An update to this article is included at the end

Microbiological Research 162 (2007) 250–256

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Microbiological Research

www.elsevier.de/micres

In silico mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites

Asheesh Shanker*, **Ashutosh Singh**, **Vinay Sharma**

Department of Bioscience and Biotechnology, Banasthali Vidyapith, Banasthali-304022, Rajasthan, India

Received 22 February 2006; received in revised form 5 May 2006; accepted 29 May 2006

KEYWORDS
Neurospora crassa;
Unigene;
Microsatellite;
Marker;
Annotation

Summary

In the present study, 3217 UniGene sequences of *Neurospora crassa* downloaded from the National Center for Biotechnology Information (NCBI) were mined for the identification of microsatellites or simple sequence repeats (SSRs). A total of 287 SSRs detected gives density of 1SSR/14.6 kb of 4187.86 kb sequences mined suggests that only 250 (7.8%) of sequences contained SSRs. Depending on the repeat units, the length of SSRs ranged from 14 to 17 bp for mono-, 14 to 48 bp for di-, 18 to 90 bp for tri-, 24 to 48 bp for tetra-, 30 for penta- and 42 to 48 bp for hexa-nucleotide repeats. Tri-nucleotide repeats were the most frequent repeat type (88.8%) followed by di-nucleotide repeats (5.9%). An attempt was also made with the help of bioinformatics approach to find out primer pairs for identified SSRs and primers were found only for 239 sequences. But, this part needs experimental validation. Annotation of SSRs containing sequences was also carried out.
© 2006 Elsevier GmbH. All rights reserved.

Intoduction

Neurospora crassa, the causative agent of an orange mold infestation in French bakeries (Perkins, 1991) is a non-pathogenic, filamentous ascomycete fungus with 43 Mb genome size that contains ~10,620 predicted protein-coding genes and has became a popular experimental model organism (Davis and Perkins, 2002) for important animal and plant pathogens. Diverse research programs centered on *Neurospora* have ranged from formal population and molecular genetics to more recent studies of development, photobiology, circadian rhythms, gene silencing, etc.

Microsatellites (simple sequence repeats (SSRs)) are short repeat motifs (1–6 bp) that are present in both protein coding and non-coding regions of DNA sequences (Gupta et al., 1996; Toth et al., 2000; Katti et al., 2001) and shows a high level of length

*Corresponding author.
E-mail address: ashomics@rediffmail.com (A. Shanker).

0944-5013/$ - see front matter © 2006 Elsevier GmbH. All rights reserved.
doi:10.1016/j.micres.2006.05.012

polymorphism due to insertion or deletion mutations of one or more repeat type (Tautz and Renz, 1984). Different taxon varies in abundance of different types of SSRs (Hancock, 1999) and these are present in greater abundance in non-coding regions than coding SSRs (Hancock, 1995). Abundance of SSR sequences in different genomes have been estimated originally via hybridization experiments (Tautz and Renz, 1984; Panaud et al., 1995) or database searches (Richard and Dujon, 1996; Toth et al., 2000). These studies were mainly based on the over-represented coding regions and limited by the partial genomic sequences available. Expressed sequence tags (ESTs), which represent the expressed part of genome also serve as source of SSRs (Liu et al., 1999).

Ideal molecular markers should be highly polymorphic, provide reproducible results and be simple to assay (Field et al., 1996). Detection of SSRs facilitates the development of SSR markers which fulfills the criteria of ideal markers therefore useful in numerous DNA- and PCR-fingerprinting experiments for strain typing of a variety of filamentous fungi and yeasts without prior knowledge of their abundance and distribution in the investigated fungal genomes (Lieckfeldt et al., 1992; Meyer et al., 2003) and are also useful across a number of related plant species (Holton et al., 2002; Thiel et al., 2003). Previously SSRs were identified in the genomes of nine phylogenetic diverse fungal genera: *Aspergillus nidulans*, *Cryptococcus neoformans*, *Encephalitozoon cuniculi*, *Fusarium graminearum*, *Magnaporthe grisea*, *N. crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Ustilago maydis* (Karaoglu et al., 2004).

Present study deals with the identification of microsatellites in UniGene sequences of *N. crassa*, which helps in the development of SSR markers and to annotate SSR containing sequences. UniGene is a database (available at NCBI) which contains sequences of well-characterized genes as well as hundreds of thousands novel EST sequences.

Materials and methods

Retrieval of UniGene sequences and detection of SSRs

A number of 3217 UniGene sequences of *N. crassa* (Ncr.seq.uniq) were downloaded from the NCBI. The harvesting of the SSRs was done using a perl script downloaded from www.missouri.edu. The minimum length of SSR was fixed at 14 bp according to criteria used by Gupta et al., 2003. The Poly A and Poly T repeats were not considered as SSRs due to their presence of mRNA/cDNA sequences at the 3′ end.

Primer designing for SSRs

A pair of primers flanking each SSR was designed using Primer3 software (www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Primer3 picks primers according to the specified parameters. In the present study, default parameters of the Primer3, viz. the optimum primer size as 20.0 (the range was 18–27), the optimum annealing temperature as 60.0 (the range was 57.0–63.0) and the range of %GC content as 20–80, were taken for primer designing.

Annotation of SSR containing sequences

Annotation of all the SSR containing sequences was determined on the basis of $\geqslant$70% similarity against non-redundant (nr) protein database entries. It was performed using program Basic Local Alignment Search tool (BLAST), variant BLASTX, available at NCBI (http://www.ncbi.nlm.nih.gov/blast). The resulting proteins obtained during similarity search were classified into their respective groups and looked for SSR in proteins.

Results

Screening of *N. crassa* sequences for SSRs

In the present study, 3217 UniGene sequences of *N. crassa* available at NCBI were searched for microsatellites with a minimum length of 14 bp. A total of 287 SSRs detected from 4187.86 kb of data screened, excluding Poly A and Poly T. Depending upon the length of the repeat unit itself (1–6 bp), the lengths of SSRs varied from 14 to 48 bp, respectively.

Frequencies of *N. crassa* SSRs with different repeat types

Only a subset of 250 sequences contains 287 SSRs, suggesting that merely 7.8% of sequences contained SSRs, which represent average density of one SSR/14.6 kb. Figure 1 shows the frequencies of SSRs with mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeat units. The most frequent repeat type found

within the UniGene sequences of *N. crassa* were tri-nucleotide repeats (88.8%) followed by di-nucleotide (5.9%) and tetra-nucleotide repeats (3.1%), respectively. Whereas, mono-, penta- and hexa-nucleotide repeats were present in equal (.7%) frequencies (Fig. 1).

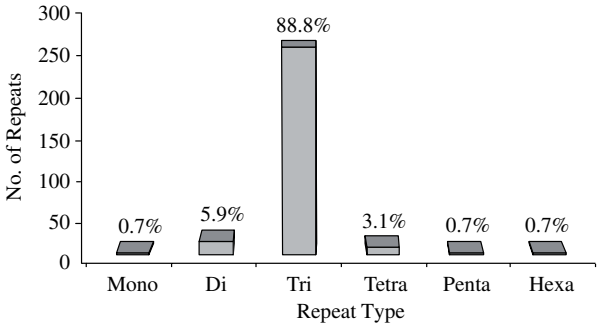

Figure 1. Frequency distribution of different repeat types identified in UniGene sequences of *N. crassa*.

The observed frequency of different repeat types comprising the SSRs is presented in Fig. 2(a)–(d) and summarized in Table 1. Identified SSRs belongs to only one type of mono-nucleotide, six different types of di-nucleotide repeats, 43 different types of tri-nucleotide repeats, nine different types of tetra-nucleotide repeats and two types of penta- and hexa-nucleotide repeats, respectively.

Among mono-nucleotide repeats, only (C)*n* repeats were found. The most abundant repeats were (CT)*n* and (CA)*n* followed by (GT)*n*, (TA)*n*, (TC)*n* in di-nucleotide and (CAG)*n* followed by (CAA)*n*, (ACA)*n*, (CCT)*n*, (GAA)*n*, (AGC)*n* repeats in tri-nucleotide repeats. Whereas tetra-, penta- and hexa-nucleotide repeats had almost equal frequency.

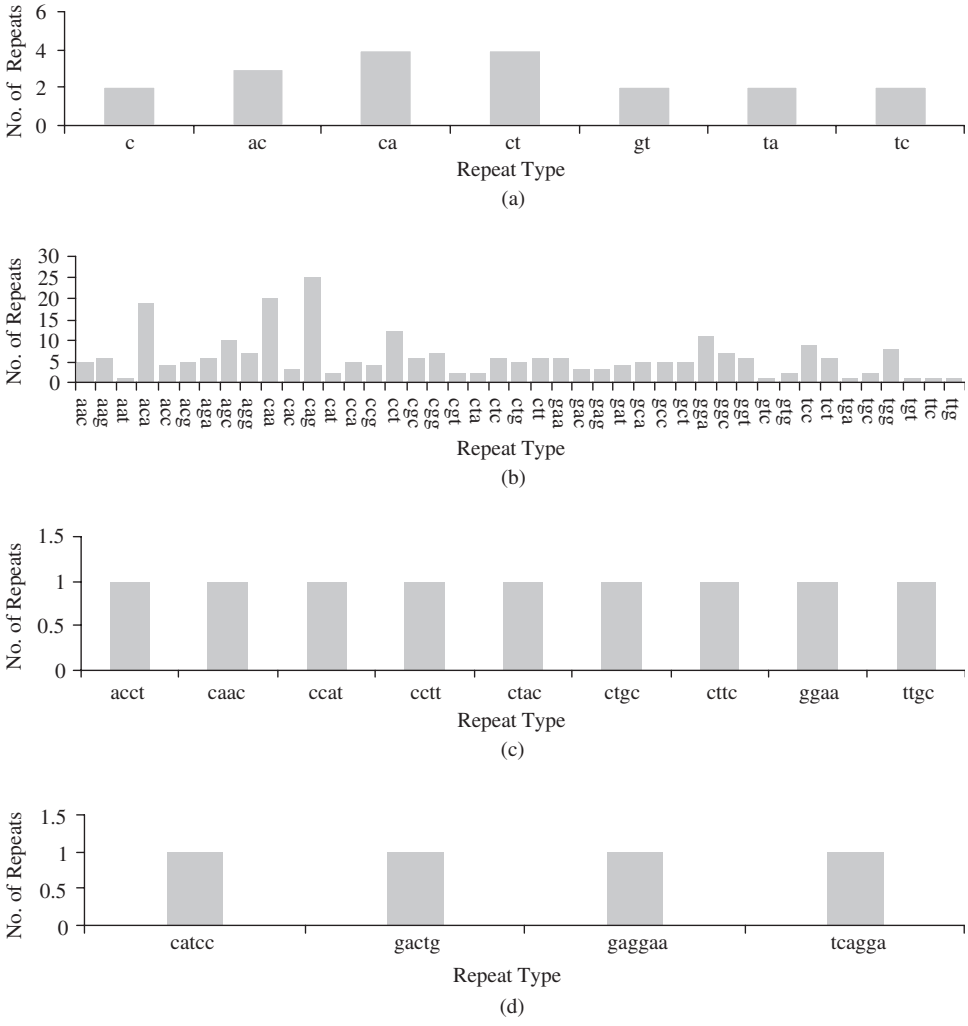

Figure 2. Frequency distribution of (a) mono- and di-, (b) tri-, (c) tetra-, (d) penta- and hexa-nucleotide repeat motifs in UniGene sequences of *N. crassa*.

Table 1. Summary of in silico mining of UniGene sequences of *N. crassa*

Parameters	Values
Total number of sequences searched	3217
Total number of SSRs identified	287
Total number of sequences containing single SSRs	220
Number of sequences containing two SSRs	24
Number of sequences containing three SSRs	5
Number of sequences containing four SSRs	1
Total number of sequences containing 287 SSRs	250
Repeat type	
Mononucleotide	2 (0.7)*
Dinucleotide	17 (5.9)
Trinucleotide	255 (88.8)
Tetranucleotide	9 (3.1)
Pentanucleotide	2 (0.7)
Hexanucleotide	2 (0.7)
Total length of sequences searched (kb)	4187.86
Density of SSRs	One/ 14.6

*Data in parentheses is the percentage value of the repeat type.

Designing of primers for SSRs

Out of 287 SSRs detected the primers could be designed only for 239 (83.3%) SSRs and 48 (16.7%) remaining sequences did not produce any acceptable primers. These 239 SSRs for which primers were designed include 10 di-, 221 tri-, 5 tetra- 1 penta- and 2 hexa-nucleotide repeats. The details of the accession numbers of UniGene sequences of *N. crassa*, repeat motif of SSRs for which primer were designed, length of repeat unit, primer sequences, annealing temperature and product size are available as Supplementary information.

Annotation of *N. crassa* sequences containing SSRs

To determine the function of SSR containing sequences, the 250 sequences from which SSRs were mined were annotated against the nr protein database available at http://www.ncbi.nlm.nih.gov. For a large number 198 (79.2%) of sequences, annotations were available (Fig. 3) of which 50 (25.25%) were predicted proteins, 99 (50%) were hypothetical proteins, 15 (7.58%) were putative proteins and 34 (17.17%) belonged to different functional classes. Only 52 (20.8%) sequences could not be assigned to any specific class due to the absence of a homolog in the protein sequence database. Further, matched proteins were searched for SSRs but no protein was found to contain SSR.

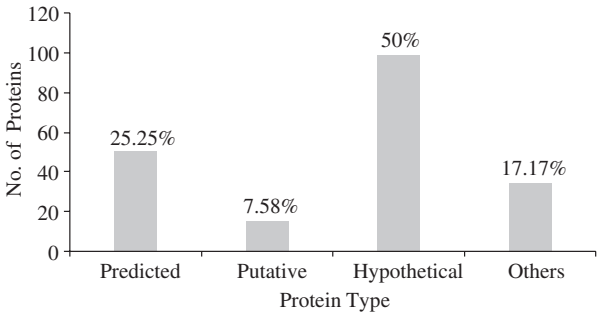

Figure 3. Distribution of SSR containing sequences according to the proteins that they encode.

Discussion

In the present study, UniGene sequences of *N. crassa* retrieved from NCBI were mined for SSRs and these SSRs were used for designing the markers. Further, all SSR containing sequences were annotated as far as possible.

N. crassa is known to have the highest SSR density in nine taxonomically different and completely sequenced fungal genomes including *A. nidulans*, *C. neoformans*, *E. cuniculi*, *F. graminearum*, *M. grisea*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *U. maydis* (Karaoglu et al., 2004). In the present study, a total of 287 SSRs were detected in UniGene sequences of *N. crassa* giving density of 1SSR/14.6 kb. However, the frequency of SSRs detected in this study was lower than the frequencies of SSRs found in fungal genomes: 1SSR/2.7 kb in *N. crassa*, 1SSR/3.3 kb in *M. grisea*, 1SSR/3.9 kb in *Saccharomyces cerevisiae*, 1SSR/4.0 kb in *Schizosaccharomyces pombe*, 1SSR/6.6 kb in *U. maydis*, 1SSR/9.6 kb in *C. neoformans*, 1SSR/12.5 kb in *F. graminearum*, 1SSR/12.5 kb in *A. nidulans* and 1SSR/15.6 kb in *E. cuniculi* (Karaoglu et al., 2004). Moreover, the frequency was also lower then EST derived SSRs reported in earlier studies on plants: 1SSR/3.4 kb in rice, 1SSR/8.1 kb in maize, 1SSR/7.4 kb in soybean, 1SSR/11.1 kb in tomato, 1SSR/6.0 kb averaged over six different species including barley, maize, rice, rye, sorghum and wheat (Varshney et al., 2002), 1SSR/1.67 kb in wheat (Morgante et al., 2002), 1SSR/14 kb in poplar and 1SSR/6 kb in *Arabidopsis* (Cardle et al., 2000). However, the density of SSRs in UniGene sequences of *N. crassa* (1SSR/14.6 kb) is higher than density (1SSR/20 kb) found in cotton (Cardle et al., 2000). This variation occurs due to difference in quantity of data

analyzed. Though due to differences in mutability and the bias in mismatch repair system, non-random distribution of SSRs occur this could lead to the fact that SSRs are over-represented in certain genomes (Harr et al., 2002). However, it is well-known that total SSR contents in fungal species are not influenced by the genome sizes (Karaoglu et al., 2004). Study on genomes of five different plant species (*Aspergillus thaliana*, rice, soybean, maize and bread wheat) shows that the densities of SSRs in transcribed regions were generally higher than in genomic DNA (Morgante et al., 2002) while our results are not consistent with these findings.

The abundance of the different repeats in the SSRs as detected in UniGene sequences of *N. crassa* was variable, so that the SSRs with different repeat motifs were not evenly distributed. These results are similar with earlier findings, which shows that the abundance of different repeats varied extensively depending upon the species examined (Toth et al., 2000). We excluded poly A and poly T repeats due to which their number is underrepresented in the study. The SSRs with tri-nucleotide repeats (88.8%) were most abundant in UniGene sequences of *N. crassa*. This abundance of tri-nucleotide repeats is in agreement with results of earlier studies in *N. crassa* genome (Karaoglu et al., 2004) and also in several crop species where tri-nucleotide repeats were found to be abundant (Scott et al., 2000; Varshney et al., 2000; Gupta et al., 2003). Di-nucleotides (5.9%) were the second abundant repeat type followed by tetra- (3.1%), mono- (0.7%), penta- (0.7%) and hexa- (0.7%) nucleotide repeats. In earlier studies on *Neurospora* genome (Karaoglu et al., 2004) AG/GA repeats were found to be predominant among di-nucleotide repeats while in our study no AG/GA repeats were found. It has been reported that $(AT)_n$ and $(CT)_n$ is the most common repeat motif in plants and insects (Lagercrantz et al., 1993). In our study $(CT)_n$ repeat were abundant while no $(AT)_n$ repeat were detected. The smaller repeat motifs were found to be predominant among SSRs identified and as the length of repeat unit increases their occurrence decreases. This may be because longer repeats have higher mutation rates, therefore less stable (Wierdl et al., 1997). Furthermore, di-nucleotide and tri-nucleotide repeat stretches found to be longer than other repeats.

The primers could be designed successfully for a very large number (239, 83.3%) of SSRs (see Supplementary table). But, it was not possible to design the primers for remaining SSRs (48, 16.7%) due to sequences flanking both ends of the SSRs was inadequate in size to design the primers. However, primer designing must be experimentally validated. The primers designed in the present study using data mining would be useful for a variety of purposes, e.g., gene tagging, genetic mapping and population studies etc. A further study is needed to experimentally prove the aforementioned part.

It has been proposed that numerous SSRs are the hot spots for recombination (Jeffreys et al., 1998; Templeton et al., 2000) especially di-nucleotide repeats are preferential sites for recombination due to their high affinity for recombination enzymes (Biet et al., 1999). SSRs may affect DNA replication (Field and Wills, 1996) and also plays important role in regulation of gene activity (Sandaltzopoulos et al., 1995). Some SSRs, found in upstream activation sequences, serve as binding sites for a variety of regulatory proteins (Lue et al., 1989; Csink and Henikoff, 1998). In addition to this, the presence of repeated sequences within proteins has been detected in all organisms examined (Marcotte et al., 1998). In the present study, similarities of SSR containing UniGene sequences were searched in nr protein database using BLAST available at NCBI (http://ncbi.nlm.nih.gov/blast) to find SSRs in proteins encoded by these sequences and to annotate the SSR containing sequences. Out of 250 SSR containing sequences, annotations were available only for 198 (79.2%) sequences and not even a single SSR were found in matched protein sequences. Sequences for which annotations were available categorized into different classes of proteins (predicted, hypothical, putative and others) and for the remaining 52 (20.8%) SSR containing sequences, no homology could be found because for a large fraction of the genomes, no obvious function has yet been assigned. In *Arabidopsis*, functions for only 57% of gene sequences were assigned, which represents a good source for annotating sequences, but is still inadequate.

UniGene sequences of *N. crassa* were systematically searched for SSRs using the "ssr_finder.pl" perl program. This approach saves both costs and time, given a sufficient amount of available UniGene sequences of *N. crassa*. The identified SSRs are useful for the development of SSR markers, which helps in genetic diversity studies and reveals variation in genomes. Annotation of SSR containing sequences provides an opportunity to examine the functional diversity of different proteins.

Acknowledgments

We are thankful to Professor Aditya Shastri, Director, Banasthali Vidyapith, for providing neces-

sary facilities and Dr. Bhumi Nath Tripathi, Department of Bioscience and Biotechnology, Banasthali Vidyapith, Banasthali, 304022, India for critical comments and suggestions on manuscript.

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.micres.2006.05.012.

References

Biet, E., Sun, J., Dutreix, M., 1999. Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. Nucl. Acids Res. 27, 596–600.

Cardle, L., Ramsay, L., Milborne, D., Macaulay, M., Marshall, D., Waugh, R., 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156, 847–854.

Csink, A.K., Henikoff, S., 1998. Something from nothing: the evolution and utility of satellite repeats. Trends in Genet. 14, 200–204.

Davis, R.H., Perkins, D.D., 2002. Neurospora: a model of model microbes. Nat. Rev. Genet. 3, 397–403.

Field, D., Eggert, L., Metzgar, D., Rose, R., Wills, C., 1996. Use of polymorphic short and clustered coding-region microsatellites to distinguish strains of Candida albicans. FEMS Immunol. Med. Microbiol. 15, 73–70.

Field, D., Wills, C., 1996. Long, polymorphic microsatellite in simple organisms. Proc. R. Soc. London B Biol. Sci. 263, 209–215.

Gupta, P.K., Balyan, H.S., Sharma, P.C., Ramesh, B., 1996. Microsatellites in plants: a new class of molecular markers. Curr. Sci. 70, 45–54.

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., Balyan, H.S., 2003. EST-SSRs for transferability, polymorphism and genetic diversity in bread wheat. Mol. Genet. Genom. 270, 315–323.

Hancock, J.M., 1995. The contribution of slippage-like processes to genome evolution. J. Mol. Evol. 41, 1038–1047.

Hancock, J.M., 1999. Microsatellites and other simple sequences, genomic context and mutational mechanisms. In: Goldstein, D.B., Schlötterer, C. (Eds.), Microsatellite, Evolution and Application. Oxford University Press, Oxford, pp. 1–9.

Harr, B., Todorova, J., Schlötterer, J., 2002. Mismatch repair driven mutational bias in *D. melanogaster*. Mol. Cell. 10, 199–205.

Holton, T.A., Christopher, J.T., McClure, L., Harker, N., Henry, R.J., 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol. Breed. 9, 63–71.

Jeffreys, A.J., Murray, J., Neumann, R., 1998. High-resolution mapping of crossovers in human sperm defines a minisatellite associated recombination hotspot. Mol. Cell 2, 267–273.

Karaoglu, H., Lee, C.M.Y., Meyer, W., 2004. Survey of simple sequence repeats in completed fungal genomes. Mol. Bio. Evol. 22, 39–49.

Katti, M.V., Ranjekar, P.K., Gupta, V.S., 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Bio. Evol. 18, 1161–1167.

Lagercrantz, U., Ellegren, H., Andersson, L., 1993. The abundance of various polymerphic microsatellite motifs differs between plants and vertebrates. Nucl. Acids Res. 21, 1111–1115.

Lieckfeldt, E., Meyer, W., Kuhls, K., Börner, T., 1992. Characterization of filamentous fungi and yeast by DNA fingerprinting and random amplified polymorphic DNA. Belg. J. Bot. 125, 226–233.

Lue, N.L., Buchman, A.R., Kornberg, R.D., 1989. Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element in vitro. Proc. of the Nat. Acad. of Sci. USA 86, 486–490.

Liu, Z.J., Tan, G., Li, P., Dunham, R.A., 1999. Transcribed dinucleotide microsatellite and their associated genes from channel catfish Ictalurus punctatus Biochem. Biophys. Res. Commun. 259, 190–194.

Marcotte, E., Pellegrini, M., Yeates, T., Eisenberg, D., 1998. A census of protein repeats. J. Mol. Biol. 293, 151–160.

Meyer, W., Castaneda, A., Jackson, S., Huynh, M., Castaneda, E., the IberoAmerican Cryptococcal Study Group, 2003. Molecular typing of IberoAmerican *Cryptococcus neoformans* isolates. Emerg. Infect. Dis. 9, 189–195.

Morgante, M., Hanafey, M., Powell, W., 2002. Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. 30, 194–200.

Panaud, O., Chen, X., McCouch, S.R., 1995. Frequency of microsatellite sequence in rice (*Oryza sativa* L.). Genome 38, 1170–1176.

Perkins, D.D., 1991. The first published scientific study of *Neurospora*, including a description of photoinduction of carotenoids. Fung. Genet. Newslett. 38, 64–65.

Richard, G.F., Dujon, B., 1996. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. Gene 174, 165–174.

Sandaltzopoulos, R., Mitchelmore, C., Bonte, E., Wall, G., Becker, P.B., 1995. Dual regulation of the Drosophila hsp26 promoter in vitro. Nucl. Acids Res. 23, 2479–2487.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., Henry, R.J., 2000. Analysis of SSRs derived from grape ESTs. Appl. Genet. 100, 723–726.

Tautz, D., Renz, M., 1984. Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. Nucl. Acids Res. 12, 4127–4138.

Templeton, A.R., Clark, A.G., Weiss, K.M., Nickerson, D.A., Boerwinkle, E., Sing, C.F., 2000. Recombinational and mutational hot spots within the human lipoprotein lipase gene. Am. J. Hum. Genet. 66, 69–83.

Thiel, T., Michalek, W., Varshney, K., Graner, A., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106, 411–422.

Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genome, survey and analysis. Genome Res. 10, 1967–1981.

Varshney, R.K., Kumar, A., Balyan, H.S., Roy, J.K., Prasad, M., Gupta, P.K., 2000. Characterization of microsatellite and development of chromosome specific STMS markers in bread wheat. Plant Mol. Biol. Rep. 18, 1–12.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., Graner, A., 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol. Biol. Lett. 7, 537–546.

Wierdl, M., Dominska, M., Petes, T.D., 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146, 769–779.

Update

Microbiological Research
Volume 163, Issue 1, 15 January 2008, Page 120

DOI: https://doi.org/10.1016/j.micres.2007.08.002

Microbiological Research **163** (2008) 120

Available online at www.sciencedirect.com
ScienceDirect

Microbiological Research

www.elsevier.de/micres

ERRATUM

Erratum to "In silico mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites" [Microbiol. Res. 162 (2007) 250–256]

Asheesh Shanker*, **Ashutosh Singh, Vinay Sharma**

Department of Bioscience and Biotechnology, Banasthali Vidyapith, Banasthali 304022, Rajasthan, India

Unfortunately, in the original article published in Microbiological Research 162/3 (2007) *Arabidopsis thaliana* has been misprinted to *Aspergillus thaliana* on page 5, line 8.

The correct sentences should be:

Study on genomes of five different plant species (*Arabidopsis thaliana*, rice, soybean, maize and bread wheat) shows that the densities of SSRs in transcribed regions were generally higher than in genomic DNA (Morgante et al., 2002) while our results are not consistent with these findings.

Reference

Morgante M, Hanafey M, Powell W. Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 2002;30:194–200.

DOI of original article: 10.1016/j.micres.2006.05.012
*Corresponding author.
E-mail address: ashomics@rediffmail.com (A. Shanker).

0944-5013/$ - see front matter © 2007 Elsevier GmbH. All rights reserved.
doi:10.1016/j.micres.2007.08.002