Journal of Scientific Research
Banaras Hindu University, Varanasi

Vol. 58, 2014 : 51-58
ISSN : 0447-9483

COMPUTATIONAL ANALYSIS OF SINGLE-NUCLEOTIDE POLYMORPHISMS AND INSERTIONS/DELETIONS IN CODING REGIONS OF THREE *FORMAESPECIALES* OF *FUSARIUMOXYSPORUM*

Shalini Rai[1], Deepak Kumar Maurya[2]
[1]National Bureau of Agriculturally Important Microorganisms (ICAR),
Kusmaur, Mau NathBhanjan (UP), India
[2]Centre of Biotechnology, University of Allahabad (UP), India
Corresponding Author : Email : rai.vns82@gmail.com,shalinimicro09@gmail.com

Abstract

Recently a public expressed sequenced tag database represents an enormous repository of sequences which makes them an attractive resource for mining of various purposes. In the present investigation, we surveyed single nucleotide polymorphism in the publically available expressed sequenced tag (EST) and transcripts sequences of three *formaespeciales* of *F. oxysporum* viz. *lycopersici* (*Fol*), *cucumerium* (*Foc*) and *melonis* (*Fom*). However, little is known about its genetics and genomic variation, including SNP variation. Single nucleotide polymorphisms (SNPs) that were polymorphic between the three *formaespeciles* of *Fusarium* were mined and 43 was identified in *Fol* followed by *Foc* (21) and *Fom* (20) and indel polymorphism, *Fol* transcripts was show 19201, *Foc*-EST shows 8741 and 6746 indel polymorphism fond in *Fom*-ESTs analysis, respectively. The average frequency of SNPs amounted to *Fom* showing highest frequency of 1.63×10^{-4} SNPs per 100 bp followed by *Foc* and *Fol* with frequencies 5.05×10^{-5} and 1.37×10^{-5} SNP/100 bp, respectively. It was found that transition of SNPs for G/C in *Fol* (1820) were higher than *Fom* (1355) and*Foc* (1251) but the transversion of SNPs for C/T was higher in *Fom* (1605) in comparison to *Foc* (1585) and *Fol* (1232). The implications of these SNPs for our understanding of the genetics, population history, ecology and evolution of this important plant pathogen species are discussed.

Keywords: Plant-Pathogens, Single nucleotide polymorphisms, Transcript Sequence

Introduction

Fusariumoxysporum is an economically important soil-borne fungal plant pathogen which causes vascular wilt disease in many agricultural crops all over world. Individual pathogenic strain within the species have a limited host range and strains with similar or identical host range are assigned to intraspecific groups, called *forma specialis* (Namiki *et al*., 1994). Some of the *formaspecialis* are further divided into subgroups, named races, on the basis of virulence to a set of different cultivars within the same plant species Armstrong *et al.,* 1981. *Fusariumoxysporum*is a plant pathogens cause severe wilts in ~80 botanical species such as tomato, watermelon, cucumber, pepper, muskmelon, beans, wheat, maize and cotton(Beckman, 1987). Wilting in tomato was caused by *Fusariumoxysporum*f. sp. *lycopersici*(*Fol*) that's like cucumber and melon was also affected by *Fusariumoxysporum*f. sp. *cucumerinum* (*Foc*) and

*Fusariumoxysporum*f. sp.*melonis* (*Fom*), respectively.To understand the evolutionary history and genomic constituents of the *formaespeciales* within *Fusariumoxysporum* requires knowledge of the phylogenetic relationships among isolates (Appel& Gordon, 1996). Molecular markers have proven powerful tools in the assessment of genetic relationships such as Restriction Fragment Length Polymorphism (RFLP), Rapid Amplified Polymorphic DNA (RAPD), Inter Simple Sequence Repeats (ISSR), Simple Sequence Repeats (SSR), Amplifies Fragment Length Polymorphisms (AFLP) and Single nucleotide polymorphism(SNP) are presently available to assess the variability and diversity.

Single nucleotide polymorphisms (SNPs) are the most common class and the smallest unit of genetic variation present in genomes (Cho *et. al.*, 1999). Single nucleotide polymorphism(SNP) markershave gained much interest, becoming the marker of choice in genetic analysis in the scientific and breeding community(Gupta *et. al*., 2001). SNP represent the most frequent type of genetic polymorphism and thus provide a high density of markers near the locus of interest. According to Syvanen (Syvanen, 2001), SNPs are highly stable, reliable and have a fine resolution. SNP marker was used as genetic markers in microbes (Adams *et al*., 2005), crop plants (Kota et al., 2008) and humans-SNP (Syvanen, 2001) that can be mined from sequence data and are useful for characterizing allelic variation, genome wide mapping and tool for marker-assisted selection (Batley *et al*., 2003, González-Martínez *et al*., 2006).The development of high throughput methods for the detection of SNPs and small indels (insertion/deletion) has led to a revolution in their use as molecular markers (Novaes *et al*., 2008).

In the present study, we investigated *in-silico* SNP mining from transcripts and EST database of the *Fusariumoxysporum*.ESTs are short DNA sequences corresponding to a fragment of a complimentary DNA (cDNA) molecule and which may be expressed in a cell at a particular given time.ESTs are currently used as a fast and efficient method of profiling genes expressed in various tissues, cell types or developmental stages and alsouseful for the discovery of novel genes, investigation of genes of unknown function, recognition of exon/intron boundaries (Adams *et. al*., 1991).Although thousands of ESTs derived SNPs identification has been done in human, animals and plants for genome analysis, their use in fungus is still in its infancy. In fact, there are only a limited number of studies on these seemingly important and introducing sets of sequences in fungal species. To accomplish this, an *in-silico*approach has been used to assess the frequency and distribution of SNPs in ESTs and transcripts sequence within three *formaespeciales* and indel variation, their types and nucleotide diversity.

Materials and methods

Sequences retrieve and Contig generation

The available ESTs of *Fom* and *Foc* were downloaded from National Center for Biotechnology Information (www.ncbi.nlm.nih.gov), whereas, annotated transcript

sequences of *Fol*were downloaded from "Fusarium Comparative Sequencing Project" (www.broadinstitute.org).EST sequenced were assembled and aligned into contigs by using CAP3 program (seq.cs.iastate.edu/). For SNP mining CAP3 uses individual sequence overlap for constructing clusters.

Indel identificationand Transition vsTransversion

Contigs were scanned to find out the candidate SNPs from these libraries with the help of SNP detection tool AutoSNP (http://hornbill.cspp.latrobe.edu.au/ snpdiscovery.html). Auto SNP takes fasta format sequences as input and provides information into two associated measurements of confidence in the validity of SNPs for each polymorphism. After the complete run Auto SNP programme, two windows were generated, first provide complete vertical alignment, highlighting SNPs and the second, lists the assembly member sequences and provide a SNP summary. Perl scripts were used for calculating number of base pairs and identifying indels (insertion/deletion) across all the three *formaespeciales*. DNA substitution ratio i.e. Transition vsTransversionratio was identified using MEGA 4.0(http://www.megasoftware.net).

Results

SNP discovery and frequency

This work was carried out for the mining SNPin the fungus species but in plant species it's quite abundant. Delmotte (Delmotte *et al*., 2011) use of SNPs Markers for characterization of *Plasmoparaviticola*, which is belongs to oomycetes and causal agent of Downy Mildew in Grapevine. This analysis leads to identification of SNPs with variation among three different *formaespeciales* of *Fusariumoxysporum*. Use of AutoSNP, we got maximum number of SNPs (43) was identified in *Fol* followed by *Foc* (21) and *Fom* (20). The higher number of SNPs in *Fol* was expected because the total size covered by transcripts sequences of *Fol* (21.7Mb) was much higher than that of ESTs of *Fom* (1.3Mb) and *Foc* (2.4Mb).To compare the SNPs count between all three *formaespeciales,* the complete length of each set of sequences were analyzed, thus, transition and transversions were calculated and depicted in Table 1.It was found that transition of SNPs for G/C in *Fol* (1820) were higher than *Fom* (1355) and*Foc* (1251) but the transversions of SNPs for C/T was higher in *Fom* (1605) in comparison to *Foc* (1585) and *Fol* (1232).Insertion and deletion shortly called indel polymorphism, *Fol* transcripts was show 19201, *Foc*-EST shows 8741 and 6746 indel polymorphism fond in *Fom*-ESTs analysis respectively. *Fom* showing highest frequency of 1.63×10^{-4} SNPs per 100 bp followed by *Foc* and *Fol* with frequencies 5.05×10^{-5} and 1.37×10^{-5} SNP/100 bp, respectively. Previous studies reported that maize (Ching *et al*., 2002) 1 coding SNP per 124 bp in 18 maize genes assayed in 36 inbreed lines. Coryell (Coryell et.al., 1999) studied SNPs in soyabean and observed that only 2 SNP found

approximately after 400 bp.Microbes have low Transcript/EST sequence so frequency obtained is very less in comparison to previously data are available in several crop plants such as beet (Schneider *et al.*, 2007), citrus (Dong *et al*., 2010), eucalyptus (Singh *et. al*., 2011) and soyabean (Rafalski, 2002) which is 0.77, 1.36, 0.61, 9.1 and 0.16, respectively.

For further analysis of identified SNP, candidate SNPs was categorized according to nucleotide substitution as either transition (C↔T or G↔A) or transversion (C↔G, A↔T, C↔A or T↔G). Comparative results of SNP and indeldiscovery in three *formae*species of *Fusarium*were listed in table 1 & Figure 1.*Fol* shows a total of 4840 bi-allelic SNPs, including 3052 transition and 1788 transversion with bias towards G↔A (1820), against C↔G (513) in table1. In*Foc* and *Fom* shows 4816 and 4478 bi-allelic SNPs including slightly higher transitions 2856 and 2940 than transversions 1960 and 1538, respectively. However, considering the individual substitutions bias, the transition type substitutions C↔T (1650) in Foc and (1585) in Fom were found to be slightly higher than transversion type substitutions (Table1). In addition, earlier studies revealed, transitions occurs at a higher frequencies than transversions such as citrus (Dong *et al*., 2010) maize (Batley *et al*., 2003) and oil palm (Riju *et al*., 2007). A high frequency transition of cytosine residue to thymine residue mutation is observed due to methylation. Moreover the recent results were observed by *Eucalyptus*transcriptome in which tranversions (14026) are found slightly higher than transitions (13666) (Singh *et al*., 1999).

In present study it was observed that indel occurred at a very low frequency (0.06indel/1000 bp) in *Fol* and continuously increased in (0.21 indel/1000 bp) *Foc* and (0.53indel/1000 bp) *Fom*. Among these three *formaespeciales*, in*Fol* guanine indel (5130) were found to be more abundant followed by cytocine (4852) while *Foc* showed cytocineindel (2363) more abundant followed by adinine (2158) and the other possible indels occurring in same fashion. For *Fom*, it was observed that cytosine involved indels (1934)are more abundant followed by guanine (1751) and so on in table1 & figure 1.This indel polymorphism occurs during DNA synthesis, repair recombination or insertion and excision of transposable elements indels may introduce that often leave a characteristic DNA foot-print of several nucleotide base.

Ratio of transition to transversion (Ts/Tv) was occurring in the range of 1.61-2.67 (Table1). Maximum Ts/Tv ratio was observed to be 2.67 for *Fom* whereas minimum Ts/Tv ratio was observed as 1.61 in *Foc* andTs/Tv ratio for *Fol* was observed to be 1.78. Recent reports of SNP analysis revealed Ts/Tv ratio < 1 was seen in regulatory genes such as endonucleases reverse transcriptase and Tcl- like transposase. Ratio of Ts/Tv was very useful to compare the genotypes of hepatitis virus C and also differences among the mitrochondrial genome of animals. The high genetic variation in

transition to transversion or indel, it was suggests that the fungus has a great capacity for adaptability and genetic change during its interaction with even this single host species. Thus SNPs can be explained as any polymorphism between two genomes that is based on single nucleotide exchange and SNPs can be helpful for variation in diversity and a comprehensive analysis among species.

Discussion

This research work has been carried out to detect several putative SNPs which can be applied not only for making genetic maps but also for exploring the markers in EST data of three *formaespeciales* of *Fusariumoxysporum.* SNPs including insertion/deletion size frequencies and transition to transversion ratio analysis can be applied in future research to facilitate gene identification. This *in-silico* analysis apparently shows that this reliable resource will be definitely contribute for functional genomics, agriculture science, and crop improvement and disease management studies. Discovered SNPs can also be used for determining pathogenic/virulence genes in *Fusariumoxysporum.* Recognition of these high-diversity areas of the transcripts and EST was focuses the direction of future work toward those regions that may have the greatest potential in elucidating the dynamics of host pathogen interactions.

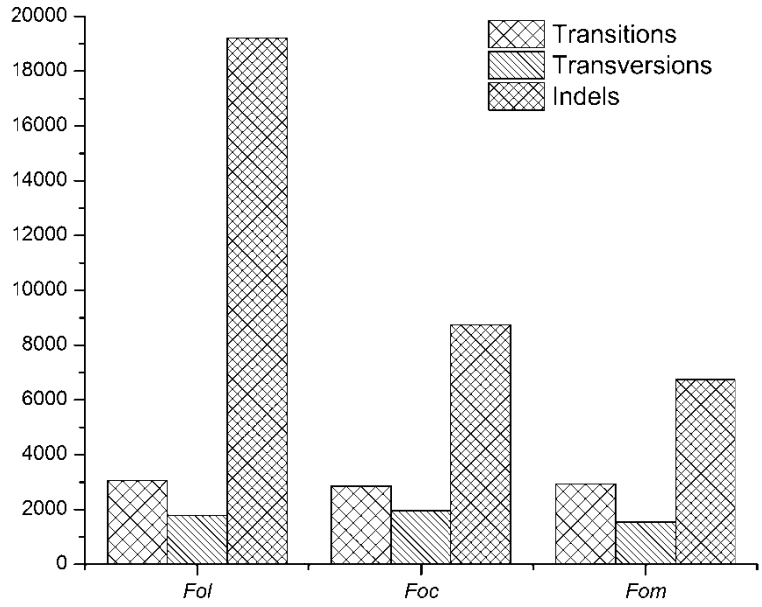

Fig. 1: Graph shows a comparative study of transitions, transversionsand indels in different *formaespeciales*of *Fusarium oxysporum*

Table 1: Summary of SNPs and indels detected in three *formaespeciales* of *fusariumoxysporum*

Results	***Fusariumoxysporum* Transcript/EST libraries**		
	lycopersici	***cucumerium***	***melonis***
Total No. of ESTs	17657	6448	3493
Total sequences analyzed	21684767	2461670	1562868
No. of contigs	873	730	319
Total SNPs detected	43	21	20
Total consensus size (bp)	313555556	41570256	12197556
Frequency of SNP per 100bp	1.37×10^{-5}	5.05×10^{-5}	1.63×10^{-4}
Transition			
G/A	1820	1251	1355
C/T	1232	1605	1585
Total	3052	2856	2940
Tranversions			
A/C	427	558	489
T/A	474	485	301
G/T	374	446	352
C/G	513	471	396
Total	1788	1960	1538
Ts/Tv ratio	1.787	1.613	2.676
Indel			
A	5036	2158	1418
G	5130	2103	1751
T	4183	2117	1643
C	4852	2363	1934
Total	19201	8741	6746
Indel frequency per 1000 bp	0.06	0.21	0.55

Acknowledgement

The first author gratefully acknowledge for the financial assistance from Indian Council of Agricultural Research (ICAR), India.

References

1. Namiki,F., Shiomi, T., Kayamura,T., Tsuge, T. (1994). Characterization of the *formaespeciales* of *Fusariumoxysporum* causing wilts of cucurbits by DNA fingerprinting with Nuclear Repetitive DNA Sequences. *Application &Environmental Microbiology***;**60, 2684–2691.
2. Armstrong, G.M., Armstrong, J.K. (1981). *Formaespeciales* and races of *Fusariumoxy sporum* causing wilt disease. In *Fusarium* : Disease, Biology, and Taxonomy, P.E. Nelson, T.A. Toussoun, and R.J. Cook, eds (University Park, PA: Pennsylvania State University Press); 391–399.
3. Beckman,C.H. The Nature of Wilt Diseases of Plants. (St. Paul, MN: *Ameri. Phytopathol. Soci.* Press). 1987.
4. Appel, D.,GordonT.R.(1996).Relationships among pathogenic and nonpathogenic isolates of *Fusariumoxysporum* based on the partial sequence of the intergenic spacer region of the ribosomal DNA. *Mol. Plant–Microbe Inter.*; 9, 125–38.
5. Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A.,Yang, W.H, Hubbell, E., Au, M., Chung, E.Y., Lashkari, D, Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W., Oefner, P.J.(1999). Genome-wide mapping with biallelic markers in Arabidopsis thaliana. *Nat Genet*.; 23, 203–207.
6. Gupta,P.K.,Roy, J.K.,Prasad,M. (2001). Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science*; 80, 524–535.
7. Syvanen, A.C. (2001). Accessing genetic variation Genotyping single nucleotide polymorphisms. *Nat Rev Genet*.; 2, 930–942.
8. Adams,R.I.,Hallen,H.E.,Pringle,A. (2005). Using the incomplete genome of the ectomycorrhizal fungus *Amanita bisporigera* to identify molecular polymorphisms in the related *Amanita phalloides*. *Mol Ecol Notes*.; 6, 218–220.
9. Kota,R.,Varshney, R.K., Prasad, M.,Zhang, H.,Stein, N.,Graner,A. (2008). EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct & IntegGenom*.;8, 223-233.
10. Batley,J.,Barker,G.,O'Sullivan, H.,Edwards,K.J.,Edwards,D. (2003). Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol*;132,84–91.
11. González-Martínez,S.C.,Ersoz,E.,Brown,G.R.,Wheeler,N.C.,Neale,D.B.(2006).DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinustaeda* L. *Genetics*;172 (31) 915-26.
12. Novaes, E., Drost, D.R., Farmerie, W.G., Pappas, G.J., Grattapaglia, D., Sederoff, R.R., Kirst, M.(2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMCGenomics.;9, 312–325.
13. Adams,M.D.,Kelly,J.M.,Gocayne,J.D.,Dubnick,M., Polymeropoulus,M.H.,Xiao,H.(1991).Links complementary DNA sequencing: expressed sequencetags and human genome projects. *Science*; 252, 1651-1656.

14. Delmotte, F., Machefer,V.,Giresse,X., Richard-Cervera,S.,Latorse, M.P., Beffa,R.(2011). Characterization of Single-Nucleotide-Polymorphism Markers for *Plasmoparaviticola*, the Causal Agent of Grapevine Downy Mildew. *Appl& Environ Microbiol*.;7861–7863.

15. Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smithm, O.S., Tingey, S., Morgante, M., Rafalski, A.J.(2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet.;3,19.

16. Coryell, V.H., Jessen, H., Schupp, J.M., Webb, D., Keim, P. (1999). Allele-specific hybridization markers for soybean. *Theor. Appl. Genet*.;10, 11291–1298.

17. Schneider, K., Kulosa, D.,Soerensen,T. R., Möhring, S., Heine, M., Durstewitz, G., Polley, A., Weber, E., Jamsari, L. J.,Hohmann, U.,Tahiro, E., Weisshaar, B., Schulz,B., Koch, G.,Jung, C., Ganal,M. (2007). Analysis of DNA polymorphisms in sugar beet (*Beta vulgaris* L.) and development of an SNP-based map of expressed genes. *Theor. Appl. Genet*.;115, 601-615.

18. Dong,J.,Qing-liangmY,E.,Fu-Sheng,W.Li,C. (2010). The Mining of citrus EST-SNP and its application in cultivar discrimination.*Agric. Sci.*China;9,179–190.

19. Singh,T.R.,GuptamA., RijumA., MahalaxmimM., Seal,A., Arunachalam, V. (2011). Computational identification and analysis of single nucleotide polymorphisms and insertions/deletions in expressed sequence tag data of Eucalyptus. *J Genet*.;90, 34–38.

20. Rafalski, J.A. (2002). Application of single nucleotide polymorphisms in crop genetics. *CurrOpin Plant Biol*.;5, 94–100.

21. Riju, A., Chandraseker, A., Arunachalam, V. (2007). Mining for single nucleotide polymorphisms and insertions/deletions in expressed sequence tag libraries of oil palm. *Bioinformation*;2, 128–131.