

DNA Microsatellites: A Review

Shikha Jain*, Ranjana Joshi, Kirti Jatwa, Avnish Sharma, S.C. Mahajan
 Department of Pharmaceutics,
 Mahakal Institute of Pharmaceutical Studies,
 Behind Air strip, Datana, Dewas Road, Ujjain, Madhya Pradesh, India
 *jain.shikhapharma@gmail.com



ABSTRACT

Microsatellites consist of tandemly repeated sequence, no more than 6 bases long. They are scattered throughout most eukaryotic genomes. The major characteristic that makes microsatellites as useful and powerful genetic tool is the extensive length polymorphism that first of all reflects allelic variation in the number of the tandemly arranged perfect repeats. Microsatellites are generally considered as the most powerful genetic marker.

A genetic marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. Genetic marker that reveal polymorphisms at the DNA level are called molecular marker.

Molecular markers are called as DNA markers, which should be thought of as signs along the DNA trail that pinpoint the location of desirable genetic traits or indicate specific genetic differences. They are responsible for various neurological diseases and hence the same cause can now be utilized for the early detection of various diseases, such as, Schizophrenia, Bipolar Disorder and Congenital generalized Hypertrichosis. These agents are widely used for forensic identification and relatedness testing, and are predominant genetic markers in this area of application.

Keywords: DNA, Microsatellites, DNA sequences

INTRODUCTION TO MICROSATELLITES

Microsatellites are DNA sequences of mono-, di, tri-, tetra- and penta- nucleotide units repeated in tandem, which are widely distributed in the genome. For example,

A-A-A-A-A-A-A-A-A would be referred to as (A) 11

G-T-G-T-G-T-G-T-G-T would be referred to as (GT) 6

C-T-G-C-T-G-C-T-G-C-T-G would be referred to as (CTG) 4

A-C-T-C-A-C-T-C-A-C-T-C-A-C-T-C would be referred to as (ACTC)4

In the literature they are often called as simple sequence repeats (SSR), short tandem repeats

(STR). Microsatellites are inherited in a Mendelian fashion.

The term Microsatellite was first coined by Lit and Luty in 1989 when analyzing the abundance and dispersion of (TG)_n in the cardiac actin gene. The existence of dinucleotide repeats- poly (C-A), poly (G-T) (i.e. an alternating sequence of cytosine and adenine, with on the opposite strand of the DNA molecule, alternating guanine and thymine) was first documented almost 15 years ago by Hamada and colleagues. Subsequent studies by Tautz and Renz have confirmed both the abundance and ubiquity of microsatellites in eukaryotes.

How to cite this article: S Jain, R Joshi, K Jatwa, A Sharma, SC Mahajan; DNA Microsatellites: A Review; PharmaTutor; 2014; 2(8); 79-86

Most microsatellites are found in non-coding DNA — DNA which does not have the "code" or instructions to synthesize protein. Consequently, they are not believed to play a significant role in cell functioning. However, there is reason to believe that a microsatellite can disrupt normal cell processes if it grows too large.

Origin of microsatellite:

According to Scientists the origin of microsatellites in genomes appears to be nonrandom, with an imbalance between the mechanisms that promote and those that prevent the microsatellites initiation. Currently, there are two non-mutually exclusive hypotheses to explain the origin of microsatellites:

1. De novo microsatellites ^[1]:

De novo microsatellites suggests that the birth of microsatellites was a consequence of the creation of a proto-microsatellite, a short region of as few as 3 or 4 repeated units within cryptically simple sequences, which are defined as a scramble of repetitive motifs lacking a clear tandem arrangement.

2. Adopted microsatellites ^[2]:

Adopted microsatellites suggests that microsatellites arise from other genomic

regions via transposable elements. The transposable elements might contain one or more sites that are predisposed to microsatellite formation and hence favor the dispersal of microsatellites in genomes.

Classification of Microsatellites ^[3]:

In literatures markers are classified according to the number of bases, i.e., short repeats (10- 30 bases) are microsatellites and longer repeats are mini satellites (between 10-100 bases). Microsatellites have been also been classified according to the type of repeated sequence presented:

1. Perfect:-The repeat sequence is not interrupted by any base not belonging to the motif (e.g. T-A-T-A-T-A-T-A-T-A-T-A-T-A-T-A)
2. Imperfect:-There is a pair of bases between the repeated motifs that does not match the motif sequence (e.g. T-A-T-A-T-A-T-A-C-T-A-T-A-T-A)
3. Interrupted:- There is a small sequence within the repeated sequence that does not match the motif sequence (e.g. T-A-T-A-T-A-C-G-T-G-T-A-T-A-T-A-T-A-T-A)
4. Composite:- The sequence contains two adjacent distinctive sequence-repeats (e.g. T-A-T-A-T-A-T-A-T-A-G-T-G-T-G-T-G-T-G-T)

(Table1)

Table 1-Classification of Microsatellites

(A) Based on the arrangement of nucleotides in the repeat motifs
· Pure or perfect or simple perfect (CA) <i>n</i> Simple imperfect (AAC) <i>n</i> ACT (AAC) <i>n</i> + 1
· Compound or simple compound (CA) <i>n</i> (GA) <i>n</i>
· Interrupted or imperfect or compound imperfect (CCA) <i>n</i> TT (CGA) <i>n</i> + 1
(B) Based on the number of nucleotides per repeat ^[4]
· Mononucleotide (A) <i>n</i>
· Dinucleotide (CA) <i>n</i>
· Trinucleotide (CGT) <i>n</i>
· Tetranucleotide (CAGA) <i>n</i>
· Pentanucleotide (AAATT) <i>n</i>
· Hexanucleotide (CTTTAA) <i>n</i> (<i>n</i> = number of variables)

(C) Based on location of SSRs in the genome
· Nuclear (nuSSRs)
· Chloroplastic (cpSSRs)
· Mitochondrial (mtSSRs)

Location of microsatellites:

The presence of SSRs in eukaryotes was verified from diverse genome regions, including 3'-UTRs, 5'-UTRs, exons and introns ^[5]. Furthermore, their localization could potentially interfere with different aspects of DNA structure, DNA recombination, DNA replication and gene expression ^[6]. Microsatellites are also commonly located in proximity of interspersed repetitive elements, such as short interspersed repeats (SINEs) and long interspersed repeats (LINEs). In promoter regions, the presence and length of SSRs could influence transcriptional activity. The microsatellites can also be present in organellar genomes, such as chloroplast and mitochondria, and nuclear DNA ^[7].

ISOLATION OF MICROSATELLITES

The method of isolation of Microsatellites can be grouped into 3 types:

- (i) Standard method: A library is screened for repeated sequences;
- (ii) Automated method: The SSR sequences are searched in sequence databases and
- (iii) Sequencing method: The whole genome or parts of the genome are sequenced using high-throughput technologies.

(i) Standard method ^[8]:

This method requires the creation of a library. There are various protocols to create and screen genomic, cDNA or PCR fragment library but the main steps can be summarized as follows:

1. Fragmentation of DNA by sonication or enzymatic digestion.
2. Ligation of the DNA fragments into a vector and transformed into *Escherichia coli*.
3. Analysis of the clones for the presence of SSR

sequences by Southern blot.
4. Sequencing of the positive clones.

Advantages:

- The number of positive clones obtained by this methodology ranges from 0.04 to 12%,
- Efficient method of isolation.

Disadvantages:

The cost of developing a microsatellite marker is high because the use of a total genomic DNA library requires the evaluation of a large number of clones to find those containing repeated sequences.

ii. Automated method ^[9]:

Isolation of Microsatellite can also be made possible through the use of public DNA databases to search for repeated sequences. Initially, database searches were performed using unspecific alignment tools, such as BLASTN. Subsequently, several computer-based software programs were developed and the SSR search became easier.

This automated approach reduces the costs associated with microsatellite marker development but is limited to species with available sequences.

iii. Sequencing method ^[10,11]:

Following the isolation of microsatellite sequences, it is necessary to develop PCR primer pairs flanking these sequences to test new loci for robust amplification, genomic copy number and sufficient polymorphism

Advantages:

- The new high-throughput sequencing technologies have allowed whole or expressed genome sequencing

- These technologies do not require the creation of libraries (total DNA or RNA can be sequenced),
- These technologies produce a huge amount of sequences quickly and because many steps have been skipped,
- These technologies have lower costs than other methods.

Mutation mechanisms and DNA repair

Although microsatellites have been extensively used in a considerable number of studies covering the most varied areas of genetics, the mutational dynamics of these genomic regions is still not well understood, although it is known that the mutation rate of microsatellite is much higher than that of other parts of the genome, ranging from 10^{-2} to 10^{-6} nucleotides per locus per generation.

Several mechanisms have been suggested to explain the high mutation rate of microsatellites, including errors during recombination, unequal crossing-over and

polymerase slippage during DNA replication or repair.

Replication slippage^[12]:

DNA slippage is a symmetrical process, where the same number of repeats is added and removed leading to either the loss of microsatellites or the insertion of a high number of repeats. The misalignment that gives rise to mutations occurs between a newly synthesized DNA strand and its complementary template strand. The two strands dissociate and re-anneal incorrectly, forming a loop, which is stable due to the repetitive nature of the sequence. If the loop is formed on the nascent strand, the resulting mutation will be a repeated expansion, while loops on the template strand result in a reduction of the repeat length. If the mutation occurs in a coding region, it could produce abnormal proteins, leading to diseases. The Huntington's disease well known example. (Figure 1)

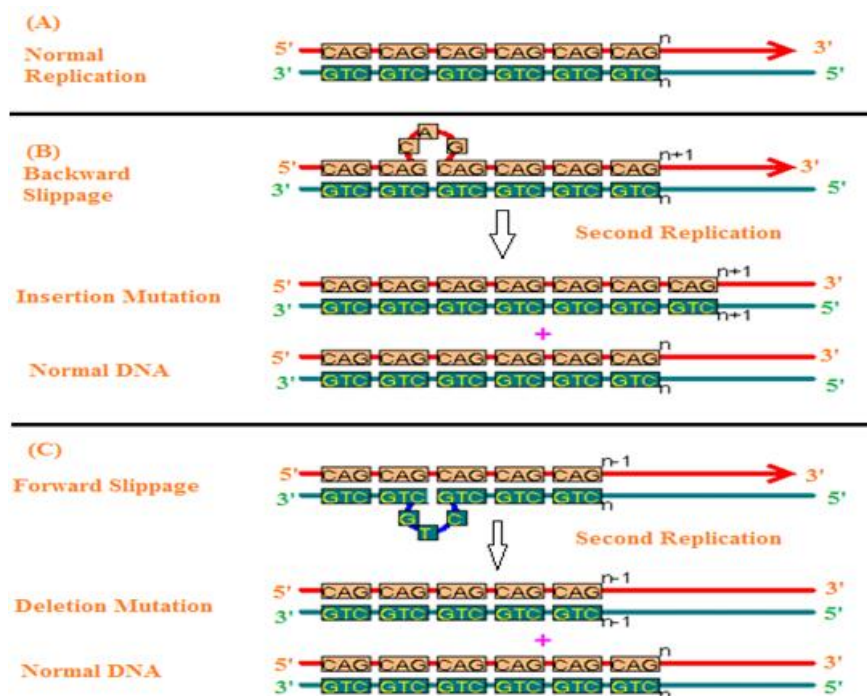


Figure 1 - The mutation caused by replication slippage. In this figure, mispairing involves only one repeat. In fact, the slippage could cause several repeats to become unpaired.

- (a) Normal replication
- (b) Backward slippage, resulting in the insertion mutation.
- (c) Forward slippage, resulting in the deletion mutation

UNEQUAL CROSSING OVER DURING MEIOSIS (RECOMBINATION) ^[13]

This mechanism is usually associated with the exchange of repeated units between homologous chromosomes, and therefore, plays a limited role in microsatellite mutation. However, this mechanism might be responsible for microsatellite multistep mutations. (Figure2)

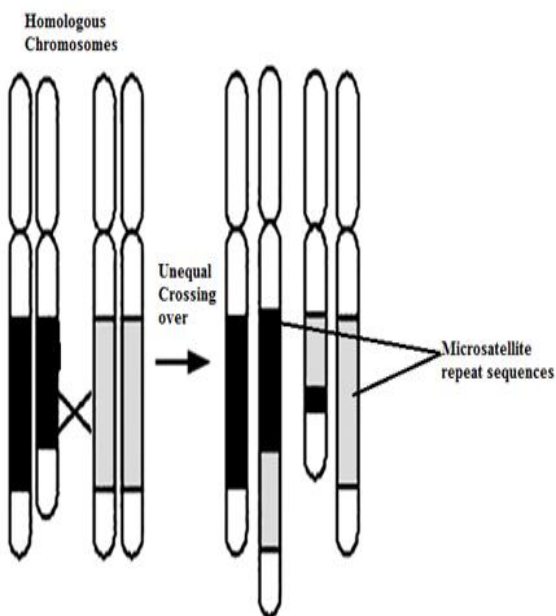


Figure 2 – Unequal crossing-over between homologous chromosomes. Black and gray regions correspond to microsatellite repeat sequences

Advantages of Microsatellites as genetic markers ^[14]:

- Low quantities of template DNA required.
- High genomic abundance
- Random distribution throughout the genome
- High level of polymorphism
- Band profiles can be interpreted in terms of loci and alleles

- Co-dominant markers
- Allele sizes can be determined with high accuracy
- Comparison across different gels possible using size standard
- High reproducibility
- Different microsatellites may be multiplexed in PCR or co-loaded in a gel
- Wide range of applications
- Amenable to automation

Disadvantages of Microsatellites as genetic markers ^[15]:

- Initial high development costs
- Heterozygotes may be misclassified as homozygotes when null-alleles occur due to mutation in the primer annealing sites
- Stutter bands may complicate accurate scoring of polymorphisms
- Underlying mutation model (infinite alleles model or stepwise mutation model) largely unknown
- Homoplasmy due to different forward and backward mutations may underestimate genetic divergence
- Time-consuming and expensive to develop

APPLICATIONS OF MICROSATELLITES

1. Forensics ^[16]:

Microsatellite loci, are widely used for forensic identification and are a predominant genetic marker in this area of application. In forensic identification cases, the goal is typically to link a suspect with a sample of blood, semen or hair taken from a criminal. Alternatively, the goal may be to link a sample found on a suspect's clothing with a victim. Relatedness testing in criminal work may involve investigating paternity in order to establish rape or incest. Another application involves linking DNA samples with relatives of a missing person. Because the lengths of microsatellites may vary from one person to the next, scientists have begun to use them to identify criminals and to determine paternity, a procedure known

as DNA profiling or "fingerprinting". The features that have made use of microsatellites attractive are due to their relative ease of use, accuracy of typing and high levels of polymorphism. The ability to employ PCR to amplify small samples is particularly valuable in this setting, since in criminal casework only minute samples of DNA may be available.

2. Diagnosis and Identification of Human Diseases ^[17]:

They serve a role in biomedical diagnosis as markers for certain disease conditions. That is, certain microsatellite alleles are associated (through genetic linkage) with certain mutations in coding regions of the DNA that can cause a variety of medical disorders.

3. Population Studies ^[18]:

By looking at the variation of microsatellites in populations, inferences can be made about population structures and differences, genetic drift, genetic bottlenecks and even the date of a last common ancestor.

It can also be very helpful in studying the inheritance of the natural characters of an individual from its ancestor.

It also reveals a significant information on the correlation of the individual's genetic constitution with respect to its ancestor from which he has been descended.

4. Conservation Biology:-

Microsatellites can be used to detect sudden changes in population, effects of population fragmentation and interaction of different populations.

5. In Mapping genomes:

They are very much helpful in mapping of genome and finding or locating the significant portion in the genome of an individual. They have found wide applications in areas such as

the widely publicized mapping of the human genome.

6. Detection of Cancer:

The rate of microsatellite expansion (that is, increase in the number of repeats) or contraction (decrease in number of repeats) in cells is increased in some types of cancers, due to defects in enzymes that correct copying mistakes in DNA. Early clinical detection of some types of colon and bladder cancers using changes in microsatellite repeats have been successful.

7. Microsatellites for Tracking Blast Resistance in Rice ^[19,20,21,22]:

Many Pi genes confer resistance to overlapping spectra of blast pathotypes, and it is often difficult to monitor for the presence of individual resistance genes and pyramid these in breeding lines using traditional phenotypic screening. Therefore, DNA markers provide a straightforward and rapid means to select for multiple blast resistance genes without performing extensive progeny testing or disease screening. DNA markers linked to several of the Pi genes have been localized on rice chromosomes, as well as markers for Pi-ta and Pi-b. Unfortunately, the majority of DNA markers for blast resistance are RFLPs, which are relatively labor intensive to analyze for use in breeding programs. Markers that can be analyzed by PCR are more amenable for breeding purposes, such as the ones developed for Pi-2 and Pi-ta .

DISEASES INDUCED BY MICROSATELLITE

These can be categorized into:

1. Diseases Involving Trinucleotide Repeats

- Huntington's disease
- Fragile X
- Myotonic dystrophy
- Spinalbulbar muscular atrophy
- Friedrich's ataxia
- Spinocerebellar ataxia

- Dentatorubral pallidolusians

2. Diseases Found By Positional Cloning

- Congenital generalized hypertrichosis
- Schizophrenia and Bipolar Disorders

CONCLUSION

Microsatellite consists of a specific sequence of DNA bases which consist of mono, di, tri tandem repeats. They are scattered through

eukaryotic genomes. Trinucleotide repeat sequence are responsible for causing various types of neurodegenerative disorder like Huntington's disease, Fragile X, Myotonic dystrophy, Spinalbulbar muscular atrophy, Friedrich's ataxia, Spinocerebellar ataxia, Dentatorubral pallidolusians. If the repeat is present in a healthy gene, a dynamic mutation may increase the repeat count and result in a defective gene.

↓ REFERENCES

1. Messier, M., Li, S.H., Stewart, C.B. The birth of microsatellites. *Nature*, 1996; 381: 483, ISSN 0028-0836.
2. Wilder, J. & Hollocher, H. Mobile elements and the genesis of microsatellites in dipterans. *Molecular Biology and Evolution*. 2011; 18(3) : 384-392, ISSN 1537-1719.
3. Andrea Akemi Hoshino, Juliana Pereira Bravo, Paula Macedo Nobile and Karina Alessandra Morelli. Microsatellites as Tools for Genetic Diversity Analysis, *Genetic Diversity in Microorganisms*, Prof. Mahmut Caliskan (Ed.), InTech, p.p 149-63. ISBN: 978-953-51-0064-5,
4. Kalia, R.K., Manoj, K.R., Sanjay, K., Rohtas, S., Dhawan, A.K. Microsatellite markers: An overview of the recent progress in plants. *Euphytica*. 2011; 177: 309–334.
5. Rajendrakumar, P, Biswal, A.K., Balachandran, S.M., Srinivasarao, K. & Sundaram, R.M. Simple sequence repeats in organellar genomes of rice: frequency an distribution in genic and intergenic regions. *Bioinformatics*. 2007; 23(1): 1–4, ISSN 1460-2059
6. Chistiakov, D.A.; Hellemans, B. & Volckaert, F.A.M. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*. 2006; 255: 1-29, ISSN 0044- 8486.
7. Kashi, Y., King, D. & Soller, M. Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics*. 1997; 13(2): 74–78, ISSN 0168-9525.
8. Mittal, N. & Dubey, A.K. Microsatellite markers - A new practice of DNA based markers in molecular genetics. *Pharmacognosy Reviews*. 2009; 3(6): 235-246, ISSN 0973-7847.
9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology*. 1990; 215(3): 403-410, ISSN 0022-2836.
10. Abdelkrim, J., Robertson, B.C., Stanton, J.A.L. & Gemmell, N.J. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*. 2009; 46(3):185-192, ISSN 0736-6205.
11. Mikheyev, A.S., Vo, T., Wee, B., Singer, M.C., Parmesan, C. Rapid Microsatellite Isolation from a Butterfly by De Novo Transcriptome Sequencing: Performance and a Comparison with AFLP-Derived Distances. *PLoS ONE*. 2010; 5(6): 112-21, ISSN 1932-6203.
12. Schlötterer, C. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 2000; 109(6) : 365–371, ISSN 0009-5915.
13. Grover, A. & Sharma, P.C. Is spatial occurrence of microsatellites in the genome a determinant of their function and dynamics contributing to genome evolution? *Current Science*. 2011; 100(6): 859-869, ISSN 0011-3891.
14. B. Pokhriyal, K. Thorat, D.A. Limaye, Y.M. Joshi, V.J. Kadam and Dubey R. Microsatellite Markers- A

- novel Tool in Molecular Genetics. International Journal of Research in Pharmacy and chemistry. 2012; 2(2): 397-412. ISSN: 2231?2781
15. Kimberly A., Selkoe and Robert J. Toonen Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers, Ecology Letters. 2006; 9:615-629.
16. woodrow.org/teachers/esi/2002/biology/projects/p3/definition.htm #APPLICATIONS.
17. Kashi Y. and Soller M. "Functional Roles of Microsatellites and Minisatellites." In:Microsatellites: Evolution and Applications. Edited by Goldstein and Schlotterer. Oxford University Press,1999.
18. Moxon E.R. and Wills C. "DNA microsatellites: Agents of Evolution?" Scientific American.1999;72-77.
19. Jia, Y., Wang, Z., Singh, P. Development of dominant rice blast Pi-ta resistance gene markers. Crop Sci. 2002;42: 2145–2149.
20. Hittalmani, S., Parco, A., Mew, T.V., Ziegler, R.S.; Huang, N. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Appl. Genet. 2000; 100: 1121–1128.
21. Nakamura, S., Asakawa, S., Ohmido, N., Fukui, K., Shimizu, N., Kawasaki, S. Construction of an 800-kb contig in the nearcentromeric region of the rice blast resistance gene Pi-ta2 using a highly representative rice BAC library. Mol. Gen. Genet. 1997; 254: 611–620.
22. Monna, L., Miyao, A., Zhong, H.S., Yano, M., Iwamoto, M., Umehara, Y., Kurata, N., Haysaka, H., Sasaki, T. Saturation mapping with subclones of YACs: DNA marker production targeting the rice blast disease resistance gene, Pi-b. Theor. Appl. Genet. 1997; 94: 170–176.