INTERNATIONAL JOURNAL OF RESEARCH IN PHARMACY AND CHEMISTRY

Available online at www.ijrpc.com

Review Article

MICROSATELLITE MARKERS – A NOVEL TOOL IN MOLECULAR GENETICS

B. Pokhriyal*, K. Thorat, DA. Limaye, YM. Joshi, VJ. Kadam and Dubey R

Bharati Vidyapeeth's College of Pharmacy, Sector 8, CBD Belapur, Navi Mumbai, India.

ABSTRACT

Amongst the different molecular markers used in molecular genetics, microsatellite markers which are the Simple Sequence Repeats (SSR) or short repetitive sequences of nucleotides of the DNA in the genome of an individual has more or less a prominent and functional role in determining the genetics of an individual. This review gives a clear understanding of these microsatellite markers, their significance in genome, isolation and characterization as well as application of these markers in the molecular genetics for assessing the genetic integrity of an individual.

Keywords: molecular markers, Microsatellite markers, simple sequence repeats.

INTRODUCTION

In the beginning of 20th century scientists have discovered that the Mendelian factors of genetic trait lies on the chromosomes and control inheritance. Later these were called as genes which were advocated to be linearly arranged on the chromosome.Combination of genes is inherited as groups and they get linked because they are closely arranged. In the absence of even one gene the phenotypes are altered. The function of gene is marked by its presence or absence. Hence, the marker may be in polymorphic forms so that a chromosome bearing a mutant can be distinguished from the normal chromosome.

In the markers the polymorphism can be detected at 3 different levels:

1) Morphological or phenotypic markers which correspond to the quantitative traits scored visually.

2) Biochemical markers which are nothing but the protein molecules or the isozymes expressed as neutral genetic markers.

3) Molecular markers which consists of the DNA sequences among which the differences can be detected and monitored in the subsequent generations.

In this report, only molecular markers have been dealt with in details[1].

What are Molecular Markers?

In recent years recombinant DNA technology and PCR technology have helped in the construction of genetic, cytogenetic and physical maps of genomes of plants and animals using molecular markers. The markers revealing variations at DNA level are referred to as the Molecular markers. A large number of genetic polymorphism occurring at the DNA sequence level can be used as the molecular markers for evaluation of the phenotypic variability.

Molecular markers must possess the following desirable properties:

1) It must be polymorphic so that the diversity can be measured

2) It should be evenly distributed throughout the genome.

3) It should be easily and rapidly detected

4) It must distinguish the homozygotes and heterozygotes[1]

Classification of Molecular Markers

Molecular markers possess unique genetic properties on the basis of techniques used for its detection.

They are classified into two major classes:

1) DNA hybridization based markers and

2) Polymerase Chain Reaction (PCR) based markers

1) DNA hybridization based markers
The DNA piece can be cloned, and allowed to hybridize with the genomic DNA which can be detected. Marker-based DNA hybridization has been widely used but the major limitation of this approach is that it requires large quantities of DNA and the use of radioactive labeled probes.
This method basically involves the following steps:
1) Isolation of genomic DNA,
2) Its digestion by the restriction enzymes,
3) Separation by gel electrophoresis and then
4) Hybridization by incubating with cloned and labeled probes.

2) PCR based markers
Polymerase Chain Reaction (PCR) is a novel technique for the amplification of the selected regions of the DNA. The advantage with PCR is that even a minute quantity of DNA can be amplified to a large extent and hence it requires only a small quantity of DNA to start with.
PCR-based markers may be divided into two types:
1) Locus non-specific markers :
e.g. a) Random amplified polymorphic DNA (RAPD);
b) Amplified fragment length polymorphism (AFLP).
Since they are locus nonspecific they doesn't give the relevant and precise information about the genome and the genes and the pattern of their occurrence and inheritance. Hence, the locus specific markers are more preferred.
2) Locus specific markers :
e.g. a) Simple sequence repeats (SSR);
b) Single nucleotide polymorphism (SNP).
Among the segregating populations, the phylogenetic markers are very useful for mapping the polygenic traits as well as Mendelian traits. The most widespread among the polygenic markers are the Variable number of tandem repeats (VNTRs) or Microsatellites. The locus-specific probe fails to cover it and shows the highly polyallelic fragment length variation. For e.g. the tandem repeat loci associated with rRNA genes which are concentrated at nucleolar organizing regions (NORs) of certain chromosomes of an individual. Similarly, most VNTRs loci get centered in pro-terminal regions of human chromosomes. Therefore, the desired density of markers is not provided. This difficulty is overcome in Microsatellite or SSR (Simple Sequence Repeats) loci[2].

What are Microsatellites?
The term Microsatellite was first coined by Lit and Lutty. These are also known as the Simple sequence repeat (SSR) or Short tandem repeat (STR). These are the stretches of DNA, consisting of tandemly repeating mono-, di-, tri-, tetra- and penta- nucleotide units, which are arranged throughout the genomes of most eukaryotic species. For example,
AAAAAAAAAAA would be referred to as $(A)_{11}$
GTGTGTGTGTGT would be referred to as $(GT)_6$
ACTCACTCACTCACTC would be referred to as $(ACTC)_4$
They are often represented as $(CA)_n$ repeat, where n is variable between alleles and CA is the most commonly found repeated units of cytosine – adenine pair in the human genome .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 which are inherited in a Mendelian fashion[3].

Why Microsatellite Markers are Preferred over other Molecular Markers?[3]
There are several widely used marker types available for molecular ecology studies, and many questions can be addressed with more than one type of marker. Microsatellites are of particular interest to ecologists because they are one of the few molecular markers that allow researchers insight into fine-scale ecological questions. Regardless of the question, a molecular marker must fundamentally be selectively neutral and follow Mendelian inheritance in order to be used as a tool for detecting demographic patterns, and these traits should always be confirmed for any marker type. The desirable traits of microsatellites compared with other marker types such as allozymes, amplified fragment length polymorphisms(AFLP), sequenced loci and single nuclear polymorphisms (SNP), focused on both practicalities and ecological considerations are as under

1) **Easy sample preparation[3]**
An ideal marker allows the use of small tissue samples which are easily preserved for future use. In contrast to allozyme methods, DNA-based techniques, such as microsatellites, use PCR to amplify the marker of interest from a minute tissue sample. The stability of DNA

compared with enzymes allows the use of simple tissue preservatives (such as 95% ethanol) for storage. In addition, because microsatellites are usually shorter in length than sequenced loci (100–300 vs. 500–1500 base pairs) they can still be amplified with PCR despite some DNA degradation (Taberlet et al. 1999)[5]. As DNA degrades, it breaks into smaller pieces and the chance of successfully amplifying a long segment is proportional to its length (Frantzen et al. 1998)[6]. This trait allows microsatellites to be used with fast and cheap DNA extraction methods, with ancient DNA, or DNA from hair and faecal samples used in non-invasive sampling (Taberlet et al. 1999).Furthermore, because microsatellites are species-specific, cross-contamination by non-target organisms is much less of a problem compared with techniques that employ universal primers (i.e. primers that will amplify DNA from any species), such as AFLP. This feature is of particular importance when working with faecal samples or species, such as scleractinian corals, in which endosymbiont contamination is practically unavoidable.

2) **High information content**[3]

Each marker locus can be considered a sample of the genome. Because of recombination, selection and genetic drift, different genes and different regions of the genome have slightly different genealogical histories. Relying on a single locus to estimate ecological traits from genetic data creates a high rate of sampling error. Thus, taking multiple samples of the genome by combining the results from many loci provides a more precise and statistically powerful way of comparing populations and individuals. Furthermore, statistical approaches to the questions of most interest to ecologists often require multiple, comparable loci (Pearse & Crandall 2004). Although AFLP, allozymes and random amplified polymorphic DNA (RAPD) techniques are also multilocus, none of them have the resolution and power of a multilocus microsatellite study (but for distinct reasons; see Sunnucks 2000). While AFLP markers can be a good alternative choice to microsatellites (Bensch & Akesson 2005). Gerber et al. (2000) showed that 159 AFLP loci provided slightly less power to determine paternity than six polymorphic microsatellite markers. Sequencing technology has advanced rapidly, but its cost still prohibits the duplication or triplication of workload by using multiple independent gene sequences in parallel (Zhang & Hewitt 2003). SNP markers hold great promise for future studies but their use in non-model organisms is still nascent (Morin et al. 2004).

3) **Co-dominant markers**[3]

Microsatellites have become so popular because they are single locus, co-dominant markers for which many loci can be efficiently combined in the genotyping process to provide fast and inexpensive replicated sampling of the genome. Microsatellite markers generally have high-mutation rates resulting in high standing allelic diversity. In species for which populations are small or recently bottlenecked, markers with lower mutation rates, such as allozymes, may be largely invariant and only loci with the highest mutation rates are likely to be informative (Hedrick 1999). A slow mutational process allows the signature of events in the distant past to persist longer. Thus, the selection of loci with high or low allelic diversity will depend on the question of interest. For example, if one is interested in a potential historical barrier to gene flow or tracing the recolonization of territory since the last ice age, markers with lower mutational rates are likely to be the most informative. [3]

Attributes of Microsatellites as Genetic markers

Some of the significant attributes of microsatellite markers which they possess are as follows:

1) Locus-specific in nature; in contrast to multi-locus markers such as mini-satellites or Random amplified polymorphic DNA (RAPDs)
2) Co-dominant transmission and therefore the heterozygotes can be distinguished from homozygotes, in contrast to Random amplified polymorphic DNA (RAPD); Amplified fragment length polymorphism (AFLP) which are binary in nature
3) Highly polymorphic and hyper-variable
4) High information content and provides considerable pattern
5) Relative abundance with uniform genome coverage
6) Higher mutation rate than standard sequences (up to 0.001gamates/generation)
7) High probability of back mutation[4].

Significance of Microsatellites In Genome[4]

As there are often many alleles present at a microsatellite locus, genotypes within pedigrees are often fully informative, in that the progenitor of a particular allele can often be identified. In this way, microsatellites can be used for determining paternity, population genetic studies and recombination mapping. It is also the only molecular marker to provide clues about which alleles are more closely related. Also, these markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number are 10 or greater.

The Significance of Microsatellite markers are as under:

- Microsatellites determine the genotype of an individual. They usually don't have any measurable effect on phenotype, and when they do mutate, may cause a change in the genotype of an individual.
- Microsatellites may act as a marker for some genetic diseases. They were previously considered to be as the "Junk" DNA which is generally found on the Non-coding regions of DNA and the variation is mostly neutral.In humans, 90% of known microsatellites are found in non-coding regions of the genome. When found in human coding regions, microsatellites are known to cause disease. Interestingly, when found in coding regions, microsatellites are usually tri-nucleotide repeats. One possible explanation is that any other type of nucleotide repeat would be too detrimental to the coding region, because it would cause a frame-shift mutation.
- Microsatellites provide a necessary source of genetic variation. The variation in microsatellite alleles in coding regions is thought to be the cause of adaptation in different environments. In other words, a short allele may be adaptive in one environment, and a long allele with many repeats may be adaptive in a different environment.
- Microsatellite variation may be a way to compensate for loss of genetic variability due to genetic drift and selection. Thus having variation within the population would ensure the survival of the population in varying environment.
- Microsatellites may help regulate gene expression and protein function.Kashi and Soller (1999) have suggested that microsatellites may have regulatory roles in gene expression. They are systematically found near coding regions. Variation in microsatellite alleles have been shown to be associated with quantitative variation in protein function and gene activity.

Mutation in Microsatellites[5]

Microsatellites owe their variability to an increased rate of mutation compared to other neutral regions of DNA. The size of the repeat unit, the number of repeats and the presence of variant repeats are all factors, as well as the frequency of transcription in the area of the DNA repeat. Interruption of microsatellites, perhaps due to mutation, can result in reduced polymorphism. The reason seems to be that their mutations occur in a fashion very different from that of "classical" point mutations (where a substitution of one nucleotide to another occurs, such as a G substituting for a C).

The mutation process in microsatellites occurs through what is known as slippage replication. If we consider the repeat units (e.g., an AC dinucleotide repeat) as beads on a chain, we can imagine that during replication two strands could slip relative positions a bit, but still manage to get the zipper going down the beads. One strand or the other could then be lengthened or shortened by addition or excision of nucleotides. The result will be a novel "mutation" that comprises a repeat unit that is one bead longer or shorter than the original.

It is estimated that microsatellites mutate 100 to 10,000 times as fast as base pair substitutions. This makes microsatellites useful for studying evolution over short time spans, whereas base pair substitutions are more useful for studying evolution over long time spans (millions of years). Microsatellite alleles mutate over time. In a population, there may exist many alleles of a single microsatellite locus. Microsatellite alleles differ in the number of repeats. For example, one allele may have 7 repeats of a CT motif, and another allele may have 8 repeats. In a population, there may exist many alleles at a single locus, with each allele having a different length. An individual who is homozygous for a locus will have the same number of repeats on both chromosomes, whereas a heterozygous individual will have different numbers of repeats on the two chromosomes.

The regions surrounding the microsatellite locus, called the flanking regions, may still have the same sequence. This is important because the flanking regions can therefore be used as PCR primers when amplifying microsatellite loci, and can be conserved across genera or sometimes even families. Below, the two lines represent the sequences on two homologous chromosomes in a diploid organism.

Homozygous
(Both strands have 7 CT repeats)
…CGTAGCCTTGCATCCTTCTCTCTCTCTCT
CTATCGGTACTACGTGG…
…CGTAGCCTTGCATCCTTCTCTCTCTCTCT
CTATCGGTACTACGTGG…
5’ flanking region microsatellite locus 3’ flanking region

Heterozygous
(One strand has 7 repeats, and the other has 8 repeats)
…CGTAGCCTTGCATCCTTCTCTCTCTCTCT
CTATCGGTACTACGTGG…
…CGTAGCCTTGCATCCTTCTCTCTCTCTCT
CTCTATCGGTACTACGTGG…

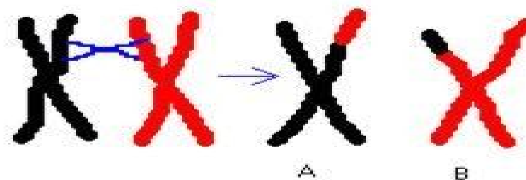
Microsatellites are useful genetic markers because they tend to be highly polymorphic. It is not uncommon to have human microsatellites with 20 or more alleles with heterozygosities. [8]

Microsatellites generally tend to occur in non-coding regions of the DNA although a few human genetic disorders are caused by (tri-nucleotide) microsatellite regions in coding regions. On each side of the repeat unit are flanking regions that consist of "unordered" DNA. The flanking regions are critical because they allow us to develop locus-specific primers to amplify the microsatellites with PCR (polymerase chain reaction). That is, given a stretch of unordered DNA 30-50 base pairs (bp) long, the probability of finding that particular stretch more than once in the genome becomes vanishingly small. This combination of widely occurring repeat units and locus-specific flanking regions are a part of strategy for finding and developing microsatellite primers. The primers for PCR will be sequences from these unique flanking regions. By having a forward and a reverse primer on each side of the microsatellite, we will be able to amplify a fairly short (100 to 500 bp, where bp means base pairs) locus-specific microsatellite region.

There are two hypotheses that explain how microsatellites mutate[5]

1. **“Polymerase slippage” or “slipped-strand mispairing”:** When the DNA replicates, the polymerase loses track of its place, and either leaves out repeat units or adds too many repeat units. The result is that the new strand has a different number of repeats as the parent strand. This is thought to explain small changes in numbers of repeats. It also explains how microsatellite loci could be generated in the first place; it is likely that sequences including two or three repeats are randomly distributed throughout the genome. Slippage could then amplify these short repeat sequences into many repeats over successive generations. Certainly, the effectiveness of the mismatch repair system would also play an important role in microsatellite mutation rate.

2. **Unequal crossing-over during meiosis:** This is thought to explain more drastic changes in numbers of repeats. In the diagram below, chromosome A obtained too many repeats during crossing-over and chromosome B obtained too few repeats.[9]



Model for Microsatellite Mutation[5]

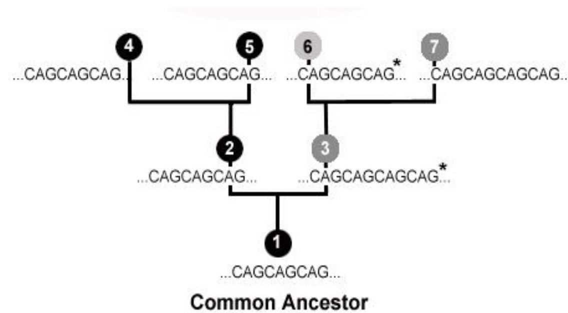
Many models have been proposed to explain the mutation in microsatellites but due to some or the other limitations or unexplained facts they have been discarded. So a model is proposed which explains most of the mutational mechanisms and facts and uncertainties in microsatellite mutation are been described in stepwise mutation model as follows:

Stepwise Mutation Model (SMM)

The idea that adding or subtracting one repeat is likely easier than adding or subtracting two or more beads is the basis for using the Stepwise Mutation Model (SMM). This model holds that when microsatellites mutate, they only gain or lose one repeat. This implies that two alleles that differ by one repeat are more closely related (have a more recent common ancestor) than alleles that differ by many repeats. In other words, size matters when doing statistical tests of population sub-structuring. An advantage of the SMM is that the difference in size conveys additional information about the phylogeny of alleles. The SMM is generally the preferred model when calculating relatedness between individuals and population sub-structuring, although there is the problem of homoplasy.

Problem: Homoplasy

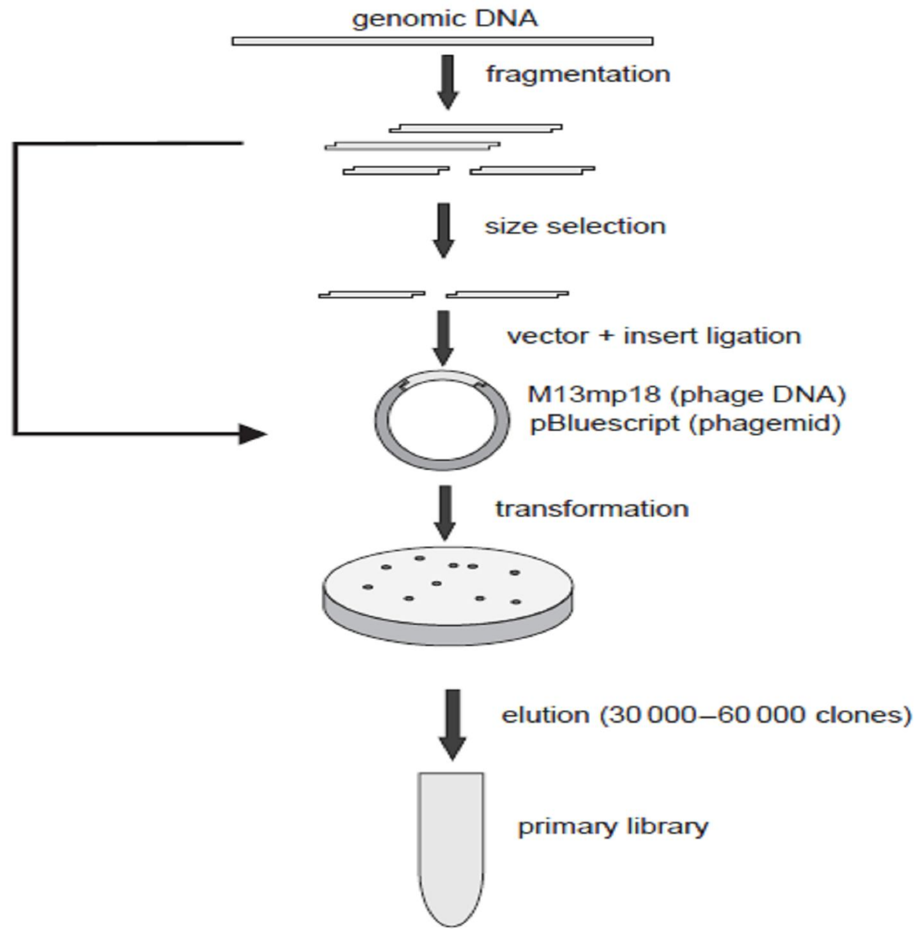
Pretend that you are studying a population and you find four individuals. Three of them have the same genotype, and one is different. This would indicate that the three with the same genotype are more closely related to each other than they are to the other. However, this is not necessarily the case. To understand why, study the phylogeny below. Asterisks indicate microsatellite mutations.

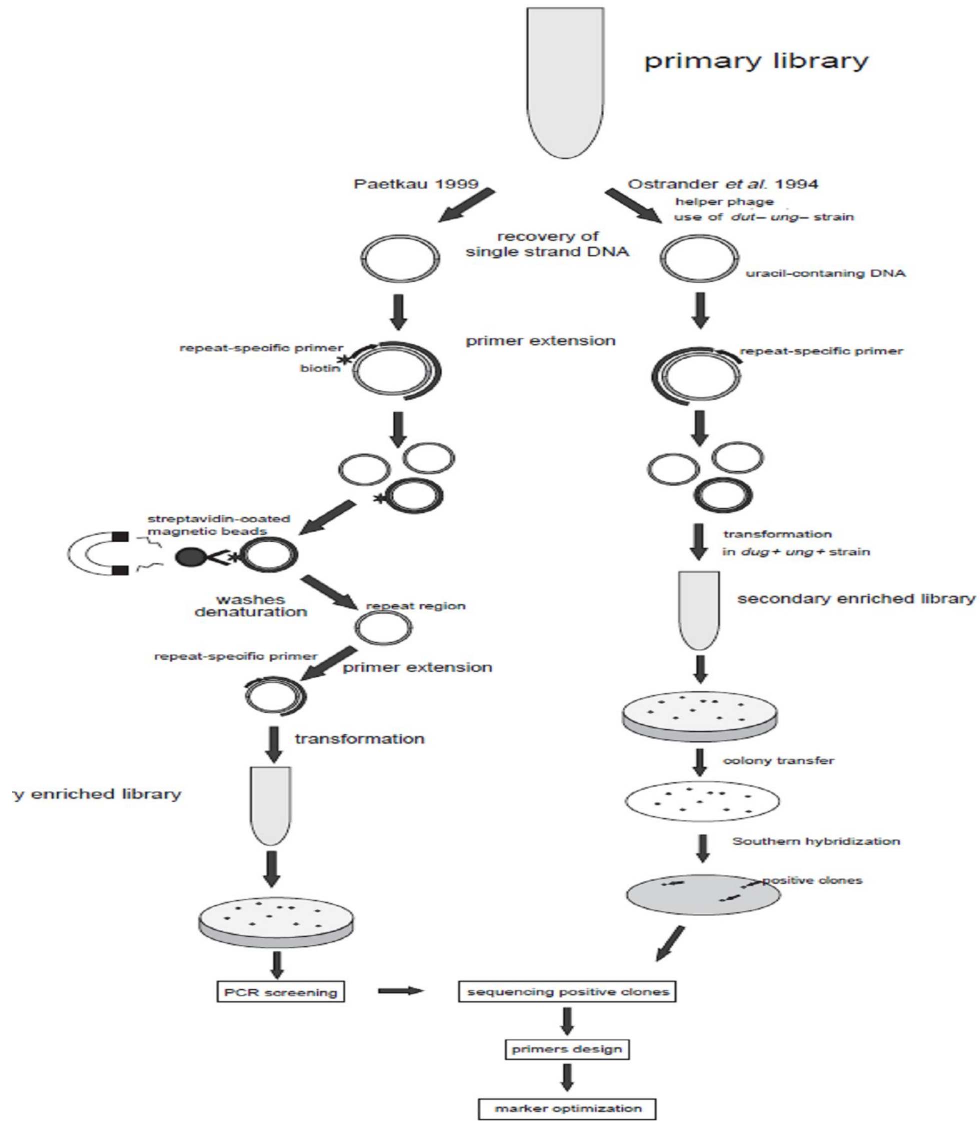


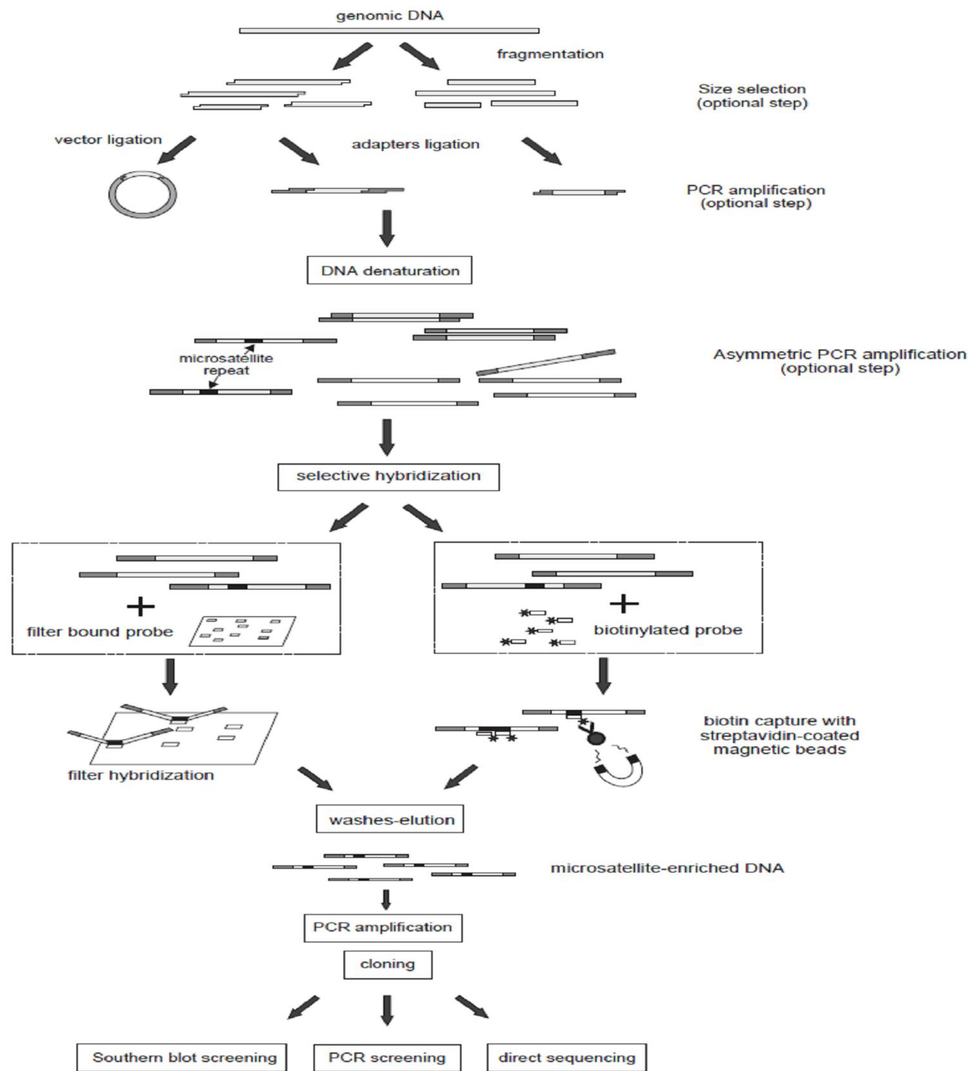
In the **fig; population 1 gave rise to two populations, 2 and 3.**
In population 3, there was a stepwise mutation, so that now there are four CAG repeats instead of three.
Population 3 gave rise to two more populations, 6 and 7.
Population 6 lost a repeat, so now it has three CAG repeats.
The problem is that populations 4, 5, and 6 have the same allele at this microsatellite locus, yet they have different evolutionary histories.We can say that their alleles are identical in state but not by descent. If a scientist were only examining this one locus, he/she would mistakenly conclude that population 6 is more closely related to populations 4 and 5 than it is to 7.
This phenomenon, when two alleles are identical in state but not identical by descent, is known as homoplasy.
In population studies, homoplasy can lead to underestimates of divergence. The only way to detect homoplasy would be to examine many other loci.
Still, homoplasy is thought to have little effect on populations over a short period of time (hundreds of generations), and stepwise mutation model is still the preferred model. [9]

Isolation, Development and Characterization of Microsatellites[6]

Developing new microsatellite markers based on the enrichment technique and primer optimization steps.
A. Extract DNA from a single tissue sample.
B. Create a DNA library:
1. Cut the genome into 500 bp fragments pieces with a restriction enzyme digest.
2. Attach ‘linker’ DNA to the ends of each fragment – linker DNA has a known sequence so that
primers can be designed to bind to them.
3. Amplify the DNA fragments using primers for the linker ends with PCR.
C. Separate out fragments with repeat sequences:
1. Mix the DNA fragments with a microsatellite probe (an oligonucleotide made of a repeatsequence of your choice) that can be recovered magnetically.
2. Promote the hybridization of probes to any complementary repeat sequences in the DNA Fragments by heating to denature the DNA and cooling slowly.
3. Hold a magnet to the tube to attract the probes (now bound to the DNA), and wash away therest of the unbound DNA with a series of rinses.
D. Sequence the fragments to find microsatellite loci:
1. Using primers for the linker DNA, amplify DNA with PCR to concentrate it.
2. Clone the DNA to prepare it for sequencing - insert it into a plasmid, inoculate bacteria with the plasmid, grow the bacteria to replicate the DNA.
3. Isolate the DNA from the bacteria.
4. Sequence microsatellite DNA in the plasmid with primers targeted to the insertion points on the plasmid.
E. Examine the sequences to find microsatellite repeats.
F. Design primers for the flanking region of the microsatellites (with help from a primer selection software program such as Primer3 which selects optimal primer sites) and have them made.
G. Attempt amplification of loci with the new primers. Use a gradient of PCR conditions in which the temperatures, times, magnesium and primer concentrations vary to find optimal conditions.
H. Use gel electrophoresis to confirm the presence of PCR products. Discard primer pairs that fail to amplify after several attempts.
I. Check for polymorphism by running the successful primer pairs on 10-20 individuals. Estimate allelic diversity and heterozygosity levels. Discard invariant loci.
J. Check for reliability. Rerun the successful primer pairs on the same individuals twice more to ensure that genotype scoring is consistently reproducible. Discard loci with unreliable amplification.
H. Order fluorescently labeled primers for the remaining loci. Complete the full screening process. Discard problematic loci.
K. Streamline the genotyping of the full dataset with the remaining loci by establishing a “multiplex” PCR protocol – primers for multiple loci (labeled with different dyes) are amplified in a single PCR reaction[3].


genomic DNA
fragmentation
size selection
vector + insert ligation
M13mp18 (phage DNA)
pBluescript (phagemid)
transformation
elution (30 000–60 000 clones)
primary library

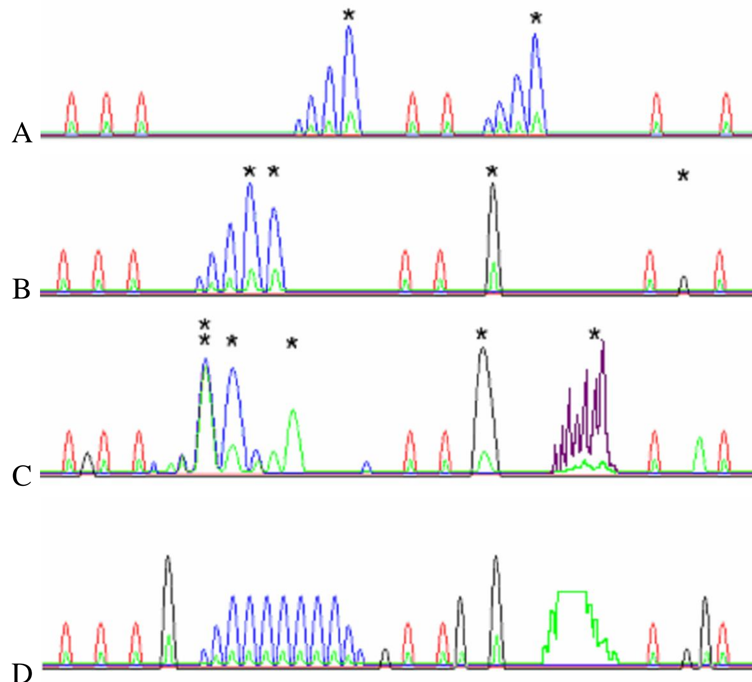

primary library
Paetkau 1999
Ostrander et al. 1994
helper phage
use of dut– ung– strain
recovery of
single strand DNA
uracil-contaning DNA
primer extension
repeat-specific primer
biotin
repeat-specific primer
streptavidin-coated
magnetic beads
transformation
in dug+ ung+ strain
secondary enriched library
washes
denaturation
repeat region
repeat-specific primer
primer extension
transformation
colony transfer
y enriched library
Southern hybridization
positive clones
PCR screening
sequencing positive clones
primers design
marker optimization



Schematic representation of selective hybridization protocols. [6]

Microsatellite Genotyping with a DNA Sequencer

A DNA sequencer is a highly precise gel electrophoresis apparatus. PCR products are loaded onto the gel and separated by size by applying a charge. A laser scans the gel to detect bands containing fluorescent dye. The primers used in the PCR reaction are tagged with different fluorescent dyes to enable this detection. The sequencer software converts the banding pattern into a plot with peaks corresponding to the width and intensity (height) of each band. The position of the peak along the x-axis corresponds to the size of the DNA product in the band measured in base pairs (BP). The height/intensity corresponds to the concentration of the DNA product, which is a consequence of the efficiency of the amplification process in PCR. One color is used for a size standard to calibrate the band positions with the size of the DNA product. An asterisk identifies all true microsatellite alleles in the figures below.



A) **An ideal output**

The two alleles of this heterozygote are even in height and easy to distinguish from the “stutter peaks” adjacent to them – during PCR some products are 1, 2 or 3 repeats short due to errors in replication (similar to step-wise mutation) and show up as evenly spaced peaks with decreasing height to the left of the true peak. Some loci show extensive stuttering and others show virtually none. The stutter bands are useful for distinguishing microsatellite products from non-specific or non-target products. Notice that there are “pull-up peaks” in the green section of the spectrum. Pull-up is due primarily to spectral overlap in the emission spectra of the dyes which the sequencer records as a false peak in a different color. This is a common artifact of the DNA sequencer’s analysis process that can create confusion in some situations.

B) **Two examples of trickier outputs**

The blue genotype is a heterozygote but the 2 alleles are only 1repeat different in size. This creates a characteristic pattern for loci with stutter in which the second peak is higher than the first because of the additive intensity of the larger allele’s first stutter peak and the smaller allele’s true peak. If the first two peaks were equal in height it would be difficult to determine if the genotype has one or two alleles. One clue is that there are 4 stutter peaks to the right of the largest peak instead of 3. The second locus in this plot is shown with black dye and has larger alleles. This locus has no stutter perhaps because there are few repeats so that the Taq polymerase doesn’t make stepwise errors. This genotype is also a heterozygote, but the larger allele is faint. Larger alleles usually show at least slightly shorter peaks because PCR is less efficient for longer products. Here, the larger allele is so small that it could be easily overlooked or mistaken for noise. If any less PCR product were loaded onto the gel it might not show up at all – in which case it would be an example of “large allele drop-out,” another common source of inflated homozygosity counts.

C) **More examples**

When multiple loci are loaded in the same gel lane for efficiency, or amplified together in one PCR “multiplex” reaction, allele peaks can overlap and be sometimes easy to miss. Here the green and blue loci share an allele size. Distinguishing the true alleles is made even more difficult due to the occurrence of green pull-up peaks. If the green allele product were less intense than the blue allele product (instead of equal as shown here), it might be mistaken for a pull up peak. Re-running the green locus separately in such cases will minimize scoring error. Here again, the blue locus shows two alleles that differ by one base pair. The heights are the same because this locus does not show strong stutter. But the blue locus does show small flanking peaks – the left side is a very faint stutter so only the largest stutter peak is visible, and the right

side might be an “A-Tail,” when the Taq adds an extra adenine nucleotide onto some copies of the product, increasing it by 1 bp. It will not be mistaken for a microsatellite allele because an extreme height difference would not occur for 2 alleles so close in size, as their amplification efficiencies should be similar. The black locus is a homozygote. Even though there is a small black peak on the left side of the plot, a smaller allele is almost never shorter than a larger allele. An additional clue is that the larger black peak is quite fat and tall -- PCR produces approximately double the amount of product when an allele is homozygous because it does not compete for the Taq with a second allele. The purple locus represents a “split peak” problem that occurs from high rates of ATailing by the Taq. The “+A” peaks occur for all of the stutter peaks, making the scoring difficult, especially for heterozygotes with closely sized alleles. Although the true allele is denoted here, a locus with alleles that look like the purple one would be too difficult to score reliably. Usually the problem can be corrected by adding an extra extension step to the PCR program that gives the Taq time to add an A-Tail to all copies of the product consistently.

D) **Unscorable loci**

The blue locus is a “stegosaur” with unacceptably high stutter. The black locus has too many non-specific artifact peaks to reliably choose the microsatellite alleles. The green locus has been overloaded on the gel or has unusually high PCR product concentration and is smearing in the lane[3].

Limitations To Utility of Microsatellites[7]

Microsatellites have proved to be versatile molecular markers, particularly for population analysis, but they are not without limitations. Despite many advantages, microsatellite markers also have several challenges and pitfalls that at best complicate the data analysis, and at worst greatly limit their utility and confound their analysis. However, all marker types have some downsides, and the versatility of microsatellites to address many types of ecological questions outweighs their drawbacks for many applications. Fortunately, many of the pitfalls common to microsatellite markers can be avoided by careful selection of loci during the isolation process. Species-specific marker isolation PCR-based marker analysis requires primer sequences that target the marker regions for amplification. In order to use the same primer sequence to amplify the same target from many individuals, the region where the primer binds must be identical, with few or no mutations causing inter-individual differences. For the gene regions commonly used as sequenced markers, primer regions are highly conserved, such that they are invariant within species and sometimes even across broad taxonomic groups. This sequence conservation necessitates only minor work to optimize a primer set for a new species. In contrast, a given pair of microsatellite primers rarely works across broad taxonomic groups, and so primers are usually developed a new for each species (Glenn & Schable 2005). However, the process of isolating new microsatellite markers has become faster and less expensive, which substantially reduces the failure rate and cost of new marker isolation in many cases (Glenn & Schable 2005).

Some of the Major limitations in utility of microsatellites are as follows

1. **Unclear mutational mechanisms[7]**

One of the challenges currently being addressed by geneticists is that the mutational processes of microsatellites can be complex (Schlo¨tterer 2000; Beck et al. 2003; Ellegren 2004). For the majority of ecological applications, it is not important to know the exact mutational mechanism of each locus, as most relevant analyses are insensitive to mutational mechanism (Neigel 1997). However, several statistics based on estimates of allele frequencies (e.g. FST and RST) rely 618 K. A. Selkoe and R. J. Toonen 2006 Blackwell Publishing Ltd/CNRS explicitly on a mutation model. Traditionally, the infinite allele model (IAM), in which every mutation event creates a new allele (whose size is independent from the progenitor allele) has been the model of choice for population genetics analyses, and because it is the simplest and most general model, continues to be widely used as a default. A model specific to microsatellites, the stepwise mutational model (SMM), adds or subtracts one or more repeat units from the string of repeats at some constant rate to mimic the process of errors during DNA replication that generates mutations, creating a Gaussian-shaped allele frequency distribution (Ellegren 2004). However, nonstepwise mutation processes are also known to occur, including point mutation and recombination events such as unequal crossing over and gene conversion (Richard & Paques 2000). While debate continues about the prevalence of non-stepwise mutation for microsatellites, the current consensus is that the frequency and effects are usually low, and

stepwise mutation appears to be the dominant force creating new alleles in the few model organisms studied to date (Eisen 1999; Ellegren 2004). Nevertheless, metrics employing the SMM tend to be highly sensitive to violations of this mutational model (e.g. loci with non-stepwise mutation or constraints on allele size) and thus metrics using the IAM are usually more robust and reliable (Ruzzante 1998; Balloux & Lugon-Moulin 2002; Landry et al. 2002). More complex and realistic mutational models that add the probability of non-stepwise mutation to the SMM are beginning to replace the SMM in common genetic analyses, and are already available in several statistical packages (e.g. Piry et al. 1999; Van Oosterhout et al. 2004).

2 Hidden allelic diversity[7]

Size-based identification of alleles (i.e. gel electrophoresis)greatly reduces the time and expense of microsatellite genotyping compared with sequencing each allele in each individual. However, this shortcut requires the assumption that all distinct alleles differ in length. In fact, alleles of the same size but different lineages can be quite common, a phenomenon termed homoplasy.
Homoplasy dampens the visible allelic diversity of populations and may inflate estimates of gene flow when mutation rate is high (Garza & Freimer 1996; Rousset 1996; Viard et al. 1998; Blankenship et al. 2002; Epperson 2005).There are two distinct types of homoplasy, detectable and undetectable. Detectable homoplasy can be revealed by sequencing alleles. For instance, point mutations will leave the size of an allele unchanged, and insertions or deletions in the flanking region might create a new allele with the same size as an existing allele. Detectable homoplasy appears to affect only a fraction of genotypes at a fraction of loci, and this bias appears to be marginal in the majority of cases (Viard et al. 1998; Adams et al. 2004; Curtu et al. 2004). Adams et al. (2004) found homoplasy was only common for compound and/or interrupted repeats. Empirical estimates of detectable homoplasy reported only a slight (1–2%) underestimation of genetic differentiation (Adams et al. 2004; Curtu et al. 2004).Undetectable homoplasy occurs when two alleles are identical in sequence but not identical by descent (i.e. they have different genealogical histories). Such non-identity occurs from the random-walk behaviour of the stepwise mutation process when there is a back-mutationto a previously existing size (e.g. an allele mutates from 5 to 6 repeats and then a copy of this allele mutates from 6 to 5 repeats) or when two unrelated alleles converge in sequence by changing repeat number in two different places in the sequence. As the SMM predicts a 50% chance of backmutation, undetectable homoplasy may be extensive when mutation rate is high, but can be accounted for in analyses (Slatkin 1995; Estoup & Cornuet 1999). In general, homoplasy is often a minimal source of bias for population genetic studies limited to populations with a shallow_ history or moderate effective population size, as the chance of homoplasy is proportional to the genetic distance of two individuals or populations (Estoup et al. 2002). However, when used for highly divergent groups, such as for phylogenetic reconstruction, high-mutation rate loci may be problematic (Estoup et al. 1995). It is important to note that undetected homoplasy plagues all marker types. When appropriate, there are several methods that can be employed to assess detectable homoplasy.

3. Problems with amplification[7]

Finding a useful DNA marker locus requires identifying a region of the genome with a sufficiently high mutation rate that multiple versions (alleles) exist in a given population, and which is also located adjacent to a low mutation rate stretch of DNA that will bind PCR primers in the vast majority (approaching 100%) of individuals of the species. If mutations occur in the primer region, some individuals will have only one allele amplified, or will fail to amplify at all (Paetkau & Strobeck 1995). In addition, primers must bind under repeatable PCR conditions so that genotyping can be performed in serial, by different workers, and by different laboratories.
Consistent amplification across all samples can only be assured by trial and error, such that at the middle or end of genotyping all the samples in a study, some loci will have to be discarded because of amplification problems. If this marker attrition is planned for in the initial isolation of microsatellite markers, the chance that amplification problems will ruin a study is minimal. However, several taxa seem more often beset by amplification problems than others, notably, bivalves, corals and some other invertebrate taxa (e.g. Hedgecock et al. 2004). A low rate of null alleles Microsatellites for ecologists can have a negligible impact on many types of analysis, although for some types of parentage analyses it can be substantial (Dakin & Avise 2004)[3].

Applications of Microsatellites[8]

Since microsatellites are widely dispersed in eukaryotic genomes, are highly variable, and are PCR based (requiring only minute amounts of starting template) they have been used in many different areas of research such as:

1) Forensics[8]

Microsatellite loci, generally known in forensic applications as Short Tandem Repeat (STR) loci, are widely used for forensic identification and relatedness testing, and are a predominant genetic marker in this area of application. They have also become the primary marker for DNA testing in forensics (court) contexts -- both for human and wildlife cases (e.g. Evett and Weir, 1998). The reason for this prevalence as a forensic marker is their high specificity. Match identities for microsatellite profiles can be very high and the probability that the evidence from the crime scene is not a match with that of the suspect is less than one in many millions in some cases. In forensic identification cases, the goal is typically to link a suspect with a sample of blood, semen or hair taken from a crime. 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) Population Studies [8,9]

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.

3) Conservation Biology [8,9]

Microsatellites can be used to detect sudden changes in population, effects of population fragmentation and interaction of different populations. Microsatellites are useful in identification of new and incipient populations. It can be useful in identifying the adaptation of an individual to the ever-changing environment and also its survival for the years in the same environment.

4) In Mapping genomes [8,9]

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 hum an genome.

5) In a Biological/Evolutionary context [8,9]

They can also be used to address questions concerning degree of relatedness of individuals or groups. For captive or endangered species, microsatellites can serve as tools to evaluate inbreeding levels. From there we can move up to the genetic structure of subpopulations and populations. They can be used to assess demographic history (e.g., to look for evidence of population bottlenecks), to assess effective population size and to assess the magnitude and directionality of gene flow between populations. Microsatellites provide data suitable for phylogeographic studies that seek to explain the concordant biogeographic and genetic histories of the floras and faunas of large-scale regions. They are also useful for fine-scale phylogenies upto the level of closely related species.

6) Diagnosis and Identification of Human Diseases[10]

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.Because microsatellites change in length early in the development of some cancers, they are useful markers for early cancer detection. Because they are polymorphic they are useful in linkage studies which attempt to locate genes responsible for various genetic disorders.

For example, in one part of chromosome number 4, CAG nucleotides are repeated many times over. They look like this CAGCAGCAGCAG....If the tri-nucleotides are

repeated too many times this would cause the person to get Huntington's disease in adult life. Other diseases that involve repeats of three nucleotides are also known to cause neurological diseases. At this time, 14 neurological disorders have been shown to result from the expansion of tri-nucleotide repeats, establishing an expanding class of diseases. Tri-nucleotide repeat diseases can be categorised into two subclasses based on the location of the trinucleotide repeats: diseases involving noncoding repeats (untranslated sequences) and diseases involving coding sequences (exonic). In general tri-nucleotide repeat disorders are either dominantly inherited or X-linked, the one exception being Friedrich's ataxia, which is Autosomal recessive (Goldstein and Schlotterer, 1999; Cummings and Zoghbi, 2000). Not all diseases are caused by a mistake in one gene. Sometimes many genes may be involved in a disease, for example, in schizophrenia. For these diseases microsatellite sequences have been used as a marker for locating the diseased region of the chromosome. This method is called positional cloning. Microsatellite markers close to the disease gene correlate with the heredity of the disease, and by analysis of these markers within families scientists can predict how the disease will be inherited (Risch, 2000).

7) Detecting Cancer[10]

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 (Yonekura et al., 2002; Moxon and Willis, 1999).

Examples of Some Diseases Involving Microsatellites Trinucleotide Repeats[11]

1) Huntington's disease

Symptoms: Late onset dementia and loss of motor control, resulting in full-blown chorea after 10-20 years. Motor disorder is often preceded or accompanied by memory deficits, cognitive decline or changes in personality. Juvenile onset is rare and patients show rigidity, bradykinesia, epilepsy, severe dementia and an accelerated disease course.

Involvement of microsatellites: CAG coding repeat in the first exon of the HD gene. Normal gene contains between 6 and 35 repeats and the affected gene from 36 to 121 repeats. Adult onset typically occurs when the repeat contains 40 - 50 units, whereas alleles containing more than 70 repeats typically result in the more severe juvenile form. The microsatellite adds a string of glutamine amino acids to the huntington protein.
Chromosome location: #4

2) Fragile X

Symptoms: Mental retardation, long and prominent ears and jaws, high-pitched speech, hyperactivity, poor eye contact, and stereotypic hand movements (e.g. hand-flapping and hand-biting). 1 in 4000 males are affected, and fewer females are affected, depending on the ratio of cells with normal X chromosome active to abnormal X active.
Involvement of microsatellites: The repeat is CGG in a non-coding region of the FMR2 gene, and normal is 6 - 53 repeats. The disease occurs if the repeat is between 60 - 200.
Chromosome location: X

3) Myotonic dystrophy

Symptoms: Congenital DM is the most severe form of this disease involving hypotonia, respiratory distress at birth and developmental abnormalities. Adult onset includes variable loss of mental function, myotonia, muscle weakness and progressive muscle wasting. Other features may include facial dysmorphology, presenile cataracts, testicular atrophy, premature balding in males, kidney failure, hyperinsulin secretion and cardiac conduction abnormalities.
Involvement of microsatellites: CTG repeats in a non-coding region of the DMPK gene. Normal is between 5 and 37, the disease may involve from 50 - 1000s of repeats.
Chromosome location: #19

4) Spinalbulbar muscular atrophy

Symptoms: Neurological degeneration leading to difficulties in speech, articulation and swallowing, muscle weakness and atrophy. Signs of mild androgen insensitivity are typically seen at adolescence.
Involvement of microsatellites: CAG repeat in first coding exon of the androgen receptor (AR) gene. Between 9 and 36 repeats is normal, and people with 38-62 repeats develop the disease.
Chromosome location: X, recessive.

5) Friedrich's ataxia
Symptoms: Ataxia, diminished tendon reflexes, loss of position and vibratory senses, dysarthria (slurred speech), cardiomyopathy, diabetes mellitus, optical atrophy, scoliosis and skeletal abnormalities. Age of onset is typically early childhood.
Involvement of microsatellites: GAA repeat in non-coding region of gene X25. Normal gene contains between 7 and 34 repeats. The disease gene has 34 to 80 repeats.
Chromosome location: #9, recessive

Examples of Diseases found by positional cloning

1) Schizophrenia and Bipolar Disorder
Symptoms: Schizophrenia is characterized by auditory and visual hallucinations and delusions; symptoms of bipolar disorder involve severe mood swings, between mania and depression.
Involvement of microsatellites: Using 388 microsatellite markers within 8 families the researchers (Bailer et al., 2002) found a susceptibility locus for both schizophrenia and bipolar disorder on chromosome 3.
Chromosome location: #3

2) Congenital generalized hyper-trichosis
Symptoms: Rare hair growth disorder, characterized by excessive hair growth on the face and upper body; sometimes called "werewolf" syndrome.
Involvement of microsatellites: Using hypervariable microsatellite markers the researchers (Figure et al., 1995) localized the gene to a part of the X chromosome.
Chromosome location: X, dominant.

3) Asthma and Bronchial Hyper-responsiveness
Symptoms: Common respiratory disorder characterized by recurrent episodes of coughing, wheezing and breathlessness.
Involvement of microsatellites: A putative asthma susceptibility gene was identified on the ADAM33 region of chromosome 20 (Van Eerdewgh et al., 2002).
Chromosome location: #2 .

Future of microsatellites
The evolutionary process of simple repeats is far fromsimple. One important implication of the complexity of microsatellite evolution is, therefore, that care needs tobe taken when using microsatellite data in populationgenetics studies. For instance, significant mutation-rateheterogeneity among loci means that it might be difficult to translate estimates of genetic distance into absolutetimescales. Similarly, directional biases in the mutationprocess have important consequences for the interpretation of differences in allele size distributions amongspecies, particularly if the character of the bias differsamong species[12]. Future mathematical models of microsatellite evolution should therefore aim to incorporate as many of the different forms of mutational heterogeneity as possible. Those who use microsatellites inpopulation genetics studies should select only the markers that are well characterized in terms of mutational properties (mutation rates, directionality,whether all alleles of equal length are identical in sequence), and, preferably, use markers that show uniformrates and patterns.Alternatively,but in many species lessrealistically, the use of many markers might compensate for heterogeneity in mutational properties among loci. Microsatellites continue to find their application inareas such as linkage mapping,paternity testing,forensics and for the inference of demographic processes. More recently,they have found most use in linkage-disequilibrium mapping studies, in which associations between markers and trait loci are searched for in population samples[13], and in hitchhiking mapping, in which genome wide screens for regions that show signs of selection are made [14].But there are also prospects for new applications. Given their high mutation rate, microsatellites offer a realistic means to study how the overall genomic mutation rate is affected by environmental factors (genetic toxicology).Elevated rates of microsatellite mutations in thegermline have been seen in animals and plants that are exposed to ionizing radiation[15, 16], and similar observations have been made for minisatellites in humans [17]. Estimating microsatellite mutation rates in samples that are exposed to different forms of radiation or toxic compounds could,when properly set in relation to data from control groups,help to make risk assessments.

CONCLUSION
Microsatellites have revolutionized the approach of molecular genetics and have proven to be an important tool in analysis and mapping of the genome, in diagnosis and treatment of diseases and disorders of genetic origin and in several intensive phylogenetic studies. It can also be used as a powerful tool in tracing back the evolutionary history. They can provide a very comprehensive data suitable for phylogeographicstudies that explain the concordant bio-geographic and genetic histories of an individual. A dedicated and continuous research in this field is needed

for the purpose which may prove to be beneficial to the mankind.

REFERENCES

1. Textbook of biotechnology, R.C Dubey, Chap 14, Pg 337-354.
2. Biotechnology, By U.Satyanarayana, Chap 53, Pg. 648-653.
3. Kimberly A, Sloe and Robert J. Tone; Microsatellites for ecologists; Ecology Letters. 2006;9:615-629.
4. Mittal N and Dubey AK. Microsatellite markers- A new practice of DNA based markers in molecular genetics. Pogo Rev. 2009;3:235-46.
5. http://www.woodrow.org/teachers/esi/2002/biology/projects/p3/evolution.htm
6. Zane L, Bargelloni L and Patarnello T, Strategies for microsatellites isolation : a review, Molecular Ecology.2002;11:1-16
7. Kimberly A, Selkoe and Robert J. ToonenMicrosatellites for ecologists: a practical guide to using and evaluating microsatellite markers, Ecology Letters. 2006;9:615-629.
8. http://www.woodrow.org/teachers/esi/2002/biology/projects/p3/definition.htm#APPLICATIONS.
9. Moxon ER and Wills C. "DNA microsatellites: Agents of Evolution?" *Scientific American*. Jan., 1999;72-77.
10. 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.
11. http://www.woodrow.org/teachers/esi/2002/biology/projects/p3/diseases.htm.
12. Amos W, Hutter CM, Schug MD and Aquadro CF. Directional evolution of size coupled with ascertainment biasfor variation in Drosophila microsatellites. Mol. Biol. Evol. 2003;20:660-662.
13. Ohashi J and Tokunaga K. Power of genome-wide linkage disequilibrium testing by using microsatellite markers .J. Hum Genet. 2003;48:487-491.
14. Schlotterer C. Hitchhiking mapping — functional genomics from the population genetics perspective. Trends Genet. 2003;19:32-38.
15. Ellegren H, Lindgren G, Primmer CR and Moller AP. Fitness loss and germline mutations in barn swallows breeding in Chernobyl. Nature. 1997;389:593-596.
16. Kovalchuk O, Kovalchuk I, Arkhipov A, Hohn B and Dubrova Y E. Extremely complex pattern of microsatellite mutation in the germline of wheat exposed to the post-Chernobyl radioactive contamination. Mutat. Res. 2003;525:93-101.
17. Dubrova YE. Human minisatellite mutation rate after the Chernobyl accident. Nature 199;380:683-686.