Proceedings of the Pakistan Academy of Sciences 49 (4): 259–267 (2012)

ISSN: 0377 - 2969 print / 2306 - 1448 online

Pakistan Academy of Sciences

Original Article

Comparative Sequence Analysis of Some *Rdr1* Resistance Genes from *Rosa multiflora*

Aneela Yasmin[1], Akhtar A. Jalbani[2]* and Thomas Debener

[1]Department of Biotechnology, Sindh Agriculture University, Tandojam, Pakistan
[2]Information Technology Center, Sindh Agriculture University, Tandojam, Pakistan
[3]Institute of Plant Genetics, University of Hannover, D-3419, Hannover, Germany

Abstract: *Rdr1* is the first monogenic dominant resistance gene discovered in genus rosa. This locus is active against a fungal disease, black spot, of roses. *Rdr1* locus consists of nine genes named as *muRdr1A* to *muRdr1I*. Functional and molecular characterization of this locus, revealed the presence of one gene, mu*Rdr1*H, active against Dort E4, race 6 of *Diplocarpon rosae.* In the present study we aimed to analyze the protein sequences of three members of this *Rdr1* gene family for their homologies, important domains, conserved regions and selection pressure to get insights of their functionality based on protein sequence. The protein sequences of one active gene, mu*Rdr1*H and two inactive genes, mu*Rdr1*C and mu*Rdr1*G, were deduced and dissected using internet based bioinformatics tools. All three proteins belong to the TIR-NBS-LRR family of resistance genes. Although these proteins carry all necessary domains and conserved regions necessary for the functionality of TIR-NBS-LRR proteins, LRR region of the proteins, found under selection pressure, indicated its role in effector identification.

Keywords: *Rdr1*, rose, black spot, TIR-NBS-LRR proteins, resistance genes

1. INTRODUCTION

The most devastating fungal diseases of roses are powdery mildew, black spot and downy mildew. Powdery mildew usually infects roses grown in greenhouses whereas the black spot is a problem for field and garden roses grown in humid and moist conditions throughout the world [1]. The present control of this disease is fungicidal sprays. On the other hand use of resistant varieties with durable genetic resistance is the safest option for the sustainable environment. Most of the cultivated roses lack natural resistance against black spot thus it has to be introgressed from wild roses [2]. Nevertheless the exploitation of natural genetic resistance requires understanding of the resistance genes in terms of diversity, genomic organization and functionality.

Rdr1 is the first single monogenic dominant resistance gene locus found active against black spot (*Diplocarpon rosae*) of roses through phytopathological methods [3]. This *Rdr1* locus was later mapped to a telomeric position of linkage group 1 in the diploid population 94/1 [4]. The construction of two large insert BAC libraries [5, 6] and sequencing of BAC clones identified the location of *Rdr1* gene within a 220 kb region [6]. The 220 kb region contains 9 copies of "resistance-gene-analogues" sequences (RGA) of the T1R-NBS-LRR type of resistance genes [7]. Out of these 9 genes one gene *muRdr1H* was found as major active gene against Dort E4, race-6 of *Diplocarpon rosae* by functional characterization [8, 9].

In TIR-NBS-LRR resistance genes, TIR motif is similar to toll/interleukin-1-receptor (TIR), the NBS is a part of a nucleotide binding (NB)-ARC domain that belongs to the STAND (signal transduction ATPases with numerous domains) family of NTPases. These proteins are proposed

Received, May 2012; Accepted November 2012
*Corresponding author: Akhtar A. Jalbani; Email: akjalbani@sau.edu.pk

to regulate signal transduction as NB domain hydrolyzes NTP and changes its conformational states [10]. The NB-ARC domain has three sub-domains conserved in NBS-LRR proteins: a P-loop NTPase fold forming a parallel β-sheet flanked by α-helices, an ARC1 consisting of a four-helix bundle, and an ARC2 adopting a winged-helix fold that is connected to the LRR domain by a short linker [10]. LRR domains contain various numbers of tandemly repeated leucine-rich motifs with a conserved core consensus of L-X-X-L-X-L-X-X-N that form a series of β-strands [11]. The arc-shaped structure of the LRR domain suggests its role in different intra and intermolecular interactions of direct recognition of pathogen effectors, regulating protein activation and signal transduction [12].

The current research was aimed to analyze protein sequences of three paralogs of *Rdr1* locus namely *muRdr1H (*AEE43932*), muRdr1C (*AEE43927) and *muRdr1G (*AEE43931). These paralogs are unique in a sense that these are the only members of this family for which 3' and 5' RACE products are available and are functionally characterized against Dort.E4. In this study we tried to get an insight in protein functionality based on their sequences and evaluated the selection pressure exerted on the genes. For this sequence variability, nucleotide substitution rates and amino acid homology check were carried out using different bioinformatics tools.

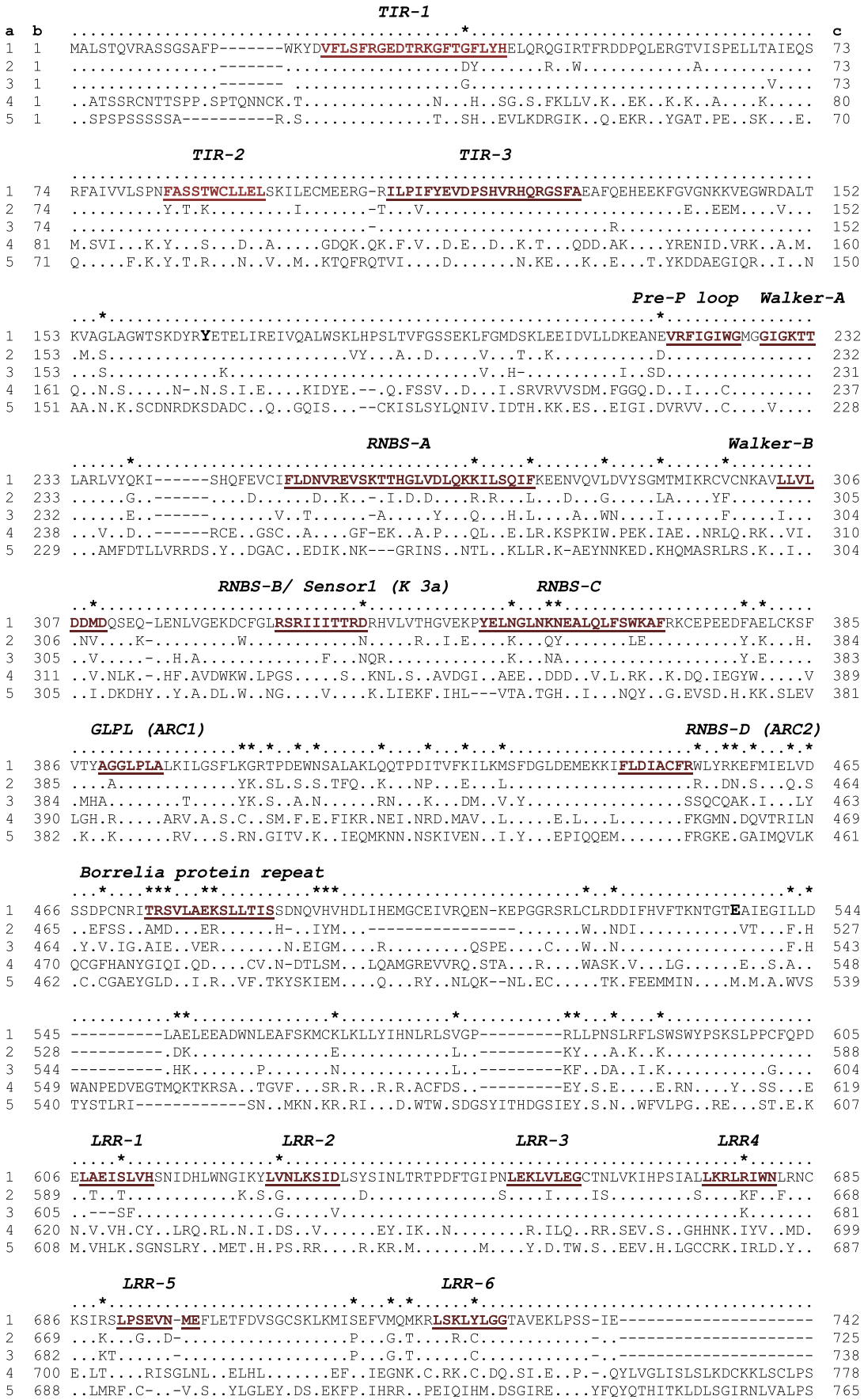
2. MATERIAL AND METHODS

RACE (Rapid Amplification of cDNA Ends) technique was used to obtain the 5' and 3' ends of RNA transcripts of *muRdr1C, G* and *H* transiently expressed in tobacco [8, 9]. Isolated and sequenced RACE products were used to get cDNA sequence from previously sequenced BAC clones [6, 9]. Protein sequences of three proteins were obtained using full length cDNA [8] in Soft-ware Bioedit Version 7.0.9 [13].The same software was used for the protein sequence editing, alignment, multiple alignment (ClustalW) and local BLAST searches to analyze and obtain conserved regions. Nucleotide sequence of *muRdr1H* (cDNA) was used for Blastx searches (http://www.ncbi.nlm.nih.gov/blast/Blast. cgi) using NCBI database to find the similar proteins to *muRdr1H*. For Neignbor-joining (NJ) analysis the homologous protein sequences were downloaded from NCBI database and phylogenetic tree was constructed in MEGA4 [14, 15]. The gene prediction and protein analysis were performed on different free internet sources as http://www.expasy.ch/, http://swissmodel.expasy.org/SWISS-MODEL.html, http://www.ebi.ac.uk/interpro/,http://linux1.softberry.com/berry.phtml whereas the ratio of synonymous (K_s) and non-synonymous (K_a) substitutions rates per synonymous/ non-synonymous site were calculated using software DnaSP v 5 [16].

3. RESULTS AND DISCUSSION

Rdr1 resistance locus in *R. multiflora* consists of a cluster of nine paralogs disease resistance genes that play important roles in innate immunity against black spot. The protein sequences of three members, *muRdr1C, muRdr1H and muRdr1G,* were deduced and subjected to sequence variability, nucleotide substitution rates and amino acid homology check. According to previous studies, among the nine R genes in this locus *muRdr1H* confers partial resistance to race 6 of the fungal pathogen *Diplocarpon rosae* [8, 9, 17]. The *muRdr1H* transcript resulted in the Open Reading Frame (ORF) of 1122 amino acids (ORF finder-NCBI), *muRdr1C* has the predicted protein of 1139 aa whereas the deduced protein of *muRdr1G* has 944 amino acids. All three proteins show the presence of TIR, NBS and LRR domains with all conserved motifs as suggested by Lukasik and Takken [18] and Meyers and colleagues [19] (Fig. 1).

The homologues of *muRdr1H* protein were identified by BLASTx searches which were carried out against the GenBank non-redundant database (http://blast.ncbi.nlm.nih.gov). The *muRdr1H* protein shares the highest identity (41%) to TIR-NBS-LRR-resistance protein of *Populus trichocarpa* (ACCESSION XP_002329162). It also shows identity to hypothetical proteins of *Vitis vinifera* (39-44%), to TIR of *Medicago truncatula* (40%; ACCESSION ABD28703), to CMR1 of *Phaseolus vulgaris* (40%; ACCESSION ABH07384) and to *N*-like protein of *N. tabacum* (39%; ACCESSION BAF95888 for resistance to Tobacco Mosaic Virus). Fig. 2 shows phylogenetic tree of different amino


TIR-1
TIR-2
TIR-3
Pre-P loop
Walker-A
RNBS-A
Walker-B
RNBS-B/ Sensor1 (K 3a)
RNBS-C
GLPL (ARC1)
RNBS-D (ARC2)
Borrelia protein repeat
LRR-1
LRR-2
LRR-3
LRR4
LRR-5
LRR-6

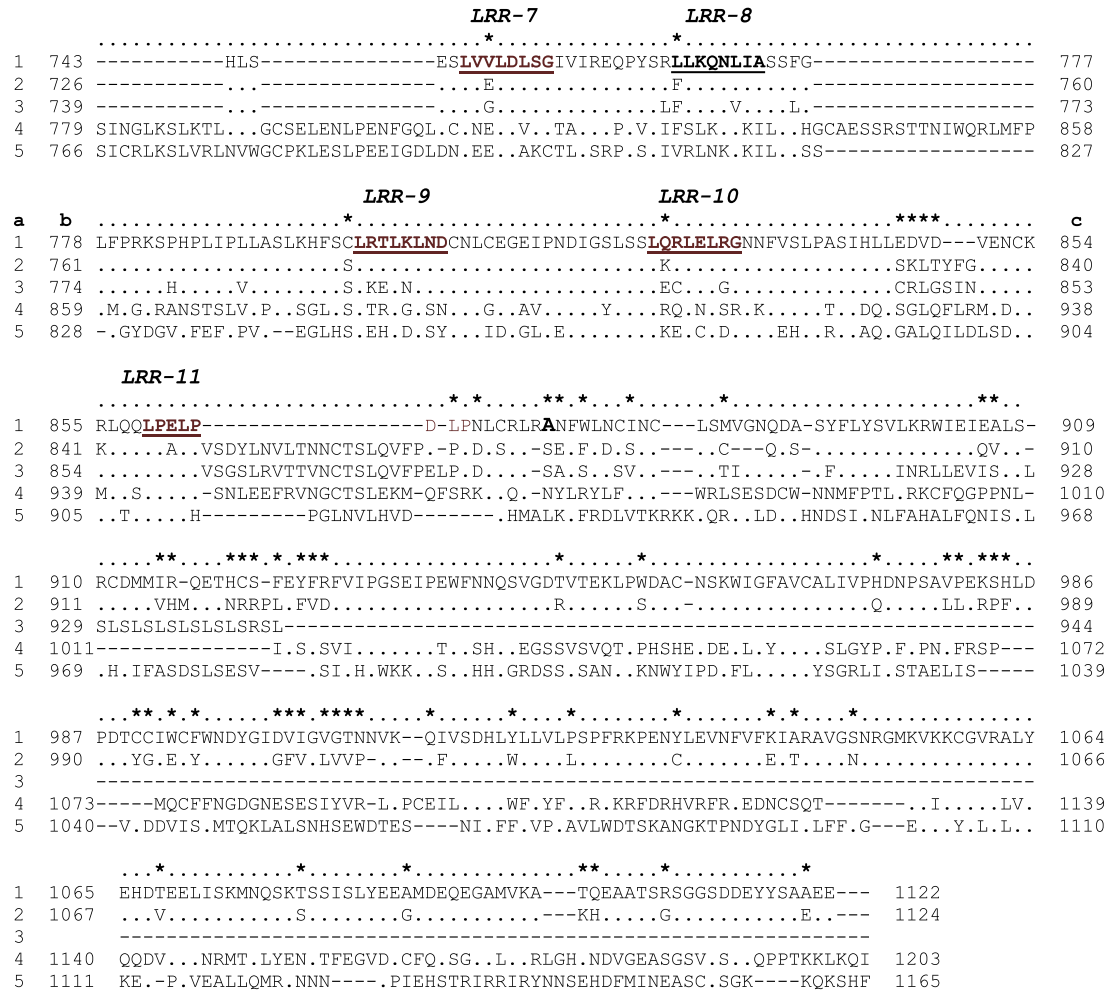


Fig. 1. Alignment of the proteins of *muRdr1H*, C and G of roses, TIR-NBS-LRR resistance protein of *Populus trichocarpa* and N like protein of *N. tabacum.* The *muRdr1H* protein and its predicted TIR, NBS and LRR domains with conserved motifs are under-lined. Stars on *muRdr1H* protein sequence represent unique amino acids of this protein when compared to proteins of *muRdr1C* and G. Within *muRdr1H* protein sequence the end amino acid of each domain is marked bold. a- Numbers in first column corresponds to *muRdr1H* (1), *muRdr1C* (2), *muRdr1G* (3), resistance protein of *Populus trichocarpa* (4) and *N* like protein of *N. tabacum* (5); b and c- sequence positions; dashes indicate gaps inserted to maintain optimal alignment.

acid sequences showing considerable identity to *muRdr1H* protein, these full length aligned protein sequences were downloaded from NCBI and tree was constructed based on the bootstrap neignbor-joining (NJ) method with the Kimura2-parameter model by MEGA4 [14, 15]. It is interesting to code here that the *CMR1* is a viral resistance gene from common bean that functions across plant families [20]. This protein sequence shows 40% identity to the *muRdr1H* protein that found to be active against black spot fungus, which suggests that the predicted NB-LRR structures and recognition of certain pathogen type lack correlation.

The kind of selection pressure exerted on different domains of the genes was estimated by calculating ratio of synonymous (K_s) and non-synonymous (K_a) rates of substitutions for each synonymous/ non-synonymous site of different parts of the ORF of three paralogs (*muRdr1H, G and C*). When K_a / K_s ratio is 1 it represents random mutation or no selection pressure operating on sequences may be for diversification ($K_a / K_s > 1$) or conservation of sequence ($K_a / K_s < 1$) [21]. Estimation of the number of synonyms and non-synonyms nucleotide per site

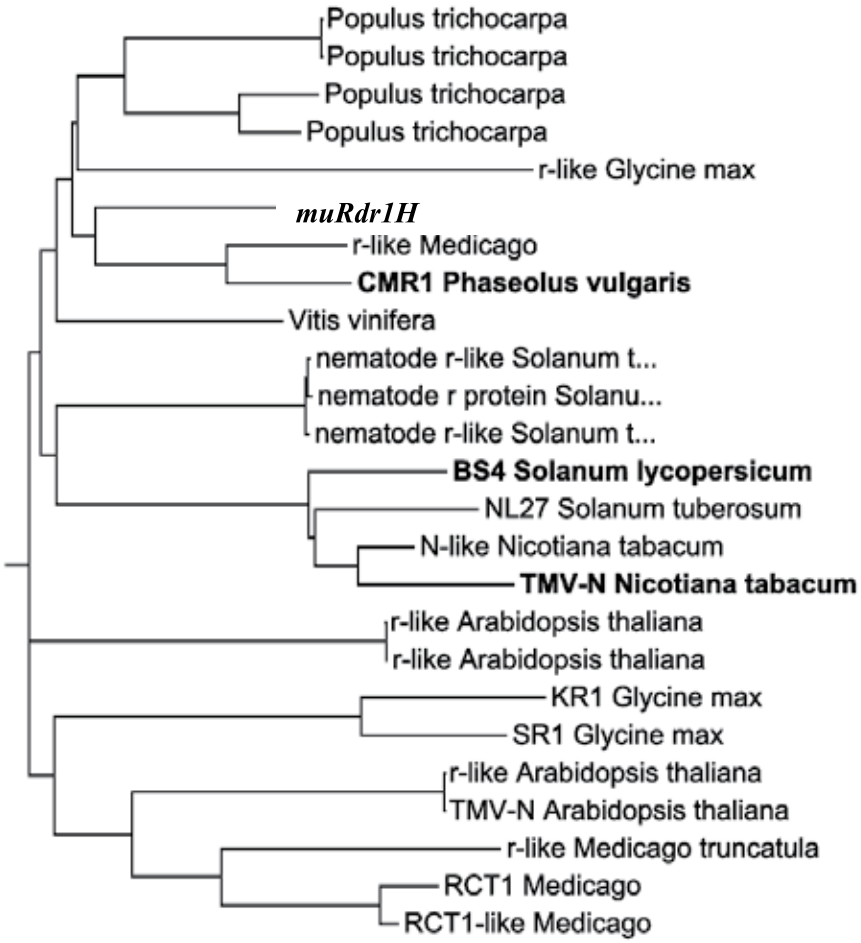


Fig. 2. NJ analysis of protein sequence showing considerable similarities to *muRdr1H*. The full length aligned protein sequences were downloaded from NCBI and analyzed in MEGA 4.0.

showed that the three genes are under conservation selection and all domains (TIR, NBS and LRR domain) are not under positive selection compared to N- and C-flanking regions to conserved LRR motifs which showed K_a / K_s ratio > 1 (Table 1). The comparison of complete LRR domains of *muRdr1H* and *muRdr1C* resulted in higher K_a values than K_s suggesting diversification in this region, further insights to data revealed that this diversification is actually the result of lower K_s for the N- and C-flanking region of xxLxLxx motifs which always shows a $K_a / K_s \geq 1$ (Table 1). Although three genes show conserved type of selection, LRR domains show the highest frequency of non-synonymous substitutions followed by NBS domain. The K_a / K_s ratios were calculated for C-terminal region of genes, results showed the diversification trend in this region. This supported the hypothesis that the parts of proteins which participate in pathogen recognition are under diversifying selection. This pattern is in agreement to previous studies that show diversifying selection acts on the LRR encoding domain of various plant disease resistance genes [11].

Comparison of *muRdr1H* protein sequence with other two (*muRdr1C* and *muRdr1G*) members of the same *Rdr1* family revealed 119 *muRdr1H* specific amino acids residues (119/ 1122; 10.6%), which are part of TIR (2 aa; 0.002%), NBS (40 aa; 3.6%), and LRR (77 aa; 6.8%) domains (Fig. 3). For RAG8, as the majority of unique amino acids are located in LRR region followed by NBS region suggesting some important role of these domains in functionality of *muRdr1H*. The deduced *Gro1-4* protein (resistance gene) showed only 16 non-conserved differences in amino acids sequence

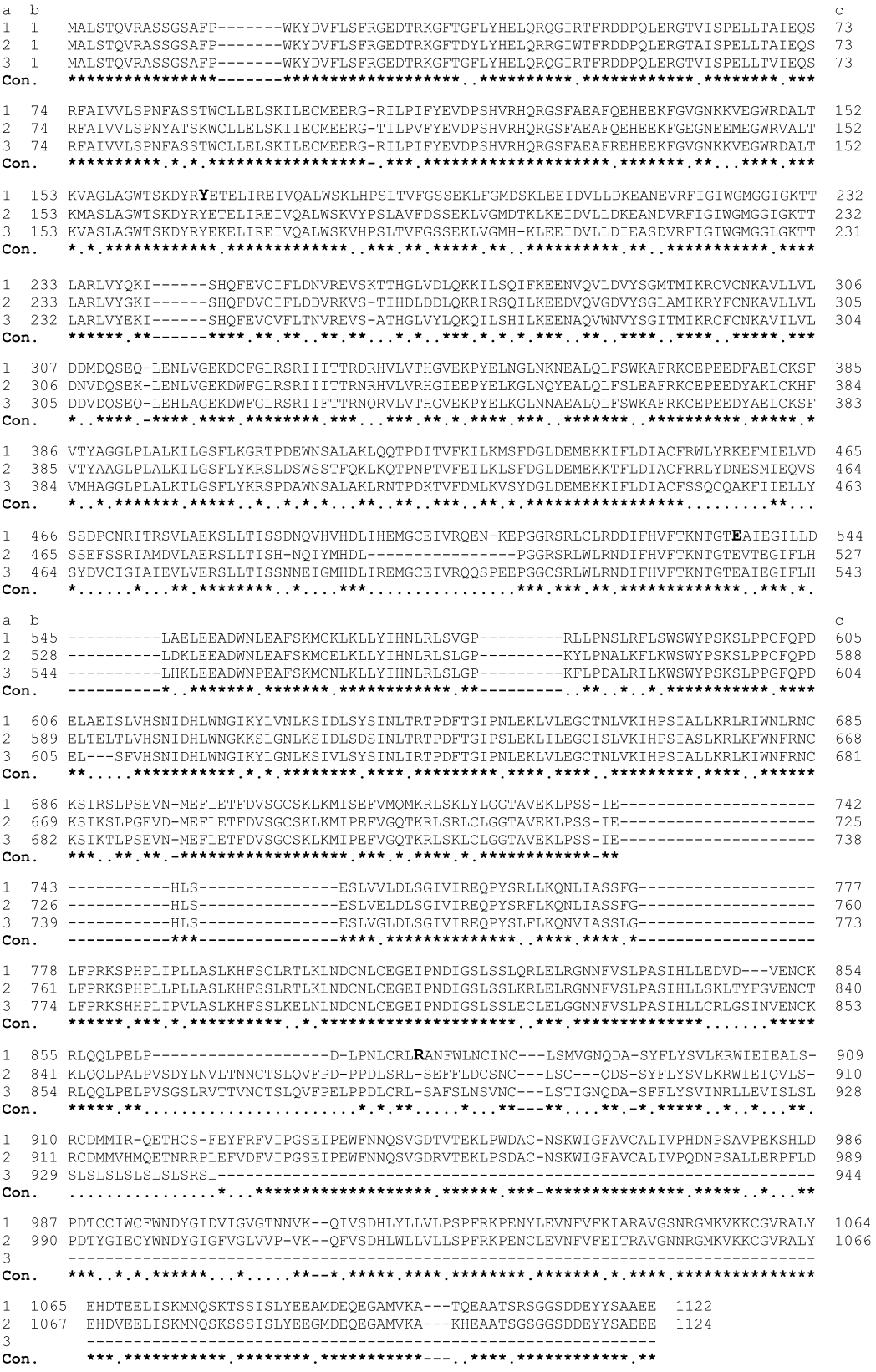


Fig. 3. Comparison of protein sequences encoded by 3-members of *Rdr1* family. Protein sequences *muRdr1H* (1), *muRdr1C* (2) and *muRdr1G* (3) were aligned to check amino acid similarities and divergence. Stars represent consensus sequence and dots show differences in amino acids. a- *muRdr1H, G & C* proteins sequences; b and c- sequence positions; dashes indicate gaps inserted to maintain optimal alignment.

Table 1. Sequence variability and nucleotide diversity in different regions of three paralogs of *Rdr1* family.

R-genes compared	Analyzed regions[1]	aa[2] (%)	Nucleotide substitutions[3] K_a	K_s	K_a / K_s
muRdr1H* vs. *muRdr1C	Complete CDS	80	0.0977	0.1041	0.92
	TIR domain	90	0.0484	0.0836	0.58
	NBS domain	76	0.1168	0.1196	0.98
	Complete LRR domain	79	0.1034	**0.0983***	1.05
	N-terminal flanking	-	0.1146	**0.1108**	1.03
	xxLxLxx motif	-	0.0722	0.0935	0.77
	C-terminal flanking	-	0.1679	**0.1156**	1.45
muRdr1H* vs. *muRdr1G	Complete CDS	75	0.0991	0.1354	0.73
	TIR domain	98	0.0079	0.0356	0.22
	NBS domain	78	0.1184	0.1853	0.64
	Complete LRR domain	66	0.1209	0.1429	0.85
	N-terminal flanking	-	0.1148	**0.0530**	2.17
	xxLxLxx motif	-	0.0717	0.1142	0.63
	C-terminal flanking	-	0.5067	0.5738	0.88
muRdr1G* vs. *muRdr1C	Complete CDS	64	0.1180	0.1497	0.79
	TIR domain	90	0.0512	0.1134	0.45
	NBS domain	71	0.1439	0.1737	0.83
	Complete LRR domain	52	0.1267	0.1508	0.84
	N-terminal flanking	-	0.0740	0.1132	0.65
	xxLxLxx motif	-	0.0793	0.1249	0.64
	C-terminal flanking	-	0.5657	**0.5029**	1.13
muRdr1H* vs. *muRdr1G &C	Complete CDS	58	0.09228	0.11032	0.83
	TIR domain	89	0.02736	0.05688	0.47
	NBS domain	67	0.10896	0.13689	0.78
	Complete LRR domain	49	0.10421	0.11089	0.94
	N-terminal flanking	-	0.10613	**0.07759**	1.40
	xxLxLxx motif	-	0.06861	0.09678	0.69
	C-terminal flanking	-	0.25922	**0.25760**	1.01

[1] Different region of R-genes is analyzed for nucleotide variability.
[2] Amino acid (aa) homology in percentage.
[3] The ratio of synonymous (K_s) and non-synonymous (K_a) substitutions rates per synonymous/ non-synonymous site were calculated using software DnaSP v 5.
* Under-lined numbers represent the exceeded value of synonymous (K_s) compared to non-synonymous (K_a) substitutions for specified region

when compared to non-functional members of the *Gro1* gene family [22]. Mutagenesis analysis of the *muRdr1H* protein or comparison of functional and non-functional orthologs could determine the essential residues necessary for pathogen recognition and/ or downstream signaling. The *Rdr1* resistance locus confers resistance to five races of *D. rosae* [23]. It is possible that the other members of this family are also functionally active against other races of *D. rosae*. Although the amino acid sequence identity of these three paralogs of *Rdr1* resistance gene cluster ranges between 58% and 80% and *muRdr1H* shares the highest overall amino acid sequence homology to *muRdr1C* (80%; Table 1) the functionality of *muRdr1C* and *muRdr1G* should be explored against different isolates of *D. rosae* and/ or some other taxonomically different pathogens.

Alignment of deduced amino acids of *muRdr1H*, C and G show that the higher degree of sequence similarity is present in N-terminal halves of proteins that harbour putative effector domains when compared to the C-terminal halves of proteins (Fig. 3) suggesting the LRR domain under selection as demonstrated for other closely related NBS-LRR proteins [24]. This kind of selection in LRR domain is exerted by single base changes, insertions, deletions and unequal exchange of meiotic recombination events for the evolution of new pathogen specificities within R-genes or between closely linked *R*-genes in a cluster [25].

4. CONCLUSION

The functionally characterized *muRdr1H* which was found active against Dort E4 and other two sequences has all conserved regions of TIR-NBS-LRR genes. *muRdr1G* and *C* were inactive when challenged by Dort E4. Due to the homology of two paralogs to mu*Rdr1*H, we suggest their functional characterization against other races of *D. rosae* to assess their functionality on molecular level. We observed less selection pressure exerted on these genes as the pathogen *D. rosae* has low mobility and races.

5. REFERENCES

1. Horst, R. K. *Compendium of Rose Diseases.* The American Phytopathological Society, St.Paul, Minn., p. 7–11 (1983).
2. Schulz, D. F., M. Linde, O. Blechert & T. Debener. Evaluation of genus *Rosa* germplasm for resistance to black spot, downy mildew and powdery mildew. *European Journal of Horticulture Sciences* 74 (1): 1–9 (2009).
3. Debener, T., R. Drewes-Alvarez & K. Rockstroh. Identification of five physiological races of black spot, *Diplocarpon rosae* Wolf, on roses. *Plant Breeding* 117: 267–270 (1998).
4. Debener, T. & L. Mattiesch. Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theoretical and Applied Genetics* 99:891–899 (1999).
5. Kaufmann, H., L. Mattiesch, H. Lörz & T. Debener. Construction of a BAC library of *Rosa rugosa* Thunb. and assembly of a contig spanning *Rdr1*, a gene conferring resistance to black spot. *Molecular Genetics and Genomics* 268:666–674 (2003).
6. Biber, A., H. Kaufmann, M. Linde, M. Spiller, D. Terefe & T. Debener. Molecular markers from a BAC contig spanning the *Rdr1* locus: a tool for marker-assisted selection in roses. *Theoretical and Applied Genetics,* doi: 10.1007/s00122-009-1197-9 (2009).
7. Terefe, D., A. Biber, A. Yasmin, H. Kaufmann & T. Debener. Comparative genomic analysis of sequences around the *Rdr1* locus in resistant and susceptible rose genotypes. *Acta Horticulture* (ISHS) 870: 197–204 (2010)
8. Yasmin, A. *Identification and Molecular Characterization of Rdr1 Resistance Gene from Roses*. PhD thesis, Univercity. of Hannover, Germany (2010).
9. Terefe-Ayana, D., A. Yasmin, T. L. Le, H. Kaufmann, A. Biber, A. Kuehr, M. Linde & T. Debener. Mining disease resistance genes in roses: functional and molecular characterization of the *Rdr1* locus. *Frontiers of Plant Sciences* 2:35, doi: 10.3389/ fpls.2011.00035 (2011).
10. Takken, F., M. Albrecht & I. Tameling. Resistance proteins: molecular switches of plant defense. *Current Opinion in Plant Biology* 9(4): 383–90 (2006).
11. Jiang, H., C. Wang, L. Ping, D. Tian & Yang. Pattern of LRR nucleotide variation in plant resistance genes. *Plant Science* 173: 253–261, doi: 10.1016/ j.plantsci.2007.05.010 (2007).
12. Padmanabhan, C., X. Zhang & H. Jin. Host small RNAs are big contributors to plant innate immunity. *Current Opinion in Plant Biology* 12:465–472, doi: 10.1016/j.pbi.2009.06.005 (2009).
13. Hall, T. A. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium*

Series 4: 95–98 (1999).
14. Tamura, K., J. Dudley, M. Nei & S. Kumar. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology of Evolution* 24:1596–1599 (2007).
15. Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111–120 (1980).
16. Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer & R. Rozas. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497 (2003).
17. Yasmin, A. *Rose and Black Spot*. VDM Verlag Dr. Müller GmbH & Co. KG (2011).
18. Lukasik, E. & F. L. M. Takken. STANDing strong, resistance proteins instigators of plant defence. *Current Opinion in Plant Biology* 12:427–436, doi 10.1016/j.pbi.2009.03.001 (2009).
19. Meyers, B. C., A. W. Dickerman, R. W. Michelmore, S. Sivaramakrishnan, B. W. Sobral & N.D. Young. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding super family. *The Plant Journal* 20:317–332 (1999).
20. Seo, Y. S., M. R. Rojas, J. Y. Lee, S. W. Lee, J. S. Jeon, P. Ronald, W. J. Lucas & R. L. Gilbertson. A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. *Proceedings of Academy of National Sciences* 102(32): 11856–11861, doi: 10.1073_pnas.0604815103 (2006).
21. Hughes, A. L. & M. Nei. Pattern of nucleotide substitution at major histo-compatibility complex class I loci reveals over dominant selection. *Nature* 335: 167-170 (1988).
22. Paal, J., H. Henselewski, J. Muth, K. Meksem, M. C. Menendez, F. Salamini, A. Ballvora & C. Gebhard. Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *The Plant Journal* 38: 285–297, doi: 10.1111/j.1365-313X.2004.02047.x (2004).
23. Von Malek, B. & T. Debener. Genetic analysis of resistance to black spot (*D. rosae*) intetraploid roses. *Theoretical and Applied Genetics* 96: 228–231 (1998).
24. Yahiaoui, N., S. Brunner & B. Keller. Rapid generation of new powdery mildew resistance genes after wheat domestication. The *Plant Journal* 47: 85–98 (2006).
25. Ellis, J., P. Dodds & T. Pryor. The generation of plant disease resistance gene specificities. *Trends in Plant Science* 5: 373–379 (2000).