

Microsatellite repeat dynamics in mitochondrial genomes of phytopathogenic fungi: frequency and distribution in the genic and intergenic regions

Sahil Mahfooz^{1*}, Pallavi Singh², Deepak K Maurya², Mahesh C Yadav³, Azram Tahoor⁴, Harmesh Sahay², Arpita Srivastava² & Anil Prakash¹

¹Department of Biotechnology and Bioinformatics, Barkatullah University, Bhopal 462 026, Madhya Pradesh, India; ²National Bureau of Agriculturally Important Microorganisms, Mau 275 101, Uttar Pradesh, India; ³National Research Centre on DNA Fingerprinting, National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012, India; ⁴Department of Wildlife Science, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India; Sahil Mahfooz - Email: sahil_mahfooz@yahoo.com; Phone: +917552491848, Fax: +917552491850; *Corresponding author

Received November 02, 2012; Accepted November 05, 2012; Published November 23, 2012

Abstract:

The frequency and distribution of microsatellites were analyzed in the 19 mitogenomes of phytopathogenic fungi covering five phyla. Our analysis revealed that in all the mitogenomes studied, the frequency and relative abundance varied, and it was neither influenced by genome size nor by GC content. SSRs were found to be differential distributed in genic and intergenic regions. An average of 5.14 (23.6%) SSRs were present in genic sequences and 21.7 (76.4%) SSRs were located in the intergenic sequences. Relative abundance of SSRs in mitogenomes was the highest in *Aspergillus tubigensis*, whereas, it was the least in *Phaeosphaeria nodurum*, the average being 0.45. Trinucleotide repeats were the most abundant motifs in the genic and intergenic regions of the mitogenomes of the phytopathogenic fungi. Among the genes, *cox1* harbors the maximum SSRs, whereas *cox3* and *nad7* contain the least. Based on the presence of SSRs in a particular gene, genetic relationships among individual organisms were also established.

Keywords: Simple sequence repeats, Comparative genomics, Mitochondria, Phytopathogenic fungi, Phylogenetics.

Background:

Microsatellites or Simple Sequence Repeats (SSRs) are the portions of the genome consisting of a sequence of repeats of a given string of nucleotides which have lengths from one to six bases. These repeats exhibit a strong level of instability, undergoing additions or deletions of repeated units, which lead to variation in the number of copies of the repeated stretches [1]. Microsatellites can be found in the protein-coding and non-coding regions of genome. Because of their abundance and inherent potential for variation, microsatellites are a valuable source of genetic markers, and are widely used in population genetics, genetic diversity, fingerprinting and forensic analysis in many organisms including bacteria, fungi, plants and human [2,3]. Apart from their applications as molecular marker, determining the abundance and density of microsatellites may

help to understand whether these sequences have any functional and evolutionary significance. Large scale genome sequencing initiatives on the growing number of organisms had provided an opportunity to evaluate the abundance and relative distribution of microsatellites in different genera. Several microsatellite-mining reports in various organisms including fungi have appeared in recent years, providing important data for the comparative analysis of microsatellite distribution [4, 5]. These reports were focused on nuclear genomes and relatively few microsatellite analyses have been conducted in organelle genomes, especially in the mitochondria.

Mitogenomes are generally characterized by low GC content, conservation in gene function, and high copy number [6] and

they can evolve at their own rate relative to the nuclear genomes of the organisms in which they occur [7]. The size and topology of the mitogenome, the number and nature of the proteins it encodes, and even the genetic code itself can vary greatly between the species [8]. The mitogenomes of fungi are generally an order of magnitude smaller than those of plants but larger than animal mitogenomes and usually contain 14 genes encoding hydrophobic subunits of respiratory chain complexes, as well as genes for the large (rnl) and small (rns) ribosomal subunits and a set of tRNAs [8].

The availability of mitochondrial genome sequences is increasing as a result of recent technological advances in molecular biology. Although, many fungal mitogenomes are publically available, no formal analyses of microsatellites in these sequences are reported. Thus, the aims of this study were, (1) to reveal various facets of the distribution and dynamics of microsatellites in the mitogenomes of phytopathogenic fungi (2) and to construct phylogenetic relationship between them. To accomplish this, an *in-silico* approach was used to analyze the frequency and distribution of the microsatellites in genic and intergenic regions of the genomes.

Methodology:

Sources of mitochondrial genomic sequences

The twenty one mitochondrial genome sequences of different phytopathogenic fungi were retrieved from National Center for Biotechnology Information and Broad Institute (www.broadinstitute.org) **Table 1 (see supplementary material)**. Files were obtained from the respective mitochondrial genome for each organism in FASTA format. Of the 19 fungal species taken, ten belongs to the phylum Ascomycota, 7 belongs to Basidiomycota and remaining two were from Chytridiomycota and Zygomycota. The frequency of the repeats motifs in genic and intergenic region were analyzed using WebSat online software [9], which is accessible through internet, requiring no programme installation. Repeats greater than twelve bases were considered as SSRs which means that there should be twelve occurrence of mono-nucleotide repeat, six occurrence of di-nucleotide repeats, four occurrence tri-nucleotide repeats, and three occurrences of tetra, penta and hexa-nucleotide repeats. All SSR were analyzed for their frequency of occurrence and relative abundance. Relative abundance was calculated as SSR per kb of sequence.

Statistical analysis

A binary data was generated on the basic of presence and absence of repeat motifs in a mitochondrial gene and analyzed using SIMQUAL route to generate Jaccard's similarity coefficient using NTSYS-PC, software version 2.1. These similarity coefficients were used to construct a dendrogram depicting genetic relationships among the species by employing the Unweighted Paired Group Method of Arithmetic Averages (UPGMA) algorithm and SAHN clustering. The robustness of the dendrogram was evaluated with a bootstrap analysis performed on the binary dataset using WINBOOT software (version. 2.0).

Evaluation of polymorphism

The Polymorphism Information Content (PIC) was measured as described by Botstein et al. [10]. PIC is defined as the probability that two randomly chosen copies of gene will be

different alleles and will be different within a population. The PIC value was calculated with the formula (described in supplementary material)

Discussion:

The recent technological inventions in genome sequencing have accelerated the decoding of organisms genetic architecture. This results in generation of huge amount of sequencing data of various organisms which can be exploited for evolutionary and comparative genomics, including phylogenomics. In the present study, we have estimated the number and frequency of SSRs in the mitogenomes of 19 phytopathogenic fungi. The relative abundance of SSRs is randomly distributed across all species **Table 2 (see supplementary material)**. This observation is significantly evident in species from the same genera. Among the three Eurotiomycetis, *A. niger* and *A. tubigensis* show comparable relative abundance to each other, whereas, *A. oryzae* exhibits much lower abundance. Similarly, two Sordariomycetes fungi, *F. graminearum* and *F. oxysporum* which varied greatly in their genome size exhibit a more similar relative abundance to each other, whereas, *F. verticillioides* showed higher. Karoglu *et al.* [11] while comparing the SSR repeats in nuclear genomes of fungi observed the same trend. The possible reason of this non-random distribution of SSRs may lies in the mechanism through which SSRs are generated. Two possible models have been proposed to explain microsatellite generation: replication slippage and unequal recombination. Replication slippage generally involves DNA polymerase pausing and dissociation [12]. Difference in mutability and the bias in repair efficiency of mismatch-repair system could also lead to over-representation of SSRs in certain genomes [13].

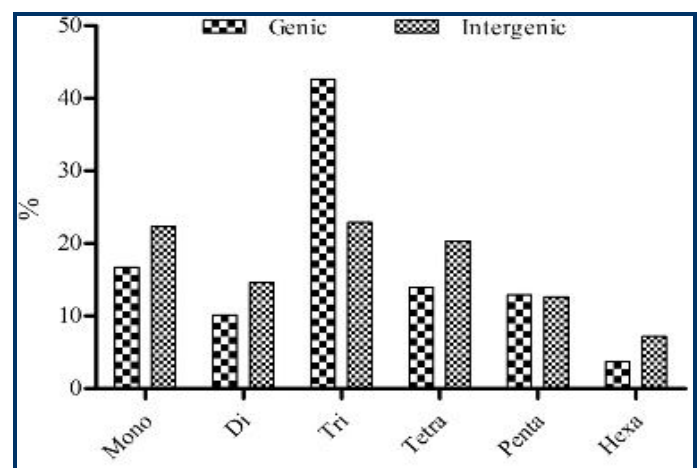


Figure 1: Graphical representation of percentage of different type of repeats found in genic and intergenic regions of mitogenomes.

While comparing the different repeat classes of SSRs in genic and intergenic regions, we observed that the trinucleotide repeats were the most abundant classes of SSRs in both genic as well as intergenic regions with a frequent occurrence of motif AAT\TTA (**Figure 1**). While comparing with other reports, we found that in the case of genic regions, this is a common trend because these trinucleotide SSRs in the coding region are translated into amino-acid repeats, which possibly contribute to the biological function of the protein [4, 5]. However, high

frequency of trinucleotide SSR in intergenic regions is rare. Rajendrakumar *et al.* [14] observed the abundance of dinucleotide SSRs in genic and intergenic regions of mitochondrial genome of rice. Our results show that the high frequency of trinucleotide repeats is a characteristic feature of mitogenomes of phytopathogenic fungi. The abundance of motif AAT\TTA may be attributed to the nucleotide composition of genome. It has been proposed that base usages of motifs are significantly correlated with the genome nucleotide composition. Percentage of AT-rich motifs will rise with the increase of genome AT content [15]. The longest repeat in the intergenic region was a mononucleotide (C) repeated 23 times in *P. noduram*. Similarly, the longest repeat in genic region was also a mononucleotide (T) which was repeated 17 times in *P. sojae*. We observed that SSRs in the genic and intergenic regions of mitogenomes were of low repeat numbers

(4-5U). The possible reason for this is the genesis of SSR which is generally through slippage and slippage can even begin at low repeat numbers. Since mitogenomes are generally conserved, it is likely that the SSRs first arise by chance substitutions that make a short repetitive sequence, which can undergo slippage if they are above some threshold size [16]. The low repeat number is also evident in the nuclear genomes of fungi where shorter numbers of repeats (5-7 U) were predominated with around 90% of all motifs [11, 17]. Cross genera conservation of motifs in a particular gene was also observed in our analysis.eg. Motif (tat)₅ was conserved in *cox 1* gene of five species. This conservation of repeat is expected because *cox 1* gene is slightly more conserved as compared to other genes and has been proposed as an appropriate region for genetic barcoding of species [18].

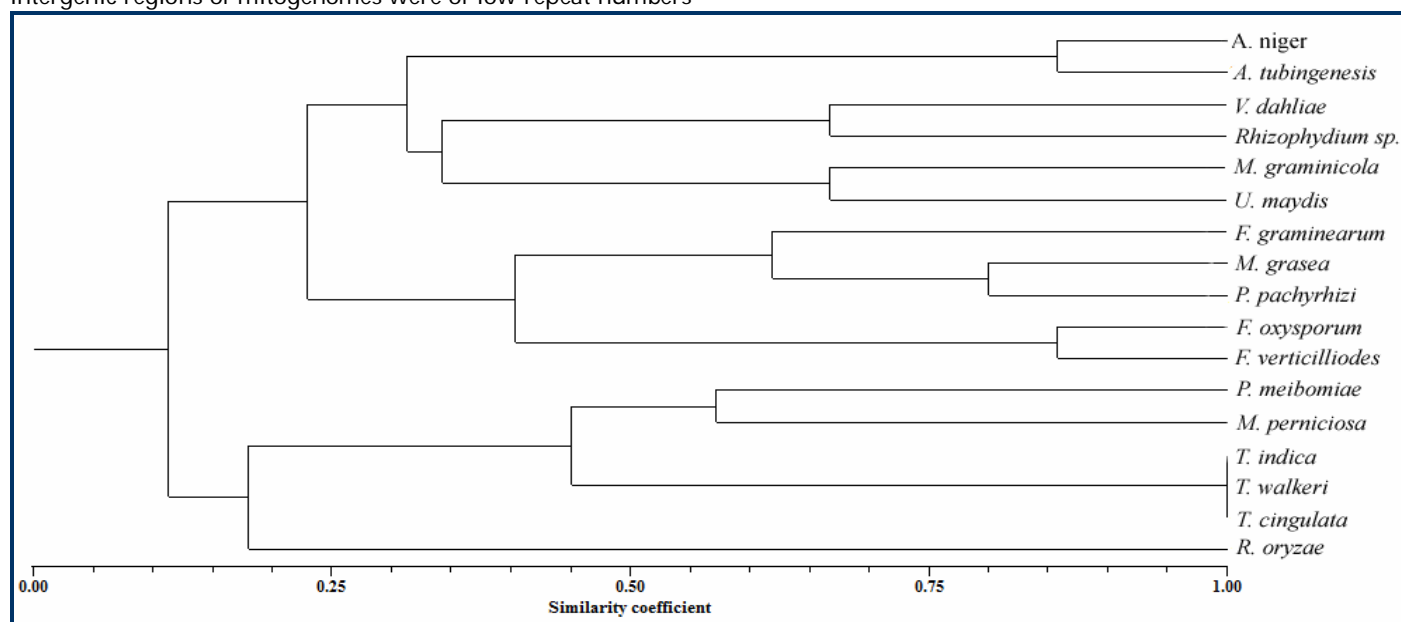


Figure 2: Dendrogram showing phylogenetic relationship between mitochondrial genome of 19 phytopathogenic fungi based on SSRs present in genes.

In our study, the presence of SSRs in specific genes was used to create genetic relationship among different species (Figure 2). The result indicates that set of genes coupled with the presence of SSR were capable of analyzing the genetic relationship. Protein-coding genes (*Cox 3* and *Nad 7*) were found to be most informative whereas ssRNA and lsRNA genes were least. Mitochondrial cytochrome genes are known to be good performer in establishing phylogenies in vertebrates also [19]. The possible explanation for the poor performance of RNA genes lies in the higher variability in these RNA genes. We observed that fungi from Ascomycota have greater proximity with Oomycota whereas, Zygomycetes were close to Basidiomycetes. This is expected because we have considered only the presence of repeat in a particular gene for estimating phylogenetic relationship. Fine resolution can be obtained by taking the presence of a particular motif or repeat number; however, this will lead to incongruent result.

Conclusion:

The study has identified the pattern of distribution of microsatellites in the genic and intergenic regions of 19 mitogenomes of phytopathogenic fungi and established

relationship between individual genomes. Most of the repeat motifs were confined to the intergenic region. Hence the occurrence of SSR is non-random in mitogenomes. Unlike nuclear genomes, trinucleotide repeat motifs were found to be the most abundant class of repeats in the intergenic regions. Genetic relationship can be obtained on the basis of SSRs present in mitochondrial protein coding genes. To the best of our knowledge, this is the first extensive study of SSR repeat dynamics in the mitogenomes of phytopathogenic fungi.

References:

- [1] Toth G *et al.* *Genome Res.* 2000 **10**: 967 [PMID: 10899146]
- [2] Datta S *et al.* *Physiol Mol Biol Plants.* 2010 **16**: 123
- [3] Mahfooz S *et al.* *FEMS Microbiol Lett.* 2012 **328**: 54 [PMID: 22171944]
- [4] Kim TS *et al.* *BMC Genomics.* 2008 **9**: 31 [PMID: 18215294]
- [5] Lawson MJ *et al.* *Genome Biol.* 2006 **7**: R14 [PMID: 16507170]
- [6] Campbell A *et al.* *Proc Natl Acad Sci USA.* 1999 **96**: 9184 [PMID: 10430917]
- [7] Ballard JW *et al.* *Mol Ecol.* 2004 **13**: 729 [PMID: 15012752]
- [8] Gray MW *et al.* *Genome Biol.* 2001 **2**: 1018 [PMID: 11423013]
- [9] WS Martins *et al.* *Bioinformatics.* 2009 **3**: 282

- [10] Botstein D *et al.* *Am J Hum Genet.* 1980 **32**: 314 [PMID: 6247908]
- [11] Karaoglu H *et al.* *Mol Biol Evol.* 2005 **22**: 639 [PMID: 15563717]
- [12] Viguera E *et al.* *EMBO J.* 2001 **20**: 2587 [PMID: 11350948]
- [13] Harr B *et al.* *Mol Cell.* 2002 **10**: 199 [PMID: 12150919]
- [14] Rajendrakumar P *et al.* *Bioinformatics.* 2007 **23**: 1 [PMID: 17077096]
- [15] Tian X *et al.* *Mol Biol Evol.* 2011 **28**: 899 [PMID: 20943830]
- [16] Bachtrog D *et al.* *Mol Biol Evol.* 1999 **16**: 602 [PMID: 10335653]
- [17] Lim S *et al.* *Genet Biol.* 2004 **41**: 1025 [PMID: 15465391]
- [18] Pacheco MA *et al.* *Mol Biol Evol.* 2011 **28**: 1927 [PMID: 21242529]
- [19] Zardoya R & Meyer A, *Mol Biol Evol.* 1996 **13**: 933 [PMID: 8752002]

Edited by P Kanguane

Citation: Mahfooz *et al.*, *Bioinformation* 8(23): 1171-1175 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Evaluation of polymorphism

Formula used to calculate PIC is as follows:

$$PIC_i = 1 - \sum_{j=1}^n P_{ij}^2$$

Where P_{ij} represents the frequency of the j th allele for the marker i , and the summation extends over n alleles.

Table 1: Taxonomic status of phytopathogenic fungi used in the study and the gene bank accession numbers of their complete mitochondrial genome

S. No	Fungi	Phylum	Class	Order	Family	GeneBank acc. no.
1.	<i>Aspergillus niger</i>	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	NC_007445
2.	<i>Aspergillus oryzae</i>	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	NC_008282
3.	<i>Aspergillus tubigenis</i>	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	NC_007597
4.	<i>Fusarium graminearum</i>	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	NC_009493
5.	<i>Fusarium oxysporum</i>	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	AY945289
6.	<i>Fusarium verticillioides</i>	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	NC_016687
7.	<i>Verticillium dahliae</i>	Ascomycota	Sadariomycetes	Hypocreales	Incertae sedis	NC_008248
8.	<i>Mycosphaerella graminicola</i>	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	NC_010222
9.	<i>Magnaporthea grisea</i>	Ascomycota	Sadariomycetes	Magnaporthales	Magnaporthaceae	Broad Inst.
10.	<i>Phaeosphaeria nodorum</i>	Ascomycota	Dothideomycetes	Pleosporales	Phaeosphaeriaceae	NC_009746
11.	<i>Phakopsora meibomia</i>	Basidiomycota	Urediniomycetes	Uredinales	Phakopsoraceae	NC_014352
12.	<i>Phakopsora pachyrhizi</i>	Basidiomycota	Urediniomycetes	Uredinales	Phakopsoraceae	NC_014344
13.	<i>Moniliophthora perniciosa</i>	Basidiomycota	Basidiomycetes	Agaricales	Marasmiaceae	NC_005927
14.	<i>Tilletia indica</i>	Basidiomycota	Ustilaginomycetes	Ustilaginales	Tilletiaceae	NC_003053
15.	<i>Tilletia walkeri</i>	Basidiomycota	Exobasidiomycetes	Tilletiales	Tilletiaceae	NC_010651
16.	<i>Trametes cingulata</i>	Basidiomycota	Agaricomycetes	Polyporales	Polyporaceae	NC_013933
17.	<i>Ustilago maydis</i>	Basidiomycota	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	NC_008368
18.	<i>Rhizophydium sp.</i>	Chytridiomycota	Chytriomycetes	Rhizophydiales	Rhizophydiaceae	NC_009880
19.	<i>Rhizopus oryzae</i>	Zygomycota	Zygomycetes	Mucorales	Mucoraceae	NC_006836

Table 2: Details of genome size, %GC, no. of genes, % coding, no. and relative abundance of SSRs in mitochondrial genomes

S. no.	Fungi	Genome size (kb)	%GC	No. of genes	% coding regions	No. of SSRs	Intergenic	Genic	Relative abundance
1.	<i>A. niger</i>	31.1	26.9	43	47	28	23 (82.0%)	5 (18%)	0.90
2.	<i>A. oryzae</i>	29.2	26.2	25	-	19	19 (100%)	-	0.65
3.	<i>A. tubigenis</i>	33.6	26.8	43	43	33	28 (84.8%)	5 (15.2%)	0.98
4.	<i>F. graminearum</i>	95.7	31.8	46	55	26	21 (80.8)	5 (19.2%)	0.27
5.	<i>F. oxysporum</i>	34.4	31.0	39	58	9	6 (66.7%)	3 (33.3%)	0.26
6.	<i>F. verticillioides</i>	53.7	32.6	50	-	26	21(80.8%)	5 (19.2%)	0.48
7.	<i>V. dahliae</i>	27.1	27.3	42	53	7	1(14.3%)	6 (85.7%)	0.26
8.	<i>M. graminicola</i>	43.9	31.9	43	52	23	21(91.3%)	2 (8.7%)	0.52
9.	<i>M. grisea</i>	34.9	28.6	-	-	26	23(88.5)	3 (11.5%)	0.74
10.	<i>P. nodorum</i>	49.8	29.4	44	47	8	8 (100%)	-	0.16
11.	<i>P. meibomia</i>	32.6	34.9	41	39	15	12(80.0%)	3 (20%)	0.46
12.	<i>P. pachyrhizi</i>	31.8	34.6	41	40	10	6 (60.0%)	4 (40.0%)	0.31
13.	<i>M. perniciosa</i>	109.1	31.9	117	50	77	60 (77.9%)	17 22.1%)	0.71
14.	<i>T. indica</i>	65.1	28.9	40	20	14	13(92.9%)	1 (7.1%)	0.22
15.	<i>T. walkeri</i>	59.4	28.8	40	22	12	11(91.7%)	1 (8.3%)	0.20
16.	<i>T. cingulata</i>	91.5	24.5	47	19	34	29 (85.3)	5 (14.7%)	0.37
17.	<i>U. maydis</i>	56.8	31.2	40	38	16	11(68.8%)	5 (31.2%)	0.28
18.	<i>Rhizophydium sp</i>	68.8	23.0	46	45	15	10(66.7%)	5 (33.3)	0.22
19.	<i>R. oryzae</i>	54.2	26.2	51	38	15	14(93.3%)	1 (6.7%)	0.28
	Total	1002.7				413	337	76	8.27
	Mean	52.7				21.7	16.6	5.14	0.43