Transcriptome-wide analysis of expressed resistance gene analogs (RGAs) in mango

Darlon V. Lantican[1]*, Cris Q. Cortaga[2], Anand Noel C. Manohar[1], Fe M. dela Cueva[3], and Maria Luz J. Sison[2]

[1]Genetics Laboratory, [2]Entomology Laboratory and [3]Plant Pathology Laboratory, Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños, College, Laguna, Philippines 4031

*corresponding author: dvlantican@up.edu.ph

Abstract

Plants have developed mechanisms to recognize pathogen and insect attacks and to activate defense response against them through the action of resistance genes (*R* genes). Hence, *R* gene analogs (RGAs) have been widely studied and efficiently used for breeding of resistant crops. Mango is an economic fruit crop largely cultivated in the (sub)tropics and thus, is constantly challenged by a myriad of insect pests and diseases. RGAs of mango from different varieties were identified and characterized via *de novo* assembly of raw transcriptome sequences currently available in biological repositories. Using bioinformatics pipelines, a core RGA database of mango with 747 protein models was established and were classified based on their conserved domains and motifs: 53 nucleotide binding site protein (NBS); 27 nucleotide binding site-leucine rich repeat proteins (NBS-LRR); 17 coiled-coil NBS-LRR (CNL); 2 toll/interleukin-1 receptor NBS-LRR (TNL); 29 coiled-coil NBS (CN); 4 toll/interleukin-1 receptor NBS (TN); 17 toll/interleukin-1 receptor with unknown domain (TX); 158 receptor-like proteins (RLP); 362 receptor-like kinases (RLK); 72 transmembrane coiled-coil domain protein (TM-CC), and 6 NBS-encoding proteins with other domains. Phylogenetic analysis broadly clustered the core RGAs into 6 major clades based on their domain classification, while TM-CC formed subclades all across the tree. The phylogenetic results suggest highly divergent functions of the RGAs which also provide insights into the mango-pest co-evolutionary arms race. The various molecular functions, biological processes, and cellular localizations of these

RGAs were functionally well-annotated through gene ontology (GO) analysis, and their expression profiles across different mango varieties were also determined. From the mango RGA transcripts, 134 unique EST-SSRs loci were identified, and primers were designed targeting these potential markers. To date, this is the most comprehensive analysis of mango RGAs which offer a trove of markers for direct utilization in resistance breeding of mango.

Keywords: mango, resistance gene analogs (RGAs), transcriptome, bioinformatics

Introduction

Mango (*Mangifera indica* L.) is a globally popular fruit crop with economical and agricultural significance, especially in the (sub)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.

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 being 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

56 immunity (P/DTI) (Jones and Dangl, 2006; Hu et al., 2018). The first layer of defense, however,
57 is often overcome by plant pests through effector proteins, leading to effector-triggered
58 susceptibility (ETS). To counteract this, plants can have a second line of defense known as
59 effector-triggered immunity (ETI) (Boller and He, 2009; Pieterse et al., 2012). In ETI, due to
60 host-pest co-evolutionary "arms race", plants acquire a family of polymorphic and diverse
61 resistance genes (*R* genes) which encode R proteins that recognize the attacker-specific
62 effectors (Cui et al., 2015; Jones and Dangl, 2006) to correspond for a gene-for-gene
63 interaction (Thompson and Burdon, 1992). This recognition results in intracellular signaling
64 and switching 'on' of plant defense genes leading to plant resistance against pest attack.

65 Collectively, the PRRs and *R* genes are called resistance gene analogs (RGAs) which
66 share conserved domains and motifs (Li et al., 2016). PRRs are predominantly categorized
67 as receptor-like kinases (RLKs) and receptor like proteins (RLPs). RLKs have an extracellular
68 sensing domain [either leucine-rich repeat (LRR) type or lysin motif (LysM) type], a
69 transmembrane (TM) domain, and an intracellular kinase domain; while RLP possess similar
70 domain structure except that it lacks an intracellular kinase domain (Tör et al., 2009). The R
71 proteins are intracellular immune receptors (effector-recognition receptors) and majority of
72 which belongs to nucleotide binding-site LRR (NBS-LRR) class (Gururani et al., 2012;
73 Neupane et al., 2018). The R proteins contain certain domains or motifs such as
74 serine/threonine kinases (STKs), LRRs, NBS, TMs, leucine-zipper (LZ), coiled-coil (CC) and
75 toll/interleukin-1 receptor (TIR) (Li et al., 2016). Depending on the domain/motif architecture
76 combinations, the subgroups of NBS-encoding proteins were designated as NBS, CC-NBS-
77 LRR (CNL), TIR-NBS-LRR (TNL), CC-NBS (CN), TIR-NBS (TN), NBS-LRR (NL), TIR with
78 unknown domain (TX), and other NBS-protein that may have chimeric domain/motif
79 architecture. Since RGAs share conserved structural features, it is possible to predict them
80 extensively to obtain a deeper insight on the underlying molecular defenses of the plant. RGAs
81 have been widely studied and efficiently used for breeding of resistant crops against insect
82 pests and diseases (Karthika et al., 2019). One way to predict RGAs in plants is through

83 bioinformatics analyses of next-generation sequencing (NGS) genomic and transcriptomic
84 data (Lantican et al., 2019; Neupane et al., 2018; Zhang et al., 2016; Karthika et al., 2019).
85 With the advent of this technology, it is possible to unravel gene networks and develop
86 molecular markers tagging economically important traits, reveal other molecular information
87 such as intron-exon boundaries and presence of transposable elements, and discover novel
88 biological processes (Goodwin et al., 2016).

89 The RGAs were first comprehensively studied on a genome-wide scale using the
90 model plant *Arabidopsis thaliana* and since then, thousands of RGAs from numerous plant
91 genomes have been identified (Meyers et al., 2003; Sekhwal et al., 2015). Despite the
92 economical and agricultural importance of mango, especially in the tropics, its whole genome
93 sequence (WGS) is yet to be available. Mango is believed to be an allotetraploid (4n=40) with
94 genome size of around 450 Mb according to the on-going WGS effort (Chagné, 2015).
95 Therefore, no detailed genome-wide analysis of mango RGAs has been performed, which is
96 vital to facilitate breeding of resistant mango varieties. With the absence of such valuable
97 resource, this study aimed to comprehensively identify and characterize the expressed RGAs
98 of mango through *de novo* assembly of mango transcriptomes currently available in biological
99 repositories. To our knowledge, this paper provides the most comprehensive identification and
100 characterization of mango RGAs to date, which offer a trove of markers for direct utilization in
101 resistance breeding of mango against insect pests and pathogens.
102

103 **Materials and Methods**

104 ***De novo transcriptome assembly***

105 To obtain the expressed RGAs in mango, raw transcriptome reads were accessed and
106 retrieved from public repository (NCBI; Table 1) and one directly from Hong et al. (2016). The
107 raw read sequences were pre-processed by removing the adapter sequences and low-quality
108 base score nucleotide sequences using Trimmomatic v0.36 (Bolger et al., 2014) with the
109 following parameters: SLIDINGWINDOW: 5:30; LEADING:5; TRAILING:5; MINLEN:85. Each

110 trimmed transcriptome FASTQ files was independently assembled *de novo* using Trinity
111 pipeline (Grabherr et al., 2011) at default parameters. Upon assembly of the contigs,
112 SuperTranscripts were then generated to provide an accurate representation of each
113 expressed genes in the transcriptome assembly (Davidson et al., 2017). TransDecoder (Haas
114 et al., 2013) was subsequently used to construct a polypeptide sequence database from the
115 coding region of the SuperTranscripts based on nucleotide composition and open reading
116 frame (ORF) length with default parameters.
117
118 ***RGA identification and classification***
119 Mango transcriptome-wide resistance gene analogs (RGA) candidates belonging to
120 Nucleotide Binding Site (NBS) and transmembrane-coiled-coil (TM-CC) containing proteins,
121 and membrane associated receptor-like kinase (RLK) and receptor-like proteins (RLP)
122 families were identified in the generated gene models from genome annotation using
123 RGAugury (Li et al., 2016), an automated RGA prediction pipeline. The input protein
124 sequences were initially filtered using BLASTp search against RGAdb database integrated in
125 the pipeline using an *e*-value cut-off of $1e^{-5}$. The domain and motif of the initial set of candidate
126 RGAs were detected using nCoils (Lupas et al., 1991), phobius (Kall et al., 2004), pfam_scan
127 (Finn et al., 2010) and InterProScan 5 (Zdobnov and Apweiler, 2001) implemented within the
128 RGAugury pipeline. A core set of RGAs were generated by concatenation and clustering (at
129 90% identity) of the identified RGAs from each polypeptide database using CD-HIT (Fu et al.,
130 2012), and de-duplication based on BLAST description. The NBS-encoding proteins
131 containing chimeric domain/motif architecture and designated by RGAugury as OTHER were
132 manually checked for their domain/motif composition using Multiple Expectation Maximization
133 for Motif Elicitation (MEME) (Bailey and Elkan, 1994) and MOTIF Search
134 (https://www.genome.jp/tools/motif/).
135
136 ***RGA characterization and annotation***

137 Gene ontology (GO) annotations such as molecular function (MF), biological
138 processes (BP), and cellular component (CC) were assigned to each protein represented in
139 the constructed mango RGA core database using BLAST2GO package (Conesa et al., 2005).
140 The homology of the predicted polypeptide sequences of each RGAs to existing entries in
141 UniProtKB/SwissProt protein database were determined using BLASTp (with e-value of $1e^{-5}$).
142 The mapped BLAST hits were then merged to InterProScan (Zdobnov and Apweiler, 2001)
143 search output to produce the GO annotations.
144
145 ***Profiling of mango RGA expression***
146 RNA-Seq by Expectation Maximization (RSEM) (Li and Dewey, 2011) was used to
147 provide insights on the expression profile of each RGA obtained from different tissues and
148 diverse mango varieties. The pre-processed RNA-seq reads were independently mapped to
149 the constructed core mango RGA reference transcriptome using Bowtie2 (Langmead and
150 Salzberg, 2012). Transcripts per million (TPM) count values were then generated by RSEM
151 from each paired-end reads. A comma-separated value (csv) file containing the RGAs and
152 TPM values for each transcriptome data set was prepared and uploaded to Heatmapper
153 (Babicki et al., 2016) for heat-map visualization of the RGA expression profiles.
154
155 ***Evolutionary analysis***
156 The FASTA amino acid sequences of the 747 core mango RGAs were used to
157 construct a phylogenetic tree. Multiple Sequence Alignment (MSA) of the RGA sequences
158 was performed using the CLUSTALW program (Thompson et al., 1994) with the following
159 parameters: Gap Opening Penalty: 10; Gap Extension Penalty: 0.2. The phylogeny of these
160 aligned sequences was reconstructed using Maximum Likelihood statistical method using IQ-
161 TREE (Nguyen et al., 2015) with best-fit substitution model selected through ModelFinder
162 (Kalyaanamoorthy et al., 2017). Based on the Bayesian Information Criterion (BIC) of the
163 models, General 'Variable Time' (VT) matrix (Mueller and Vingron, 2004) with empirical amino
164 acid frequencies (+F) and FreeRate (+R5) rate heterogeneity across sites (Yang 1995;

165 Soubrier et al., 2012) was used to generate the tree. The resulting phylogenetic tree was
166 validated with 1000 replicates of ultrafast bootstrapping (Hoang et al., 2017) and visualized
167 using FigTree (v1.4.4) (Rambaut, 2018).
168
169 ***Identification and design of EST-SSRs***
170 The FASTA file corresponding to the RGA transcripts selected for analysis was
171 uploaded into GMATA (Genome-wide Microsatellite Analyzing Toward Application) Software
172 Package (Wang & Wang, 2016). SSR loci within the data set were identified using default
173 parameters [Min-length (nt) = 2; Max-length (nt) = 6; Min. repeat-times = 5]. Consequently,
174 oligonucleotide primer pairs were designed at regions flanking the identified SSRs using the
175 Primer3 (Untergasser et al., 2012; Koressaar and Remm, 2007) integrated within the package.
176 Default parameters were also used in primer design [Min. amplicon size (bp) = 120; Max.
177 amplicon size (bp) = 400; Optimal annealing T_m (°C) = 60; Flanking sequence length = 400;
178 Max. template length = 2000].
179
180 **Results and Discussion**
181 ***Identification of mango RGAs***
182 Raw transcriptome sequences from diverse mango cultivars were retrieved,
183 assembled, and the RGAs were identified using RGAugury, an integrative NGS bioinformatics
184 pipeline that efficiently predicts RGAs in plants (Li et al., 2016). In general, this pipeline works
185 by identifying first the domains and motifs related to RGAs such as NBS, LRR,
186 serine/threonine and tyrosine kinase (STTK), CC, LysM, TM, and TIR. Based on the presence
187 of combinations of these domains and motifs, the RGA candidates are then identified and
188 categorized into four major families: NBS-encoding, TM-CC, and membrane associated RLP
189 and RLK. The number of RGAs predicted from the sequence reads archive (SRA) files ranged
190 from 56 to 454 (Table 1). In total, 7,391 RGAs were identified from the assembled
191 transcriptome sequences (Table 1; Supplemental File 1). These RGAs were further clustered
192 at 90% homology creating 2,892 clusters, followed by automated selection of a representative

193 protein sequence per cluster. The representative protein sequences were described using
194 BLAST and BLAST2GO, and proteins with duplicated BLAST description were removed
195 retaining only unique/non-redundant protein type per RGA domain to generate 747 core RGAs
196 with unique functions (Table 2, Supplemental Table 1 and Supplemental File 2). The core
197 RGA proteins were classified as follows: 53 nucleotide binding site protein (NBS); 27
198 nucleotide binding site-leucine rich repeat proteins (NBS-LRR); 17 coiled-coil NBS-LRR
199 (CNL); 2 toll/interleukin-1 receptor NBS-LRR (TNL); 29 coiled-coil NBS (CN); 4 toll/interleukin-
200 1 receptor NBS (TN); 17 toll/interleukin-1 receptor with unknown domain (TX); 158 receptor-
201 like proteins (RLP); 362 receptor-like kinases (RLK); 72 transmembrane coiled-coil domain
202 protein (TM-CC); and 6 NBS-encoding proteins with other domains (Table 2).

203 Upon comparison to other important agricultural crops, as reported by Li et al. (2016),
204 it appears that mango harbors more RGAs (747) than papaya (467) but has lower compared
205 to cacao (1,281), tomato (1,003), corn (1,062), rice (1,652) and banana (892) (Table 2).
206 Eventually, the RGA content and characteristics of a plant has been frequently associated
207 with crop resistance (Sekhwal et al., 2015). Like most crops, RLKs constitute the largest group
208 of RGAs in mango comprising almost 50% of the RGAs characterized, followed by RLPs, then
209 the NBS-encoding proteins (Table 2, Fig. 1). In dicots such as mango, all NBS-encoding
210 proteins are present (i.e., NBS, CNL, TNL, CN, TN, NL, TX, and other NBS-encoding proteins)
211 (Table 2) but in monocots such as corn, rice, banana, the TNL protein is usually absent (Table
212 2) (Tarr and Alexander, 2009). It is hypothesized that after the divergence of dicots and
213 monocots, the TNL genes might have been lost from the monocot lineage (Zhang et al., 2016).
214 Six mango RGA models showed chimeric domain/motif architecture and were classified as
215 'other' NBS-encoding proteins (Table 2). These were manually checked using the MEME
216 platform and appear to be NBS-encoding proteins containing TIR sequences and other
217 uncharacterized domain (Supplemental Figure 1).

218

219 ***Mango resistance (R)/defense proteins and expression profiles***

220 BLAST analysis of the core mango RGA database revealed homology to a number of
221 well-known resistance (R) /defense proteins against pathogens and insects (Table 3). Among
222 these R proteins include RRS proteins (putative WRKY transcription factors), which confer
223 resistance against *Colletotrichum higginsianum* and *Ralstonia solanacearum* (Narusaka et al.,
224 2009; Saucet et al., 2015); RPP proteins (including At4g19530 and At4g19520 proteins) which
225 confer resistance against downy mildew (*Peronospora parasitica*) (Parker et al., 1997;
226 Sinapidou et al., 2004; Bittner-Eddy et al., 2000; Wan et al., 2019); the LRK10 proteins (all of
227 which are of RLK domains) which provide resistance against leaf rust (*Puccinia triticina*) (Shiu
228 and Bleecker, 2003; Feuillet et al., 1998); the RGA-blb and R1A-10 proteins which provide
229 resistance against the devastating late blight disease (*Phytophtora infestans*) (Lokossou et
230 al., 2010; Kuang et al., 2005); the RLM proteins which confer resistance against the fungal
231 pathogen *Leptosphaeria maculans* (Staal et al., 2008; 2006); the Pik-2 and RGA5 proteins
232 which confer resistance against blast disease (*Magnaporthe oryzae*) (Zhai et al., 2011;
233 Okuyama et al., 2011); and many homologues of TMV resistance protein N which are
234 responsible for tobacco mosaic virus (TMV) resistance (Dinesh-Kumar et al., 2000). The core
235 mango RGA also contained numerous sequences with homology to various RPS proteins,
236 including RPM1 and TAO1 proteins, which are important R proteins against the bacterium
237 *Pseudomonas syringae*, a widely studied biotrophic pathogen (Eitas et al., 2008; Mackey et
238 al., 2002; Warren et al., 1998; Kim et al., 2009).

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

248 confers broad-spectrum resistance against many pathogens such as strains *of Pseudomonas*
249 *syringae* and *Hyaloperonospora arabidopsis*, and full immunity to several races of the
250 pathogen *Albugo candida*, the causal agent of white rust disease (Borhan et al., 2008; Aboul-
251 Soud et al., 2009). This R protein is also reported to play a role in the response to UV stress
252 (Piofczyk et al., 2015). Predicted mango RGAs homologous to the R-like proteins DSC 1 and
253 2 (DOMINANT SUPPRESOR OF CAMTA3) were also found which act as guard of CAMTA3,
254 a negative regulator of immunity, during infection of pathogens (Lolle et al., 2017). Three NBS-
255 encoding RGAs in mango showed homology to SUMM2 (SUPPRESSOR OF *mkk1 mkk2* 2)
256 which acts as an R protein that becomes activated when the MEKK1-MKK1/MKK2-MPK4
257 cascade in the basal defense response is disrupted by the pathogen effector HopAI1 (Zhang
258 et al., 2012; Kong et al., 2012).

259 Other putative resistance proteins abundantly found in mango showed homology to
260 various DRL-coded (Uniprot/SwissProt ID) potential R proteins from *Arabidopsis thaliana*
261 which have not been fully studied (Supplemental Table 1). Homologues to CHS and CHL
262 proteins were also found in mango RGAs (Supplemental Table 1) which are important R
263 proteins against chilling/low temperature and also confer resistance against bacterial infection
264 such as *Pseudomonas syringae* (Zbierzak et al., 2013). Aside from resistance to pathogens,
265 R/defense proteins against insects were also identified by filtering the gene ontology results
266 with GO term “defense response to insect” (GO:0002213). The RGAs found include the
267 At4g11170 putative resistance protein, RPP5, DSC1 and 2, TRANSPORT INHIBITOR
268 RESPONSE 1 (TIR1), and CORONATINE INSENSITVE-1 (COI-1). TIR1 is an F-box protein
269 and a major receptor of the plant hormone auxin and mediate in the auxin/indole acetic acid
270 (IAA)-regulated signaling pathways involved in plant growth and development, as well as
271 defense response against pathogens and insect pests (Dharmasiri et al., 2005). COI-1 is a
272 jasmonic acid-isoleucine (JA-Ile) coreceptor and also an F-box protein involved in the JA
273 signaling pathway which regulate plant defense against insects and pathogens, wound
274 healing, and other vegetative and reproductive developmental functions (Xie et al., 1998). The
275 plant hormone JA is conjugated to isoleucine (Ile) as a consequence of insect/pathogen

276 attack, and JA-Ile binding with COI-1 stimulates degradation of JAZ coreceptor repressor
277 proteins, thereby promoting the expression of JA-responsive genes (Pieterse et al., 2012).
278 Further filtering of RGAs containing the GO term “response to insect” (GO:0009625) revealed
279 17 RGAs that putatively play vital roles in the defense pathways against insects (Table 4), all
280 of which were characterized to possess RLK/RLP domains. The complete and detailed
281 BLASTp analysis and description of the RGAs are provided in Supplemental Table 1.

282 Of all the mango varieties analyzed in this study, the three varieties that showed the
283 lowest expression of RGAs were SRR8449851 (CHOKANAN), SRR3359450 (DASHEHARI),
284 and SRR1562177 (KENT) (Fig. 2; Supplemental Table 2). On the other hand, the following
285 showed the highest expressions of the RGAs characterized: SRR2162953 (pooled RNA from
286 different varieties), SRR2162919 (THAI EVERBEARING), SRR2162878 (BURMA),
287 SRR2162889 (M. CASTURI “PURPLE”), SRR2159471 (TOMMY ATKINS), SRR2162929
288 (TURPENTINE), SRR2162836 (AMIN ABRAHIMPUR), SRR2162908 (NEELUM) (Fig. 2;
289 Supplemental Table 2). The relatively high expression of RGAs in these RNA-seq reads could
290 be explained by the fact that they are pooled RNA samples from different varieties or from
291 different plant tissues.

292

293 ***Gene ontology and functional annotation of mango RGAs***

294 Gene ontology (GO) analysis and functional annotation of the core RGA database
295 provided a broader knowledge on the major roles of these proteins and their cellular
296 localizations. The molecular functions of the core RGAs are primarily associated with protein
297 and nucleotide binding, and kinase activity (Fig. 3) as they are known extra- and intracellular
298 binding receptors and relays defense signaling in the cells through a cascade of kinase
299 activities (Rodriguez et al., 2010). The absence of an intracellular kinase in RLPs suggests
300 that it relies/interact with other domains such as RLK-type receptors to relay the signalling
301 from extracellular matrix to the intracellular component (Tör et al., 2009). Meanwhile, the
302 biological processes of the RGAs are highly diverse, distributing the RGA sequences to a wide
303 array of GO IDs/terms. The RGAs are mainly involved in protein (auto)phosphorylation during

304 signal transductions in the cell, defense/immune responses to different biotic (insect pests and
305 diseases) and abiotic (water deprivation, ozone, UV stress, etc.) stresses, and in various
306 developmental processes (i.e., from embryonic to floral/pollen development) (Fig. 4).
307 Numerous RGAs were also described to be involved in hormone-mediated signaling pathways
308 and systemic acquired resistance which include the plant hormones brassinosteroids, abscisic
309 acid (ABA), salicylic acid (SA), auxin, jasmonic acid (JA), ethylene, and gibberellic acid (GA).
310 These plant hormones create a crosstalk to modulate defense signaling pathways and activate
311 systemic resistance against a broad spectrum of pathogens and insect pests. As an example
312 of this crosstalk, SA antagonizes JA to activate immune response against biotrophic
313 pathogens (Williams, 2011). Likewise, JA antagonizes SA to activate two branches of immune
314 response which also antagonizes each other: the ERF branch against necrotrophic pathogens
315 (synergistic interaction with ethylene) and the MYC branch against insects (synergistic
316 interaction with ABA) (Pieterse et al., 2012). In the cell, majority of the mango RGAs are
317 localized in the plasma membrane, plasmodesma, cytosol, and nucleus which are primary
318 sites for recognition of pathogen/insect invasion and effector proteins. In these cellular
319 components, the RGAs functions to convert extracellular stimuli into intracellular responses to
320 activate defense cascades and counteract the attack. In general, the results of GO analysis
321 and functional annotation show that plant immune response is a consequence of the activity
322 of a wide range of plant hormones, resistance/defense-related proteins, biochemical and
323 developmental processes, etc. that aim to suppress pest attack thereby improving plant
324 defense. The complete and detailed GO analysis and annotation of each RGAs are provided
325 in Supplemental Table 1 and Supplemental Table 3.

326

327 ***Evolutionary relationships among mango RGAs***

328 Using the evolutionary best-fit model selected according to Bayesian Information
329 Criterion (BIC) (Supplemental File 3), a maximum likelihood phylogenetic tree was constructed
330 using the 747 core mango RGA proteins to investigate their evolutionary relationships and

331 diversity (Fig. 6; Supplemental File 4). It showed that clustering of RGAs based on domain
332 classification is very apparent but mingled with subclades from other RGA domains. As shown
333 in the tree, the mango RGAs have become highly diversified and are broadly clustered into 6
334 major clades with many subclades (Fig. 6). RGAs having TM-CC domain didn't emerge as a
335 separate cluster but formed subclades all across the tree suggesting that it shares similarity
336 to all other RGA classes. The clade 1 is composed mainly of NBS-encoding proteins and other
337 domains such as TM-CC and RLP. Notably, the clade 1 appeared to be the most distantly
338 related RGA group among other clades. The clade 2 is predominantly composed of RLP
339 proteins with subclades from all other RGA domains. This clade appears to be a sister clade
340 to clade 3, which contains some RLP, TM-CC, and most especially, numerous RLK proteins
341 which formed a major nested subclade indicating their close relationship with RLPs. The
342 clades 4 and 5 appeared to be sister clades and are composed mainly of NBS-encoding
343 proteins with subclades from all other RGA domains and most especially, from TM-CC
344 domains. Rody et al. (2019) also observed that TM-CC sequences in sugarcane form
345 widespread subclades across the phylogenetic tree and showed close relationship with NBS-
346 encoding proteins. The clade 6 is composed predominantly of RLKs and shared close
347 relationship to some other RGAs. In this clade, NBS-encoding proteins that shared close
348 similarity related to RLKs were observed to form a major nested subclade. The NBS and RLKs
349 are among the largest and diversified group of RGAs in plants (Hulbert et al., 2001; Meyers et
350 al., 2003; Tör et al., 2009, Li et al., 2016), which also holds true in mango based on this study.
351 The distribution of RGAs in the phylogenetic tree may indicate that their functions are highly
352 divergent and form clusters that may be related to functions but not necessarily in protein
353 sequence (Chang et al., 2002; Zhang et al., 2016). This could be the case in mango RGAs,
354 especially the NBS-encoding proteins which are broadly distributed into 3 major clades (clades
355 1, 4 and 5) and 1 major subclade (nested in clade 6), as well as in RLKs which formed a
356 major clade (clade 6) and a major subclade (nested in clade 3).

357 Many factors have been attributed to the evolutionary pattern and diversity of RGAs in
358 plants. One of which is the well-known co-evolutionary "arms race" model between the host

359 plant and associated pests and diseases which drives the selective forces to overcome each
360 other (Anderson et al., 2010; Edger et al., 2015; Boller and He, 2009). Other selection
361 pressures such as climatic conditions (e.g. rainfall, temperatures, humidity, etc.) that favors
362 pest and disease development have been correlated also to the diversity of RGAs (Sela et al.,
363 2009). Whole-genome duplications (WGDs) and genomic reorganizations that occurred
364 during ancient times, especially in angiosperms, have been associated to the expansion of
365 RGA families and in creating new gene functions (Perazzolli et al., 2014, Chagné, 2015;
366 Michelmore and Meyers, 1998). It is likely that in the evolutionary history of mango, it had
367 undergone at least one round of polyploidization i.e., allopolyploidy which is a type of WGD
368 via hybridization followed by genome doubling (Glover et al., 2016; Soltis and Soltis, 2009).
369 These are some factors that could have influenced the complex phylogenetic structure of
370 RGAs of mango, given also that the plant's tropical growing condition provides a thriving
371 environment to a plethora of insect pests and diseases.
372
373 ***RGA-derived EST-SSRs of mango***
374 Using the RGA transcripts (Supplemental File 5), a total of 151 EST-SSR (Expressed
375 Sequence Tags – Simple Sequence Repeat) loci were identified. Among these, 134 loci
376 yielded unique potential markers. Mostly di-repeat motifs (2-mer) were observed, accounting
377 for 64.24% of the markers, followed by tri-repeats (3-mer) at 30.46%. The remaining 5.3% is
378 composed of tetra-, hexa-, and penta-repeat motifs. AT repeat is the most represented motif
379 among the di-repeats at 12.5%, followed by TC and TA at 11.26% each (Fig. 7). Meanwhile,
380 AG/CT is the most abundant paired repeat motif among the SSR loci analyzed at 16.56% (Fig.
381 8). Data on the distribution of *R* gene motif-linked SSRs (Fig. 9) revealed that RLK has the
382 most number of EST-SSRs designed with 67 markers, followed by RLP and TM-CC motifs at
383 34 and 18 markers, respectively. This distribution corresponds directly to the amount of *R*
384 genes per category identified in the whole mango RGA transcriptome.
385 Similar *in silico* approaches in EST-SSR marker development were performed by other
386 studies focusing on other plants, such as bamboo (Cai et al., 2019), mint (Kumar et al., 2015),

387 castor (Thatikunta et al., 2016), Ethiopian potato (Gadissa et al., 2018) and *Elymus* species
388 (Zhang et al., 2019). The EST-derived SSR markers generated in this study can be used for
389 functional analysis of *R* genes, once verified by wet lab experiments. Moreover, these markers
390 may also be used in genotyping, genetic diversity, linkage mapping, gene-based association
391 studies and marker-assisted selection in mango breeding. The sequences of the designed
392 forward and reverse primers derived from mango RGAs are found in Supplemental Table 4.
393

394 **Summary and Conclusions**

395 In this paper, 747 core mango RGAs were identified and broadly categorized into four
396 major families based on their conserved structural features: NBS-encoding proteins (155
397 proteins), RLPs (158), RLKs (362) and TM-CC (72). In addition, the mango RGAs have been
398 functionally well-annotated revealing the involved biological processes, molecular functions
399 and cellular components. This provides an important insight to the overall functional response
400 of mango against insect pests and diseases. Moreover, the evolutionary relationships,
401 diversity, and expression profiles across different mango varieties were presented as an
402 invaluable reference in the design of effective framework for the actual mango resistance
403 breeding programs. A marker resource, comprising of 134 RGA-linked EST-SSR markers,
404 was also developed to aid in the discovery of associated RGA/s to a specific disease or insect
405 pest through population genetics analyses (e.g., QTL mapping, association mapping). The
406 basic information from the mango RGAs and the linked EST-SSR markers established in this
407 paper will facilitate in the development of outstanding mango varieties with resistance to
408 various pests and diseases.
409

410 **Acknowledgements**

411 The authors would like to thank the Department of Science and Technology-Philippine
412 Council for Agriculture, Aquatic and Natural Resources Research and Development (DOST-
413 PCAARRD) for funding the project entitled “Characterization of ‘Carabao’ and other Mango
414 Varieties with Resistance to Fruit Fly and Anthracnose”. Also, the University of the Philippines

Los Baños for the core-funded project, “Genome-wide analysis of resistance gene analogs (RGAs) and development of RGA-linked DNA markers in Philippine Important Crops”.

Conflict of interest

The authors declare no conflict of interest.

References

Aboul-Soud, M. A., Chen, X., Kang, J. G., Yun, B. W., Raja, M. U., Malik, S. I., & Loake, G. J. (2009). Activation tagging of ADR2 conveys a spreading lesion phenotype and resistance to biotrophic pathogens. *New phytologist*, *183*(4), 1163-1175.

Acevedo, F. E., Rivera-Vega, L. J., Chung, S. H., Ray, S., & Felton, G. W. (2015). Cues from chewing insects—the intersection of DAMPs, HAMPs, MAMPs and effectors. *Current Opinion in Plant Biology*, *26*, 80-86.

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010). Plants versus pathogens: an evolutionary arms race. *Functional plant biology*, *37*(6), 499-512.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., & Wishart, D. S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic acids research*, *44*(W1), W147-W153.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 1994, 2, 28–36.
Bally, I. S. (2006). Mangifera indica (mango). *Species profiles for pacific island agroforestry*, 1-25.

Bittner-Eddy, P. D., Crute, I. R., Holub, E. B., & Beynon, J. L. (2000). RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. *The Plant Journal*, *21*(2), 177-188.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Boller, T., and He, S. Y. (2009). Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science*, *324*(5928), 742-744.

Borhan, M. H., Gunn, N., Cooper, A., Gulden, S., Tör, M., Rimmer, S. R., & Holub, E. B. (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. *Molecular Plant-Microbe Interactions*, *21*(6), 757-768.

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. *Frontiers in plant science*, *10*, 50.

Chagné, D. (2015). Whole genome sequencing of fruit tree species. In *Advances in botanical research* (Vol. 74, pp. 1-37). Academic Press.

Chang, J. H., Tai, Y. S., Bernal, A. J., Lavelle, D. T., Staskawicz, B. J., & Michelmore, R. W. (2002). Functional analyses of the Pto resistance gene family in tomato and the identification of a minor resistance determinant in a susceptible haplotype. *Molecular plant-microbe interactions*, *15*(3), 281-291.

Choi, H. W., & Klessig, D. F. (2016). DAMPs, MAMPs, and NAMPs in plant innate immunity. *BMC plant biology*, *16*(1), 232.

Conesa, A., Götz, S., García-Gómez, J. M., 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.

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annual Review of Plant Biology, 66(1), 487–511. doi:10.1146/annurev-arplant-050213-040012

Dangl, J. L., & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature*, *411*(6839), 826–833. doi:10.1038/35081161

Davidson, N. M., Hawkins, A. D., & Oshlack, A. (2017). SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome biology*, *18*(1), 148.

Dharmasiri, N., Dharmasiri, S., & Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature*, *435*(7041), 441-445.

Dinesh-Kumar, S. P., Tham, W. H., & Baker, B. J. (2000). Structure–function analysis of the tobacco mosaic virus resistance gene N. *Proceedings of the National Academy of Sciences*, *97*(26), 14789-14794.

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., ... & Hofberger, J. A. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, *112*(27), 8362-8366.

Eitas, T. K., Nimchuk, Z. L., & Dangl, J. L. (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.

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 molecular biology*, *37*(6), 943-953.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., ... & Holm, L. (2010). The Pfam protein families database. *Nucleic acids research*, *38*(suppl 1), D211-D222.

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.

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 genetics*, *19*(1), 92.

Glover, N. M., Redestig, H., & Dessimoz, C. (2016). Homoeologs: what are they and how do we infer them?. *Trends in plant science*, *21*(7), 609-621.

Goodwin, S., Mcpherson, J. D., & Mccombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333-351.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, *29*(7), 644.

Gururani, M.A.; Venkatesh, J.; Upadhyaya, C.P.; Nookaraju, A.; Pandey, S.K.; Park, S.W. Plant disease resistance genes: Current status and future directions. *Physiol. Mol. Plant Pathol.* 2012, *78*, 51–65.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), 1494.

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. *The Plant Journal*, *67*(6), 980-992.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, *35*(2), 518-522.

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.

Hu, Q., Min, L., Yang, X., Jin, S., Zhang, L., Li, Y., ... & Lindsey, K. (2018). Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. *Plant physiology*, *176*(2), 1808-1823.

Hulbert, S. H., Webb, C. A., Smith, S. M., & Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annual review of phytopathology*, *39*(1), 285-312.

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. Nature, 444(7117), 323–329.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, *338*(5), 1027-1036.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, *14*(6), 587-589.

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. *Scientia Horticulturae*, *248*, 81-88.

566
567 Kim, S. H., Kwon, S. I., Saha, D., Anyanwu, N. C., & Gassmann, W. (2009). Resistance to the
568 Pseudomonas syringae effector HopA1 is governed by the TIR-NBS-LRR protein RPS6 and
569 is enhanced by mutations in SRFR1. *Plant physiology*, *150*(4), 1723-1732.
570
571 Kong, Q., Qu, N., Gao, M., Zhang, Z., Ding, X., Yang, F., ... & Zhang, Y. (2012). The MEKK1-
572 MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-
573 activated protein kinase kinase kinase in Arabidopsis. *The Plant Cell*, *24*(5), 2225-2236.
574
575 Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design
576 program Primer3. *Bioinformatics*, *23*(10), 1289-1291.
577
578 Kuang, H., Wei, F., Marano, M. R., Wirtz, U., Wang, X., Liu, J., ... & Lobst, S. (2005). The R1
579 resistance gene cluster contains three groups of independently evolving, type I R1
580 homologues and shows substantial structural variation among haplotypes of Solanum
581 demissum. *The Plant Journal*, *44*(1), 37-51.
582
583 Kumar, B., Kumar, U., & Yadav, H. K. (2015). Identification of EST–SSRs and molecular
584 diversity analysis in Mentha piperita. *The Crop Journal*, *3*(4), 335-342.
585
586 Langmead, B., Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat*
587 *Methods* 9, 357–359.
588 Lantican, D. V., Strickler, S. R., Canama, A. O., Gardoce, R. R., Mueller, L. A., & Galvez, H.
589 F. (2019). De novo genome sequence assembly of dwarf coconut (Cocos nucifera L.'Catigan
590 Green Dwarf') provides insights into genomic variation between coconut types and related
591 palm species. *G3: Genes, Genomes, Genetics*, *9*(8), 2377-2393.
592
593 Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data
594 with or without a reference genome. *BMC bioinformatics*, *12*(1), 323.
595
596 Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., & You, F. M. (2016). RGAugury: a pipeline for
597 genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC genomics*, *17*(1),
598 852.
599
600 Lokossou, A. A., Rietman, H., Wang, M., Krenek, P., van der Schoot, H., Henken, B., ... &
601 Jacobsen, E. (2010). Diversity, distribution, and evolution of Solanum bulbocastanum late
602 blight resistance genes. *Molecular Plant-Microbe Interactions*, *23*(9), 1206-1216.
603
604 Lolle, S., Greeff, C., Petersen, K., Roux, M., Jensen, M. K., Bressendorff, S., ... & Petersen,
605 M. (2017). Matching NLR immune receptors to autoimmunity in camta3 mutants using
606 antimorphic NLR alleles. *Cell host & microbe*, *21*(4), 518-529.
607
608 Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein
609 sequences. *Science*, 1162-1164.
610
611 Mackey, D., Holt III, B. F., Wiig, A., & Dangl, J. L. (2002). RIN4 interacts with Pseudomonas
612 syringae type III effector molecules and is required for RPM1-mediated resistance in
613 Arabidopsis. *Cell*, *108*(6), 743-754.
614
615 Maria Cristina Suarez Rodriguez, Petersen, M., & Mundy, J. (2010). Mitogen-activated protein
616 kinase signaling in plants. *Annual review of plant biology*, *61*, 621-649.
617
618 Masle, J., Gilmore, S. R., & Farquhar, G. D. (2005). The ERECTA gene regulates plant
619 transpiration efficiency in Arabidopsis. *Nature*, *436*(7052), 866-870.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell*, *15*(4), 809-834.

Michelmore, R. W., & Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome research*, *8*(11), 1113-1130.

Müller, T., & Vingron, M. (2004). Modeling amino acid replacement. *Journal of Computational Biology*, *7*(6), 761-776.

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.

Neupane, S., Andersen, E. J., Neupane, A., & Nepal, M. P. (2018). Genome-Wide Identification of NBS-Encoding Resistance Genes in Sunflower (Helianthus annuus L.). *Genes*, *9*(8), 384.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*(1), 268-274.

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., ... & 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. *The Plant Journal*, *66*(3), 467-479.

Parker, J. E., Coleman, M. J., Szabò, V., Frost, L. N., Schmidt, R., van der Biezen, E. A., ... & Jones, J. D. (1997). The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. *The Plant Cell*, *9*(6), 879-894.

Perazzolli, M., Malacarne, G., Baldo, A., Righetti, L., Bailey, A., Fontana, P., ... & Malnoy, M. (2014). Characterization of resistance gene analogues (RGAs) in apple (Malus× domestica Borkh.) and their evolutionary history of the Rosaceae family. *PloS one*, *9*(2).

Pieterse, C. M., Van der Does, D., Zamioudis, C., Leon-Reyes, A., & Van Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annual review of cell and developmental biology*, *28*, 489-521.

Piofczyk, T., Jeena, G., & Pecinka, A. (2015). Arabidopsis thaliana natural variation reveals connections between UV radiation stress and plant pathogen-like defense responses. *Plant Physiology and Biochemistry*, *93*, 34-43.

Pitzschke, A. (2013). Agrobacterium infection and plant defense—transformation success hangs by a thread. Frontiers in Plant Science, 4(December), 1–12.

Rambaut, A. (2018) FigTree v1.4.4. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed February 7 2020).

Rody, H. V., Bombardelli, R. G., Creste, S., Camargo, L. E., Van Sluys, M. A., & Monteiro-Vitorello, C. B. (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.

Sánchez-Rodríguez, C., Estévez, J. M., Llorente, F., Hernández-Blanco, C., Jordá, L., Pagán, I., ... & Molina, A. (2009). The ERECTA receptor-like kinase regulates cell wall–mediated

675 resistance to pathogens in Arabidopsis thaliana. *Molecular plant-microbe interactions*, *22*(8),
676 953-963.
677
678 Saucet, S. B., Ma, Y., Sarris, P. F., Furzer, O. J., Sohn, K. H., & Jones, J. D. (2015). Two
679 linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial
680 effector AvrRps4. *Nature communications*, *6*(1), 1-12.
681
682 Sekhwal, M. K., Li, P., Lam, I., Wang, X., Cloutier, S., & You, F. M. (2015). Disease resistance
683 gene analogs (RGAs) in plants. *International journal of molecular sciences*, *16*(8), 19248-
684 19290.
685
686 Sela, H., Cheng, J., Jun, Y., Nevo, E., & Fahima, T. (2009). Divergent diversity patterns of
687 NBS and LRR domains of resistance gene analogs in wild emmer wheat
688 populations. *Genome*, *52*(6), 557-565.
689
690 Shiu, S. H., & Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family
691 and receptor-like proteins in Arabidopsis. *Plant physiology*, *132*(2), 530-543.
692
693 Sinapidou, E., Williams, K., Nott, L., Bahkt, S., Tör, M., Crute, I., ... & Beynon, J. (2004). Two
694 TIR: NB: LRR genes are required to specify resistance to Peronospora parasitica isolate Cala2
695 in Arabidopsis. *The Plant Journal*, *38*(6), 898-909.
696
697 Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review*
698 *of plant biology*, *60*, 561-588.
699
700 Soubrier, J., Steel, M., Lee, M. S., Der Sarkissian, C., Guindon, S., Ho, S. Y., & Cooper, A.
701 (2012). The influence of rate heterogeneity among sites on the time dependence of molecular
702 rates. *Molecular Biology and Evolution*, *29*(11), 3345-3358.
703
704 Staal, J., Kaliff, M., Bohman, S., & Dixelius, C. (2006). Transgressive segregation reveals two
705 Arabidopsis TIR-NB-LRR resistance genes effective against Leptosphaeria maculans, causal
706 agent of blackleg disease. *The Plant Journal*, *46*(2), 218-230.
707
708 Staal, J., Kaliff, M., Dewaele, E., Persson, M., & Dixelius, C. (2008). RLM3, a TIR domain
709 encoding gene involved in broad-range immunity of Arabidopsis to necrotrophic fungal
710 pathogens. *The Plant Journal*, *55*(2), 188-200.
711
712 Tarr, D. E. K., & Alexander, H. M. (2009). TIR-NBS-LRR genes are rare in monocots: evidence
713 from diverse monocot orders. *BMC research notes*, *2*(1), 197.
714
715 Thatikunta, R., Sankar, A. S., Sreelakshmi, J., Palle, G., Leela, C., Rani, C. V. D., ... &
716 Dudhe, M. Y. (2016). Utilization of in silico EST–SSR markers for diversity studies in castor
717 (Ricinus communis L.). *Physiology and molecular biology of plants*, *22*(4), 535-545.
718
719 Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity
720 of progressive multiple sequence alignment through sequence weighting, position-specific
721 gap penalties and weight matrix choice. *Nucleic acids research*, *22*(22), 4673-4680.
722
723 Thompson, J. N., and Burdon, J. J. (1992). Gene-for-gene coevolution between plants and
724 parasites. *Nature*, *360*(6400), 121-125.
725
726 Tör, M., Lotze, M. T., & Holton, N. (2009). Receptor-mediated signalling in plants: molecular
727 patterns and programmes. *Journal of experimental botany*, *60*(13), 3645-3654.
728

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115-e115.

Wan, L., Essuman, K., Anderson, R. G., Sasaki, Y., Monteiro, F., Chung, E. H., ... & Nishimura, M. T. (2019). TIR domains of plant immune receptors are NAD+-cleaving enzymes that promote cell death. *Science*, *365*(6455), 799-803.

Wang, X., & Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in plant science*, *7*, 1350.

Warren, R. F., Henk, A., Mowery, P., Holub, E., & Innes, R. W. (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. *The Plant Cell*, *10*(9), 1439-1452.

Williams, M.E. (June 16, 2011). Introduction to Phytohormones. Teaching Tools in Plant Biology: Lecture Notes. The Plant Cell (online), doi/10.1105/tpc.110.tt0310.

Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M., & Turner, J. G. (1998). COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. *Science*, *280*(5366), 1091-1094.

Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics*, *139*(2), 993-1005.

Zbierzak, A. M., Porfirova, S., Griebel, T., Melzer, M., Parker, J. E., & Dörmann, P. (2013). A TIR–NBS protein encoded by A rabidopsis C hilling S ensitive 1 (CHS 1) limits chloroplast damage and cell death at low temperature. *The Plant Journal*, *75*(4), 539-552.

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9), 847-848.

Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., ... & Pan, Q. (2011). The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytologist*, *189*(1), 321-334.

Zhang, X., Liang, P., & Ming, R. (2016). Genome-wide identification and characterization of nucleotide-binding site (NBS) resistance genes in pineapple. *Tropical Plant Biology*, *9*(3), 187-199.

Zhang, Z., Wu, Y., Gao, M., Zhang, J., Kong, Q., Liu, Y., ... & 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.

Zhang, Z., Xie, W., Zhao, Y., Zhang, J., Wang, N., Ntakirutimana, F., ... & 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 biology*, *19*(1), 235.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., 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.

783 Table 1. List of source transcriptome sequences and their predicted number of RGAs.
784

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	**7391**

Fr: Fruit; Fl: Flower, L: Leaf, S: Seed; B: Buds; St: Stem; R: Roots

794 Table 2. The distribution of core RGAs of mango and its comparison to other major crops.

RGA codes		Mango (*Mangifera indica*)[a]	Cacao (*Theobroma cacao*)[b]	Papaya (*Carica papaya*)[b]	Tomato (*Solanum lycopersicum*)[b]	Corn (*Zea mays*)[b]	Rice (*Oryza sativa*)[b]	Banana (*Musa acuminata*)[b]
NBS	NBS	53	16	11	64	37	47	11
CC-NBS-LRR	CNL	17	173	10	70	52	222	50
TIR-NBS-LRR	TNL	2	14	4	19	0	0	0
CC-NBS	CN	29	13	5	8	23	38	7
TIR-NBS	TN	4	3	1	7	2	3	2
NBS-LRR	NL	27	66	21	91	49	196	44
TIR with unknown domain	TX	17	8	5	12	3	3	3
Other NBS-encoding	Other	6	0	3	1	1	0	0
RLP	RLP	158	167	27	86	63	118	107
RLK	RLK	362	640	299	512	671	867	530
TM-CC	TM-CC	72	181	81	133	161	158	138
Total RGAs		**747**	**1,281**	**467**	**1,003**	**1,062**	**1,652**	**892**

795 [a] inferred from this paper based on transcriptome data; [b] from Li et al (2016) based on genome-wide RGA analysis

796

797 Table 3. Some mango RGAs with homology to known Resistance (*R*)/defense proteins
798 against pathogens and insects with homology to the entries in Uniprot/SwissProt database
799 based on BLASTp hits.

Mango *R*/defense protein	RGA domain	e-Value	R/defense protein name/description	Resistance/ defense response to
Zill_394_c0_g1.p2	NBS	2.41E-42	Putative disease resistance protein At4g11170	Pathogen and insect
SRR1448302_13791_c0_g1.p1	NL	3.90E-36		
SRR1448302_13356_c0_g1.p2	TN	2.76E-31		
SRR2159471_30596_c0_g1.p1	TX	1.18E-41		
SRR2162908_17959_c0_g1.p1	OTHER	7.44E-33		
SRR2162889_655_c0_g2.p3	NBS	1.22E-16	Disease resistance protein RRS1; Disease resistance protein RCH2; Probable WRKY transcription factor 52; Resistance to Colletotrichum higginsianum 2 protein; Resistance to Ralstonia solanacearum 1 protein	Pathogen
SRR2162878_10232_c0_g1.p1	NBS	8.95E-35	Disease resistance protein RRS1B; Probable WRKY transcription factor 16; WRKY DNA-binding protein 16	Pathogen
Zill_10069_c0_g1.p1	NBS	1.55E-20	TMV resistance protein N	Pathogen
Zill_15696_c0_g1.p1	NL	6.92E-48		
Zill_16228_c0_g1.p1	TN	2.52E-65		
Zill_4100_c0_g1.p1	TX	2.37E-31		
SRR1448302_8769_c0_g1.p1	CN	5.82E-43		
SRR1448302_10123_c0_g2.p1	TNL	0		
SRR2159471_13808_c0_g1.p1	OTHER	1.55E-53		
SRR3359450_14591_c0_g1.p1	RLP	9.35E-39		
Zill_7264_c0_g1.p1	TX	4.74E-39	Disease resistance protein TAO1; Protein TARGET OF AVRB OPERATION 1	Pathogen
SRR1448302_10424_c0_g1.p1	CN	2.98E-43	Disease resistance protein SUMM2; Disease resistance protein At1g12280; Protein SUPPRESSOR OF MKK1 MKK2 2	Pathogen
SRR1448302_12885_c0_g1.p1	NBS	1.28E-10		
SRR2162929_3892_c0_g1.p1	NL	1.04E-98		
Zill_3443_c0_g2.p1	NBS	5.51E-13	Disease resistance protein RFL1; RPS5-like protein 1; pNd13/pNd14	Pathogen
SRR2162919_15595_c0_g2.p2	CN	9.65E-21		
SRR2162929_6841_c0_g1.p1	NL	1.80E-28		
SRR5966306_7909_c0_g1.p1	TX	1.09E-21	Disease resistance protein RPS6; Resistance to Pseudomonas syringae 6	Pathogen
SRR8449858_33757_c0_g1.p1	NBS	1.44E-17	Disease resistance protein RPS6; Resistance to Pseudomonas syringae 6	Pathogen
SRR1448302_7186_c0_g1.p1	NBS	3.86E-36	Disease resistance protein RPS5; Resistance to Pseudomonas syringae protein 5; pNd3/pNd10	Pathogen
SRR2162889_11315_c0_g1.p1	CNL	0		
SRR2162953_17628_c0_g1.p1	NL	2.01E-42		
Zill_10085_c0_g1.p1	NBS	9.29E-55	Disease resistance protein RPS2; Resistance to Pseudomonas syringae protein 2	Pathogen
SRR2162908_5437_c0_g1.p1	NL	4.04E-62		
SRR2162919_3439_c1_g1.p1	CNL	9.91E-78		
SRR2162919_4996_c0_g1.p2	RLP	5.55E-14		
Zill_5583_c0_g1.p1	CN	3.98E-21	Disease resistance protein RPM1; Resistance to Pseudomonas syringae protein 3	Pathogen
Zill_10742_c0_g1.p1	NBS	1.36E-23		
Zill_30931_c0_g1.p1	NL	1.80E-50		
SRR1448302_9780_c0_g1.p1	CNL	4.05E-96		
SRR2159471_6253_c0_g1.p1	TX	2.23E-28	Disease resistance protein RPP5; Protein RECOGNITION OF PERONOSPORA PARASITICA 5	Pathogen and insect
SRR3319054_18081_c0_g1.p1	OTHER	1.42E-35		
Zill_4463_c0_g1.p2	RLP	4.77E-07	Probable disease resistance protein At4g19530	Pathogen
SRR2162919_8948_c0_g2.p1	TN	3.95E-30	Probable disease resistance protein At4g19520	Pathogen

SRR1448302_1069_c0_g1.p1	RLP	9.17E-55	Disease resistance protein RPP2B	Pathogen
SRR2159471_12447_c0_g2.p1	NL	5.12E-101		
SRR2162953_222_c0_g1.p2	NL	7.60E-45	Disease resistance protein RPP2A	Pathogen
Zill_19434_c0_g1.p1	NBS	1.93E-53	Disease resistance protein RPP13; Resistance to Peronospora parasitica protein 13	Pathogen
Zill_10742_c0_g1.p3	NL	2.70E-20		
SRR2159471_23664_c0_g1.p1	CN	3.57E-29		
SRR2736811_8634_c0_g1.p1	CNL	4.58E-170		
SRR2162889_14411_c0_g1.p1	NBS	2.47E-24	Probable disease resistance protein RPP1; Protein RECOGNITION OF PERONOSPORA PARASITICA 1	Pathogen
SRR2162953_10793_c0_g1.p1	TX	2.59E-33		
SRR3288569_10876_c0_g3.p1	CN	2.07E-19	Probable disease resistance RPP8-like protein 2	Pathogen
SRR2162929_192_c0_g1.p3	TX	4.57E-38	Disease resistance protein RLM3; Protein RESISTANCE TO LEPTOSPHAERIA MACULANS 3	Pathogen
SRR2162878_1969_c0_g1.p1	RLP	5.14E-23	Disease resistance protein RML1B; Protein RESISTANCE TO LEPTOSPHAERIA MACULANS 1B	Pathogen
SRR2162919_4480_c1_g1.p1	NBS	2.22E-20		
SRR2162953_20108_c0_g1.p1	TX	3.93E-32		
SRR2162878_10232_c0_g1.p2	TX	3.23E-26	Disease resistance protein RML1A; Protein RESISTANCE TO LEPTOSPHAERIA MACULANS 1A	Pathogen
SRR2159471_14993_c0_g1.p1	OTHER	1.15E-32	Disease resistance-like protein DSC2; Protein DOMINANT SUPRESSOR OF CAMTA3 NUMBER 2	Pathogen and insect
SRR2159471_10112_c0_g1.p1	TX	5.42E-43		
Zill_8229_c0_g1.p2	NBS	9.39E-19	Disease resistance-like protein DSC1; Protein DOMINANT SUPPRESSOR OF CAMTA3 NUMBER 1	Pathogen and insect
SRR1448302_13356_c0_g1.p1	TX	1.50E-47		
SRR2159471_16556_c0_g1.p1	RLP	5.16E-27		
SRR2162836_1731_c0_g2.p1	NL	1.04E-101		
SRR2162889_14475_c0_g2.p1	TN	2.61E-56		
SRR2162919_11945_c0_g1.p1	TNL	0		
SRR2162953_14082_c0_g1.p2	OTHER	1.51E-29		
SRR2736811_2638_c0_g1.p1	CNL	6.55E-123		
SRR3192873_14562_c0_g2.p1	OTHER	0		
SRR2159471_15747_c0_g1.p2	NBS	4.36E-19	Disease resistance protein Pik-2; Pik-2 blast resistance protein	Pathogen
SRR2159471_15747_c0_g1.p1	NBS	2.00E-26	Disease resistance protein RGA5; Os11gRGA5; SasRGA5	Pathogen
Zill_11941_c0_g1.p1	NBS	5.81E-29	Putative disease resistance protein RGA4; RGA4-blb	Pathogen
SRR1956775_13322_c0_g1.p1	CN	4.06E-28		
SRR2162908_1538_c0_g1.p1	CNL	4.56E-115		
SRR3288569_1324_c0_g1.p1	NL	3.37E-158		
Zill_2822_c0_g1.p1	CN	3.84E-91	Putative disease resistance protein RGA3; Blight resistance protein B149; RGA1-blb	Pathogen
Zill_2317_c0_g1.p5	NBS	1.35E-26		
SRR1448302_19774_c0_g1.p1	CNL	0		
SRR1448302_12653_c0_g1.p1	NL	1.12E-77		
Zill_40171_c0_g1.p1	NBS	2.32E-24	Disease resistance protein RGA2; Blight resistance protein RPI; RGA2-blb	Pathogen
Zill_5557_c0_g1.p1	NL	2.55E-22		
Zill_12068_c0_g1.p2	RLP	1.75E-05		
SRR1956775_2968_c1_g2.p1	CN	4.43E-57		
Zill_11208_c0_g1.p1	CN	3.21E-88	Putative disease resistance protein RGA1; RGA3-blb	Pathogen
Zill_2928_c3_g1.p1	NBS	1.21E-50		
Zill_4058_c0_g1.p1	NL	8.99E-92		
SRR2159471_1162_c2_g3.p1	CNL	5.33E-138		
SRR5966284_6355_c0_g1.p2	NBS	7.60E-35	Putative late blight resistance protein homolog R1A-10	Pathogen
SRR1448302_30603_c0_g1.p1	NL	0		Pathogen

SRR3359450_19530_c0_g1.p1	NBS	6.35E-94	Disease resistance RPP13-like protein 4; Disease resistance protein ZAR1; Protein HOPZ-ACTIVATED RESISTANCE 1	
SRR2159471_29188_c0_g1.p1	NBS	1.87E-64	Putative disease resistance RPP13-like protein 3	Pathogen
SRR2162836_25848_c0_g1.p1	CN	1.77E-26		
SRR2162836_22262_c0_g1.p1	NL	5.61E-100		
Zill_8566_c0_g1.p1	NBS	2.82E-22	Putative disease resistance RPP13-like protein 2	Pathogen
SRR2159471_14284_c0_g1.p1	CNL	1.84E-167		
SRR3192873_20475_c0_g1.p1	RLP	5.06E-42		
SRR3319054_34638_c0_g1.p1	NL	1.14E-20		
Zill_11317_c0_g1.p1	NBS	8.38E-21	Putative disease resistance RPP13-like protein 1	Pathogen
Zill_2317_c0_g1.p2	RLP	8.51E-48		
SRR1448302_10269_c0_g2.p1	CN	7.02E-63		
SRR1448302_1339_c0_g1.p1	CNL	4.16E-158		
SRR1448302_2626_c0_g1.p1	NL	0		
SRR2162836_19327_c0_g1.p1	TM-CC	4.13E-25		
SRR2159471_25754_c0_g1.p1	RLK	1.28E-92	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.5; Probable receptor-like serine/threonine-protein kinase LRK10L-2.5	Pathogen
SRR1448302_15170_c0_g1.p1	RLK	5.94E-71	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.4; Probable receptor-like serine/threonine-protein kinase LRK10L-2.4	Pathogen
SRR3319054_22570_c0_g1.p1	RLK	7.38E-41	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.2; Probable receptor-like serine/threonine-protein kinase LRK10L-2.2	Pathogen
SRR2159471_9027_c0_g2.p1	RLK	5.60E-151	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.1; Probable receptor-like serine/threonine-protein kinase LRK10L-2.1	Pathogen
SRR2162836_4545_c0_g1.p1	RLK	0	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.5; Probable receptor-like serine/threonine-protein kinase LRK10L-1.5	Pathogen
SRR1448302_13187_c0_g1.p1	RLK	1.62E-159	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.3; Probable receptor-like serine/threonine-protein kinase LRK10L-1.3	Pathogen
Zill_2156_c0_g1.p1	RLK	4.69E-14	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2; Probable receptor-like serine/threonine-protein kinase LRK10L-1.2	Pathogen
Zill_5127_c0_g2.p1	RLK	1.43E-111	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.1; Probable receptor-like serine/threonine-protein kinase LRK10L-1.1	Pathogen
Zill_13036_c0_g1.p1	RLK	1.86E-104	Rust resistance kinase Lr10; Probable receptor-like serine/threonine-protein kinase LRK10	Pathogen
SRR2162878_1544_c0_g2.p1	RLP	1.72E-119	LRR receptor-like serine/threonine-protein kinase ERECTA; Protein QUANTITATIVE RESISTANCE TO PLECTOSPHAERELLA 1; Protein QUANTITATIVE RESISTANCE TO RALSTONIA SOLANACEARUM 1; Protein TRANSPIRATION EFFICIENCY 1	Pathogen
SRR3288569_12948_c0_g1.p2	RLK	5.39E-133		
SRR2162878_2968_c0_g1.p1	RLK	2.12E-122	Serine/threonine-protein kinase EDR1; MAPKK kinase EDR1; Protein ENHANCED DISEASE RESISTANCE 1; AtEDR1; Serine/threonine/tyrosine-protein kinase 10	Pathogen
SRR2162889_24778_c0_g2.p1	NBS	1.15E-17	Disease resistance protein ADR2; Protein ACTIVATED DISEASE RESISTANCE 2; Protein WHITE RUST RESISTANCE 4	Pathogen
Zill_1197_c0_g1.p1	RLP	2.34E-85	Protein TRANSPORT INHIBITOR RESPONSE 1; Weak ethylene-insensitive protein 1	Pathogen and insect
SRR2162908_7086_c0_g1.p1	RLP	1.27E-144	Coronatine-insensitive protein 1; COI-1; F-box/LRR-repeat protein 2; AtCOI1; AtFBL2	Pathogen and insect

800

801 Table 4. Other mango RGAs with homology to proteins in Uniprot/SwissProt database based
802 on BLASTp hits and gene ontology analysis with “response to insects” annotation.
803

Mango contig ID	RGA domain	e-Value	Description
SRR2162908_12681_c0_g1.p1	RLK	0	Calmodulin-binding receptor kinase CaMRLK;
SRR8449851_23663_c0_g1.p1	RLP	6.40E-66	Calmodulin-binding receptor-like kinase; AtCaMRLK; Protein MATERNAL EFFECT EMBRYO ARREST 62
SRR5966284_6489_c0_g1.p1	RLK	4.04E-54	Cysteine-rich receptor-like protein kinase 18; Cysteine-rich RLK18
SRR2162953_2879_c0_g1.p2	RLK	7.43E-20	Putative cysteine-rich receptor-like protein kinase 43; Cysteine-rich RLK43
SRR5966306_10141_c0_g1.p1	RLK	3.68E-52	Cold-responsive protein kinase 1
SRR2162836_4413_c0_g3.p1	RLK	0	Putative receptor protein kinase ZmPK1
SRR2162836_3219_c0_g1.p1	RLK	0	Probable LRR receptor-like serine/threonine-protein kinase RFK1; Receptor-like kinase in flowers 1
SRR8449851_14359_c0_g1.p1	RLK	0	Probable LRR receptor-like serine/threonine-protein kinase RKF3; Receptor-like kinase in flowers 3
SRR2159471_39050_c1_g1.p1	RLK	0	Probable receptor-like protein kinase At1g11050
SRR3192873_14159_c0_g1.p1	RLK	0	C-type lectin receptor-like tyrosine-protein kinase At1g52310
SRR8449858_41086_c0_g1.p1	RLK	9.52E-66	Probable LRR receptor-like serine/threonine-protein kinase At1g53420
SRR1561197_1147_c0_g1.p1	RLK	8.21E-87	Probable LRR receptor-like serine/threonine-protein kinase At1g56130
SRR3192873_4970_c0_g1.p1	RLK	0	Probable LRR receptor-like serine/threonine-protein kinase At1g07650
SRR1448302_14204_c1_g1.p1	RLK	0	Probable leucine-rich repeat receptor-like serine/threonine-protein kinase At3g14840
Zill_8965_c0_g1.p1	RLK	1.47E-47	Probable LRR receptor-like serine/threonine-protein kinase At1g53430
SRR2159471_11990_c0_g1.p1	RLK	8.55E-62	Probable LRR receptor-like serine/threonine-protein kinase At5g45780
Zill_7246_c0_g2.p2	RLK	2.96E-117	Probable LRR receptor-like serine/threonine-protein kinase At1g56140

804

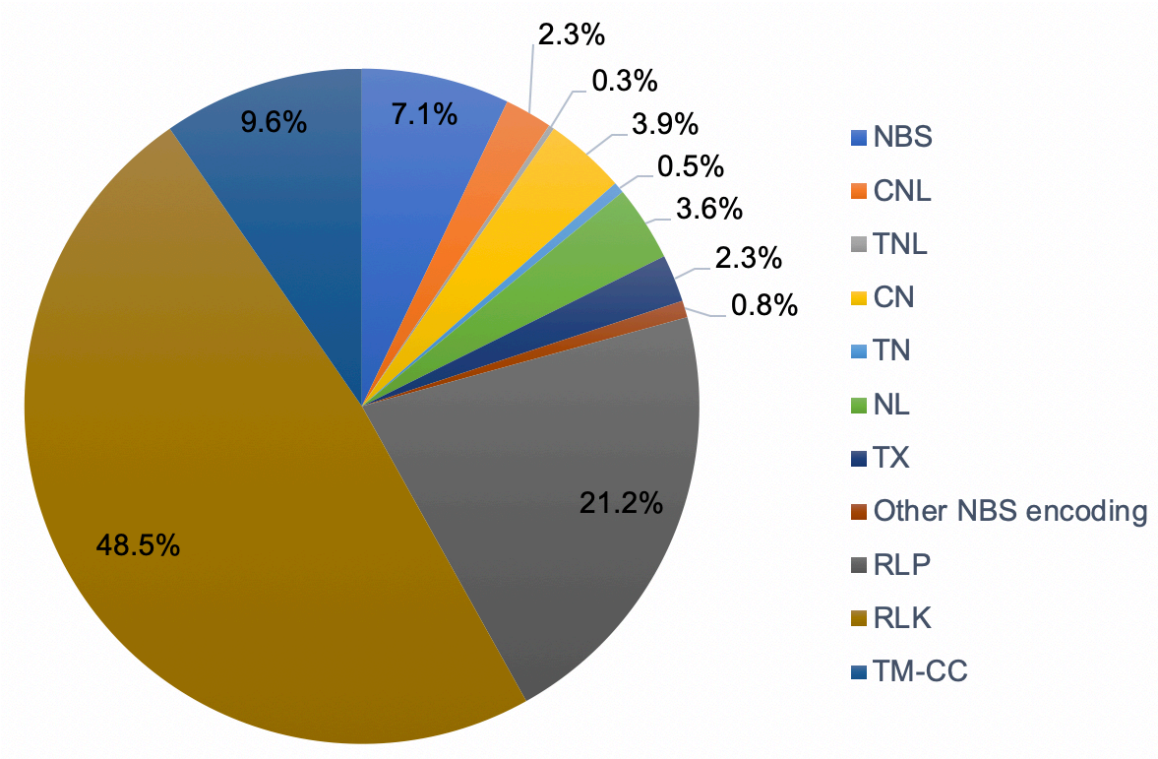

Figure 1. Distribution of the classifications based on conserved domains and motifs of core resistance gene analogs (RGA) detected in mango transcriptome *de novo* assemblies.

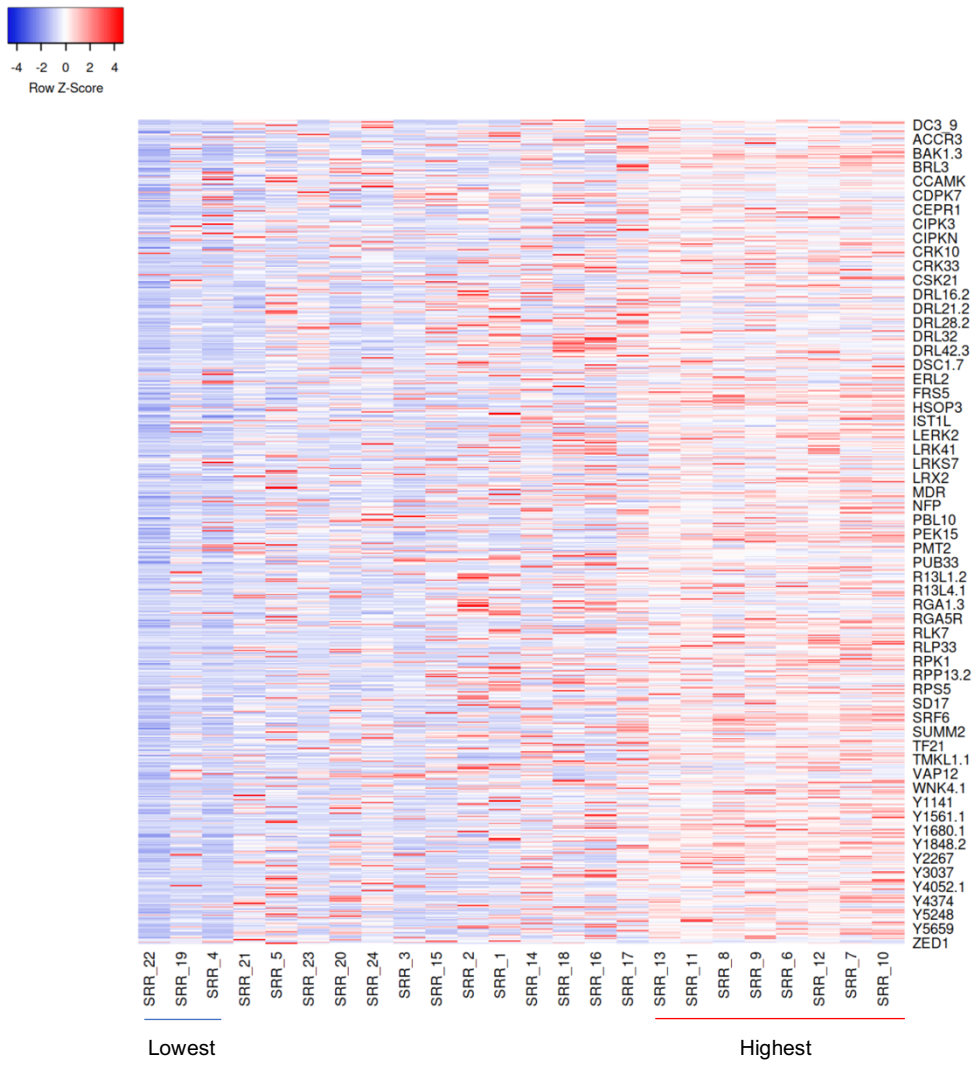

Figure 2. Resistance gene analogs (RGAs) expression profile from mango transcriptomes visualized as heatmap. Count values were normalized within samples as transcripts per million (TPM) using RSEM. Y-axis correspond to RGA gene IDs. X-axis correspond to RNA-seq raw read samples: *SRR_1. Hong et al., 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)*

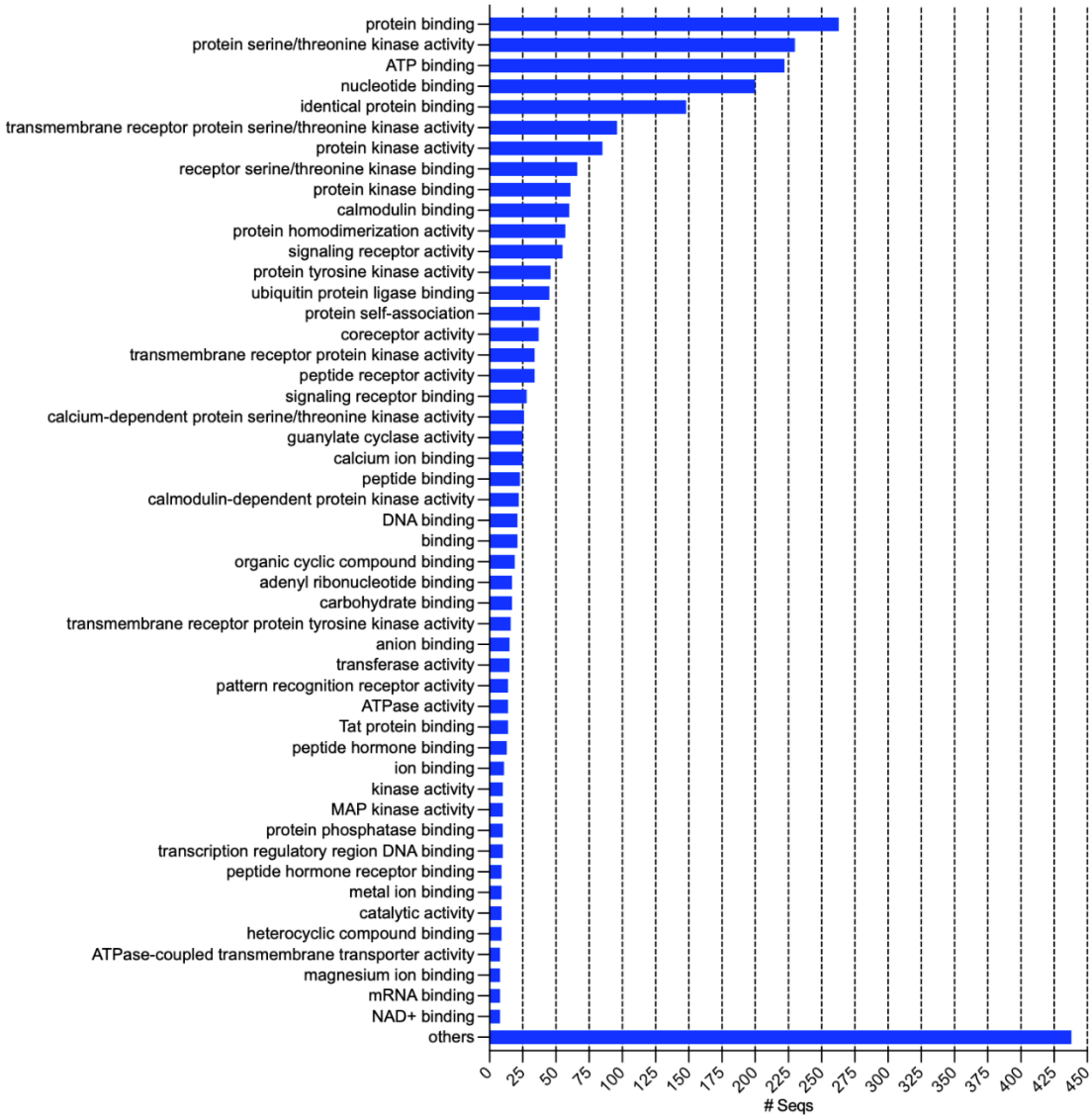

Figure 3. Gene ontology (GO) distribution of the expressed resistance gene analogs (RGAs) in mango based on molecular function (MF).

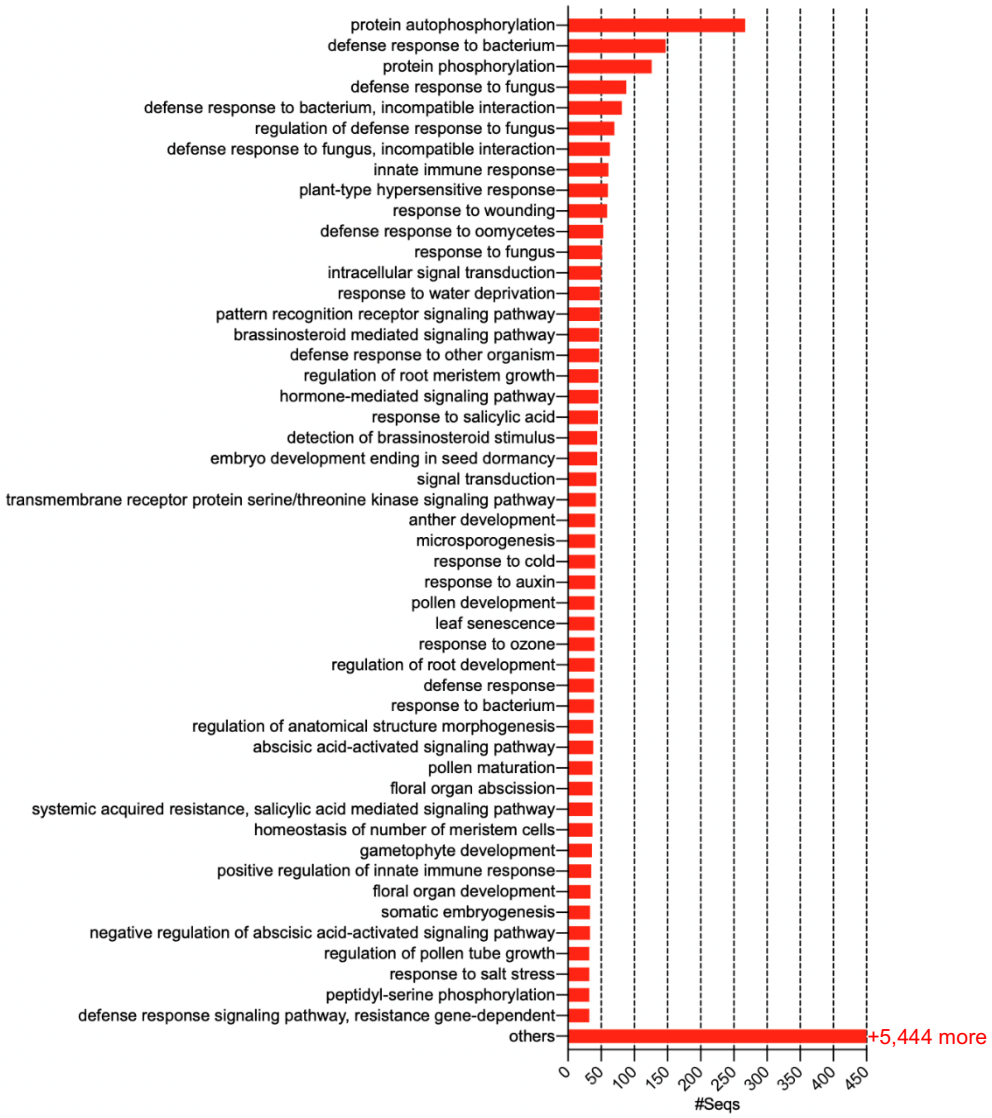

Figure 4. Gene ontology (GO) distribution of the expressed resistance gene analogs (RGAs) in mango based on biological processes (BP).

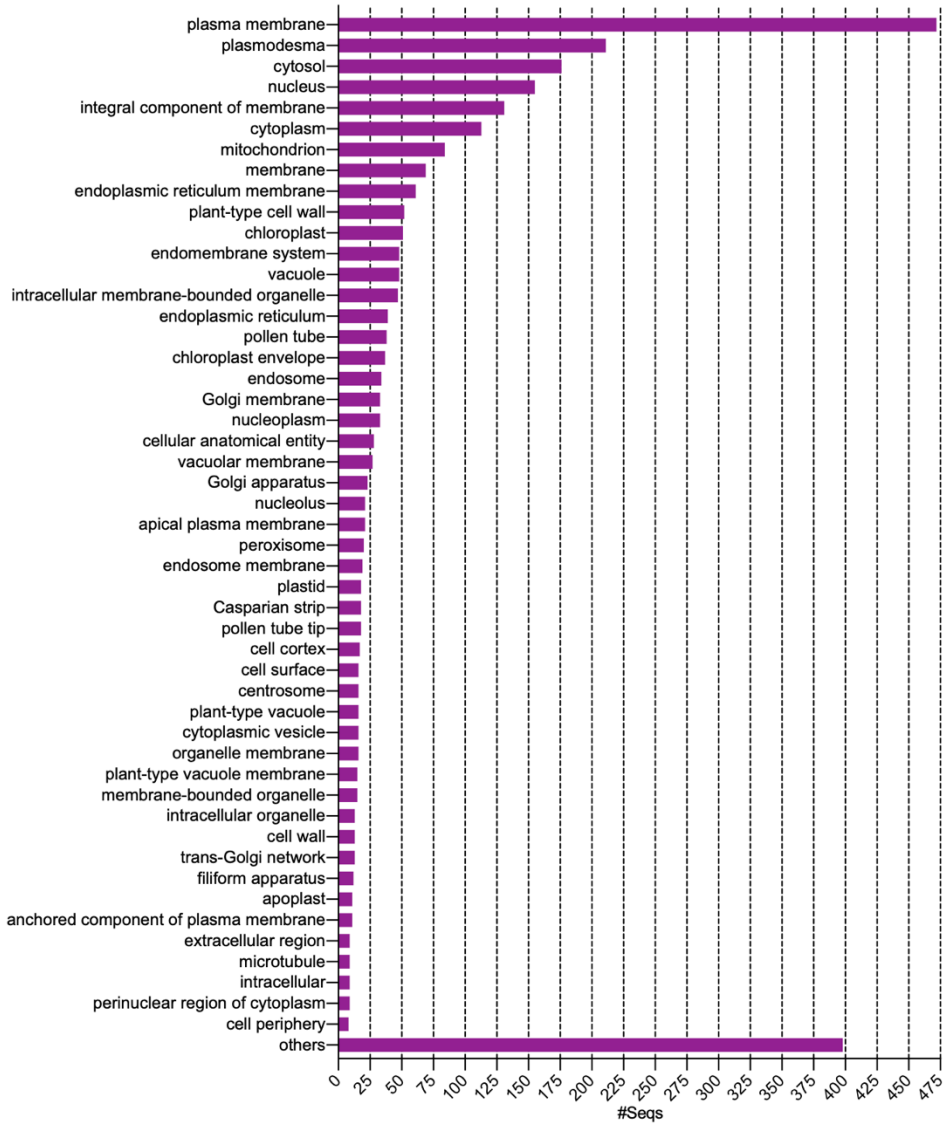

Figure 5. Gene ontology (GO) distribution of the expressed resistance gene analogs (RGAs) in mango based on cellular component (CC).

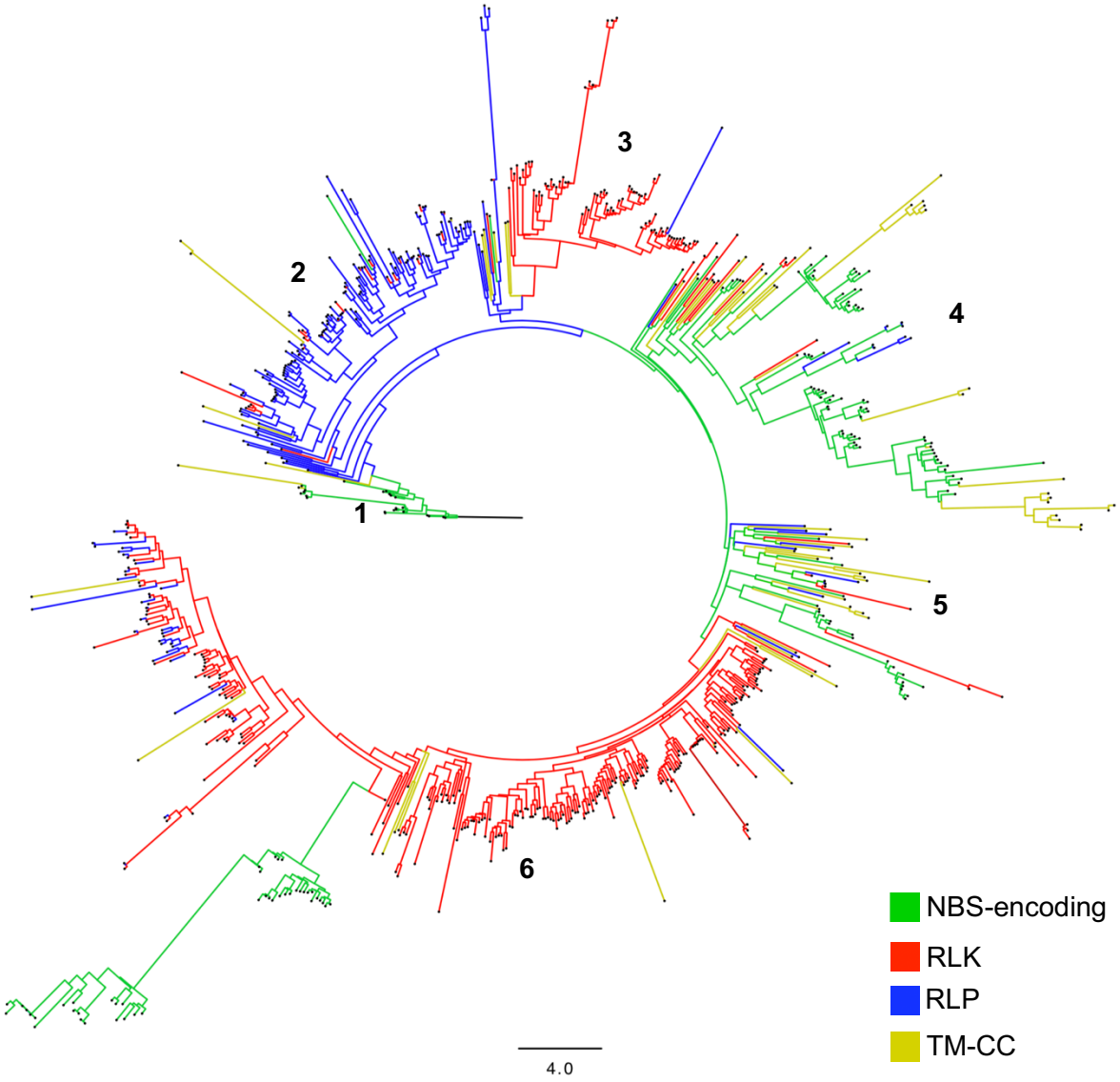

Figure 6. Maximum likelihood phylogenetic tree constructed using IQ-TREE from the sequence alignment of the core resistance gene analogs (RGAs) generated from *de novo* transcriptome assemblies of mango. Best-fit model was selected according to Bayesian Information Criterion (BIC) using ModelFinder (Kalyaanamoorthy et al., 2017). General 'Variable Time' (VT) matrix (Mueller and Vingron, 2004) with empirical amino acid frequencies (+F) and FreeRate (+R5) rate heterogeneity across sites (Yang 1995; Soubrier et al. 2012) was used to generate the tree, tested with 1000 replicates of ultrafast bootstrapping (Hoang et al. 2017). The numbers correspond to the clades.

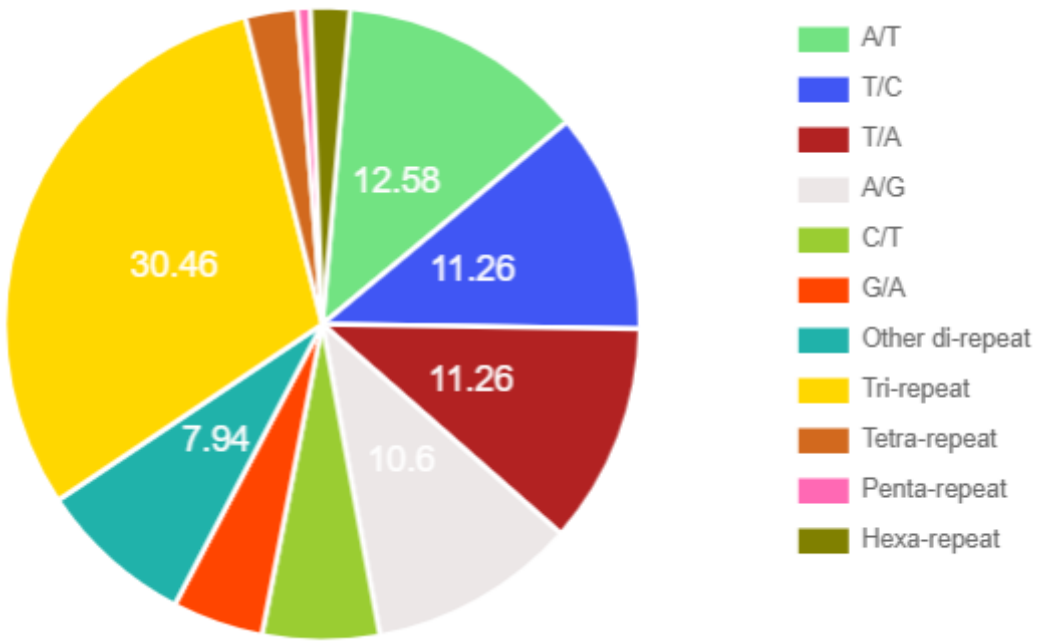

Figure 7. Distribution of the different repeat motifs that were detected in the mango RGA transcriptome.

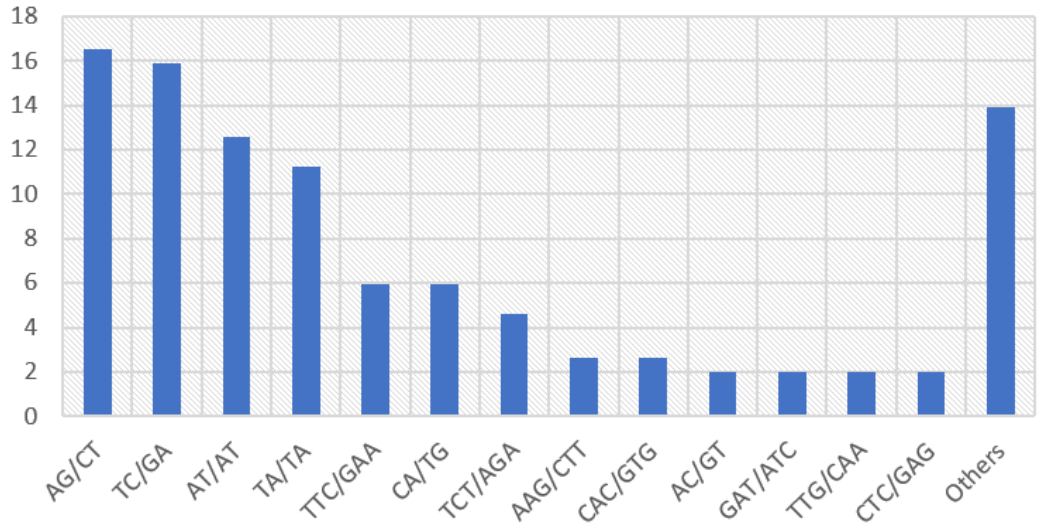

Figure 8. Paired-repeat SSR motif distribution of the mango RGA transcriptome.

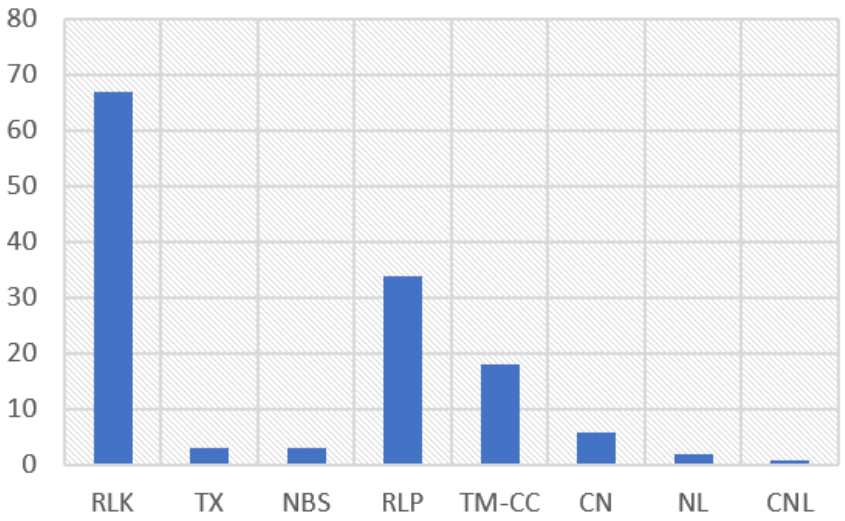

Figure 9. Distribution of EST-SSR markers designed across the different RGA domains identified in *de novo* mango transcriptomes.