Identification and Genotyping of SNPs in \textit{RKMI} and \textit{RKM4} Genes of \textit{Sordaria fimicola}

Iqra Mobeen$^*$, Rabia Arif$,^*$ Maimoona Ilyas$^1$, Siu Fai Lee$^2$ and Muhammad Saleem$^1$

$^1$Molecular Genetics Research Laboratory, Department of Botany, University of the Punjab, Lahore 54590, Pakistan  
$^2$Department of Biological Sciences, Macquarie University and CSIRO Black Mountain Laboratories, Building 101, Clunies Ross Street, ACT 2601, Australia

**ABSTRACT**

Single nucleotide polymorphisms (SNPs) are one of the most common and abundant class of molecular markers present in the genome of many organisms. The current study represents the first attempt to investigate the natural variations in the \textit{RK-MTases} genes; Ribosomal N-lysine methyltransferase1 (\textit{RKMI}) and Ribosomal N-lysine methyltransferase4 (\textit{RKM4}) in \textit{Sordaria fimicola} using SNP markers. A total seven SNPs in the \textit{RKMI} gene and nine in \textit{RKM4} gene were identified. A subset of SNPs were unique in SFS strains and others were fixed in the NFS strains of \textit{S. fimicola}. These polymorphisms might be adaptive in stressful environmental conditions. Genotyping of eight SNPs of \textit{RK-MTases} genes of \textit{S. fimicola} was accomplished by designing allele specific primers via amplification refractory mutation system–PCR (ARMS-PCR) yielding amplicons of different sizes. This study concluded that SNP markers are an efficient and informative marker system in \textit{S. fimicola}. Most of the studied SNPs are non-synonymous substitutions, which might underpin functional differences in their protein products.

**INTRODUCTION**

From last two decades, single nucleotide polymorphisms (SNPs) have become the most popular molecular marker system to study polymorphisms in natural populations of numerous organisms (Váši et al., 2008; Costes et al., 2009; Ljungqvist et al., 2010; Guichoux et al., 2017; Fischer et al., 2018). SNPs are the most abundant type of molecular marker and can be identified in animals, plants and as well as fungi. Their abundance make them ideal for the study of inheritance of genomic regions including exonic and intronic regions (Berger et al., 2001; Wicks et al., 2001; Stickney et al., 2002).

SNP is one of the simplest and most common forms of polymorphism which arises due to the substitution of one nucleotide with the other nucleotide (Shastry, 2002). Due to the environmental stress, gene conversion, and deficiency in the DNA repair mechanisms. These variations are driving force of species evolution and adaptation (Lamb et al., 1998; Hoffmann and Hercus, 2000; Saleem et al., 2001).

Owing to their abundance, SNPs are present in the frequency of approximately one in every kilobase in the human genome (Brookes, 1999). These simplest forms of genetic variation are more in the non-coding regions, with less deleterious effects. Those SNPs, which do not change the encoded amino acids are known as synonymous substitutions and are usually not involved in natural selection (Kimura, 1983). In contrast, SNPs that alter the encoded amino acids are recognized as non-synonymous substitutions and are more likely to be under natural selection. SNPs can be observed between individuals in a population, may change the promotor activity, influence the DNA and pre mRNA conformation as well as change the phenotypic expression (Lamb et al., 1998; Hoffmann and Hercus, 2000; Saleem et al., 2001).

SNPs are present twice as frequent in non-coding and intergenic regions than in coding regions of the genome (Zhao et al., 2003). Moreover, genome wide studies depict that SNPs of non-coding regions are physically associated with functional regions of genome (Kim et al., 2007). These days, automated next generation sequencing make the SNPs detection and genotyping straightforward (Kaiser et al., 2016). In addition, current SNP based studies generally need to bear high upfront costs in SNP discovery (Chen et al., 2008; Lai et al., 2007) and then genotype them in target organisms (Van Oorsouw et al., 2007; Van Tassell et al., 2008).

For low and medium throughputs SNPs genotyping PCR is most commonly used (Chuang et al., 2008). There are many PCR methods available for this purpose but...
selecting the suitable one is critical key factors that taken into account include the nature of polymorphism, number and type of samples, and availability of appropriate instrument for the desired sensitivity and throughput levels (Hamajima et al., 2002).

A very simple and cost-effective method for SNP genotyping is tetra-primer amplification refractory mutation system–PCR (ARMS-PCR), which requires only a PCR reaction followed by gel electrophoresis (Ye et al., 2001). We applied ARMS-PCR approach to study genetic variation in the fungus Sordaria fimicola, targeting SNPs in RKM1 and RKM4 genes.

MATERIALS AND METHODS

Sub-culturing of fungi

S. fimicola strains (S1, S2, S3, N5, N6 and N7) were sub-cultured on potato dextrose agar (PDA) media under sterile conditions, which were provided by Molecular Genetics Laboratory, University of the Punjab, Lahore. These strains were originally collected from “Evolution Canyon” in Israel (S1, S2 and S3 strains were collected from the South Facing Slope (SFS) and the N5, N6, N7 strains from the North Facing Slope (NFS) of the “Evolution Canyon”. The fungal cultures were incubated at 20°C in a refrigerator for 7-9 days and were harvested for DNA extraction.

DNA extraction and PCR amplification of RKM1 and RKM4 genes

DNA extraction from all studied strains of S. fimicola was carried out by using modified DNA extraction protocol of Pietro et al. (1995). Four primer pairs (two for each gene) were used to amplify both R-K-MTases genes. The primer pairs used for amplification of RKM1 gene were; RKM1F1 (5' -GGAGAAGAAGCAGTATCTTAGT-3'), RKM1R1 (5'-GCAATCCATATCCAGAGAC-3') and RKM1F2 (5'-TCATGGCATTAGTATGG-3'), RKM1R2 (5'-TAATGTGCTTCCGTGGG-3').

For RKM4 gene, primer pairs were; RKM4F1 (5' -AGAGATACCCGAAAACATTAGT-3'), RKM4R1 (5'-CAGTTAGAGCTGTAAGTAA-3') and RKM4F2 (5'-GAAGAAAGGTGTACAACA-3'), RKM4R2 (5'-GGACGTGGAGACAGCTT-3'). The PCR reaction volume was 15µl, which contained; 10µl 2X Amp Master Mix (GeneAll), 1µl forward primer, 1µl reverse primer (100µM each), 2µl DNA sample (1 in 10 dilution of the g-DNA stock) and 1µl dd H2O. Touch down PCR conditions were used to amplify the R-K-MTases genes. The stage 1 included the 15 cycles with initial denaturation at 95°C for 3 min, second denaturation for 30 sec, annealing at Tm+10°C for 45 sec and elongation at 72°C for 60 sec. The stage 2 contained 25 cycles with denaturation at 95°C for 30 sec, annealing at Tm-5°C for 45 sec and elongation at 72°C for 60 sec. The termination stage contained elongation at 72°C for 5 min, stop reaction at 4°C for 15 min and final hold at 23°C until removed from thermal cycler. 1.0% agarose gel electrophoresis was carried out to resolve the PCR products, stained with ethidium bromide and visualized under UV light in Gel Documentation System (Syngene).

Sequencing of genes and sequence analysis for SNPs

Amplicons were sequenced at Macrogen Korea and sequences were edited using the BioEdit program. Multiple sequence alignment was carried out for both methyltransferase genes (RKM1 and RKM4) separately using the Clustal Omega online tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) to identify SNPs using S. cerevisiae sequence as reference.

Designing of primers for SNP sites

For SNP genotyping, a total of eight specific primers based on SNP sites for both RKM1 and RKM4 genes were designed; rkm1F1 (specific for SFS strains), rkm1F2 (specific for all strains of S. fimicola), rkm1F3 (specific for SFS strains), rkm1F4 (specific for all strains of S. fimicola) and rkm4F1, rkm4F2, rkm4F4 (specific for SFS strains), rkm4F3 (specific for NFS strains). The reverse primer RKM1R1 (5'-ACAAATCCATATCCAGAGAC-3'), (specific for RKM1 gene) was used in combination with the forward primers specific to the SNPs of RKM1 gene. Likewise, reverse primer RKM4R1 (5'-CAGTTAGAGCTGTAAGTAA-3'), (specific for RKM4 gene) was used in combination with the forward primers of SNPs of RKM4 gene. An additional mismatch at third base towards the 3’ end was deliberately introduced in each SNP primer, a G was substituted with a T and a C substituted with an A and vice versa (Table I).

ARMS PCR conditions for amplification of SNP sites

To differentiate between the target SNPs in the RKM1 and RKM4 genes ARMS-PCR conditions were used. For amplification of four SNPs of RKM1 gene of S. fimicola, four forward primers (rkm1F1, rkm1F2, rkm1F3, rkm1F4) along with reverse primer rkm1R1 (specific to RKM1 gene) were used. Likewise, for four SNPs of RKM4 gene of S. fimicola, four forward primers (rkm4F1, rkm4F2, rkm4F3, rkm4F4) along with reverse primer rkm4R1 (specific to RKM4 gene) were used. The 20µl PCR reaction mixture for SNPs of RKM1 gene contained 2µl DNA (1 in 10 dilution of the g-DNA stock), 1µl rkm1F1 primer, 1µl rkm1F2 primer, 1µl rkm1F3 primer, 1µl rkm1F3 primer, 1µl rkm1R1 primer (100µM each), 10µl 2X Amp Master
Mix (GeneAll) and 3µl ddH2O. The reaction mixture for SNPs of RKM4 gene was prepared in the same way as for SNPs of RKM1. The PCR conditions for SNPs of both genes consisted of 5 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 62.5 °C, 40 s at 72 °C, and a final 7 min extension at 72 °C (Yang et al., 2017).

**Table I. Oligonucleotide primers used in the current study.**

<table>
<thead>
<tr>
<th>Name of primer</th>
<th>Oligonucleotide sequence</th>
<th>Position in gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>RKM1F1</td>
<td>5'-GTAAAAGCACTACTTCAGT-3'</td>
<td>15-34</td>
</tr>
<tr>
<td>RKM1R1</td>
<td>5'-ACAAATGGGATGAGAGAG-3'</td>
<td>395-414</td>
</tr>
<tr>
<td>RKM1F2</td>
<td>5'-TAAATAGTCTCCTCGGTTG-3'</td>
<td>844-864</td>
</tr>
<tr>
<td>RKM1R2</td>
<td>5'-AGATAAGAAGAACTTTTG-3'</td>
<td>1662-1681</td>
</tr>
<tr>
<td>RKM4F1</td>
<td>5'-AGAAATTGACTGTTAGAAGG-3'</td>
<td>16-35</td>
</tr>
<tr>
<td>RKM4R1</td>
<td>5'-CAGTTAGAGTCAAGGTAA-3'</td>
<td>656-675</td>
</tr>
<tr>
<td>RKM4F2</td>
<td>5'-GAACTGCTAGAAGATATTG-3'</td>
<td>702-727</td>
</tr>
<tr>
<td>RKM4R2</td>
<td>5'-GGACGTTGACAGAGCTTTT-3'</td>
<td>1391-1410</td>
</tr>
<tr>
<td>rkm1F1</td>
<td>5'-GTGAATCCACTAAGACT-3'</td>
<td>A 20-37</td>
</tr>
<tr>
<td>rkm1F2</td>
<td>5'-AAAGAGTGGTTGGAAATTT-3'</td>
<td>C 256-273</td>
</tr>
<tr>
<td>rkm1F3</td>
<td>5'-TTTTATGGTCGCACC-3'</td>
<td>T 428-445</td>
</tr>
<tr>
<td>rkm1F4</td>
<td>5'-GAAACTGTCCTGACAC-3'</td>
<td>A 485-502</td>
</tr>
<tr>
<td>rkm4F1</td>
<td>5'-TAACGGGTATACCTGG-3'</td>
<td>T 33-50</td>
</tr>
<tr>
<td>rkm4F2</td>
<td>5'-TAGCGACGTCCTTCG-3'</td>
<td>A 266-243</td>
</tr>
<tr>
<td>rkm4F3</td>
<td>5'-GATTCCACTTGCTGTA-3'</td>
<td>A 424-441</td>
</tr>
<tr>
<td>rkm4F4</td>
<td>5'-ATGGTTGCTTTGAGG-3'</td>
<td>G 502-519</td>
</tr>
</tbody>
</table>

**Note:** Bold underlined nucleotides are showing additional mismatches, where G substituted with T and C substituted with A and vice versa. Highlighted nucleotides are showing SNPs.

**RESULTS AND DISCUSSION**

Molecular markers have become a popular tool for observing polymorphism in plants, animals and fungi. Among all marker systems, SNPs are the most prevalent molecular marker for describing genetic variation in natural populations. These are useful for observing genetic variation, population genetic structure and reconstructing the evolutionary history of species (Banke and McDonald, 2005; Coates et al., 2009; Fischer et al., 2017).

To the best of our knowledge, it is the first time SNPs identified and genotyped in the RK-MTases genes in S. fimicola. A lot of work has been carried out on SNPs of plants and humans but a very few or negligible studies are done on fungal SNPs. In the present study, SNPs genotyping was carried out by performing allele specific PCR conditions to observe polymorphisms in RKM1 and RKM4 methyltransferase genes of S. fimicola. The RKM1 and RKM4 regions of S. fimicola were amplified with target-specific primers by using touchdown PCR conditions and the product sizes were 1320bp and 900bp respectively.

The results of multiple sequence alignment showed that the RKM1 and RKM4 regions for six strains of S. fimicola and S. cerevisiae were identical except for polymorphic sites. These polymorphic sites are due to substitution of single nucleotide and hence termed as single nucleotide polymorphism (SNP). Total seven SNP sites for RKM1 region and nine SNP sites for RKM4 region of S. fimicola were identified (Supplementary Figs. 1 and 3) but genotyping of total eight SNPs were carried out for both RK-MTases in this study (Figs. 1 and 2). All observed SNPs of RKM1 region for all strains of S. fimicola were identical except for two SNPs, which are unique for SFS strains (Fig. 1). In RKM4 region, some SNPs are present in SFS strains but not present in NFS strains and vice versa. For example, SNP at 35th position in the RKM4 region is present only in SFS strains but not present in NFS strains (Fig. 2).

**Fig. 1.** Multiple sequence alignment of RKM1 region of NFS and SFS strains of S. fimicola with respect to the Saccharomyces cerevisiae. Symbol (*) showing conserved sites, space and highlighted regions showing SNPs. Four primers specific to SNP sites are; rkm1F1, rkm1F2, rkm1F3, rkm1F4.
Fig. 2. Multiple sequence alignment of RKM4 region of S. fimicola strains with respect to S. cerevisiae in order to observe single nucleotide polymorphism (SNP). Symbol (*) is showing fully conserved sites, space and highlighted regions showing SNPs. Four primers specific to SNP sites are; rkm4F1, rkm4F2, rkm4F3, rkm4F4.

For RKM4 region, the SNP site at 50th nucleotide position is unique for S1 and S3 strains where A substituted with G. The SNP site at 284th position is unique for NFS strains (T substituted with A) and absent in SFS strains. The SNPs at 442 and 520 nucleotide position, in which G substituted with A and T with C are present only in SFS strains, but not in NFS strains (Fig. 2).

In RKM1 gene, at first SNP of SFS strains, T substituted with A at second base of codon, resulted in change of ATC codon into AAC, which changed the Isolucine (I) into asparagine (N). At second SNP of SFS and NFS strains, substitution at third base of the codon did not change the encoded amino acid. At 3rd SNP of SFS strains, replacement of A with G, resulted in the change of codon from ATA to GTA and amino acid from isoleucine (I) to valine (V). At 4th SNP of SFS and NFS strains, substitution at first base of codon occurred and codon changed from TCT to CCT and encoded amino acid from serine (S) to proline (P) (Fig. 1 and Supplementary Fig. 2). First and third SNPs of SFS strains showed conservation among the groups of strongly similar properties. These are shown by symbol (:) in the amino acid sequence of RKM1 protein in Supplementary Figure 2.

In RKM4 gene, at first polymorphic site of SFS strains, T substituted with A at second base of codon, resulted in change of ATC codon into AAC, which changed the Isolucine (I) into asparagine (N). At second polymorphic site in NFS strains, T replaced with A at first base of the codon, where TTT converted into ATT and changed the amino acid from phenylalanine (F) to isoleucine (I). In NFS strains at 4th polymorphic site, G substituted with A at third base of the codon (ATG-ATA), resulted into the change of methionine (M) into isoleucine (I) (Fig. 2 and Supplementary Fig. 4).

A number of studies have been carried out on SNPs in different genes as well as whole genome studies by next generation sequencing to observe polymorphism in different plants and fungi. Sun et al. (2013) identified three SNP rich genomic regions and observed polymorphisms in rice false smut Ustilaginoidea virens. Whole genome scan for SNP identification was carried out in Soybean. Likewise, Trick et al. (2009) and Park et al. (2010) observed SNPs in the whole genome of Brassica and Li et al. (2009) reported SNPs in candidate genes controlling morphological traits of leaves and flowering time. Lopez et al. (2000) reported SNPs in candidate genes delta 12 fatty acid desaturase and in fatty acid desaturase 2A in Arachis hypogaea L.

ARMs has become a standard technique that was first described by Newton and colleagues in 1989. It allows the discrimination of alleles that differ by as little as 1bp. In order to genotype the SNPs, the critical part of ARM-PCR is to design the primers. A single mismatch at 3rd end is not sufficient to avoid non-specific binding, so an extra mismatch was introduced at the 3rd base pair at 3rd end to allow specific binding (Wang et al., 2010; Medrano and de oliveiro, 2014).

In the current study, eight SNP sites for both RKM1 and RKM4 regions (4 for each) were amplified using SNPs site specific primers. Four forward primers; rkm1F1, rkm1F2, rkm1F3 and rkm1F4 were used in combination with reverse primer RKM1R1 to amplify the four SNP sites of the RKM1 region. Four SNP sites having 650bp, 520bp, 240bp and 200bp were amplified by ARM-PCR respectively, shown in schematic diagram in Figure 3a. Likewise, four SNPs primers (rkm4F1, rkm4F2, rkm4F3 and rkm4F4) in combination with reverse primer RKM4R1 used to amplify the SNPs of RKM4 region and amplicons obtained as; 600bp, 450bp, 250bp and 200bp respectively, shown in schematic diagram in Figure 3b.

ARM-PCR is allele specific PCR, which is much reliable and reproducible. It does not require restriction
digestion and sequencing of PCR product. It only requires the separation of different DNA fragments by using agarose gel electrophoresis. Hence, it is a cost-effective technique for genotyping of SNPs in coding and non-coding regions of plants, animals, and fungal genome.

CONCLUSION

SNP is a reliable, efficient, and highly reproducible molecular marker to observe polymorphisms in coding regions as well as in non-coding regions. In this study, we successfully identified and genotyped SNPs in the fungus *S. fimicola* via ARMS-PCR. SNPs specific to the SFS strains and the NFS strains of *S. fimicola* were found, some of which were non-synonymous substitutions, which might have an important role in evolution and adaptive values in their respective environmental conditions.

FUNDING ACKNOWLEDGEMENT

This research is not funded by any governmental or non-governmental organization, institution etc.

Supplementary material

There is supplementary material associated with this article. Access the material online at: https://dx.doi.org/10.17582/journal.pjz/20190902090906

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES


Identification and Genotyping of SNPs in RKM1 and RKM4 Genes of Sordaria fimicola

Iqra Mobeen1*, Rabia Arif1*, Maimoona Ilyas1, Siu Fai Lee2 and Muhammad Saleem1

1Molecular Genetics Research Laboratory, Department of Botany, University of the Punjab, Lahore 54590, Pakistan
2Department of Biological Sciences, Macquarie University and CSIRO Black Mountain Laboratories, Building 101, Chutes Ross Street, ACT 2601, Australia

*Corresponding author: iqramobeen905@gmail.com; phdgenetics@gmail.com

Pakistan J. Zool., pp 1-7, 2021. DOI: https://dx.doi.org/10.17582/journal.pjz/20190902090906

Supplementary Material
Supplementary Fig. 1. Multiple sequence alignment of RKM1 region of NFS and SFS strains of *S. fimicola* with respect to the *Saccharomyces cerevisiae*.

Note: Symbol (*) showing conserved sites, space and highlighted regions showing SNPs.

Supplementary Fig. 2. Multiple sequence alignment of amino acid sequence of RKM1 protein of different strains of *S. fimicola* with respect to the *S. cerevisiae* amino acid sequence to observe the genetic variations. Symbol (*) showing fully conserved sites, symbol (:) depicting conservation between groups of strongly similar properties, space and highlighted regions showing polymorphic sites.
Supplementary Fig. 3. Multiple sequence alignment of RKM4 region of *S. fimicola* strains with respect to *S. cerevisiae* in order to observe single nucleotide polymorphism (SNP). Note: Symbol (*) is showing fully conserved sites, space and highlighted regions showing SNPs.

Supplementary Fig. 4. Multiple sequence alignment of amino acid sequence of RKM4 protein of different strains of *S. fimicola* with respect to the reference strain *S. cerevisiae* amino acid sequence to spot genetic diversity. Symbol (*) showing conserved sites, space and highlighted regions showing polymorphic sites.