SHORT COMMUNICATION

Identification and evidence of positive selection upon resistance gene analogs in cotton (*Gossypium hirsutum* L.)

Antonios Zambounis[1] · Ioannis Ganopoulos[2] · Apostolos Kalivas[3] ·
Athanasios Tsaftaris[1,4] · Panagiotis Madesis[4]

Abstracts Upland cotton (*Gossypium hirsutum* L.) is an important fiber crop species, which is intensively plagued by a plethora of phytopathogenic fungi such as *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*) causing severe wilt disease. Resistance gene analogs (*RGAs*) are the largest class of potential resistance (*R*) genes depicting highly conserved domains and structures in plants. Additionally, *RGAs* are pivotal components of breeding projects towards host disease resistance, serving as useful functional markers linked to *R* genes. In this study, a cloning approach based on conserved *RGAs* motifs was used in order to amplify 38 *RGAs* from two upland cotton cultivars differing in their *Fov* susceptibility. Besides, we assessed the phylogenetic expansion and the evolutionary pressures acting upon 127 *RGA* homologues, which were previously deposited in GenBank along with the 38 *RGAs* from this study. A total of 165 *RGAs* sequences were clustered according to their BLAST(P) similarities in ten paralogous genes groups (PGGs). These *RGAs* exhibited intensive signs of positive selection as it was revealed by inferring various maximum likelihood analyses. The results showed robust signs of positive selection, acting in almost all PGGs across the phylogeny. The evolutionary analysis revealed the existence of 42 positively selected residue sites across the PGG lineages, putatively affecting their ligand-binding specificities. As *RGAs* derived markers are in close linkage to *R* genes, these results could be used in ongoing breeding programs of upland cotton.

Keywords Fungal diseases · Molecular breeding · Non-synonymous nucleotide substitution

✉ Antonios Zambounis
   antbio@yahoo.gr

✉ Panagiotis Madesis
   pmadesis@certh.gr

[1] Laboratory of Genetics and Plant Breeding, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, P.O. Box 261, 54124 Thessaloníki, Greece

[2] Laboratory of Forest Genetics and Tree Breeding, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, P.O. Box 238, 54006 Thessaloníki, Greece

[3] Plant Breeding and Phytogenetic Resources Institute, Hellenic Agricultural Organization "Demeter", 5700 Thermi, Thessaloníki, Greece

[4] Institute of Applied Biosciences, CERTH, 570 01 Thermi, Thessaloníki, Greece

Introduction

Upland cotton (*Gossypium hirsutum* L.) is one of the most important cultivated fiber crops worldwide. However, upland cotton is susceptible to a wide range of fungal pathogens, which may have detrimental effects in its yield and product quality (Zambounis et al. 2012a). *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*), an ubiquitous soil inhabiting fungus causing extensive vascular wilt, is among the most destructive fungal diseases of upland cotton (Zambounis et al. 2007). Despite the fact that the genome of upland cotton has been recently sequenced (Li et al. 2015), identification of functional molecular markers, which could be linked to loci related to resistance genes has been hampered due to intraspecific nature of the genome (Tan et al. 2003). Complete resistant upland cotton cultivars against *Fov* isolates are not reported till date. Nevertheless, numerous efforts have been previously made for assisting disease resistance breeding programs (Fang et al.

2014; Hinchliffe et al. 2005; Niu et al. 2011), mainly by means of cloning segments of putative disease resistance (*R*) genes known as resistance gene analogues (*RGA*s).

Employment of molecular breeding approaches allows breeders to meet the challenge of developing disease resistant crops more accurately and faster (Ameline-Torregrosa et al. 2008). Most of these approaches are focused on the family of genes containing NBS (nucleotide-binding site)/LRRs (Leucine-Rich Repeats)-domains which constitute an expanded group of plant receptor genes linked directly to defense reactions (Dangl et al. 2013; Debener and Byrne 2014). In turn, LRR domains evolve novel ligand-binding specificities under strong positive selection with a wealth of different mechanisms, ranging from point mutations, variations in their repeat numbers and tandem gene duplications (Yang et al. 2008; Zambounis et al. 2012b). As a consequence, the modes of evolution of NBS-LRR *R* genes and of their *RGA* counterparts are an intricate process in plants (Sekhwal et al. 2015).

The aim of this study, based on the above evidence, was: (1) to amplify novel *RGA* clones using degenerate primers designed upon conserved motifs of known *R* genes from two upland cotton cultivars, with differences in their *Fov* susceptibility, (2) to explore whether it is likely that positive selective signatures are employed as an intensive evolutionary force, acting on a combined *RGA* dataset comprising these *RGA* clones, along with the *RGA*s that were previously deposited in GenBank from upland cotton. The results of this study, allow us to gain insights into the evolutionary profiles of these *RGA*s in upland cotton, hypothesizing that various episodes of positive selection might contribute in the acquisition of novel pathogen recognition repertoires. In parallel, these results might provide a crucial foundation for the disease resistance breeding programs in upland cotton.

Materials and methods

Upland cotton seeds of the "EMERALD" (Deltapine, Narrabri, NSW, Australia) and the "LACTA" (Bayer Crop Science) cultivars, partially resistant and fully susceptible, respectively to *Fov* (Zambounis et al. 2012a) were planted in plug trays containing sterile Promix soil and grown in growth chambers with a 16-h photoperiod at 25 °C and 70 μmol m^{-2} s^{-1} light. Genomic DNA was extracted from young leaves using the DNeasy Plant Mini Kit (Qiagen Inc., Chatsworth, CA, USA).

A pair of degenerate primers (P-loop_dom:5′-GGTG GTATGGGGNAARACNACNYT-3′; GLPL_dom: 5′-CTTG AAAGCVAAWGGNARNCC-3′) was designed with some modifications (Shen et al. 1998) in order to amplify coding regions spanning the most conserved NBS domains (the

P-loop and the GLPL) of cotton *RGA*s. These two degenerate primers were particularly designed based on the above conserved domains of the *I2* gene of tomato, which confers resistance to race 2 of *Fusarium oxysporum* f. sp. *lycopersici* (Simons et al. 1998) and of the *Arabidopsis* *RPS2* gene, which encodes a protein that specifically recognizes *Pseudomonas syringae* pv. tomato strains (Axtell et al. 2001).

All PCR amplifications were performed in 50 μl volumes containing 0.2 mM of each dNTP, 0.5 μM of each degenerate primer, 1 U of Dy-NAzyme II DNA polymerase (Finnzymes, Espoo, Finland), 2.5 mM of MgCl$_2$ and 100 ng of template genomic DNA. The initial denaturation step was 95 °C for 5 min, followed by 30 cycles of 95 °C for 1 min, 48 °C for 1 min, 72 °C for 2 min, with a final extension step at 72 °C for 10 min. PCR products were electrophoresed on a 1.2 % agarose gel and the 530 bp respective bands were selectively recovered and purified by precipitation with 0.1 volume of 3 M sodium acetate and 2.5 volumes of absolute ethanol to remove excess dNTPs and primers.

The PCR products were cloned into the pGEM-T Easy vector (Promega) and a total of 49 putative recombinant colonies were obtained and sequenced (I.M.B.B., Crete, Greece). The BLAST(hX) program was employed to confirm their homologies with known *RGA*s and R genes from upland cotton. Sequences with internal stop codons were excluded from the analysis. Thus, there remained 26 sequences from the EMERALD (*Gh_EM*) and 12 from the LACTA (*Gh_LA*). These *RGA* sequence data, 38 in total, were deposited in the Genbank under accession numbers KU680752 through KU680789.

Additionally, based on their NB-ARC domain profiles (PF00931.18/CL0023) and keyword searches, 127 partial *RGA*s from upland cotton (*Gh_GB*) previously deposited in GenBank were mined. Using their amino acid sequence IDs, we also retrieved the respective ORF nucleotide sequences using a python script. Therefore, our final *RGA* dataset consisted of 165 sequences in total. All these sequences had amino acid length up to 100 residues, no internal stop codons, and BLAST hits with known *R* genes of upland cotton. They shared valid structures and functional domains (LRRs or NB-ARC domain motifs) as it was observed by searches against both the Pfam and InterPro protein databases (Jones et al. 2014).

The phylogenetic relationships among these 165 *RGA* amino acid sequences were revealed by performing a Muscle alignment (Edgar 2004) and a tree reconstruction employing the RAxML software (Stamatakis 2014) with a gamma model of rate heterogeneity. The above analyses were performed using the Geneious R7 platform (Kearse et al. 2012). The *RGA*s sequences were consequently clustered using the MCL (Markov Cluster Algorithm)

approach with an inflation value of 2.00 (Enright et al. 2002). The clustering process utilized the all-against-all BLAST(P) output as a distance metric and finally ten unique paralogous genes groups (PGGs) were assigned. The validity of these PGGs was confirmed based on the overall similarity throughout the coding sequences, the existence of no or few gaps across the aligned sequences at each PGG, and a 50 % amino acid identity between the sequences in each PGG.

The eight out of the ten PGGs, which contained up to three *RGAs*, were independently evaluated for positive selective signatures (the remaining two PGGs, which contained only two *RGAs* each were not evaluated) using the maximum likelihood programs (CODEML and CODEMLSITES) inferring in the PAML software package (Yang 2007). CODEML estimates variable selective pressures among the branches of the phylogeny, whilst CODEMLSITES tests for site-specific codon substitution models permitting the identification of selective signatures among individual amino acids. In order to compare the rates of non-synonymous/synonymous nucleotide substitution mutations (omega values), NJ trees and nucleotide respective alignments were performed (Lynn et al. 2004). Nucleotide coding sequences were initially aligned by a perl script in order to maintain the gaps and to exclude any columns in the alignments that had more than three gaps (Zambounis et al. 2012b). Amino acid alignments were performed using the MUSCLE software (Edgar 2004), whilst the constructions of the respective evolutionary trees were carried out by employing the MEGA 5 software (Tamura et al. 2011). The Poisson correction was employed for all distances calculations along with a gamma distribution among the sites and the rate variations (Zambounis et al. 2012b). In both programs, comparisons among the various selective models were performed by employing likelihood ratio tests. The Bayesian approach allowed further detection of codon residues subjected to positive selection by posterior probabilities.

Results and discussion

Yield of upland cotton can be substantially decreased by harmful fungi, such as by *Fov* virulent isolates, which might have devastating effects on the vascular system and is a major threat worldwide (Zambounis et al. 2012a). Early recognition is the first crucial step of defense reactions against fungi attacks in plants. It is being commonly mediated by a plethora of rapidly evolving protein receptors, which are containing ligand-binding and signal transduction domains, such as LRRs and NB-ARC domains, respectively (Perazzoli et al. 2014; Sekhwal et al. 2015). The development through molecular breeding

of fungal resistant genotypes is crucial towards the substantial establishment of fungal diseases management strategies (Ganopoulos et al. 2011), especially in upland cotton with a complex allotetraploid genome (Fang et al. 2014) being recently available (Li et al. 2015).

In this study, 38 *RGAs* were cloned and sequenced from two upland cotton cultivars using degenerate primers being designed upon conserved motifs of known *R* genes. These *RGAs* were overlapping the coding regions of P-loop and GLPL motifs flanking the RNBS-A, the Kinase-2, the RNBS-B and the RNBS-C motifs of known *R* genes in upland cotton genome. Since, no introns were present between these conserved motifs, only one expected size, approximately 530 bp in length, was amplified, cloned and sequenced. These experiments have been performed without the upland cotton genome being available at that time and thus, the degenerate primers were designed upon conserved regions of known functional *R* genes. Besides, in upland cotton there do not exist known *R* genes whose function is confirmed at the moment. Furthermore, our analysis has been conducted in two upland cotton cultivars, which were different than the one whose whole genome has been sequenced. These two cultivars were also chosen as they are different at their susceptibility against *Fov*.

BLAST(X) searches of these *RGAs* revealed that their conserved NB-ARC domains were highly homologous with those of the annotated *R* genes in the genome of upland cotton. *RGAs* are in general useful candidates and quite informative components for any disease resistance breeding approach. They are also useful for the accurate characterization of *R* genes (Chen et al. 2015). These coding sequences, which were identified by PCR-based approaches, were only partial regions of the respective *R* genes without any solid functional annotations in the genome and thus, mapping them against the genome was technically intricate. However, all these *RGA* sequences were found to share either Pfam hits with the NB-ARC domain (PF00931.18) or InterProScan hits with the LRR-containing proteins (PTHR23155) when they were compared with the annotated *R* genes from the upland cotton genome. In the near feature, we have the intention to explore with RNA-seq data and to confirm by RT-PCR approaches their expression levels after challenging both tested cultivars with *Fov* isolates.

These *RGA* sequences were also ideal candidates for applying evolutionary approaches in order to gain insights into the positive selective pressures acting upon them. They were sharing conserved domains and features overlapping coding regions and functional domains of *R* genes. Thus, we mined an ample number of previously deposited *RGA* sequences of upland cotton from GenBank. For the phylogenetic and evolutionary analyses, our final dataset comprised of the 38 *RGAs* from this study, in addition to

Table 1 Overview of CODEML and CODEMLSITES programs statistics upon the evolutionary analyses in upland cotton *RGAs*

RGAs PGGs	Number of clustered *RGAs* sequenses (composition of clusters)	CODEML analysis			CODEMLSITES analysis			
		Proportion of branches under positive selection	Positive selection acting on anchestral (A) and (or) terminal (T) branches	Highest omega values in branches of the evolutionaty trees	Statistical significance (P)	Total number of sites[a]	Positively selected sites	Positively selected sites[b]
*Gh*_RGA-1	62 (*Gh*_GB, *Gh*_EM, *Gh*_LA)	6.8	A/T	∞	<0.01	447	6	2I, 7Y, 9A, 12H, 16R, 17L
*Gh*_RGA-2	38 (*Gh*_GB, *Gh*_EM, *Gh*_LA)	29.7	A/T	∞	< 0.9	252		
*Gh*_RGA-3	5 (*Gh*_GB)	12.5	T	288.1	<0.05	258	2	25L, 39Y
*Gh*_RGA-4	6 (*Gh*_GB, *Gh*_EM)	10	A	∞	<0.05	486	3	34Q, 42H, 44D
*Gh*_RGA-5	13 (*Gh*_GB)	12.5	A/T	∞	<0.001	510	9	17S, **31 K**, 40 V, 104L, 117A, 121G, 138L, **168F, 169A**
*Gh*_RGA-6	13 (*Gh*_GB)	16.7	T	1.88	<0.01	237	15	**2E, 3 K**, 8 K, **28D**, 30C, 42P, 45E, 50 N, **54E**, 55I, 58G, **62 N, 69 V**,70A, **72 N**
*Gh*_RGA-7	5 (*Gh*_GB)	42.9	T	3.52	<0.01	507	6	48L, 52D, 99 V, 116E, 152P, 154P
*Gh*_RGA-8	19 (*Gh*_GB)	13.9	T	∞	<0.9	495	1	8L
*Gh*_RGA-9	2 (*Gh*_EM)	NT						
*Gh*_RGA-10	2 (*Gh*_GB)	NT						

Likelihood ratio test among M8 and M7 models

[a] Lengths are refering to the alignments feeding the CODEMLSITES program

[b] Posterior Bayesian probabilities >0.95 are shown in lightface, and >0.99 values are shown in boldface
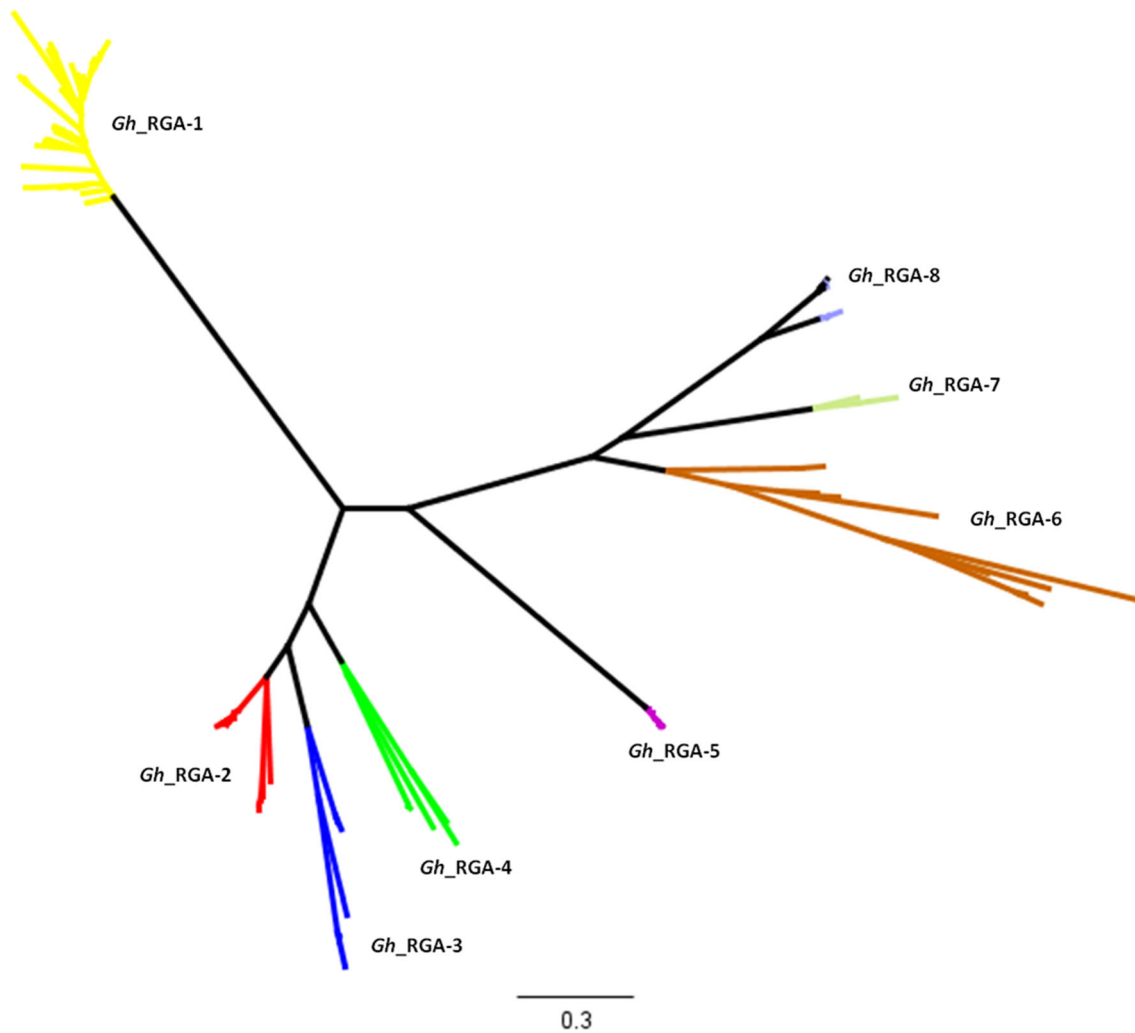
Fig. 1 RAxML phylogenetic reconstruction of the 161 upland cotton *RGAs* sequences which were aligned using the Muscle software (Edgar 2004). In various colors the eight PGGs which were assigned inferring the MCL-based clustering approch

the 127 *RGA*s, which were previously deposited in Gen-Bank. All-against-all BLAST(P) searches (*E* value cutoff $<10^{-10}$) were applied in order to assess the extent to which these 165 *RGAs* genes were clustered using the MCL approach (similarity cutoff of 50 %), allowing their classification into respective PGGs. They were assigned in 10 PGGs (named as *Gh*_RGA) in total. Eight of them contained from three up to 62 *RGA*s and 161 *RGA*s in total (Table 1). The majority (88 %) of *RGA*s were clustered in five different PGGs (*Gh*_RGA-1, *Gh*_RGA-2, *Gh*_RGA-5¸ *Gh*_RGA-6, *Gh*_RGA-8). In contrast, two *Gh*_RGA PGGs contained only two *RGA*s (Table 1); these *RGA*s were excluded from the further phylogenetic and evolutionary analysis.

In order to decipher the expansion of the 161 *RGA*s, which were clustered in the eight PGGs, a RAxML phylogenetic analysis was performed (Fig. 1). The largest portion of the phylogenetic clades contained only a limited number of *RGA*s, whilst the backbone topology was adequately resolved with up to 320 branches (Fig. 1). The overall pairwise identity among these 161 amino acid sequences was 50.1 %, an indication of a high divergence among them.

We also investigated whether accelerated evolution and signatures of positive selection might have also contributed to the divergence of these *RGAs*. We subjected separately each of the eight PGGs to several successive tests, which could be used for seeking evidence for positive selection (Table 1). Statistically significant evidence of positive selection (omega values higher then 1.00) were detected in all eight PGGs, whereas the highest proportions of branches under positive selection were reported in *Gh*_RGA-2 (29.7 %) and *Gh*_RGA-7 (42.9 %) (Table 1). These positive selection signatures were observed rather unequally across the entire tree lineages, acting with a non-general rule. They were evident mostly at terminal branches in

seven PGGs, and in four PGGs at their ancient branches (Table 1). We assume that recent episodes of positive selection events have been occurred, overlapping similar more ancient events among these *RGA*s. Constantly with our results, similar existences of positive selective pressures acting on plant *RGA*s have been recently reported (Khan et al. 2015; Perazzolli et al. 2014).

Finally, using the CODEMLSITES program and various selection models, we tested whether positive selection was acting at individual amino acid sites. Extensive signs of positive selection, with posterior Bayesian probabilities under the implementation of the most stringent M8 model, were observed in seven PGGs acting mostly towards the N-termini regions (Table 1). Only *Gh*_RGA-2 did not reflect any evidence of positive selection, perhaps because of the low divergence among the 38 *RGA*s comprising this PGG. The 42 positively selected sites are depicted in Table 1. Previously, comparative analyses of *R* genes at *Rosaceae* species revealed that the solvent-exposed amino acid residues of the LRRs domains were significantly variable with intensive positive selective pressures acting upon them (Perazzolli et al. 2014). Similar results were also observed in a genome wide analysis of the *RGA* in cherry (Zambounis et al. 2016). Such evidences are interpreted as an additional indication of the active involvement of these regions in fungal pathogens recognition pathways (Zambounis et al. 2012b), which is accurately consistent with the host-pathogen co-evolution processes leading to acquisition of disease resistance specificities (Mondragón-Palomino et al. 2002).

The overall findings of this study might facilitate the development of informative *RGA*-derived molecular markers being associated with resistant phenotypes and being particularly useful for identifying *Fov* resistant upland cotton cultivars. In the near feature, we have the intention to perform a more deeply genome wide analysis targeting the annotated *R* genes, particularly these of the NBS-LRRs family, in older to decipher their expansion and the evolutionary signatures acting upon them in the genome of upland cotton.

References

Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146(1):5–21

Axtell MJ, McNellis TW, Mudgett MB, Hsu CS, Staskawicz BJ (2001) Mutational analysis of the *Arabidopsis RPS2* disease resistance gene and the corresponding *Pseudomonas syringae avrRpt2* avirulence gene. Mol Plant Microbe Interact 14(2):181–188

Chen JY, Huang JQ, Li NY et al (2015) Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium* wilt. BMC Plant Biol 15:148

Dangl JL, Horvath DM, Staskawicz BJ (2013) Pivoting the plant immune system from dissection to deployment. Science 341(6147):746–751

Debener T, Byrne DH (2014) Disease resistance breeding in rose: current status and potential of biotechnological tools. Plant Sci 228:107–117

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:1–19

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584

Fang H, Zhou H, Sanogo S, Lipka AE, Fang DD, Percy RG, Hughs SE, Jones DC, Gore MA, Zhang J (2014) Quantitative trait locus analysis of Verticillium wilt resistance in an introgressed recombinant inbred population of Upland cotton. Mol Breed 33(3):709–720

Ganopoulos IV, Kazantzis K, Chatzicharisis I, Karayiannis I, Tsaftaris AS (2011) Genetic diversity, structure and fruit trait associations in Greek sweet cherry cultivars using microsatellite based (SSR/ISSR) and morpho-physiological markers. Euphytica 181(2):237–251

Hinchliffe DJ, Lu Y, Potenza C, Segupta-Gopalan C, Cantrell RG, Zhang JF (2005) Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. Theor Appl Genet 110:1074–1085

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649

Khan AM, Khan AA, Azhar MT, Amrao L, Cheema HM (2015) Comparative analysis of resistance gene analogues encoding NBS-LRR domains in cotton. J Sci Food Agr 96:530–538

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, Liang X, Huang G, Percy RG, Liu K, Yang W, Chen W, Du X, Shi C, Yuan Y, Ye W, Liu X, Zhang X, Liu W, Wei H, Wei S, Huang G, Zhang X, Zhu S, Zhang H, Sun F, Wang X, Liang J, Wang J, He Q, Huang L, Wang J, Cui J, Song G, Wang K, Xu X, Yu JZ, Zhu Y, Yu S (2015) Genome sequence of cultivated Upland cotton (*Gossipium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33(5):524–553

Lynn DJ, Lloyd AT, Fares MA, O'Farrelly C (2004) Evidence of positively selected sites in mammalian α-defensins. Mol Biol Evol 21(5):819–827

Mondragón-Palomino M, Meyers BC, Michelmore RW, Gaut BS (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res 12(9):1305–1315

Niu C, Lu Y, Yuan Y, Percy RG, Ulloa M, Zhang J (2011) Mapping resistance gene analogs (RGAs) in cultivated tetraploid cotton using RGA-AFLP analysis. Euphytica 181:65–76

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):e83844

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

Shen KA, Meyers BC, Islam-Faidi MN, Chen DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleoride primers map to clusters of resistance genes in lettuce. Mol Plant Microbe Interact 11(8):815–823

Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, Van der Lee T, Bleeker M, Onstenk J, de Both M, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P (1998) Dissection of the *fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10(6):1055–1068

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

Tan H, Callahan FE, Zhang XD, Karaca M, Saha S, Jenkins JN, Creech RG, Ma DP (2003) Identification of resistance gene analogs in cotton (*Gossypium hirsutum* L.). Euphytica 134:1–7

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591

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

Zambounis A, Paplomatas E, Tsaftaris AS (2007) Intergenic spacer-RFLP analysis and direct quantification of Australian *Fusarium oxysporum* f. sp. *vasinfectum* isolates from soil and cotton infected tissues. Plant Dis 91:1564–1573

Zambounis A, Kalamaki M, Tani E, Paplomatas E, Tsaftaris A (2012a) Expression analysis of defense-related genes in cotton (*Gossypium hirsutum*) after *Fusarium oxysporum* f. sp. *vasinfectum* infection and following chemical elicitation using a salicylic acid analog and methyl jasmonate. Plant Mol Biol Rep 30(1):225–234

Zambounis A, Elias M, Sterck L, Maumus F, Gachon CMM (2012b) Highly dynamic exon shuffling in candidate pathogen receptors…What if brown algae were capable of adaptive immunity? Mol Biol Evol 29(4):1263–1276

Zambounis A, Ganopoulos I, Avramidou E, Aravanopoulos F, Tsaftaris A, Madesis P (2016) Evidence of extensive positive selection acting on cherry (*Prunus avium* L.) resistance gene analogs (RGAs) Aust J Crop Sci (**in press**)