in the evolutionary adaptation of these plant species from various selective pressures imposed by a wide array of biotic and abiotic stresses. Specifically, mango exhibited the highest gene expansion on the RGA orthogroup (OG0000002) containing RPP13-like protein 1 (with 85 orthologous genes) as compared to papaya (three orthologs), tomato (24 orthologs), and cacao (41 orthologs) (Supplemental File 6).

The highest number of orthologous genes was observed in peach (1,792 RGAs) while the lowest was observed in papaya (392 RGAs) (Table 3). Pairwise comparison of shared orthologs revealed that mango has the highest number of orthologous RGAs with orange (274 orthologs) whereas the lowest was observed in banana (187 orthologs) (Figure 6b). Mango consists of 65 species-specific RGA orthogroups (396 orthologous genes; 22.3% of total mango RGAs), which is higher as compared to that of papaya (0 orthogroup), banana (15 orthogroups), tomato (22 orthogroups), corn (26 orthogroups), *Arabidopsis* (48 orthogroups), and cacao (23 orthogroups) but lower than rice (79 orthogroups), peach (89 orthogroups), and orange (89 orthogroups) (Table 3; Figure 6a). Among these, mango species-specific orthogroups with the most RGA members are probable disease resistance protein At4g27220 (OG000041; 50 orthologs) and disease resistance-like protein DSC1 (OG0000057; 38 orthologs) (Supplemental Files 6 and 7). The lineage-specific occurrence and expansion of these RGA orthogroups in mango suggest its possible role in its defense and immunity responses against various biotic stresses.

One of the major drives in the expansion of the *R* gene families in major crops is the occurrence of gene duplication events during the course of evolution (Perazzolli *et al.* 2014; Chagné 2015; Michelmore and Meyers 1998). Here, gene duplication events among the orthogroups were also investigated to provide further insights on the evolutionary dynamics of mango RGAs in relation to other plant species (Figure 7). Based on the analysis, mango RGAs had undergone 1,005 gene duplication events, which was lower than that of peach (1,157 gene duplications) but higher than rice, corn, cacao, tomato, papaya, orange, banana, and *Arabidopsis*. Furthermore, mango shares 26 gene duplication events with orange while 16 in cacao and papaya. Among the mango-specific gene duplication events, orthogroups OG0000002 and OG0000057 were observed to constitute the most RGA members arising from a common ancestral gene (Supplemental Table 5). As previously demonstrated in other plants, RGA duplication events are clearly their adaptive strategy due to the functional importance of these genes in triggering proper host defense response (Rizzon *et al.* 2006; Yu *et al.* 2015). Also, it is believed that the amplification of *R* genes is an indication of the

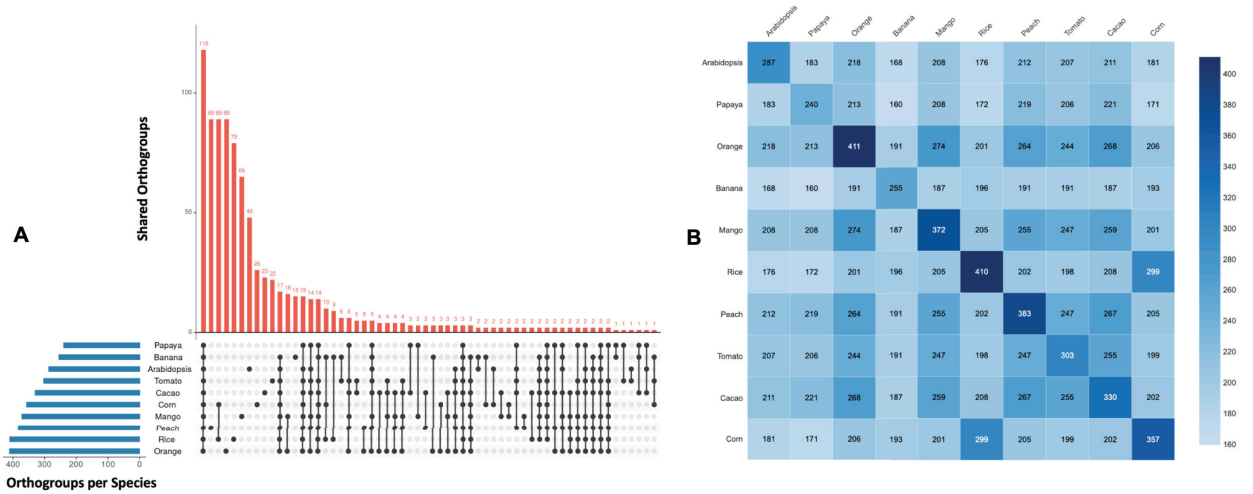

Figure 6. UpSetR plot showing the intersection of RGA orthogroups across different plant species analyzed (A). Heatmap showing the pairwise species overlaps of the RGA orthologs (B).

adaptive value of a particular species, thus contributing to its increased fitness (Holub 2001; Panchy *et al.* 2016). Thus, the lineage-specific amplification of the mango-specific orthogroups containing RPP13-like protein 1 and disease resistance-like protein DSC1, which arise from gene duplication events might have played a role in the evolutionary adaptation of mango during the course of selective pressure caused by various biotic stresses.

SUMMARY AND CONCLUSION

In this study, transcriptomic data (from diverse mango varieties) and genomic data (from the Alphonso variety) were utilized in order to obtain a comprehensive analysis of mango RGAs. From the transcriptomic assembly, a core (deduplicated) mango RGA database with 747 protein models was established. Meanwhile, 1,775 RGAs were identified in the mango WGS and broadly categorized into four major families based on their conserved structural features, *i.e.* 642 NBS-encoding proteins, 133 RLPs, 917 RLK proteins, and 83 TM-CC proteins. In addition, the mango RGAs have been functionally well-annotated revealing the involved BPs, MFs, and CCs. This provides an important insight to the overall functional response of mango against insect pests and diseases. Moreover, the evolutionary relationships, diversity, and expression profiles across different mango varieties were presented as an invaluable reference in the design of an effective framework for the actual mango resistance breeding programs. A marker resource, comprising of 134 RGA-linked EST-SSR markers, was also developed to aid in the discovery of associated RGA/s to a specific disease or insect pest through population genetics analyses (*e.g.* QTL mapping, association mapping). The evolutionary dynamics of mango RGAs were also investigated in comparison with the identified RGAs from other important plant species. The comparative RGA analysis revealed that mango has 65 species-specific RGA families and 1,005 RGA gene duplication events, and shares most

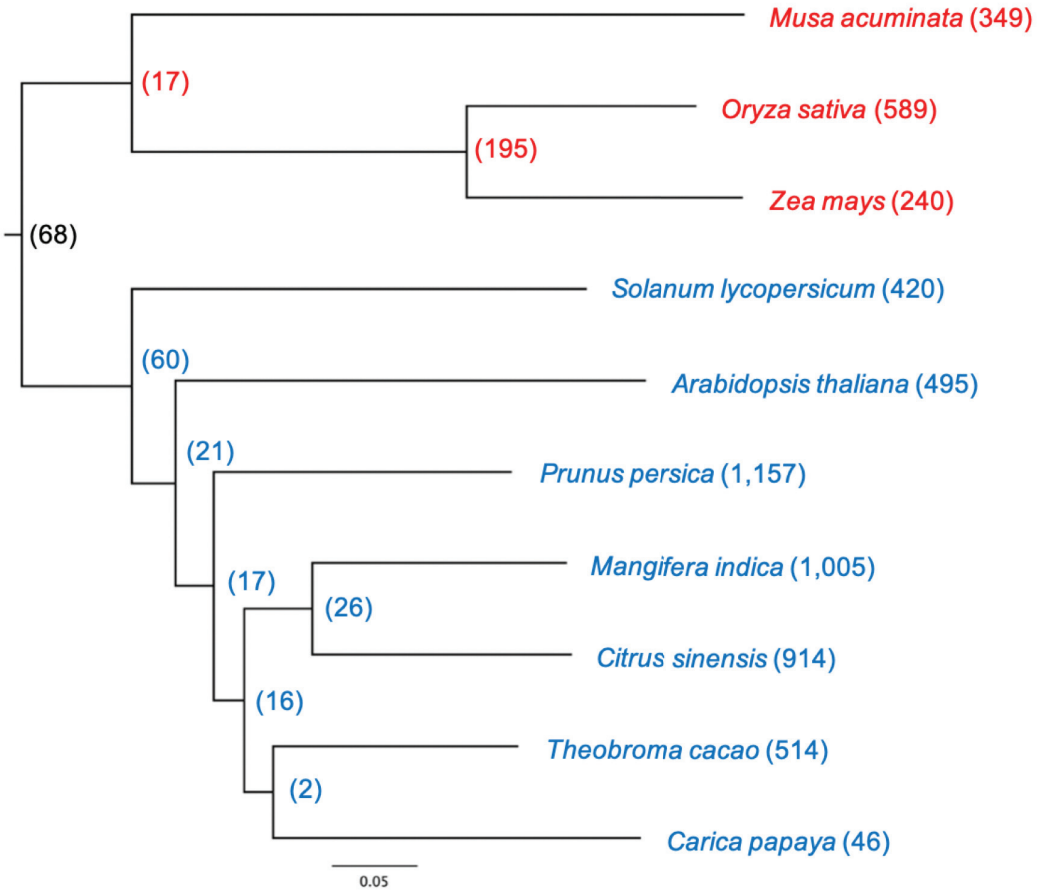

Figure 7. RGA duplications events over the branches of the STAG (Emms and Kelly 2018) species tree inferred from all conserved orthogroups implemented using OrthoFinder2 (Emms and Kelly 2019). The numbers after each node or species name are the number of gene duplication events with at least 50% support that occurred on the branch leading to the species. Scientific names in red fonts are monocot species while in blue fonts are dicot species.

orthologs with orange among other analyzed plant species. The comprehensive information from the mango RGAs and the linked EST-SSR markers established in this paper will facilitate in the development of outstanding mango varieties with resistance to various pests and diseases.

ACKNOWLEDGMENTS

The authors would like to thank the Department of Science and Technology –Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development for funding the project entitled "Characterization of 'Carabao' and other Mango Varieties with Resistance to Fruit Fly and Anthracnose." Also acknowledged is the University of the Philippines Los Baños for the core-funded project entitled "Genome-wide analysis of RGAs and development of RGA-linked DNA markers in Philippine Important Crops."

STATEMENT ON CONFLICT OF INTEREST

The authors declare no conflict of interest.

NOTES ON APPENDICES

The complete supplemental materials of the study are accessible at https://doi.org/10.6084/m9.figshare.12714425.v1

REFERENCES

ABOUL-SOUD MA, CHEN X, KANG JG, YUN BW, RAJA MU, MALIK SI, LOAKE GJ. 2009. Activation tagging of ADR2 conveys a spreading lesion phenotype and resistance to biotrophic pathogens. New Phytol. 183(4): 1163–1175. doi:10.1111/j.1469-8137.2009.02902.x

ACEVEDO FE, RIVERA-VEGA LJ, CHUNG SH, RAY S, FELTON GW. 2015. Cues from chewing insects – the intersection of DAMPs, HAMPs, MAMPs and effectors. Curr Opin Plant Biol 26: 80-86. doi:10.1016/j.pbi.2015.05.029

ANDERSON JP, GLEASON CA, FOLEY RC, THRALL PH, BURDON JB, SINGH KB. 2010. Plants versus pathogens: an evolutionary arms race. Funct Plant Biol 37(6): 499–512. doi:10.1071/FP09304

BABICKI S, ARNDT D, MARCU A, LIANG Y, GRANT JR, MACIEJEWSKI A, WISHART DS. 2016. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res 44(W1): W147–W153. doi:10.1093/nar/gkw419

BALLY ISE. 2006. *Mangifera indica* (mango), ver. 3.1. Species Profiles Pacific Isl Agrofor 1–25.

BITTNER-EDDY PD, CRUTE IR, HOLUB EB, BEYNON JL. 2000. RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J 21(2): 177–188. doi:10.1046/j.1365-313X.2000.00664.x

BOLGER AM, LOHSE M, USADEL B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–2120. doi:10.1093/bioinformatics/btu170

BOLLER T, HE SY. 2009. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. Science 324(5928): 742–744. doi:10.1126/science.1171647

BORHAN MH, GUNN N, COOPER A, GULDEN S, TÖR M, RIMMER SR, HOLUB EB. 2008. WRR4 encodes a TIR-NB-LRR protein that confers broad-spectrum white rust resistance in *Arabidopsis thaliana* to four physiological races of *Albugo candida*. Mol Plant-Microbe Interact 21(6): 757–768. doi:10.1094/MPMI-21-6-0757

CAI K, ZHU L, ZHANG K, LI L, ZHAO Z, ZENG W, LIN X. 2019. Development and characterization of EST-SSR markers from RNA-Seq data in *Phyllostachys violascens.* Front Plant Sci 10: 50. doi:10.3389/fpls.2019.00050

CHAGNÉ D. 2015. Whole Genome Sequencing of Fruit Tree Species. Adv Bot Res 74: 1–37. doi:10.1016/bs.abr.2015.04.004

CHANG JH, TAI YS, BERNAL AJ, LAVELLE DT, STASKAWICZ BJ, MICHELMORE RW. 2002. Functional analyses of the *Pto* resistance gene family in tomato and the identification of a minor resistance determinant in a susceptible haplotype. Mol Plant-Microbe Interact 15(3): 281–291. doi:10.1094/MPMI.2002.15.3.281

CHOI HW, KLESSIG DF. 2016. DAMPs, MAMPs, and NAMPs in plant innate immunity. BMC Plant Biol 16(1): 232. doi:10.1186/s12870-016-0921-2

CONESA A, GÖTZ S, GARCÍA-GÓMEZ JM, TEROL J, TALÓN M, ROBLES M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18): 3674–3676. doi:10.1093/bioinformatics/bti610

CUI H, TSUDA K, PARKER JE. 2015. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annu Rev Plant Biol 66(1): 487–511. doi:10.1146/annurev-arplant-050213-040012

DANGL JL, JONES JDG. 2001. Plant pathogens and integrated defense responses to infection. Nature 411(6839): 826–833. doi:10.1038/35081161

DAVIDSON NM, HAWKINS ADK, OSHLACK A. 2017. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biol 18(1): 148. doi:10.1186/s13059-017-1284-1

DHARMASIRI N, DHARMASIRI S, ESTELLE M. 2005. The F-box protein TIR1 is an auxin receptor. Nature 435(7041): 441–445. doi:10.1038/nature03543

DINESH-KUMAR SP, THAM WH, BAKER BJ. 2000. Structure-function analysis of the tobacco mosaic virus resistance gene N. Proceedings of the National Academy of Sciences 97(26): 14789–14794. doi:10.1073/pnas.97.26.14789

EDGER PP, HEIDEL-FISCHER HM, BEKAERT M, ROTA J, GLÖCKNER G, PLATTS AE, HOFBERGER JA. 2015. The butterfly plant arms-race escalated by gene and genome duplications. Proceedings of the National Academy of Sciences 112(27): 8362–8366. doi:10.1073/pnas.1503926112

EITAS TK, NIMCHUK ZL, DANGL JL. 2008. Arabidopsis TAO1 is a TIR-NB-LRR protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector AvrB. Proceedings of the National Academy of Sciences 105(17): 6475–6480. doi:10.1073/pnas.0802157105

EMMS DM, KELLY S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol 20(1): 1–14. doi:10.1186/s13059-019-1832-y

EMMS DM, KELLY S. 2018. STAG: Species Tree Inference from All Genes. bioRxiv 267914. doi:10.1101/267914

EMMS DM, KELLY S. 2017. STRIDE: Species Tree Root Inference from Gene Duplication Events. Mol Biol Evol, msx259–msx259

FEUILLET C, REUZEAU C, KJELLBOM P, KELLER B. 1998. Molecular characterization of a new type of receptor-like kinase (wlrk) gene family in wheat. Plant Mol Biol 37(6): 943–953. doi:10.1023/A:1006062016593

FINN RD, MISTRY J, TATE J, COGGILL P, HEGER A, POLLINGTON JE, GAVIN OL, GUNASEKARAN P, CERIC G, FORSLUND K, HOLM L. 2009. The Pfam protein families database. Nucleic Acids Res 38(Suppl 1): D211–D222. doi:10.1093/nar/gkp985

FU L, NIU B, ZHU Z, WU S, LI W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23): 3150–3152. doi:10.1093/bioinformatics/bts565

GADISSA F, TESFAYE K, DAGNE K, GELETA M. 2018. Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. BMC Genet 19(1): 92. doi:10.1186/s12863-018-0682-z

GLOVER NM, REDESTIG H, DESSIMOZ C. 2016. Homoeologs: What Are They and How Do We Infer Them? Trends Plant Sci 21(7): 609–621. doi:10.1016/j.tplants.2016.02.005

GOODWIN S, MCPHERSON JD, MCCOMBIE WR. 2016. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet 17(6): 333–351. doi:10.1038/nrg.2016.49

GRABHERR MG, HAAS BJ, YASSOUR M, LEVIN JZ, THOMPSON DA, AMIT I, ADICONIS X, FAN L, RAYCHOWDHURY R, ZENG Q, CHEN Z. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29(7): 644. doi:10.1038/nbt.1883.Trinity

GURURANI MA, VENKATESH J, UPADHYAYA CP, NOOKARAJU A, PANDEY SK, PARK SW. 2012. Plant disease resistance genes: current status and future directions. Physiol Mol Plant Pathol 78: 51–65. doi:10.1016/j.pmpp.2012.01.002

HAAS BJ, PAPANICOLAOU A, YASSOUR M, GRABHERR M, BLOOD PD, BOWDEN J, COUGER MB, ECCLES D, LI B, LIEBER M, MACMANES MD. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8(8): 1494. doi:10.1038/nprot.2013.084

HANCOCK JM, SIMON M. 2005. Simple sequence repeats in proteins and their significance for network evolution. Gene 345(1): 113–118. doi:10.1016/j.gene.2004.11.023

HIRUMA K, NISHIUCHI T, KATO T, BEDNAREK P, OKUNO T, SCHULZE-LEFERT P, TAKANO Y.

2011. *Arabidopsis* ENHANCED DISEASE RESISTANCE 1 is required for pathogen-induced expression of plant defensins in nonhost resistance, and acts through interference of MYC2-mediated repressor function. Plant J 67(6): 980–992. doi:10.1111/j.1365-313X.2011.04651.x

HOANG DT, CHERNOMOR O, VON HAESELER A, MINH BQ, VINH LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35(2): 518–522. doi:10.1093/molbev/msx281

HOLUB E. 2001. The arms race is ancient history in history in *Arabidopsis*, the wildflower. Nat Rev Genet 2(7): 516–527.

HONG K, GONG D, ZHANG L, HU H, JIA Z, GU H, SONG K. 2016. Transcriptome characterization and expression profiles of the related defense genes in postharvest mango fruit against *Colletotrichum gloeosporioides*. Gene 576(1): 275–283. doi:10.1016/j.gene.2015.10.041

HU Q, MIN L, YANG X, JIN S, ZHANG L, LI Y, MA Y, QI X, LI D, LIU H, LINDSEY K. 2018. Laccase GhLac1 modulates broad-spectrum biotic stress tolerance *via* manipulating phenylpropanoid pathway and jasmonic acid synthesis. Plant Physiol 176(2): 1808–1823. doi:10.1104/pp.17.01628

HULBERT SH, WEBB CA, SMITH SM, SUN Q. 2001. RESISTANCE GENE COMPLEXES : Evolution and Utilization. Annu Rev Phytopathol 39(1): 285–312. doi:10.1146/annurev.phyto.39.1.285

JONES DT, TAYLOR WR, THORNTON JM. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8(3): 275–282.

JONES JDG, DANGL JL. 2006. The plant immune system. Nature 444(7117): 323–329. doi:10.1038/nature05286

KÄLL L, KROGH A, SONNHAMMER ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338(5): 1027–1036. doi:10.1016/j.jmb.2004.03.016

KALYAANAMOORTHY S, MINH BQ, WONG TKF, VON HAESELER A, JERMIIN LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14(6): 587–589. doi:10.1038/nmeth.4285

KARTHIKA R, PRASATH D, ANANDARAJ M. 2019. Transcriptome-wide identification and characterization of resistant gene analogs (RGAs) of ginger (*Zingiber officinale* Rosc.) and mango ginger (*Curcuma amada* Roxb.) under stress induced by pathogen. Sci Hortic (Amsterdam) 248: 81–88. doi:10.1016/j.scienta.2019.01.003

KASHI Y, KING DG. 2006. Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22(5): 253–259. doi:10.1016/j.tig.2006.03.005

KHAN A, MATHELIER A. 2017. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18(1): 287. doi:10.1186/s12859-017-1708-7

KIM SH, KWON S IL, SAHA D, ANYANWU NC, GASSMANN W. 2009. Resistance to the *Pseudomonas syringae* effector hopA1 is governed by the TIR-NBS-LRR Protein rps6 and is enhanced by mutations in srfr1. Plant Physiol 150(4): 1723–1732. doi:10.1104/pp.109.139238

KONG Q, QU N, GAO M, ZHANG Z, DING X, YANG F, LI Y, DONG OX, CHEN S, LI X, ZHANG Y. 2012. The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in *Arabidopsis*. Plant Cell 24(5): 2225–2236. doi:10.1105/tpc.112.097253

KORESSAAR T, REMM M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10): 1289–1291. doi:10.1093/bioinformatics/btm091

KUANG H, WEI F, MARANO MR, WIRTZ U, WANG X, LIU J, SHUM WP, ZABORSKY J, TALLON LJ, RENSINK W, LOBST S. 2005. The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. Plant J 44(1): 37–51. doi:10.1111/j.1365-313X.2005.02506.x

KUMAR B, KUMAR U, YADAV HK. 2015. Identification of EST–SSRs and molecular diversity analysis in *Mentha piperita*. Crop J 3(4): 335–342. doi:10.1016/j.cj.2015.02.002

LANGMEAD B, SALZBERG SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359. doi:10.1038/nmeth.1923

LANTICAN DV, STRICKLER SR, CANAMA AO, GARDOCE RR, MUELLER LA, GALVEZ HF. 2019. *De novo* genome sequence assembly of dwarf coconut (*Cocos nucifera* L. 'Catigan Green Dwarf') provides insights into genomic variation between coconut types and related palm species. G3 Genes, Genomes, Genet. 9(8). doi:10.1534/g3.119.400215

LI B, DEWEY CN. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12(1): 323.

doi:10.1186/1471-2105-12-323

LI P, QUAN X, JIA G, XIAO J, CLOUTIER S, YOU FM. 2016. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC Genomics 17(1): 852. doi:10.1186/s12864-016-3197-x

LI YC, KOROL AB, FAHIMA T, NEVO E. 2004. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21(6): 991–1007. doi:10.1093/molbev/msh073

LIU F, HU Z, LIU W, LI J, WANG W, LIANG Z, WANG F, SUN X. 2016. Distribution, function and evolution characterization of microsatellite in *Sargassum thunbergii* (Fucales, Phaeophyta) transcriptome and their application in marker development. Sci Rep 6(1): 1–9. doi:10.1038/srep18947

LOKOSSOU AA, RIETMAN H, WANG M, KRENEK P, VAN DER SCHOOT H, HENKEN B, HOEKSTRA R, VLEESHOUWERS VG, VAN DER VOSSEN EA, VISSER RG, JACOBSEN E. 2010. Diversity, distribution, and evolution of *Solanum bulbocastanum* late blight resistance genes. Mol Plant-Microbe Interact 23(9): 1206–1216. doi:10.1094/MPMI-23-9-1206

LOLLE S, GREEFF C, PETERSEN K, ROUX M, JENSEN MK, BRESSENDORFF S, RODRIGUEZ E, SØMARK K, MUNDY J, PETERSEN M. 2017. Matching NLR Immune Receptors to Autoimmunity in camta3 Mutants Using Antimorphic NLR Alleles. Cell Host Microbe 21(4): 518–529. doi:10.1016/j.chom.2017.03.005

LUPAS A, VAN DYKE M, STOCK J. 1991. Predicting coiled coils from protein sequences. Science 1162–1164. doi:10.1126/science.252.5009.1162

MACKEY D, HOLT BF, WIIG A, DANGL JL. 2002. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. Cell 108(6): 743–754. doi:10.1016/S0092-8674(02)00661-X

MASLE J, GILMORE SR, FARQUHAR GD. 2005. The ERECTA gene regulates plant transpiration efficiency in *Arabidopsis*. Nature 436(7052): 866–870. doi:10.1038/nature03835

MEYERS BC, KOZIK A, GRIEGO A, KUANG H, MICHELMORE RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15(4): 809–834. doi:10.1105/tpc.009308

MICHELMORE RW, MEYERS BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8(11): 1113–1130. doi:10.1101/gr.8.11.1113

MÜLLER T, VINGRON M. 2001. Modeling amino acid replacement. J Comput Biol *7*(6): 761–776. doi:10.1089/10665270050514918

NARUSAKA M, SHIRASU K, NOUTOSHI Y, KUBO Y, SHIRAISHI T, IWABUCHI M, NARUSAKA Y. 2009. RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. The Plant Journal 60(2): 218–226. doi:10.1111/j.1365-313X.2009.03949.x

NEUPANE S, ANDERSEN EJ, NEUPANE A, NEPAL MP. 2018. Genome-wide identification of NBS-encoding resistance genes in sunflower (*Helianthus annuus* L.). Genes (Basel) 9(8): 384. doi:10.3390/genes9080384

NGUYEN LT, SCHMIDT HA, VON HAESELER A, MINH BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32(1): 268–274. doi:10.1093/molbev/msu300

OKUYAMA Y, KANZAKI H, ABE A, YOSHIDA K, TAMIRU M, SAITOH H, FUJIBE T, MATSUMURA H, SHENTON M, GALAM DC, UNDAN J. 2011. A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. Plant J 66(3): 467–479. doi:10.1111/j.1365-313X.2011.04502.x

PANCHY N, LEHTI-SHIU M, SHIU SH. 2016. Evolution of gene duplication in plants. Plant Physiol 171(4): 2294–2316. doi:10.1104/pp.16.00523

PARKER JE, COLEMAN MJ, SZABÒ V, FROST LN, SCHMIDT R, VAN DER BIEZEN EA, MOORES T, DEAN C, DANIELS MJ, JONES JD. 1997. The *Arabidopsis* downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. Plant Cell 9(6): 879–894. doi:10.1105/tpc.9.6.879

PERAZZOLLI M, MALACARNE G, BALDO A, RIGHETTI L, BAILEY A, FONTANA P, VELASCO R, MALNOY M. 2014. Characterization of resistance gene analogues (RGAs) in apple (*Malus* x *domestica* Borkh.) and their evolutionary history of the Rosaceae family. PLoS One 9(2). doi:10.1371/journal.pone.0083844

PIETERSE CMJ, VAN DER DOES D, ZAMIOUDIS C, LEON-REYES A, VAN WEES SCM. 2012. Hormonal Modulation of Plant Immunity. Annu Rev Cell Dev Biol 28: 489–521. doi:10.1146/annurev-cellbio-092910-154055

PIOFCZYK T, JEENA G, PECINKA A. 2015. *Arabidopsis thaliana* natural variation reveals connections

between UV radiation stress and plant pathogen-like defense responses. Plant Physiol Biochem 93: 34–43. doi:10.1016/j.plaphy.2015.01.011

PITZSCHKE A. 2013. *Agrobacterium* infection and plant defense-transformation success hangs by a thread. Front Plant Sci 4: 1–12. doi:10.3389/fpls.2013.00519

PRAMOD S, PERKINS A, WELCH M. 2014. Patterns of microsatellite evolution inferred from the *Helianthus annuus* (Asteraceae) transcriptome. J Genet 93(2): 431–442. doi:10.1007/s12041-014-0402-z

RAMBAUT A. 2018. FigTree v. 1.4.4. Retrieved on 07 Feb 2020 from http://tree.bio.ed.ac.uk/software/figtree/

RIZZON C, PONGER L, GAUT BS. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLoS Comput Biol 2(9). doi:10.1371/journal.pcbi.0020115

RODRIGUEZ MCS, PETERSEN M, MUNDY J. 2010. Mitogen-activated protein kinase signaling in plants. Annual Review of Plant Biology 61: 621–649.

RODY HVS, BOMBARDELLI RGH, CRESTE S, CAMARGO LEA, VAN SLUYS MA, MONTEIRO-VITORELLO CB. 2019. Genome survey of resistance gene analogs in sugarcane: genomic features and differential expression of the innate immune system from a smut-resistant genotype. BMC Genomics 20(1): 809. doi:10.1186/s12864-019-6207-y

SÁNCHEZ-RODRÍGUEZ C, ESTÉVEZ JM, LLORENTE F, HERNÁNDEZ-BLANCO C, JORDÁ L, PAGÁN I, BERROCAL M, MARCO Y, SOMERVILLE S, MOLINA A. 2009.The ERECTA receptor-like kinase regulates cell wall-mediated resistance to pathogens in *Arabidopsis thaliana.* Mol Plant-Microbe Interact 22(8): 953–963. doi:10.1094/MPMI-22-8-0953

SAUCET SB, MA Y, SARRIS PF, FURZER OJ, SOHN KH, JONES JDG. 2015. Two linked pairs of *Arabidopsis* TNL resistance genes independently confer recognition of bacterial effector AvrRps4. Nat Commun 6(1): 1–12. doi:10.1038/ncomms7338

SEKHWAL MK, LI P, LAM I, WANG X, CLOUTIER S, YOU FM. 2015. Disease resistance gene analogs (RGAs) in plants. Int J Mol Sci 16(8): 19248–19290. doi:10.3390/ijms160819248

SELA H, CHENG J, JUN Y, NEVO E, FAHIMA T. 2009. Divergent diversity patterns of NBS and LRR domains of resistance gene analogs in wild emmer wheat populations. Genome 52(6): 557–565. doi:10.1139/G09-030

SHIU SH, BLEECKER AB. 2003. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. Plant Physiol 132(2): 530–543. doi:10.1104/pp.103.021964

SINAPIDOU E, WILLIAMS K, NOTT L, BAHKT S, TÖR M, CRUTE I, BITTNER-EDDY P, BEYNON J. 2004. Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. Plant J 38(6): 898–909. doi:10.1111/j.1365-313X.2004.02099.x

SOLTIS PS, SOLTIS DE. 2009. The Role of Hybridization in Plant Speciation. Annu Rev Plant Biol 60: 561–588. doi:10.1146/annurev.arplant.043008.092039

SOUBRIER J, STEEL M, LEE MS, DER SARKISSIAN C, GUINDON S, HO SY, COOPER A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol 29(11): 3345–3358. doi:10.1093/molbev/mss140

STAAL J, KALIFF M, BOHMAN S, DIXELIUS C. 2006. Transgressive segregation reveals two *Arabidopsis* TIR-NB-LRR resistance genes effective against *Leptosphaeria maculans*, causal agent of blackleg disease. Plant J 46(2): 218–230. doi:10.1111/j.1365-313X.2006.02688.x

STAAL J, KALIFF M, DEWAELE E, PERSSON M, DIXELIUS C. 2008. RLM3, a TIR domain encoding gene involved in broad-range immunity of *Arabidopsis* to necrotrophic fungal pathogens. Plant J 55(2): 188–200. doi:10.1111/j.1365-313X.2008.03503.x

TARR DEK, ALEXANDER HM. 2009. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2(1): 197. doi:10.1186/1756-0500-2-197

THATIKUNTA R, SANKAR AS, SREELAKSHMI J, PALLE G, LEELA C, RANI CV, SHANKAR VG, LAVANYA B, REDDY PN, DUDHE MY. 2016. Utilization of in silico EST-SSR markers for diversity studies in castor (*Ricinus communis* L.). Physiol Mol Biol Plants 22(4): 535–545. doi:10.1007/s12298-016-0367-x

THOMPSON JD, HIGGINS DG, GIBSON TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22): 4673–4680. doi:10.1093/nar/22.22.4673

THOMPSON JN, BURDON JJ. 1992. Gene-for-gene coevolution between plants and parasites. Nature 360(6400): 121–125. doi:10.1038/360121a0

TÖR M, LOTZE MT, HOLTON N. 2009. Receptor-mediated signalling in plants: molecular patterns and programmes. J Exp Bot 60(13): 3645–3654. doi:10.1093/jxb/erp233

UNTERGASSER A, CUTCUTACHE I, KORESSAAR T, YE J, FAIRCLOTH BC, REMM M, ROZEN SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res 40(15): e115. doi:10.1093/nar/gks596

WAN L, ESSUMAN K, ANDERSON RG, SASAKI Y, MONTEIRO F, CHUNG EH, NISHIMURA EO, DIANTONIO A, MILBRANDT J, DANGL JL, NISHIMURA MT. 2019. TIR domains of plant immune receptors are NAD+-cleaving enzymes that promote cell death. Science 365(6455): 799–803. doi:10.1126/science.aax1771

WANG X, WANG L. 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci 7: 1350. doi:10.3389/fpls.2016.01350

WANG P, LUO Y, HUANG J, GAO S, ZHU G, DANG Z, GAI J, YANG M, ZHU M, ZHANG H, YE X, GAO A, TAN X, WANG S, WU S, CAHOON EB, BAI B, ZHAO Z, LI Q, WEI J, CHEN H, LUO R, GONG D, TANG K, ZHANG B, NI Z, HUANG G, HU S, CHEN Y. 2020. The genome evolution and domestication of tropical fruit mango. Genome Biol 21: 60

WARREN RF, HENK A, MOWERY P, HOLUB E, INNES RW. 1998. A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene RPS5 partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell 10(9): 1439–1452. doi:10.1105/tpc.10.9.1439

WILLIAMS ME. 2011. Introduction to Phytohormones. Teaching Tools in Plant Biology: Lecture Notes. The Plant Cell (online). doi/10.1105/tpc.110.tt0310.

WU YC, RASMUSSEN MD, BANSAL MS, KELLIS M. 2014. Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. Genome Res 24: 475–486.

XIE DX, FEYS BF, JAMES S, NIETO-ROSTRO M, TURNER JG. 1998. COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. Science 280(5366): 1091–1094. doi:10.1126/science.280.5366.1091

YANG Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39(3): 306–314.

YANG Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics 139(2): 993–1005.

YU J, KE T, TEHRIM S, SUN F, LIAO B, HUA W. 2015. PTGBase: an integrated database to study tandem duplicated genes in plants. Database 2015. doi:10.1093/database/bav017

ZBIERZAK AM, PORFIROVA S, GRIEBEL T, MELZER M, PARKER JE, DÖRMANN P. 2013. A TIR-NBS protein encoded by *Arabidopsis* Chilling Sensitive 1 (CHS1) limits chloroplast damage and cell death at low temperature. Plant J 75(4): 539–552. doi:10.1111/tpj.12219

ZDOBNOV EM, APWEILER R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9): 847–848. doi:10.1093/bioinformatics/17.9.847

ZHAI C, LIN F, DONG Z, HE X, YUAN B, ZENG X, WANG L, PAN Q. 2011. The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. New Phytol 189(1): 321–334. doi:10.1111/j.1469-8137.2010.03462.x

ZHANG X, LIANG P, MING R. 2016. Genome-Wide Identification and Characterization of Nucleotide-Binding Site (NBS) Resistance Genes in Pineapple. Trop Plant Biol 9(3): 187–199. doi:10.1007/s12042-016-9178-z

ZHANG Z, WU Y, GAO M, ZHANG J, KONG Q, LIU Y, BA H, ZHOU J, ZHANG Y. 2012. Disruption of PAMP-induced MAP kinase cascade by a *Pseudomonas syringae* effector activates plant immunity mediated by the NB-LRR protein SUMM2. Cell Host Microbe 11(3): 253–263. doi:10.1016/j.chom.2012.01.015

ZHANG Z, XIE W, ZHAO Y, ZHANG J, WANG N, NTAKIRUTIMANA F, YAN J, WANG Y. 2019. EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. BMC Plant Biol 19(1): 235. doi:10.1186/s12870-019-1825-8

ZIPFEL C, KUNZE G, CHINCHILLA D, CANIARD A, JONES JD, BOLLER T, FELIX G. 2006. Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts *Agrobacterium*-Mediated Transformation. Cell 125(4): 749–760. doi:10.1016/j.cell.2006.03.037