

Analysis of simple sequence repeats (SSRs) dynamics in fungus *Fusarium graminearum*

Rajender Singh*, Sonia Sheoran, Pradeep Sharma, Ravish Chatrath

Directorate of Wheat Research, Post Box 158, Karnal - 132001, India; Rajender Singh - Email: rajenderkhokhar@yahoo.com; *Corresponding author

Received July 26, 2010; Accepted January 06, 2011; Published February 15, 2011

Abstract:

The abundance and inherent potential for variations in simple sequence repeats (SSRs) or microsatellites resulted in valuable source for genetic markers in eukaryotes. We describe the organization and abundance of SSRs in fungus *Fusarium graminearum* (causative agent for Fusarium head blight or head scab of wheat). We identified 1705 SSRs of various nucleotide repeat motifs in the sequence database of *F. graminearum*. It is observed that mononucleotide repeats (62%) were most abundant followed by di- (20%) and trinucleotide repeats (14%). It is noted that tetra-, penta- and hexanucleotide repeats accounted for only 4% of SSRs. The estimated frequency of Class I SSRs (perfect repeats ≥ 20 nucleotides) was one SSR per 124.5 kb, whereas the frequency of Class II (perfect repeats >10 nucleotides and <20 nucleotides) was one SSR per 25.6 kb. The dynamics of SSRs will be a powerful tool for taxonomic, phylogenetic, genome mapping and population genetic studies as SSR based markers show high levels of allelic variation, codominant inheritance and ease of analysis.

Keywords: SSR, microsatellites, *Fusarium graminearum*, EST

Background:

Fusarium head blight or head scab of wheat is a global problem in the humid and subhumid wheat growing areas of world [1]. *Fusarium* head blight has been associated with upto 17 organisms, of which *Fusarium graminearum* is the principle pathogen responsible for head blight in many countries including India [2]. In addition to causing yield losses, strains of *F. graminearum* are known to produce trichothecene mycotoxins which pose a serious threat to human and animal health and food safety [3].

Simple sequence repeats (SSRs) or microsatellites have been proven to be the markers of choice during the last decade in plant research because of their hypervariability and ease of detection. SSR markers have been developed for many species of plants, animals and fungi from genomic DNA through the construction of SSR enriched libraries. This approach is labor intensive and time consuming. However in recent years, with the establishment of several sequencing projects in crop plants, animals and microorganisms resulted in a wealth of DNA sequence information. This sequence data for expressed sequence tags (ESTs), genes and cDNA clones can be downloaded from various databases in public domain and by using computer programs these can be scanned for identification of SSRs, referred as EST-SSRs or genic microsatellites. Microsatellite sequences obtained through *in silico* mining have more or less the same utility and potential comparative with those derived from a genomic library. However, the negligible cost of *in silico* mining and high abundance of microsatellites in different sequence resources make this approach extremely attractive for the generation of microsatellite markers. SSR provides a powerful tool for taxonomic, phylogenetic and population genetic studies because of its highly polymorphic nature. The polymorphism in SSRs is generally believed to be the result of DNA polymerase slippage and unequal recombination [4]. The information on abundance and distribution of SSRs may also help in understanding their relevance in gene function or genome evolution. The main objective of this study was to analyze the abundance and distribution of different classes of SSRs in the EST database *Fusarium graminearum*, which may help in understanding the evolution and diversity analysis.

Methodology:

Dataset:

The genomic sequences (433 contigs of 36.22 Mb) of *Fusarium graminearum* available in Fusarium comparative database of Broad Institute of MIT and Harvard, Cambridge (http://www.broadinstitute.org/annotation/genome/fusarium_graminearum/) were used for the study.

SSR analysis:

Perfect mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs with a repeat of ≥ 6 times were identified using the software WebSat (SSR finder program)[5]. The sequences from each contigs were downloaded from Fusarium comparative database and entered in the WebSat software. As the program can process 150,000 characters, the longer sequences were divided into two or more parts and then processed for SSR analysis. The output generated by the program highlight the SSR sequences in yellow color.

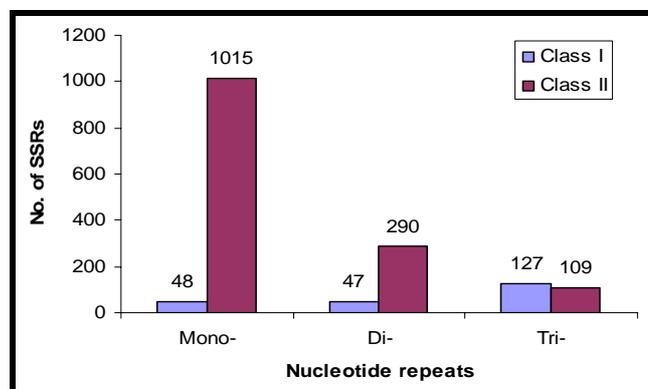


Figure 1: Frequency of Class I and Class II SSRs of mono-, di- and trinucleotide repeats.

Discussion:

Genomic sequence data of 36.22 Mb size of *F. graminearum* assembled into 433 contigs and further assembled into 31 scaffolds or supercontigs from Broad Institute (<http://www.broadinstitute.org/>) was used to search for mono-, di-, tri-, tetra-, penta- and hexanucleotide motifs with a repeat of ≥ 6 times. A total number of 1705 SSRs were identified from the EST database *Fusarium graminearum* (Table 1 see supplementary material). Chromosome 1 possessed highest number of SSRs (611) and chromosome 3 had the least number of SSRs (318). Ten SSRs were identified in the contigs not mapped to any chromosome. Mononucleotide repeats were the most abundant (1063) repeats in all the chromosomes accounting 62% of SSRs. Next to mononucleotides, dinucleotides (20%) were predominant followed by trinucleotides (14%). Tetra-, penta- and hexanucleotide repeats were the least frequent repeats accounting 4% of SSRs. The density of SSRs was found to be one SSR per 21.2 kb.

Among the mononucleotides, polyA and polyT were more abundant repeats with a frequency of 492 and 439 (Table 2 see supplementary material), respectively. PolyG and polyC repeats were rare representing 5.9% and 6.5% of mononucleotide repeats. The number of repeat units ranged from 10 to 41 among mononucleotides, but majority of repeats had 10-12 repeat units. Twelve types of dinucleotide repeat motifs (Table 3 see supplementary material) were found in the genome. The AT/TA dinucleotide repeat motif was the most predominant while the CG/GC repeat motif was rare. Among trinucleotide repeats, 53 different types of repeat motifs were identified and the CTT repeat motif was predominant in *Fusarium graminearum* genome. Tetra-, penta- and hexanucleotide repeats were least frequent repeats in the genome, tetranucleotide repeats occur more in number (34) followed by penta- (21) and hexanucleotide repeats (14). The genome possessed 29 different types of tetra-, 20 types of penta- and 14 types of hexanucleotide repeats. The number of repeat units in di-, tri-, tetra-, penta- and hexanucleotides ranged from 6 to 46, but the majority of SSRs (70%) had six to seven repeat units. Some of the highly repeated sequences identified were (AG)₂₈, (AAG)₃₁, (GAA)₄₆, (GTATG)₁₈, (GAAGAG)₂₁, (TGAAGA)₂₂ and (CCCTAA)₂₃.

SSRs were categorized into two groups based on length of SSR tracts and their potential as informative genetic markers: Class I SSRs contain perfect repeats ≥ 20 nucleotides in length and Class II contain perfect repeats >10 nucleotides and <20 nucleotides in length. Out of 1705 SSRs, 291 repeats were categorized as Class I SSRs. 54% of trinucleotide repeats were Class I SSRs, followed by dinucleotide (14%) and mononucleotide repeats (4.5%) (Figure 1). All tetra-, penta- and hexanucleotide repeats were Class I SSRs. The estimated frequency of Class I SSRs was one SSR per 124.5 kb, whereas the frequency of Class II was one SSR per 25.6 kb.

Conclusion:

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.

The *in silico* mining of EST database of *F. graminearum* provided a rich source of SSRs which can be used for taxonomic, evolutionary and population genetic studies. Role of microsatellites in regulation of gene expression and in the evolution of gene regulation [6, 7] are well documented. The implications of excess numbers of short iterated repeats could be extremely important not only for genomic stability, but also for the evolution of additional genomic features such as codon usage [8].

In general, microsatellites show a decrease in abundance with increasing repeat length [9]. However, more than the expected number of long microsatellite repeats were also reported [10]. The rationale for Class I and Class II categories of SSRs is that longer perfect repeats (Class I) are highly polymorphic as evidenced by the experimental data originally reported from human [11] and then confirmed by studies in many other organisms, including rice [12]. Microsatellites in Class II tended to be less variable, representing sites where SSR expansion may occasionally occur but its probability is limited due to a smaller chance of slipped-strand impairing over the shorter SSR template [11, 13]. The microsatellites identified in this study could be used for the development of genome specific markers for evolutionary studies in *F. graminearum*.

Acknowledgement:

The financial support for Agri Bioinformatics Promotion Program provided by Bioinformatics Initiative Division, Department of Information Technology, Ministry of Communications & Information Technology, Government of India, New Delhi is gratefully acknowledged. The authors are thankful to Project Director, DWR, Karnal for providing facilities for this work.

References:

- [1] J Gilbert & A Tekauz, *Can J Plant Pathol.* (2000) **22**:1
- [2] MS Saharan *et al. Proc Indian Natl Sci Acad.* (2004) **B70**:255
- [3] A Bottalico & G Perrone, *Eur J Plant Pathol.* (2002) **108**:611
- [4] YC Li *et al. Mol Ecol.* (2002) **11**:2453 [PMID: 12453231]
- [5] WS Martins *et al. Bioinformatics* (2009) **3**:282 [PMID: 19255650]
- [6] YC Li *et al. Mol Biol Evol.* (2004) **21**:991
- [7] TS Huang *et al. Biochem Biophys Res Commun.* (2003) **300**: 901 [PMID: 12559958]
- [8] D Field & C Wills, *Proc Nat Acad Sci.USA* (1998) **95**:1947
- [9] A Grover *et al. Mol Genet Genomics.* (2007) **277**: 469 [PMID: 17237941]
- [10] D Dieringer & C Schlotterer, *Genome Res.* (2003) **13**: 2242 [PMID: 14525926]
- [11] JL Weber, *Genomics* (1990) **7**:524 [PMID: 1974878]
- [12] YG Cho *et al. Theor Appl Genet.* (2000) **95**:713
- [13] S Temnykh *et al. Genome Res.* (2001) **11**:1441 [PMID: 11483586]

Edited by P Kanguane

Citation: Singh *et al.* Bioinformatics 5(10): 402-404 (2011)

Supplementary material:

Table 1: Distribution of EST-SSRs in different chromosomes identified from EST database of *Fusarium graminearum*

Chromosome	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
Chromosome 1	366	133	87	18	6	1	611
Chromosome 2	239	68	49	3	4	5	368
Chromosome 3	201	56	50	4	3	5	319
Chromosome 4	248	79	50	9	8	3	397
Unmapped	9	1	0	0	0	0	10
Total	1063	337	236	34	21	14	1705

Table 2: Frequency of mononucleotide repeats in different chromosomes of *Fusarium graminearum* genome

Chromosome	A	T	G	C	Total
Chromosome 1	163	155	27	21	366
Chromosome 2	111	102	9	17	239
Chromosome 3	95	75	14	17	201
Chromosome 4	116	105	13	14	248
Unmapped	7	2	0	0	9
Total	492	439	63	69	1063

Table 3: Different types of di-, tri-, tetra-, penta- and hexa-nucleotide repeat motifs

Repeats	Motifs
Di-nucleotide	AC, AG, AT, CA,CG,CT, GA, GC, GT, TA, TC, TG
Tri-nucleotide	AAC, AAG, AAT, ACA, ACC, AGA, AGC, AGG, ATA, ATC, ATG, ATT, CAA, CAC, CAG, CAT, CCA, CCT, CGA, CGT, CGG, CTA, CTC, CTG, CTT, GAA, GAC, GAG, GAT, GCA, GCC, GCT, GGA, GGC, GGT, GTA, GTC, GTG, GTT, TAC, TAG, TAT, TCA, TCC, TCG, TCT, TGA, TGC, TGT, TGG, TTA, TTC, TTG
Tetra-nucleotide	AATC, ACTC, AGGT, AGTG, ATAC, ATAG, CAAG, CACT, CAGG, CATA, CATC, CATT, CTAT, CTGG, CTTT, GATG, GGAT, GTGA, TAGA, TAGG, TATC, TAGT, TCAA, TCCA, TCTT, TGAC, TTCC, TTCT, TTGC
Penta-nucleotide	ACACG, AGACC, CTTTT, GATGC, GCAAG, GTAAA, GTCTC, TGGAC, TGGTA, TTGCG, AGCAC, AGTCT, ATATC, CTCTG, GCATC, GTATG, GTTTG, TATGC, TCTGG, TGCTC
Hexa-nucleotide	AGAGCC, CTTCGA, GATAAG, TCAGGA, TCTGGC, TGAAGA, AAGCCC, ACTTAG, CAAAGC, CCCTAA, GAAGAG, GCAGAG, GTTGCC, TCATCC