

PROTEIN MODELING OF YELLOW RUST DISEASE IN WHEAT

SHANILA EMMANUEL AZIZ¹, RAISA BANO¹, MOHAMED EZZ ZAYED²,
MOHAMED SOLIMAN ELSHIKH², MOHAMMAD HAROON KHAN³,
ZUBEDA CHAUDHRY⁴ AND HAMID RASHID⁵

¹Capital University of Science and Technology, Islamabad, Pakistan

²Department of Botany and Microbiology, College of Science, King Saud University,
P.O. Box 2455, Riyadh 11451, Saudi Arabia,

³Department of Bioinformatics, Hazara University, Mansehra, Pakistan

⁴Department of Botany, Hazara University, Mansehra, Pakistan

⁵Department of Biosciences, COMSATS Sahiwal Campus, Sahiwal, Pakistan

*Corresponding author: e-mail: drhamidrashid58@gmail.com / hamidrashid@ciitsahiwal.edu.pk

Abstract

Wheat production in Pakistan is affected by yellow rust disease caused by a fungus *Puccinia striiformis*. There is a need to broaden the genetic basis of wheat by identifying new resistance genes. The present study was aimed to identify an alternate resistance gene for yellow rust disease in wheat caused by *Puccinia striiformis*. Genome sequence was compared with databases and similar gene was identified for disease resistance in rye plant. Structural analysis of RGA1 gene (resistance gene in wheat) was carried out using different bioinformatics tools and an alternative gene having same structure was identified on the basis of structural and sequence homology. Rye plant is the proposed plant for the alternate new resistance gene. The result of pairwise alignment of RGA1 gene in wheat and gene of rye plant is 94.2% with accession # DQ494535. The secondary structures of both the genes was compared and found similar to each other. These comparisons between the wheat resistance gene and gene from rye plant depict structural similarities between the two genes. Results of RGA1 gene's structural analysis in wheat is as follow: Helices: 59, Extended sheets: 30, Turns: 12, Coils: 13 and for alternate resistance genes in Rye is as follow: Helices: 52, Extended sheets: 30, Turns: 14, Coils: 17. As structures are similar, the alternate identified gene could be used for resistance in wheat.

Key words: Wheat, Protein modelling, Rust resistance, Bioinformatics.

Introduction

Being the largest grain crop of Pakistan, wheat is a major staple food of people of the country (Government of Pakistan, 2004; Zafar *et al.*, 2015). The major wheat growing centers, producing 15% of world's wheat include Pakistan, India, and Bangladesh which feed more than a billion of the world's people (www.guardian.co.uk/environment/2009/mar/19/rust-fungus-global-wheat-crops). In Pakistan, wheat is cultivated on an area of 8 million hectares and its annual production was about 18.47 million tons in 2001-2002. Flourmills are one of the major food processing industries in Pakistan (Ahmed, 2005). Pakistan is much below in respect of international average wheat yield and the same is stagnant for the last few years while the population has increased drastically thereby widening the gap between demand and supply of basic staple food (Afzal *et al.*, 2007).

Wheat Rusts are the common diseases in many parts of the world but the frequency of epidemics (Wellings *et al.*, 2012) and yield loss caused by these diseases are different in different parts of the world. Various methods are available to control wheat rusts. One of them is by growing resistant wheat varieties which are economically as well as environmentally safe for protection of wheat against rusts (Johnson *et al.*, 1996).

The rusts of wheat are still the most important diseases on a global basis. The importance of these diseases is due to their wide distribution, their capacity to mutate and become virulent on previously resistant cultivars, their rapid rate of disease increase (epidemic potential), and their ability to remain viable after dispersal over great distances. Stripe rust of wheat is important in the northern areas of India, Pakistan and Nepal, in

Southeast Afghanistan, and at the higher elevations in the south (Roelfs & Bushnell, 1985).

There are three main types of rust diseases that occur on wheat viz., stem rust, leaf rust and yellow rust. Each of these diseases is caused by a particular species of the "rust" fungus, *Puccinia* (Marsalis & Goldberg, 2006; Safdar *et al.*, 2013). Stripe rust, caused by *Puccinia striiformis* F. sp. *tritici* (Pst), is one of the most destructive diseases of wheat (*Triticum aestivum* L.) worldwide (Yin *et al.*, 2009; Chakraborty *et al.*, 2010). Genetic resistance is the only economic and practical control of rust diseases (Pathan & Park, 2006). Encouraging wheat breeders to broaden the genetic basis of resistance of the individual cultivars may expand the genetic basis of resistance in a cultivar (Cantu *et al.*, 2011). This can be achieved by accumulating genes from diverse sources which may increase the expected lifetime of the cultivars resistance (Heisey *et al.*, 1997).

Computational analysis provided us with techniques such as sequence alignment, homology modeling, comparative genomics etc. A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species which is the use of molecular systematic to construct phylogenetic trees.

The current study was aimed to propose an alternate resistance gene for wheat against wheat rust disease that would be more effective to produce desirable results. Thus through a series of computational analysis steps, we obtained a resistant gene that will work in collaboration with the already existing genes to produce more positive dominant effect against *Puccinia striiformis* rust attack in wheat crop.

Methodology: The aim of our work is to feature out a new resistance gene against yellow rust disease in wheat.

Retrieval of gene sequence: To find out alternate resistance gene, the sequences of wheat's own resistance genes were retrieved from National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov>). The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Translation of gene sequence: These sequences were analyzed for their structure and function using different online Bioinformatics tools. The nucleotide sequences were translated to amino acid sequences using Expasy Translate tool (<http://au.expasy.org/tools/dna.html>). Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

Primary structure of translated protein: The Protein sequences generated from the translating tools were then analyzed for their primary structures on Swiss.pdb viewer (<http://au.expasy.org/spdbv/text/main.htm>). Swiss-PdbViewer is an application that provides a user friendly interface that allows analysis of several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.

Secondary structure prediction of protein: Secondary structures were predicted by using Hierarchical Neural Network tool (HNN) (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html). CPH Model was used to predict the tertiary structures of the proteins which is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions.

Tertiary structure prediction of protein: Protein modeling was done by using GENEIOUS software. BLAST was used in which query sequence were compared with other sequences in the database. On the basis of alignment and minimum e-value most similar sequence was identified as when structures are similar, functions are also similar.

Results

According to literature review RGA1 is the gene (www.ncbi.nlm.nih.gov/nucleotide/DQ494534) which provides protection against fungal attack to wheat plant in yellow rust disease.

Retrieval of gene sequence: Nucleotide sequence of RGA1 was retrieved from NCBI database with accession number DQ494534.

RGA gene sequence:

```
gggctgggcaaacccacgctgcacaagtattctacaatgtacctgaaattcaaa
agcacttaatttgatgctatgggtgcatctctga
caactttgatgtgattctctggctacaagaatgtgaagcagcttctctgagaa
gaaggatgatggcacaagaaccgctccaagaagaagaaaaccactg
gataaactcaggatttggcagcgggcaaggtacctcttactggatgatgt
ctggaataaagaggtttataaatggaacagcttaaggcccgacttcaacatggt
ggcatgggtagcgtagtcttccaccccaa
```

Translation of RGA1 gene: The nucleotide sequence of RGA1 gene was translated and compared with the translated sequence present on the NCBI Database (www.ncbi.nlm.nih.gov/nucleotide/DQ494534). Every region of DNA has six possible reading frames, three in each direction (left or right). Normally only one reading frame is used in translating a gene (in Eukaryotes), and this is often the longest open reading frame.

Validation of translation: The protein sequence obtained by Expasy tool was compared with protein structure obtained by other translating tool (www.attotron.com/cybertory/analysis/trans.htm) in order to find out which frame is translated in actual protein. Results show that the first reading frame in 5' to 3' direction is the original protein encoded. This translated sequence was similar to the RGA1 protein sequence obtained before, present at NCBI Database.

Translated sequence:

```
GLGKTTLAQVIYNVPEIQKHFNMLWVCISDNFDV
DSLATRIVEAASPEKKDDGKTKET
APKKKKTPLDKLQDLVSGQRYLLVLDDVWNKEVY
KWKQLKARLQHGGMGSVVFPTP
```

Primary structure of RGA1 gene: In order to generate tertiary structure of protein, it is important to be familiar with its primary and secondary structure. For this purpose protein sequence was first viewed in Swiss.pdb (<http://au.expasy.org/spdbv/text/main.htm>) which generates its primary structure. It shows molecules composed of groups like amino acids, containing atoms whose coordinates are taken directly from PDB file (Fig. 1).

Secondary structure prediction: Secondary structure prediction is a set of techniques in Bioinformatics that are used to predict the local secondary structures of proteins based only on knowledge of their primary structure i.e., amino acid sequence respectively. This prediction was done with the help of Hierarchical Neural Network tool (HNN) (Fig. 2).

Tertiary structure prediction: Next was a straight move on tertiary structure prediction of the query protein. The tertiary structure of a protein is its three-dimensional structure, as defined by the atomic coordinates. It is largely determined by the Protein's primary structure or sequence of amino acids of which it is composed. Efforts to predict tertiary structure from primary structure are known as protein structure prediction. The primary structure was modeled through CPH model-3.0 which shows an abstract structure of the RGA1 protein. 3-D structure was predicted on the basis of e-value, where minimum e-value was shown by 1Z6T protein as shown in Fig. 3.

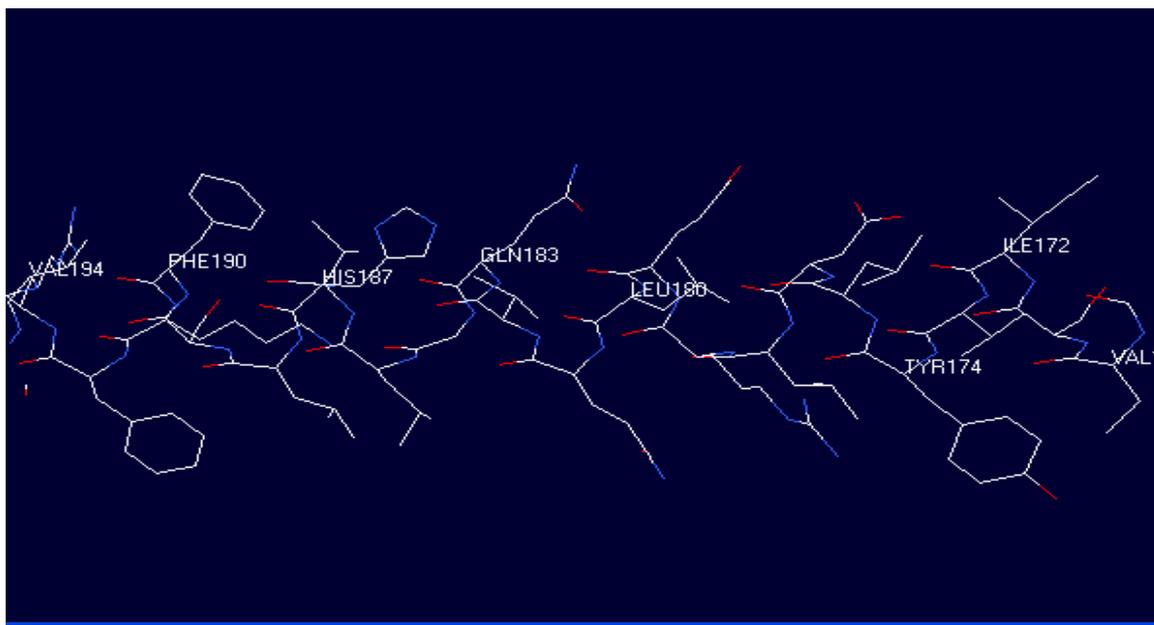


Fig. 1. Primary arrangement of RGA1 gene in wheat.

```

GLGKTTLAQVIYNVPEIQKHFNLMLWVCISDNFDVDSLATRIVEAASPEKKDDGKTKETAPKKKKTPLDKL
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
QDLVSGQRYLLVDDVWNKEVYKWKQLKARLQHGGMGSVVFPPT
Hhhhcceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
    
```

Fig. 2. Color distinction shows helices, beta sheets, strands, and extended strands, and each percentage depicts the intensity in the RGA1 protein sequence).

```

Searching for template ...

Round 1. All hits:
entry: 1Z6T chain: A score: 42 E: 1e-04
entry: 1Z6T chain: D score: 42 E: 1e-04
entry: 1Z6T chain: B score: 42 E: 1e-04
entry: 1Z6T chain: C score: 42 E: 1e-04
entry: 2A5Y chain: B score: 30 E: 0.71
entry: 3ECC chain: A score: 30 E: 0.74
entry: 2A5Y chain: C score: 30 E: 0.79
entry: 3EC2 chain: A score: 28 E: 3.1
entry: 2IPC chain: B score: 26 E: 8.8
entry: 2IPC chain: D score: 26 E: 8.8
entry: 2IPC chain: A score: 26 E: 8.9
entry: 2IPC chain: C score: 26 E: 8.9
    
```

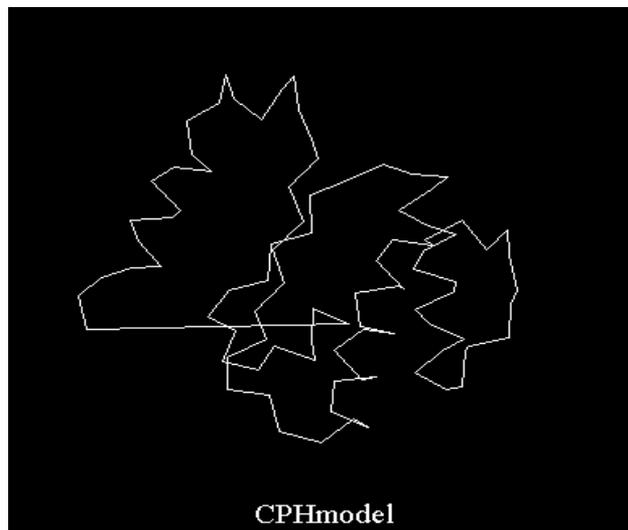


Fig. 3. Predicted structure on the basis of existing template of 1Z6T gene.

The clear image of RGA1 protein was viewed by using Rasmol (Fig. 4a). It showed clear interpretation of 3D structure of protein. As mentioned before, 3D structure of RGA1 gene is predicted on the basis of minimum e-value which was shown by 1Z6T gene as the structure of RGA1 matches with the 4 chains of 1Z6T gene (Fig. 4a). A BLAST search enables a researcher to compare a query sequence with a database of sequences and identify sequences that resemble the query sequence above/at a certain threshold. The query sequence of RGA1 was blast against different species in order to identify a new gene that is similar in structure to RGA1

gene. Similar gene was selected on the basis of pair wise alignment and minimum e-value. Resistance gene from Rye plant shows 94.2% sequence similarity with 2.37e-144 e-value (Fig. 4b). With the help of its Accession number DQ494535, nucleotide sequence of that gene was retrieved from NCBI database. Its protein structure was analyzed showing high similarity and was assumed that it could perform similar function as RGA1 gene of wheat plant. Therefore, gene sequence from Rye plant is proposed as the new resistance gene which if placed in wheat plant, can impart maximum resistance against fungal attack in case of yellow rust disease.



Fig. 4a. (3-D structure of RGA1 gene).

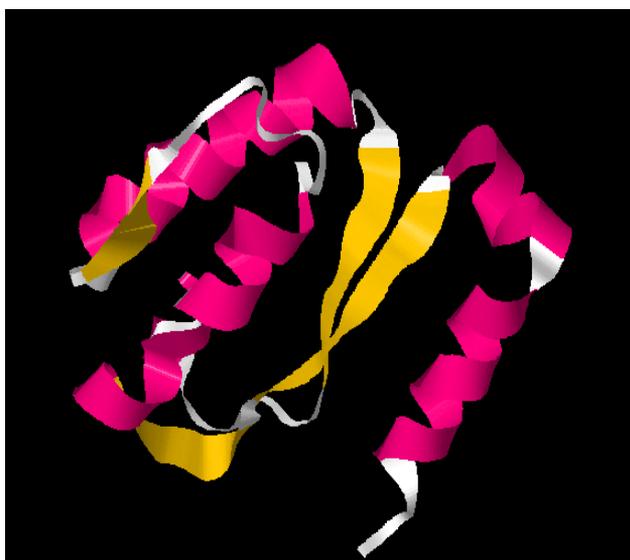


Fig. 4b. (3-D structure of alternate resistance gene).



Fig. 4c. (3-D structure of 1Z6T gene).

Discussion

Two forms of resistant genes against Yellow Rust disease in wheat were identified i.e., RGA1 and RGA2. RGA1 gene comprises 343 base pairs having accession number DQ494534 at NCBI database, encodes RGA1 protein (<http://www.ncbi.nlm.nih.gov/nucore/DQ494534>). RGA2 gene consists of 316 base pairs having an accession number AY015491 (www.ncbi.nlm.nih.gov/nucore/AY015491). Coding sequence of RGA2 gene is not determined. Along with resistance genes of wheat, six virulence genes of *Puccinia striiformis* were also identified which are vir 1, vir 2, vir 3, vir 4, vir 5 and vir 6 (Amsaleg *et al.*, 2002). These virulence Genes expressed in germinated urediniospores of *P. striiformis*, were identified by Expressed Sequence Tag analysis. Secondary structure shows the arrangement of amino acids in α -helices, β -sheets and Coils. Secondary structure obtained for RGA1 gene and alternate Gene (Rye) are very much similar in their secondary structures. The number of occurrence of helices in RGA1 are 37 in a sequence whereas number of occurrence of helices in alternate are 39 (two amino acids in alternate are more than RGA1 gene.). The software results clearly show the occurrences of Helices, B-sheets and coils in proteins sequences (Fig. 5a & 5b). Similarly occurrences of β -sheets in RGA1 gene are 24 and alternate gene has 22. Numbers of occurrences for Coils in RGA1 gene is 53 and for alternate genes are 52. These minor differences in their amino acid arrangement showed much similarity between them. The comparison of both genes is illustrated in Table 1.

3D-Structure of the Protein is predicted on the basis of e-value. 3D structure of RGA1 is predicted on the basis of minimum e-value shown by 1Z6T gene as the structure of RGA1 matches with the 4 chains of 1Z6T gene. 1Z6T gene is apoptotic-protease activating factor which is involved in Programmed cell death (Fig. 4a & 4c).

BLAST is used to find the alternate resistant gene. Query sequence is compared with other sequences in the database. On the basis of alignment and minimum e-value most similar sequence is identified. When structures are similar, functions are also similar. As a result we identified the most similar gene which is almost similar in structure and function. If we observe the structure of protein of the alternate gene, we may come across slight differences in the arrangement of amino acids which has no effect on the overall structure of alternate gene's protein (Fig. 6).

The newly identified resistance sources could be of great importance for enhancing the genetic base of resistance of bread wheat to Yellow Rust. Previously all the genes were inserted into the crop through recombinant DNA technology which was carried out in wet lab. The new alternate resistance gene has been identified by using Bioinformatics tools/approaches. The rapidly emerging field of bioinformatics promises to lead to advances in understanding basic biological processes. Bioinformatics has transformed the discipline of biology from a purely lab-based science to an information science as well.

Acknowledgement

The Authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding the work through the research group project RGP-VPP-062.

References

- Afzal, S.N., M.I. Haque, M.S. Ahmedani, S. Bashir and A.R. Rattu. 2007. Assessment of Yield losses caused by *Puccinia striiformis* triggering stripe rust in the most common wheat varieties. *Pak. J. Bot.*, 39(6): 2127-2134.
- Amsaleg, R.C., V.C. de Pope, Y. Brygoo and C. Levis. 2002. Characterization of a length polymorphism in the two intergenic spacers of the ribosomal RNA gene of *Puccinia striiformis* F. sp. *tritici*, the causal agent of the wheat yellow rust. *Mycological Research*, 106: 918-924.
- Cantu, D., M. Govindarajulu, A. Kozik, M. Wang, X. Chen, K.K. Kojima, J. Jurka, R.W. Michelmore and J. Dubcovsky. 2011. Next Generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE*, 6(8).
- Chakraborty, S., J. Luck, G. Hollaway, G. Fitzgerald and N. White. 2010. Rust-proofing wheat for a changing climate. *International Journal of Plant Breeding*.
- Heisey, P.W., M. Smale, D. Byerlee and E. Souza. 1997. Wheat rusts and the costs of genetic diversity in the Punjab of Pakistan. *Amer. J. Ag. Econ.*, 79: 726-737.
- Johnson, R., K. Manninger, M. Csoz and A. Mesterhazy. 1996. Vulnerability of yellow rust resistance in Hungarian wheats. In: (Ed.): Anon. *Proc. 9th Eur. and Meiterr. Cereal Rusts & Powdery Mildew Conference*, Lunteren. Wageningen. Wageningen Press. pp. 307-308.
- Marsalis, M.A. and N.P. Goldberg. 2006. Extension agronomist, leaf, stem and stripe rust diseases of wheat. *Extension Plant Pathologist. Guide A-415*: 1-8.
- Munir Ahmed. 2005. Post harvest technology for various crops of Pakistan: Current Status, Employment Generation and Prospects. *Science, Technology and Development; A Quarterly Journal*, 24(2).
- Pathan, A.K. and R.F. Park. 2006. Evaluation of seedling and adult plant resistance to leaf rust in European wheat cultivars. *Euphytica*, 149: 327-342.
- Roelfs, A.P. and W.R. Bushnell. 1985. The Cereal Rusts. *Academic Press Inc.*, 2: 261-264.
- Safdar, S.F., H. Rashid and Z. Chaudhry. 2013. In silico Identification of alternate resistance gene against Yellow Rust disease in Wheat. *Pak. J. Bot.*, 45(S1): 347-352.
- Wellings, C.R., A.B. Lesley and X.M. Chen. 2012. Resistance to stripe rust in wheat: Pathogen biology driving resistance breeding. *Durable Disease Resistance Symposium*. Sharma, I. *Disease Resistance in Wheat*. Edition I. New York, NY, CAB International. 63-83.
- Yin, C., X. Chen, X. Wang, Q. Han, Z. Kang and S.H. Hulbert. 2009. Generation and analysis of expression sequence tags from haustoria of the wheat stripe rust fungus *Puccinia striiformis* F. sp. *tritici*. *BMC Genomics*, 10: 626.
- Zafar, S., M.Y. Ashraf., M. Niaz., A. kausar and J.hussain. 2015. Evaluation of wheat genotypes for salinity tolerance using physiological indices as screening tool. *Pak. J. Bot.*, 47(2): 397-405.

(Received for publication 4 January 2016)