6

Microsatellites as Tools for Genetic Diversity Analysis

Andrea Akemi Hoshino[1], Juliana Pereira Bravo[2],
Paula Macedo Nobile[1] and Karina Alessandra Morelli[3]
[1]*Universidade Estadual de Campinas, SP,*
[2]*Universidade Estadual Paulista, SP,*
[3]*Fundação Oswaldo Cruz, RJ,*
Brazil

1. Introduction

Powerful tools for the analysis of genetic biodiversity are molecular markers, which are based on DNA sequence polymorphisms. Indeed, DNA sequences determine the diversity of organisms, and therefore, the techniques used to evaluate DNA polymorphisms directly measure the genetic diversity. Because molecular markers show Mendelian inheritance, it is possible to trace the fingerprint of each organism and determine the evolutionary history of the species by phylogenetic analysis, studies of genetic relationship, population genetic structures and genetic mapping.

According to technical principles, there are three classes of molecular markers: (i) nucleic acid hybridization based on complementary bases, e.g., restriction fragment length polymorphisms (RFLPs), (ii) Polymerase Chain Reaction (PCR) based on DNA amplification, e.g., random amplification of polymorphic DNAs (RAPD), amplified fragment length polymorphisms (AFLP), microsatellites or simple sequence repeats (SSRs) and (iii), single nucleotide polymorphisms (SNPs). The first technique, RFLP, has been decreasingly used due to the difficulties involved in manipulating high throughput sampling and the third technique, SNPs, represents high costs related to large-scale genotyping. However, the cost-effective PCR-based techniques have been largely used.

With the advent of PCR technology in the mid 1980s (Mullis & Faloona, 1987; Saiki et al., 1985), new perspectives have evolved for molecular biology fields that have largely impacted several applied purposes, e.g., diagnostics, plant and animal breeding programs, forensics and others. Microsatellites were detected in eukaryote genomes almost thirty years ago and they are the most promising PCR-based markers. Microsatellites are tandemly repeated motifs of variable lengths that are distributed throughout the eukaryotic nuclear genome in both coding and noncoding regions (Jarne & Lagoda, 1996). They also appear in prokaryotic and eukaryotic organellar genomes, e.g., chloroplast (Powell et al., 1995) and mitochondria (Soranzo et al., 1999).

Due to the high mutation rate of microsatellites, they are potentially the most informative molecular marker with the advantage of easy and low-cost detection by PCR. Moreover, the

bottleneck caused by the high cost and laborious approaches to isolate microsatellite loci has been overcome by new sequencing technologies. Large databases of genomic and EST sequences, that could be screened using bioinformatics tools, are now available and many published loci could be transferred from related species. Another great advantage of microsatellites is their co-dominant feature. Unlike RAPD and AFLP, which are dominant markers that detect only the presence or absence of a locus, microsatellite markers detect both homozygote and heterozygote genotypes.

A search using the Web of Science facility (Thomson Reuters, http://webofknowledge.com) confirmed that the microsatellite has been the most used molecular marker to address genetic diversity (Table 1). This marker has been applied for the germplasm conservation, phylogenetic analyses, plant and animal breeding programs, constructing linkage maps, mapping economically important quantitative traits and identifying genes responsible for desired traits.

Science Category	Molecular marker				
	Microsatellite or SSR	SNP	RAPD	AFLP	RFLP
Biochemistry Molecular Biology	1178	185	83	96	38
Evolutionary Biology	989	33	18	78	11
Ecology	989	23	17	49	16
Genetics Heredity	1134	493	124	131	32
Biodiversity Conservation	405	4	3	8	3
Total	4690	1.269	925	668	531

Table 1. A search using the Web of Science facility for the number of articles published in the last five years whose title contains the name of one of the markers: microsatellite or SSR, SNP, RAPD, AFLP or RFLP.

2. Identification and features

Microsatellites are DNA sequences of mono-, di-, tri-, tetra- and pentanucleotide units repeated *in tandem*, which are widely distributed in the genome (Powell et al., 1996). Litt & Luty first used the term "microsatellites" in 1989 when analyzing the abundance and dispersion of $(TG)_n$ in the cardiac actin gene. Microsatellites were originally designed to research degenerative and neurology diseases in humans but showed great applicability for other species.

Many authors classified the markers according to the number of bases, i.e., short repeats (10-30 bases) are microsatellites and longer repeats are minisatellites (between 10-100 bases). Microsatellites have been also been classified according to the type of repeated sequence presented: (i) perfect, when showing only perfect repetitions, e.g., $(AT)_{20}$, (ii) imperfect repeats, when the repeated sequence is interrupted by different nucleotides that are not repeated, e.g., $(AT)_{12}GC(AT)_8$, and (iii) composite, when there are two or more different motifs *in tandem*, e.g., $(AT)_7(GC)_6$. The composite repeats can be perfect or imperfect. The sequences of di-, tri- and tetranucleotide repeats are the most common choices for molecular genetic studies (Selkoe & Toonen, 2006).

In addition to their co-dominant feature, i.e., the identification of all alleles of a given locus, microsatellites can also be amplified using polymerase chain reaction (PCR) in stringent

conditions that usually only permit the amplification of single loci, thus facilitating data integration (Bravo et al., 2006). Furthermore, microsatellites are widely distributed throughout the genome, highly polymorphic and transferable between species. These features provide the foundation for their successful application in a wide range of fundamental and applicable fields (Chistiakov et al., 2006).

The presence of SSRs in eukaryotes was verified from diverse genome regions, including 3´-UTRs, 5´-UTRs, exons and introns (Rajendrakumar et al., 2007). Furthermore, their localization could potentially interfere with different aspects of DNA structure, DNA recombination, DNA replication and gene expression as illustrated by Chistiakov et al. (2006). The transposable elements might contain one or more sites that are predisposed to microsatellite formation and enables SSRs dispersion throughout the genome (Bhargava & Fuentes, 2010). Microsatellites are also commonly located in proximity of interspersed repetitive elements, such as short interspersed repeats (SINEs) and long interspersed repeats (LINEs). Kashi et al. (1997) reported that 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. Powell et al. (1995) provided experimental evidence of length variation in the mononucleotide repeats of the chloroplast genome of angiosperms, and polymorphisms within these regions might be used to study both intraspecific and interspecific variability. Soranzo et al. (1999) was the first to show length variation at a mitochondrial SSR locus in conifers.

Knowledge of the complete genome sequence of many species in the public domain now permits the determination of SSR frequencies at the whole genome level, decreases the economic limitations and accelerates the process of SSR analysis. The accessibility and data analysis of microsatellite content in whole genome sequences would also facilitate comprehensive studies on the direct role of microsatellites in genome organization, recombination, gene regulation, quantitative genetic variation and the evolution of genes (Katti et al., 2001). The density analyses of SSRs in fully sequenced eukaryotic genomes showed a higher density in mammals and the initial analysis of the human genome sequence concluded that approximately 3% of all DNA is represented by SSRs. The human genome is estimated to contain on an average 10-fold more microsatellites than plant genomes (Powell et al., 1996). The analyses of microsatellite distribution in the genomes of many species revealed that compared with *Drosophila*, *Arabidopsis*, *Caenorhabditis elegans* and yeast, human chromosomes 21 and 22 are rich in mono- and tetranucleotide repeats. *Drosophila* chromosomes have higher frequencies of di- and trinucleotide repeats and, surprisingly, the *C. elegans* genome contains less SSRs per million base pairs of sequence than the yeast genome (Katti et al., 2001).

3. Isolation and analysis

3.1 Isolation

Since the first studies using microsatellites were performed, the methods of SSR loci isolation have been improved and several protocols were published. There are published reviews concerning this topic (Weising et al., 2005; Zane et al., 2002), but with the recent development of technology and evolution of methodology, new methods and modifications have been proposed.

The published microsatellite isolation protocols can be grouped into three types: (i) the standard method, where a library is screened for repeated sequences; (ii) the automated method, where the SSR sequences are searched in sequence databases and (iii) the sequencing method, where the whole genome or parts of the genome are sequenced using high-throughput technologies. Each of these methods was modified and optimized to many species and conditions, generating a large number of protocols. Here, we will present an overview of the commonly used protocols.

3.1.1 Standard method

This method requires the creation of a library. There are various protocols to create and screen a genomic, cDNA or PCR fragment library [revised by Mittal & Dubey (2009) and Weising et al. (2005)], but the main steps can be summarized as follows:

1. The DNA is fragmented by sonication or enzymatic digestion.
2. The DNA fragments are ligated into a vector and transformed into *Escherichia coli*.
3. The clones are analyzed for the presence of SSR sequences by Southern blot. Then, the positive clones are sequenced.

The number of positive clones obtained by this methodology ranges from 0.04 to 12%, with the lowest yields occurring in birds (Zane et al., 2002). These protocols are efficient; however, 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. Ito et al. (1992) proposed the use of a biotinylated oligonucleotide to screen the plasmids of a restriction fragment library. The oligonucleotide and plasmid interact to form a triple helix, and the positive clones could be recovered using streptavidin coated magnetic beads. Subsequently, the microsatellite-enriched plasmids are purified and transformed into *E. coli*. However, this technique is limited to sequence motifs that are capable of triple helix formation (such as GA- and GAA-repeats).

Another technique to increase the number of positive clones or enrich the libraries relies on the extension of the library of single-stranded genomic DNA using repeat specific primers. For example, Paetkau (1999) amplified genomic libraries using biotinylated oligonucleotides, which were complementary to the microsatellite sequence, as primers. The single-stranded biotinylated sequences were recovered with streptavidin bound to magnetic particles, made double-stranded and transformed into *E. coli*. In this case, the enrichment efficiency was 100% for the dinucleotide $(CA)_{18}$. However, the enrichment efficiency depends on the size of the genomic library.

The most popular enrichment methods for SSR sequences are based on hybridization selection (Weising et al., 2005). Therefore, the following steps are added after DNA fragmentation:

1. The DNA fragments are ligated to adapters and amplified by PCR.
2. The PCR products are hybridized to microsatellite sequences that are attached to nylon membranes or biotin, and the hybrid sequences are eluted from the membrane or recovered via streptavidin-coated magnetic beads.
3. The selected PCR products are ligated into a vector and transformed into *E. coli*.

Researchers using hybridization selection have reported up to 80% of clones containing a microsatellite. Using two rounds of amplification and hybridization with biotin/streptavidin, Kandpal et al. (1994) generated a high enrichment efficiency of approximately 90% for CA repeats.

Yue et al. (2009) described another method to enrich microsatellite libraries. These authors applied a duplex-specific nuclease to normalize a pool of cDNA prior to cloning and generated 30 times more positive clones as compared with direct sequencing methods. Recently, Santana et al. (2009) and Malausa et al. (2011) applied pyrosequencing to enriched DNA libraries of many species and demonstrated that this methodology is more rapid, effective and economical than others.

3.1.2 Automated method

Microsatellite identification and development is also 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 (Altschul et al., 1990). Subsequently, several computer-based software programs were developed and the SSR search became easier. Mittal & Dubey (2009) reported a list of programs, their applications and references. Because microsatellites located in expressed sequences are more conserved and gene related, many studies have described and applied EST-SSRs [as reviewed by Varshney et al. (2005) for plants].

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

3.1.3 Sequencing method

The new high-throughput sequencing technologies have allowed whole or expressed genome sequencing (Abdelkrim et al., 2009; Mikheyev et al., 2010). These technologies do not require the creation of libraries (total DNA or RNA can be sequenced), produce a huge amount of sequences quickly and because many steps have been skipped, have lower costs than other methods.

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. Arthofer et al. (2011) reviewed published research concerning microsatellite isolation and showed that approximately half of all loci were lost due to inconsistent PCR amplification, multicopy status in the genome or monomorphism, regardless of the isolation strategy used. Moreover, these authors demonstrated the applicability of high-resolution melting (HRM) analyses to screen candidate loci for marker development, reducing the costs of traditional tests.

3.2 Analyses

In microsatellite loci analyses, variations in the amplification product size are related to the number of repeated motifs and would indicate the polymorphism level of that specific locus

in a population. There are many protocols to amplify and detect microsatellite loci variation. Weising et al. (2005) described the most frequently used methods.

The protocol choice depends on the availability of equipment and reagents and the desired accuracy of the polymorphism detection. Agarose gels stained with ethidium bromide are easy to handle and are one of the cheapest protocols but do not allow precise fragment size determination. However, one of the most accurate methods requires an automated sequencer and fluorescent-labeled primers. The combined use of multiplex reactions (with primers labeled with different fluorochromes) with capillary DNA sequencers allow high-precision genotyping and high-throughput.

Regardless of the electrophoretic technique chosen to determine the banding pattern of the amplified fragments, the next step is statistical analysis. Molecular markers with known band sizes are usually added to electrophoresis gels to estimate the fragment size.

There are several methods and computer programs that can be used in data analysis, depending on the final application. Excoffier & Heckel (2006), Labate (2000) and Weising et al. (2005) reviewed many of them and summarized their main applications. Several statistical analyses are based on genetic distances, and as a first step, the pairwise similarity is quantified. Most commonly, the similarity index is calculated from band sharing data and the complement to this index is the genetic distance between the samples (Weising et al., 2005). When large number of samples are involved, it is difficult to interpret genetic distances. In these instances, the use of ordination, clustering and dendrograms condenses the differences into fewer characters and permits the visualization of these entries in a multidimensional space (Weising et al., 2005).

Unfortunately, most of the computer programs use a specific data file format, but there are several that can read or write data from, or to, other file formats. It is essential to avoid the limitation of a single program or having to reformat the data manually. Excoffier & Heckel (2006) identified two conversion programs considered as starting points for formatting input data files: Convert (Glaubitz, 2004) and Formatomatic (Manoukis, 2007). These programs can create input files for several other formats.

A critical point in data analyses is that most computer programs conceal the mathematical complexities from the user, but they rely on crucial assumptions that should be taken in account for the correct interpretation of the results (Excoffier & Heckel, 2006).

4. Transferability

Microsatellites are transferable because their flanking regions are highly conserved across taxa, allowing cross-species amplification, i.e., primers developed in one species can be used in others of the same genus or family, especially for vertebrates, such as fishes, reptiles and mammals (Peakall et al., 1998; Rico et al., 1996). The transferability of SSRs derived from EST databases (EST-SSR) is greater than that of SSRs derived from enriched genomic DNA libraries. The EST-SSRs originate from expressed regions, and therefore, they are more conserved across a number of related species than non-coding regions (Varshney et al., 2005).

Many researchers have studied the transferability of SSRs. Zhao et al. (2011) showed the high transferability (86%) of *Brachypodium* SSR markers to *Miscanthus sinensis.* Moreover, 18

(31%) of the transferable markers produced perfect polymorphic and easy-scoring bands; consequently, this study confirms the significance of *Brachypodium* as a model plant for *Miscanthus*. Faria et al. (2010) used *Eucalyptus* EST databases to develop, select and conduct a detailed characterization of a novel set of 20 microsatellite markers that are polymorphic and transferable across 6 of the major *Eucalyptus* plant species. The primers were developed from more conserved transcribed regions; therefore, the transferability and polymorphism of these microsatellites likely extended to the other 300 or more species within the same subgenus *Symphyomyrtus*, further highlighting their applied value for *Eucalyptus* genetics and breeding. Pépin et al. (1995) showed that an estimated 40 per cent of the microsatellites isolated from cattle were useful to study the caprine genome and characterize economically important genetic loci in this species. Moreover, bovine microsatellites were shown to be useful tools for the study of the genetic diversity of Artiodactyla. Dawson et al. (2010) developed primer sets for 33 polymorphic loci that are highly useful in the study of passerine, shorebirds and other non-passerine birds and for genotyping in species belonging to the Passeridae and Fringillidae families.

5. Evolution and mutation models

Microsatellites have a wide variety of applications in life sciences. In addition, these markers are related to several human neurodegenerative diseases and have demonstrated roles in regulating transcription and expression of various genes. Despite the great interest in the functions of these sequences and their applicability as molecular markers, knowledge about the mutational and evolutionary dynamics of microsatellites is still controversial.

The methods used for studying evolution of microsatellites involve pedigree analysis, sequence structure analysis of the alleles within species, sequence comparison of orthologous loci in related species and analysis of microsatellite instability through cloning and maintenance of sequence *in vivo* (Ellegren, 2004). More recently, data from complete genomes coupled with bioinformatics analysis has helped researchers to understand the distribution and variability of microsatellites in genomes.

5.1 Origin of microsatellite

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 (Bhargava & Fuentes, 2010). Currently, there are two non-mutually exclusive hypotheses to explain the origin of microsatellites:

- *De novo* microsatellites (Messier et al., 1996) - 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. Proto-microsatellites were originated from base substitutions or indel events; the latter is supported by the observation that insertions tend to copy adjacent bases (Buschiazzo & Gemmell, 2006). Once a 'proto-microsatellite' is initiated, maintenance and multiplication is favored by its propensity to undergo strand slippage during replication and, depending primarily on the repeat motif, its capacity to form unusual DNA conformations and participate in recombination and transposition events. The number of repeat units correlates

positively with the mutability of the microsatellite, but the minimum repeat number required for strand slippage or other mechanisms of microsatellite mutation is debatable (Jentzsch et al., 2008).

- Adopted microsatellites (Wilder & Hollocher, 2001) – 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. Transposable elements can be divided into two main classes based on their mechanisms of movement: class I (retrovirus-like transposons) and class II (so called cut and paste transposons). Both of these elements can leave traces of their presence and movement during the transposition process across DNA sequences, which resemble microsatellites, especially poly A arrays. A poly A tail is added to the 3' end of class I retrotransposons after mRNA transcription, which then gets inserted together with the transposed sequence into the new position. Retrotransposons can also contain other microsatellite-like stretches, dinucleotide and tetranucleotide repeats, within their sequences. Class II transposons preferentially insert into certain DNA sequences, which can be either inverted repeats or tandem repeat sequences. This suggests a reciprocal association in which microsatellites act as 'retroposition navigator sequences,' while retrotransposons generate more microsatellites during their dispersion throughout the genome. An example of a retrotransposon-mediated microsatellite genesis in humans is the origin of A/T rich microsatellites with motifs ranging from one to six nucleotides in length from Alu elements (Jentzsch et al., 2008).

5.2 Evolutionary dynamics

Microsatellites are highly mutable as compared with point mutations in coding genes and mutation rates range from 10^{-6} to 10^{-2} events per locus per generation. These rates are highly affected by multiple factors, which influence both the probability of mutations generation and the repair efficiency of these mutations. Mutation mechanisms, DNA repair, structure and characteristics of microsatellite, genomic and individual context and selective biological influences are factors that interact and control the evolutionary dynamics of microsatellites.

5.2.1 Mutation mechanisms and DNA repair

Currently, two mechanisms have been proposed as mutation models in microsatellites: (i) replication slippage and (ii) unequal crossing over during meiosis. The mechanism of DNA replication slippage is most widely observed in microsatellites.

Replication slippage - DNA slippage is a symmetrical process, where the same number of repeats are added and removed. This process inevitably leads to either the loss of microsatellites or the insertion of a high number of repeats (Schlötterer, 2000). The misalignment that gives rise to mutations occurs between a newly synthesized DNA strand and its complementary template strand. The two strands dissociate and reanneal 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 (Jentzsch et al., 2008).

Unequal crossing over during meiosis (recombination) - 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 (Grover & Sharma, 2011).

The relative rates of point mutations and slippage might be altered by changes in the efficiency of MMR (mismatch repair) and proofreading. Failure of the MMR system during replication results in a 10^{-3}-fold increase in microsatellite instability (Strand et al., 1993). These proteins govern the balance between enrichment and prevention of microsatellites within genomes. In a given species, MMR proteins play a role in the mutational variability among alleles, loci and individuals, and because they are driven by selective forces, are certainly the cause of differential allele distributions between species. Moreover, the proteins involved in MMR can vary in number and nature among eukaryotes, suggesting variability in their intrinsic efficiency (Buschiazzo & Gemmell, 2006).

5.2.2 Structure characteristics of microsatellite, genomics and biological contexts

Mutation rates might vary greatly among loci and alleles of the same locus depending of the structure of the microsatellite (length and number of repeat units, interruptions within the motif, motif nucleotide composition, motif length and flanking sequences).

Length of microsatellite - mutation increases with the increasing number of repeat units and this is presumable because the more repeating units, the more opportunities for replication slippage to occur. Therefore, loci with a large number of repeats are more polymorphic (Ellegren, 2004).

Interruptions - point mutations and other interruptions within the repeat reduce the mutation rate. Any mutation within the repeated region that causes an interruption will split the original repeat into two shorter units, which would increase locus stability by reducing the substrate for polymer slippage (Bhargava & Fuentes, 2010).

Motif nucleotide composition - repeats with certain motifs have a heightened propensity to form secondary structures and alter DNA structure. Secondary structures, such as hairpins, quadruplex structures, H-DNA or sticky DNA are intermediate DNA hybrid forms that increase the likelihood of strand misalignment and subsequent polymerase slippage. A conformational change in the DNA structure, such as Z-DNA formed by long AC tracts, will affect both polymerases and repair enzymes (Jentzsch et al., 2008).

Motif length - dinucleotide repeats have the highest mutation rate, followed by tri- and tetranucleotide repeats. Shorter microsatellites motifs allow more opportunities for misalignment, whereas motifs longer than three nucleotides require higher dissociation energy and are thus less likely to generate enough single-stranded DNA to form a stable loop. Moreover, motif length can affect MMR efficiency. If the loop is too large, the efficiency of MMR is reduced (Buschiazzo & Gemmell, 2006; Jentzsch et al., 2008).

Flanking sequence - the mutability of microsatellites greatly depends on the genomic constitution of their flanking sequences. Large scale mutation of a sequence that contains or flanks a microsatellite will modify the genomic context of the microsatellite and may change the mutability of the locus (Bhargava & Fuentes, 2010).

The influence of genomic context on the mutation rate of microsatellites becomes clear when the effect of the mutations has a high probability of being disadvantageous and is strongly

counteracted by selection, i.e., where the distribution of microsatellites in coding regions is observed (Buschiazzo & Gemmell, 2006; Jentzsch et al., 2008).

The mode of reproduction, metabolic rate and generation time could also influence the mutational dynamics of microsatellites at the species level. It has been reported that the sex and age of some organisms could also influence the mutation rate of microsatellites. For example, men have more cell divisions than women for the production of gametes and it is therefore expected that the microsatellites would undergo more mutations per generation (Buschiazzo & Gemmell, 2006)

5.3 Mutation models

A mutation model of microsatellites evolution is needed for the estimation of population parameters, such as number of migrants, population structure and effective population size. A wide range of models has been proposed to explain the mutational dynamics of microsatellites and some of them are discussed below. For more details see Balloux & Lugon-Molin (2002) and Bhargava & Fuentes (2010).

Infinite Allele Model (IAM) – This model was first described by Kimura & Crow (1964) and assumes that every mutation results in the creation of a new allele. This model does not allow for homoplasy; identical alleles share the same ancestry and are identical-by-descent, which best describes the unusual dynamics of compound/complex microsatellites.

K-Allele model (KAM) – This model was developed by Crow & Kimura (1970) and assumes that there are K possible allelic states and any allele has a constant probability of mutating towards any of the other K-1 allelic states. This model treats all alleles as equivalents with the potential to mutate from one allele to any other allele and allows homoplasy, which is more suitable for data where the mutation pattern is unknown.

Stepwise mutation model (SMM) – This model was developed by Otha & Kimura (1973) and assumes that each mutation creates a novel allele by either adding or deleting a single microsatellite repeated unit. Alleles of very different sizes will be more distantly related than alleles of similar sizes (memory of allele size); this model is often used when estimating relatedness between individuals and population sub-structuring, except when homoplasy is present.

Two-phase-model (TPM) – This model was developed by Di Rienzo et al. (1994) and is an extension of the SMM, which allows for infrequent multistep mutations; one-step mutations are more likely to occur and follow the SMM, whereas the magnitude of multistep mutations follows a truncated geometric distribution.

6. Applications

Due to all of the previously discussed features, microsatellites have been a class of molecular markers chosen for diverse applications. In this review, the SSR applications will be summarized into four categories: (i) genome mapping and marker-assisted selection, (ii) genetic diversity and individual identification, (iii) population and phylogenetic relationships and (iv) bioinvasion and epidemiology.

6.1 Genome mapping and marker-assisted selection

Genome mapping includes genetic, comparative, physical and association mapping. Genetic mapping is one of the major research fields in which microsatellites have been applied because they are highly polymorphic and require a small amount of DNA for each test. Linkage maps are known as recombination maps and define the order and distance of loci along a chromosome on the basis of inheritance in families or mapping populations (Chistiakov et al., 2006). Association mapping links a locus to a phenotypic trait and comparative mapping aligns chromosome fragments of related species based on genetic mapping to trace the history of chromosome rearrangements during the evolution of a species (Wang et al., 2009a). However, in physical mapping, markers anchor large pieces of DNA fragments, such as bacterial artificial clones (BACs), and provide the actual physical distance between the markers (Wang et al., 2009a). Apotikar et al. (2011) constructed a SSR-based skeleton linkage map of two linkage groups of sorghum in a population of 135 recombinant inbred lines derived from a cross between IS18551 (resistant to shoot fly) and 296B (susceptible to shoot fly) varieties. The authors found 14 markers that were mapped to each linkage group and three quantitative trait loci (QTL) governing more than one trait (pleiotropic QTLs). The identification of genomic regions/QTLs that influence resistance can help breeders to introgress them into the breeding lines using the linked molecular markers. Baranski et al. (2010) analyzed the flesh color and growth related traits in salmonids with 128 informative microsatellite loci, distributed across all 29 linkage groups, in individuals from four F2 families. Chromosomes 26 and 4 presented the strongest evidence for significant QTLs that affect flesh color, while chromosomes 10, 5 and 4 presented the strongest evidence for significant QTLs that affect growth traits (length and weight). These potential QTLs provide a starting point for further characterization of the genetic components underlying flesh color and growth in salmonids and are strong candidates for marker-assisted selection.

The use of the markers to indicate the presence of a gene (trait) is the basis for marker-assisted selection (MAS). Therefore, the construction of high-density and high-resolution genetic maps is necessary to select for markers that are tightly linked to the target locus (gene) (Chistiakov et al., 2006). Once a linkage is established between a locus and the gene of interest, the inheritance of the gene can be traced, which could greatly enhance the efficiency of breeding programs (Wang et al., 2009a).

6.2 Genetic diversity and individual identification

Genetic diversity refers to any variation in nucleotides, genes, chromosomes or whole genomes of organisms (Wang et al., 2009a). Genetic diversity can be assessed among different accessions/individuals within same species (intraspecific), among species (interspecific) and between genus and families (Mittal & Dubey, 2009).

Even crops with advanced studies in genomics (e.g., rice, corn, soybean and apple) have been recently evaluated by SSRs to access the genetic diversity. As mentioned previously, large-scale screening requires low-cost technologies. In a recent publication, Ali et al. (2011) evaluated the genetic and agro-morphological diversity of rice (*Oryza sativa*) among subpopulations and their geographic distribution. A selection of 409 Asian landraces and cultivars were chosen from 79 countries representing all of the major rice growing regions of

the world. This rice diversity panel with the accompanying genetic and phenotypic information provides a valuable foundation for association mapping and understanding the basis of both genotypic and phenotypic differences within and between subpopulations.

Microsatellite markers have also been used for plants with poor genomic knowledge. For example, in an interspecific analysis, Hoshino et al. (2006) evaluated 76 accessions of 34 species from nine *Arachis* sections and showed that this germplasm bank possessed high variability, even when a species was represented by few accessions. This information was used to maintain *Arachis* genetic diversity during the storage and conservation process. Beatty & Provan (2011) published research utilizing intraspecific analysis through SSR markers. These authors assessed the genetic diversity of glacial and temperate plant species, respectively *Orthilia secunda* (one-sided wintergreen) and *Monotropa hypopitys* (yellow bird's nest). In this case, microsatellites were extremely useful to evaluate biogeographical distributions and the impact of changes in the species ranges on total intraspecific diversity. These authors concluded the following: "given that future species distribution modeling suggests northern range shifts and loss of suitable habitat in the southern parts of the species' current distribution; extinction of genetically diverse rear edge populations could have a significant effect in the range wide intraspecific diversity of both species, but particularly in *M. hypopitys*".

The great variability detected by microsatellites could be used to identify a person, a cultivar or a population. A set of SSR markers could be selected for each species/situation to distinguish one cultivar/genotype from all others. This practice is employed to protect the intellectual property rights of new varieties by commercial companies (Wang et al., 2009a). It is also used in paternity testing, when a progeny inherits one allele from the male parent and another allele from the female parent (Chistiakov et al., 2006). The genotypic profile is highly discriminating, which suggests that a random individual would have a low probability of matching a given genotype and if only a few potential parents are being consider, paternity could be determined by exclusion (Weising et al., 2005).

"Assignment tests" (assignment of an individual to the population) can be used in forensics, conservation biology and molecular ecology. An interesting example is the study of Primmer et al. (2000), which used this approach to identify a case of fishing competition fraud. The assignment of the SSR genotype of the suspect fish to its most likely original population indicated a high level of improbability that the fish originated from Lake Saimaa (where the competition occurred). When this evidence was presented, the offender confessed purchasing the salmon at a local fish shop and criminal charges were laid.

6.3 Population and phylogenetic relationships

Microsatellite markers can be used to determine the population structure within and among populations (Wang et al., 2009a). Evaluations of population differentiation permit the estimation of the migration rate between populations, assuming that these populations are in equilibrium (e.g., no selection, identical mutation rates and generation time) (Weising et al., 2005). In plants, migration rates correspond with the gene flow through seeds and pollen (Weising et al., 2005). Microsatellite markers are a powerful system for revealing inter or intraspecific phylogenetic relationships, even in closely related species (Wang et al., 2009a). Phylogenetic relationships reflect the relatedness of a group of species based on a calculated

genetic distance in their evolutionary history. Genomic SSRs, specifically EST-SSR markers, are the best choice for cross-species phylogenetics (Mittal & Dubey, 2009). However, the high incidence of homoplasy increases evolutionary distances and might undermine the confidence of the phylogenetic hypotheses, compromise the accuracy of the analysis and limit the depth of the phylogenetic inference (Jarne & Lagoda, 1996). Another problem with SSR-based phylogenetic inference is that primer transferability might not work well in all taxa and even when it is possible to amplify, the sequences might not be similar enough to permit a confident orthology assessment. Flanking regions of microsatellites have also been used in phylogenetic relationships between species and families because they evolve more slowly than repeated sequences (Chistiakov et al., 2006).

Microsatellites have been used successfully in some phylogenetic cases. Using EST-SSR markers derived from *Medicago*, cowpea and soybean, the genetic diversity of the USDA *Lespedeza* germplasm collection was assessed and its phylogenetic relationship with the genus *Kummerowia* was clarified (Wang et al., 2009b), despite the fact that phylogenetic analysis with morphological reexamination provides a more complete approach to classify accessions in plant germplasm collection and conservation. Orsini et al. (2004) used a set of 48 polymorphic microsatellites derived from *Drosophila virilis* to infer phylogenetic relationships in the *D. virilis* clade and found results consistent with previous studies (*D. virilis* and *D. lummei* were the most basal group of the species). Furthermore, these authors detected differentiations between *D. americana texana*, *D. americana americana* and *D. novamexicana* that were previously supported by FST analyses and a model-based clustering method for multilocus genotype data. Rout et al. (2008) assessed the phylogenetic relationships of Indian goats using 17 microsatellite markers. Breeds were sampled from their natural habitats, covering different agroclimatic zones. Analyses showed that the results of the microsatellite analysis were consistent with mitochondrial DNA data, which classifies Indian goat populations into distinct genetic groups or breeds. The phylogenetic and principal component analysis showed the clustering of goats according to their geographical origin. The authors concluded that although the goat breeding tracts overlapped and spread countrywide, they still maintain genetic distinctions while in their natural habitats.

In the scope of biodiversity conservation and evolutionary genetics, microsatellites have been used to contribute accurate information on issues of population dynamics, demography and ecological/biological factors intrinsic to species and populations. As examples of this approach, we can cite Palstra et al. (2007) and Becquet et al. (2007). Palstra et al. (2007) examined the population structure and connectivity of Atlantic salmon (*Salmo salar*) from Newfoundland and Labrador, which are regions where populations of this species are relatively pristine. Using the genetic variation of 13 microsatellite loci from samples (n=1346) collected from a total of 20 rivers, the connectivity at several regional and temporal scales was verified, and the hypothesis that the predominant direction of the gene flow is from large into small populations was tested. However, this hypothesis was rejected by evidence that the temporal scale in which gene flow is assessed affects the directionality of migration. Whereas large populations tend to function as sources of dispersal over contemporary timescales, such patterns are often changed and even reversed over evolutionary and coalescent-derived timescales. Furthermore, these patterns of population structure vary among different regions and are compatible with demographic and life-

history attributes. No evidence for sex-biased dispersal underlying gene flow asymmetry was found. These results are inconsistent with generalizations concerning the directionality of the gene flow in Atlantic salmon and emphasize the necessity of detailed regional study, if such information is to be meaningfully applied in conservation and management of salmonids.

Becquet et al. (2007) used 310 microsatellite markers genotyped in 78 common chimpanzees and six bonobos, allowing a high-resolution genetic analysis of chimpanzee population structure. These chimpanzees have been traditionally classified into three populations: western, central and eastern. While the morphological or behavioral differences are small, genetic studies of mitochondrial DNA and the Y chromosome have supported the geography-based designations. The findings showed that the populations seem to be discontinuous and provided weak evidence for gradients of variation reflecting hybridization among chimpanzee populations. In addition, the results demonstrated that central and eastern chimpanzees are more closely related to each other in time than to western chimpanzees.

6.4 Bioinvasions and epidemiology

The analysis of genetic diversity, population structure and demographic inferences using microsatellite has been useful to elucidate the processes of bioinvasion, understand the epidemiological patterns and aid in controlling and eradicating diseases.

The characterization of the genetic structure of invasive populations is important because genetically variable populations tend to be more successful as invaders than those that are relatively genetically homogeneous, and genetic data might provide an important tool to resource managers concerned with invasion risk assessments and predictions. To examine the invasion genetics of the Eurasian spiny water flea, *Bythotrephes longimanus*, which is a predacious zooplankter with increased range in Europe that is rapidly invading inland waterbodies throughout North America's Great Lakes region, Colautti et al. (2005) employed microsatellite markers. Three populations where *B. longimanus* has been historically present (Switzerland, Italy and Finland), a European-introduced population (the Netherlands) and three North American populations (Lakes Erie, Superior and Shebandowan) were sampled. Consistent with a bottleneck during colonization, the average heterozygosities of the four European populations were higher than the three North American populations. The pairwise FST estimated among North American populations was not significantly different from zero and was much lower than that among European populations. This result is consistent with a scenario of higher gene flow among North American populations. The assignment tests identified several migrant genotypes in all introduced populations (the Netherlands, Erie, Superior and Shebandowan), but rarely in native ones (Switzerland, Italy and Finland). A large number of genotypes from North America were assigned to Italian populations, suggesting a second invasion in the region of northern Italy that was previously unidentified. These results support a bottleneck in the invasion of North American populations that has been largely offset by the gene flow from multiple native sources and among introduced populations.

Microsatellites have also been chosen to evaluate the genetic variability and dynamics of the invasion of *Ambrosia artemisiifolia*, an aggressive North American annual weed, found

particularly in sunflower and cornfields. Besides its economic impact on crop yield, this plant represents a major health problem because of strongly allergenic pollen. The results of Genton et al. (2005) suggested that the French invasive populations include plants from a mixture of sources. The reduced diversity in populations distant from the original introduction area indicated that ragweed range expansion probably occurred through sequential bottlenecks from the original populations and not from subsequent new introductions.

Understanding the epidemiology of the disease is related to knowledge about the basic biology of the organisms involved. Population genetic studies can provide information about the taxonomic status of species, the spatial limits of populations and the nature of the gene flow among populations. Examples of the important results in this approach are Pérez de Rosas et al. (2007) and Fitzpatrick et al. (2008). Pérez de Rosas et al. (2007) examined the genetic structure in populations of the Chagas disease vector, *Triatoma infestans*. Levels of genetic variability (assessed by microsatellites) were compared in populations of *T. infestans* from areas with different periods after insecticide treatment and from areas that never received treatment. These authors found that genetic drift and limited gene flow appear to have generated a substantial degree of genetic differentiation among the populations of *T. infestans* and the microgeographical analysis supports the existence of subdivision in *T. infestans* populations. Levels of genetic diversity in the majority of *T. infestans* populations from insecticide-treated localities were similar or higher than those detected in populations from areas without treatment. This study supports the hypothesis of vector population recovery from survivors of the insecticide-treated areas, and therefore highlights the value of population genetic analyses in assessing the effectiveness of the Chagas disease vector control programs. Fitzpatrick et al. (2008) investigated the identity of silvatic *Rhodnius* (vector of Chagas' disease) using sequencing and microsatellites and whether silvatic populations of *Rhodnius* are isolated from domestic populations in Venezuela. Sequencing confirmed the presence of *R. prolixus* in palms and that silvatic bugs can colonize houses. The analyses of microsatellites revealed a lack of genetic structure between silvatic and domestic ecotopes (non-significant FST values), which is indicative of unrestricted gene flow. These results demonstrate that silvatic *R. prolixus* presents an unquestionable threat to the control of Chagas disease in Venezuela.

7. Limitations

Although microsatellites have many advantages over other molecular markers, all data sets might include some errors and genotyping errors remain a subject in population genetics because they might bias the final conclusions (Bonin et al., 2004). Microsatellite genotyping errors result from many variables (reagent quality, *Taq* polymerase error or contamination), as reviewed by Pompanon et al. (2005), and the primary consequence is the misinterpretation of allele banding patterns.

Microsatellite markers are mainly limited by:

1. Null alleles: locus deletion or mutations in the annealing primer site prevent locus amplification and heterozygous identification and lead to erroneous estimations of allele frequencies and segregation rates. Primer redesign might resolve this problem.

2. Homoplasy: alleles identical in state (length) but not by descent are homoplasic alleles (Jarne & Lagoda, 1996). They can be identical in length but not in sequence or identical in length and sequence but with different evolutionary history (Anmarkrud et al., 2008). Because homoplasy is disregarded, the actual divergence between populations is underestimated. Sequencing could be used to identify differences in sequences, but differences in evolutionary history can only be identified by mutations documented in known pedigrees.
3. Linkage disequilibrium: deviations from the random association of alleles in a population, which are primarily caused by population substructuring and high levels of inbreeding (Weising et al., 2005). It is especially problematic for population studies and paternal exclusion. Computer programs or an offspring analysis could detect the problems.

The error impact depends on the data application. In population genetic analyses, homoplasy is not a significant problem (Estoup et al., 2002), except for hypermutable markers that have increased slippage rates (Anmarkrud et al., 2008). However, error rates as low as 0.01 per allele resulted in a rate of false paternity exclusion exceeding 20%, making even modest error rates strongly influential (Hoffman & Amos, 2005). There are a lot of informatics tools that account for genotyping errors (listed by Pompanon et al., 2005). These authors also proposed a protocol for estimating error rates that should be used to attest the reliability of published genotyping studies.

8. Perspectives

The utilization of microsatellites has been demonstrated by a large number of studies applying this marker and by the variety of areas that apply microsatellites for several purposes. Furthermore, novel technologies have enabled the development of markers for previously neglected species through the generation of new sequences and a more refined search in databases.

Nevertheless, there are some bottlenecks that need to be overcome as they hamper the best and widespread use of SSR data, e.g., an exchangeable data format to allow users to access different kinds of analyses and computer programs easily (Excoffier & Heckel, 2006) and the best understanding about microsatellite evolution and mutation mechanisms.

9. Acknowledgment

We are grateful to the Brazilian funding agencies (CAPES, CNPq and FAPESP) for their financial support. PMN is the recipient of a FAPESP postdoctoral fellowship (2010/08238-0).

10. References

Abdelkrim, J.; Robertson, B.C.; Stanton, J.A.L. & Gemmell, N.J. (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques,* Vol.46, No.3, (March 2009), pp. 185-192, ISSN 0736-6205.

Ali, M.L.; McClung, A.M.; Jia, M.H.; Kimball, J.A.; McCouch, S.R. & Eizenga, G.C. (2011). A Rice Diversity Panel Evaluated for Genetic and Agro-Morphological Diversity

between Subpopulations and its Geographic Distribution. *Crop Science*, Vol.51, No.5, (September–October 2011), pp. 2021-2035, ISSN 1435-0653.

Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, Vol.215, No.3, (October 1990), pp. 403-410, ISSN 0022-2836.

Anmarkrud, J.A.; Kleven, O.; Bachmann, L. & Lifjeld, J.T. (2008). Microsatellite evolution: Mutations, sequence variation, and homoplasy in the avian microsatellite locus HrU10. *BMC Evolutionary Biology,* Vol.8, (May 2008), pp. 138, ISSN 1471-2148.

Apotikar, D.B.; Venkateswarlu, D.; Ghorade, R.B.; Wadaskar, R.M.; Patil, J.V. & Kulwal, P.L. (2011). Mapping of shoot fly tolerance loci in sorghum using SSR markers. *Journal of Genetics*, Vol.90, No.1, (April 2011), pp. 59-66, ISSN 0973-7731.

Arthofer, W.; Steiner, F.M. & Schlick-Steiner, B.C. (2011). Rapid and cost-effective screening of newly identified microsatellite loci by high-resolution melting analysis. *Molecular Genetics and Genomics*,Vol.286, No.3-4, (August 2011), pp. 225-235, ISSN 1617-4623.

Balloux, F. & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology,* Vol.11, No.2, (February 2002), pp. 155–165, ISSN 1365-294X.

Baranski, M.; Moen, T. & Vage D.I. (2010). Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, Vol.42, (June 2010), pp. 17, ISSN 1297-9686.

Beatty, G.E. & Provan, J. (2011). Comparative phylogeography of two related plant species with overlapping ranges in Europe, and the potential effects of climate change on their intraspecific genetic diversity. *BMC Evolutionary Biology,* Vol.11, (January 2011), pp. 29, ISSN 1471-2148.

Becquet, C.; Patterson, N.; Stone, A.C.; Przeworski, M. & Reich, D. (2007). Genetic Structure of Chimpanzee Populations. *PLoS Genetics,* Vol.3, No.4, (April 2007), pp. 617-626, ISSN 1553-7404.

Bhargava, A. & Fuentes, F.F. (2010). Mutational Dynamics of Microsatellites. *Molecular Biotechnology*, Vol.44, No.3, (March 2010), pp. 250–266, ISSN 1559-0305.

Bonin, A.; Bellemain, E.; Bronken Eidesen, P.; Pompanon, F.; Brochmann, C. & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, Vol.13, No.11, (November 2004), pp. 3261–3273, ISSN 1365-294X.

Bravo, J.P.; Hoshino, A.A.; Angelici, C.M.L.C.D.; Lopes C.R. & Gimenes, M.A. (2006). Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some Arachis section species of the genus Arachis. *Genetics and Molecular Biology*, Vol.29, No.3, pp. 516-524, ISSN 1415-4757.

Buschiazzo, E. & Gemmell, N.J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays*, Vol.28, No.10, (September 2006), pp. 1040–1050, ISSN 1521-1878.

Chistiakov, D.A.; Hellemans, B. & Volckaert, F.A.M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, Vol.255, (May 2006), pp.1-29, ISSN 0044-8486.

Colautti, R.I.; Manca, M. & Viljanen, M. (2005). Invasion genetics of the Eurasian spiny waterflea: evidence for bottlenecks and gene flow using microsattelites. *Molecular Ecology*, Vol.14, No.7, (June 2005), pp. 1869–1879, ISSN 1365-294X.

Crow, J. & Kimura, M. (1970). *An introduction to population genetics theory*. Burgess Publishing Company, ISBN 0-8087-2901-2, Minneapolis, United States of America.

Dawson, D.A.; Horsburgh, G.J.; Küpper, C.; Stewart, I.R.; Ball, A.D.; Durrant, K.L.; Hansson, B.; Bacon, I.; Bird, S.; Klein, A.; Krupa, A.P.; Lee, J.W.; Martín-Gálvez, D.; Simeoni, M.; Smith, G.; Spurgin, L.G. & Burke, T. (2010). New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility – as demonstrated for birds. *Molecular Ecology Resources,* Vol.10, No.3, (May 2010), pp. 475–494, ISSN 1755-0998.

Di Rienzo, A.; Peterson, A.C.; Garza, J.C.; Valdes, A.M & Slatkin, M. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.91, No.8, (April 1994), pp. 3166–3170, ISSN 0027-8424.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics,* Vol.5, No.6, (June 2004), pp. 435–445, ISSN 1471-005.

Estoup, A.; Jarne, P. & Cornuet, J.M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, Vol.11, No.9, (September 2002), pp. 1591–1604, ISSN 1365-294X.

Excoffier, L. & Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics,* Vol.7, No.10, (October 2006), pp. 745-758, ISSN 1471-0056.

Faria, A.D.; Mamani, E.M.C.; Pappas, M.R.; Pappas Jr, G.J. & Gratapaglia D. (2010). A Selected Set of EST-Derived Microsatellites, Polymorphic and Transferable across 6 Species of Eucalyptus. *Journal of Heredity*, Vol.101, No.4, (March 2010), pp. 512-520, ISSN 1465-7333.

Fitzpatrick, S.; Feliciangeli, D.; Sanchez-Martin, M., Monteiro, F.A. & Miles, M.A. (2008). Molecular genetics reveal that silvatic Rhodnius prolixus do colonise rural houses. *PLoS Neglected Tropical Diseases*, Vol.2, No.4, (April 2008), pp. 1-16, ISSN 1935-2735.

Genton, B.J.; Shykoff, J.A. & Giraud, T. (2005). High genetic diversity in French invasive populations of common ragweed, Ambrosia artemisiifolia, as a result of multiple sources of introduction. *Molecular Ecology*, Vol.14, No.14, (December 2005), pp. 4275–4285, ISSN 1365-294X.

Glaubitz, J.C. (2004). CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, Vol.4, No.2, (June 2004), pp. 309–310, ISSN 1755-0998.

Grover, A. & Sharma, P.C. (2011). Is spatial occurrence of microsatellites in the genome a determinant of their function and dynamics contributing to genome evolution? *Current Science*, Vol.100, No.6, (March 2011), pp. 859-869, ISSN 0011-3891.

Hoffman, J.I. & Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, Vol.14, No.2, (February 2005), pp. 599–612, ISSN 1365-294X.

Hoshino, A.A.; Bravo, J.P.; Angelici, C.M.L.C.D.; Barbosa, A.V.G.; Lopes, C.R. & Gimenes M.A. (2006). Heterologous microsatellite primer pairs informative for the whole genus Arachis. *Genetics and Molecular Biology*, Vol.29, No.4, pp.665-675, ISSN 1415-4757.

Ito, T.; Smith, C.L. & Cantor, C.R. (1992). Sequence-specific DNA purification by triplex affinity capture. *Proceedings of the National Academy of Sciences*, Vol.89, No.2, (January 1992), pp. 495–498, ISSN 1091-6490.

Jarne, P. & Lagoda, P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution,* Vol.11, No.10, (October 1996), pp. 424-429, ISSN 0169-5347.

Jentzsch, V.I.M.; Bagshaw, A.; Buschiazzo, E.; Merkel, A. & Gemmell, J.E. (2008). Evolution of microsatellite DNA. In: *Encyclopedia of Life Sciences (ELS)*, 1-12, John Wiley & Sons Ltd., ISBN 9780470015902, Chichester, United Kingdom.

Kandpal, R.P.; Kandpal, G. & Weissman, S.M. (1994). Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of the National Academy of Sciences*, Vol.91, No.1, (January 1994), pp. 88–92, ISSN 1091-6490.

Kashi, Y.; King, D. & Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics,* Vol.13, No.2, (February 1997), pp. 74–78, ISSN 0168-9525.

Katti, M.V.; Ranjekar, P.K. & Gupta, V.S. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology Evolution,* Vol.18, No.7, (July 2001), pp. 1161–1167, ISSN 1537-1719.

Kimura, M. & Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, Vol.49, No.4, (April 1964), pp. 725–738, ISSN 1943-2631.

Labate, J.A. (2000). Software for population genetic analyses of molecular marker data. *Crop Science*, Vol.40, No.6, (November 2000), pp. 1521-1528, ISSN 1435-0653.

Litt, M. & Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, Vol.44, No.3, (March 1989), pp. 397-401, ISSN 0002-9297.

Malausa, T.; Gilles, A.; Meglécz, E.; Blanquart, H.; Duthoy, S.; Costedoat, C.; Dubut, V.; Pech, N.; Castagnone-Sereno, P.; Délye, C.; Feau, N.; Frey, P.; Gauthier, P.; Guillemaud, T.; Hazard, L.; Le Corre, V.; Lung-Escarmant, B.; Malé, P.-J. G.; Ferreira, S. & Martin, J.F. (2011). High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources*, Vol.11, No.4, (July 2011), pp. 638–644, ISSN 1755-0998.

Manoukis, N.C. (2007). FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. *Molecular Ecology Notes*, Vol.7, No.4, (July 2007), pp. 592–593, ISSN 1755-0998.

Messier, M.; Li, S.H. & Stewart, C.B. (1996). The birth of microsatellites. *Nature*, Vol.381, (June 1996), pp. 483, ISSN 0028-0836.

Mikheyev, A.S.; Vo, T.; Wee, B.; Singer, M.C. & Parmesan, C. (2010). Rapid Microsatellite Isolation from a Butterfly by De Novo Transcriptome Sequencing: Performance and

a Comparison with AFLP-Derived Distances. *PLoS ONE,* Vol.5, No.6, (June 2010), pp. e11212, ISSN 1932-6203.

Mittal, N. & Dubey, A.K. (2009). Microsatellite markers - A new practice of DNA based markers in molecular genetics. *Pharmacognosy Reviews*, Vol.3, No.6, (July-December 2009), pp. 235-246, ISSN 0973-7847.

Mullis, K. & Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods in Enzymology,* Vol.155, pp. 335-350, ISSN 0076-6879.

Orsini, L.; Huttunen, S. & Schlotterer, C. (2004). A multilocus microsatellite phylogeny of the Drosophila virilis group. *Heredity*, Vol.93, No.2, (June 2004), pp. 161-165, ISSN 0018-067X.

Otha, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, Vol.22, No.2, (October 1973), pp. 201–204, ISSN 0016-6723.

Paetkau, D. (1999). Microsatellites obtained using strand extension: an enriched protocol. *BioTechniques,* Vol.26, No.4, (April 1999), pp. 690-697, ISSN 0736-6205.

Palstra, F.P.; O'Connell, M.F. & Ruzzante, D.E. (2007). Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology*, Vol.16, No.21, (October 2007), pp. 4504–4522, ISSN 1365-294X.

Peakall, R.; Gilmore, S.; Keys, W.; Morgante, M. & Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Molecular Biology and Evolution,* Vol.15, No.10, (October 1998), pp. 1275–1287, ISSN 0737-4038.

Pépin, L.; Amigues, Y.; Lépingle, A.; Berthier, J.L.; Bensaid, A. & Vaiman, D. (1995). Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heredity,* Vol.74, No.1, (January 1995), pp. 53–61, ISSN 0018-067X.

Pérez de Rosas, A.R.; Segura, E.L. & García, B.A. (2007). Microsatellite analysis of genetic structure in natural *Triatoma infestans* (Hemiptera: Reduviidae) populations from Argentina: its implication in assessing the effectiveness of Chagas' disease vector control programmes. *Molecular Ecology*, Vol.16, No.7, (April 2007), pp. 1401-1412, ISSN 1365-294X.

Pompanon, F.; Bonin, A.; Bellemain, E. & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, Vol.6, No.11, (November 2005), pp. 847-846, ISSN 1471-005.

Powell, W.; Machray, G.C. & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science,* Vol.1, No.7, (July 1996), pp. 215-222, ISSN 1360-1385.

Powell, W.; Morgante, M.; Andre, C.; McNicol, J.W.; Machray, G.C.; Doyle, J.J.; Tingey, S.V. & Rafalski, J.A. (1995). Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology,* Vol.5, No.9, (September 1995), pp. 1023-1029, ISSN 0960-9822.

Primmer, C.R.; Koskinen, M.T. & Piironen, J. (2000). The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London*, Vol.267, (March 2000), pp. 1699–1704, ISSN 09628452.

Rajendrakumar, P.; Biswal, A.K.; Balachandran, S.M.; Srinivasarao, K. & Sundaram, R.M. (2007). Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics*, Vol.23, No.1, (January 2007), pp. 1–4, ISSN 1460-2059.

Rico, C.; Rico, I. & Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proceedings Biological Sciences,* Vol. 263, No.1370, (May 1996), pp. 549–557, ISSN 0962-8452.

Rout, P.; Joshi, M.; Mandal, A.; Laloe, D.; Singh, L. & Thangaraj, K. (2008). Microsatellite-based phylogeny of Indian domestic goats. *BMC Genetics*, Vol.9, (January 2008), pp. 11, ISSN 1471-2164.

Saiki, R.K.; Scharf, S.J.; Fallona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A. & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* Vol.230, No.4732, (December 1985), pp. 1350-1354, ISSN 1095-9203.

Santana, Q.; Coetzee, M.; Steenkamp, E.; Mlonyeni, O.; Hammond, G.; Wingfield, M. & Wingfield, B. (2009). Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques*, Vol.46, No.3, (March 2009), pp. 217-223, ISSN 0736-6205.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, Vol.109, No.6, (June 2000), pp. 365–371, ISSN 0009-5915.

Selkoe, K.A. & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, Vol.9, No.5, (May 2006), pp. 615–629, ISSN 1461-0248.

Soranzo, N.; Provan, J. & Powell, W. (1999). An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome,* Vol.42, No.1, (February 1999), pp. 158–161, ISSN 0831-2796.

Strand, M.; Prolla, T.A.; Liskay, R.M. & Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, Vol.365, (September 1993), pp. 274–276, ISSN 0028-0836.

Varshney, R.K.; Graner, A. & Sorrels, M.E. (2005). Genetic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, Vol.23, No.1, (January 2005), pp. 48-55, ISSN 0167-7799.

Wang, M.L.; Barkley, N.A. & Jenkins, T.M. (2009a). Microsatellite Markers in Plants and Insects. Part I: Applications of Biotechnology. *Genes, Genomes and Genomics*, Vol.3, No.1, (December 2009), pp. 54-67, ISSN 1749-0383.

Wang, M.L.; Mosjidis, J.A.; Morris, J.B.; Chen, Z.B.; Barkley, N.A. & Pederson, G.A. (2009b). Evaluation of *Lespedeza* germplasm genetic diversity and its phylogenetic relationship with the genus *Kummerowia*. *Conservation Genetics*, Vol.10, No.1, (February 2009), pp. 79-85, ISSN 1566-0621.

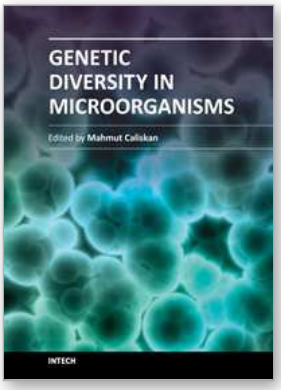
Weising, K.; Nybom, H.; Wolff, K. & Kahl, G. (2005). *DNA Fingerprinting in Plants: Principles, Methods and Applications,* CRC Press, ISBN 0-8493-1488-7, Boca Raton, United States.

Wilder, J. & Hollocher, H. (2001). Mobile elements and the genesis of microsatellites in dipterans. *Molecular Biology and Evolution*, Vol.18, No.3, (March 2001), pp.384-392, ISSN 1537-1719.

Yue, G.H.; Zhu, Z.Y.; Wang, C.M. & Xia, J.H. (2009). A simple and efficient method for isolating polymorphic microsatellites from cDNA. *BMC Genomics*, Vol.10, (March 2009), pp. 125, ISSN 1471-2164.

Zane, L.; Bargelloni, L. & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, Vol.11, No.1, (January 2002), pp. 1-16, ISSN 1365-294X.

Zhao, H.; Yu, J.; You, F.M.; Luo, M. & Peng, J. (2011). Transferability of Microsatellite Markers from *Brachypodium distachyon* to *Miscanthus sinensis*, a Potential Biomass Crop. *Journal of Integrative Plant Biology,* Vol.53, No.3, (Março 2011), pp. 232–245, ISSN 1744-7909.



Genetic Diversity in Microorganisms

Edited by Prof. Mahmut Caliskan

ISBN 978-953-51-0064-5
Hard cover, 382 pages
Publisher InTech
Published online 24, February, 2012
Published in print edition February, 2012

Genetic Diversity in Microorganisms presents chapters revealing the magnitude of genetic diversity of microorganisms living in different environmental conditions. The complexity and diversity of microbial populations is by far the highest among all living organisms. The diversity of microbial communities and their ecologic roles are being explored in soil, water, on plants and in animals, and in extreme environments such as the arctic deep-sea vents or high saline lakes. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in microorganisms. The purpose of the book is to provide a glimpse into the dynamic process of genetic diversity of microorganisms by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of microbial phylogeny, genetic diversity, and molecular biology.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Andrea Akemi Hoshino, Juliana Pereira Bravo, Paula Macedo Nobile and Karina Alessandra Morelli (2012). Microsatellites as Tools for Genetic Diversity Analysis, Genetic Diversity in Microorganisms, Prof. Mahmut Caliskan (Ed.), ISBN: 978-953-51-0064-5, InTech, Available from: http://www.intechopen.com/books/genetic-diversity-in-microorganisms/microsatellites-as-tools-for-genetic-diversity-analysis

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821