Microsatellites: Simple Sequences with Complex Evolution

By Hans Ellegren
Polymorphism in Microsatellites

- High rates of mutation
- Mutations lead to the alteration of microsatellite length through the addition or deletion of repeat units
- By studying these differences in microsatellite lengths, researchers can make inferences about when two populations diverged, and about gene flow between populations.
Where are you, Microsatellite?

- Microsatellites are normally found in non-coding regions of DNA such as introns and intergenic sequences.

- Are there any trends in the location/density of microsatellites in the genome?
  - Base composition of the sequence influences microsatellite density.
  - There appear to be regional differences in microsatellite density. In mice and humans microsatellites are more frequent at the end of chromosome arms.
  - Microsatellites have been found to be more common on the X chromosome than in non-sex related chromosomes.
  - Commonly found near SINES and LINES, and in some cases near retrotransponon-like elements.
  - In prokaryotes microsatellites are rare, but some long microsatellites act as “switches” for virulence factors and are NOT evolutionarily neutral.
They account for 3% of the genome, though the proportion of microsatellite DNA found in the genome varies among species.

Usually microsatellite density is higher in larger genomes, but this is not always true. For example rodents are thought to have unusually large quantities of microsatellites.

AND in plants the larger the genome the less the microsatellites. (AT)_n repeats are more frequent than other repeat sequences.

Prokaryotes hardly have any.
Mutations

- **Stepwise Mutation Model**
  - Mutations shorten or lengthen microsatellites by 1 repeat unit at some fixed rate
  - Problems: Multiple step changes have been observed, and microsatellites appear to demonstrate an upper size limit

- **Balance Model**
  - Point mutations break up long microsatellite sequences, and other mutations lengthen short ones, thereby holding microsatellite length in equilibrium

- Keep in mind that these models are important for making assumptions about mutation rate and directionality, which will ultimately effect estimation of genetic differences between populations.
Microsatellite Evolution

Mechanism I: Replication slippage
- During the replication process of a microsatellite sequence DNA polymerase pauses, and temporarily dissociates from the DNA.
- The terminal end of the newly synthesized DNA separates and anneals with another section of DNA, forming a small DNA “loop”.
- After synthesis is complete, the mismatch repair system corrects the mistake; adding or deleting a repeat unit.
  - If the loop formed on the nascent strand the microsatellite would increase in length
  - If the loop formed in the template strand, the microsatellite would decrease in length

Replication slippage occurs during DNA amplification for PCR reactions, and show up as small peaks known as stutter bands.

Mechanism II: Unequal cross over or gene conversion.
- Occurs in minisatellites, but not a lot of evidence of it occurring in microsatellites

![Diagram of DNA replication process with slippage](image)
Using pedigrees to study microsatellite mutations

- Have samples of DNA from parents and offspring, and observe when children's DNA displays non-mendelian inheritance of microsatellite length.
- Lets you observe new mutation events

**Problems:**
- Can't always be sure of parentage
- Can't detect mutations that change a microsatellite from one parental form to another
- Aren't always sure from whom a mutant allele has originated

**Picture doesn't show a true pedigree analysis, but you can see how they compare differences in microsatellite length.**
What did they find?

The mutation process is incredibly variable among loci, types of repeats, and test organisms.

- Multiple-step mutation events have been found to account for anywhere between 11-63% of all microsatellite mutations in humans.
- In other organisms, studies of particular loci have found multi-step mutations events account for between 5-75% of mutations.
- No strong evidence that show whether gains or losses of repeat units occur more frequently
- Some studies have noted that contraction events occur more frequently in long alleles in eukaryotes, and that bacteria seem to show a trend towards reduction of length (maybe explains low number of microsatellites in prokaryotes)

How much does this heterogeneity of the observed mutational character affect our understanding of the mutational mechanism?
Using DNA sequencing data to study microsatellite mutations

- Make within species microsatellite sequence characterizations and compare the various alleles
- Look at loci where divergence in the DNA sequence was a result of a speciation event
- Pedigree studies look at new mutations, and sequencing looks at the history of mutations

Problems: Difficult to determine precise order of mutation events.
Microsatellite mutations are often the same in structure by not necessarily by descent.

\[ (CA)_{27} = (CA)_{28} -1ru \text{ or } (CA)_{26} +1ru \]
What did they find?

- Short repeat sequences are the start place for microsatellite growth
  - Base substitutions can form a repeat (e.g. from paper A-G transition GTATGT to GTGTGT)
  - Insertion mutations of an adjacent sequence

- Long repeat units are broken up by point mutations

- Microsatellite alleles can vary within a species not only by repeat length but also by sequence
Mutation Rate

- Mutation rate varies among loci, alleles, and species.
- Microsatellite length is the greatest known influence of mutation rates; the longer the sequence, the greater the chance of mutation.
- Some studies have found that greater allele size leads to higher mutation rates.
- Mutation rates are generally found not to be affected by gender.
But, a lot of variation in mutation rates is still unaccounted for.

- The study of orthologous loci of related species link individual locus differences with mutation rate variations.

One hypothesis is that flanking sequences could affect mutation rates.

Another hypothesis implicates differences in local transcription-coupled repair, point-mutation rates, chromatin structure, and “regional sequence context”

What do these very individualistic differences in mutation rate imply for microsatellite use in genetic studies?
A new approach to microsatellite mutation study

- Most studies have examined germline cell divisions because most mutations are thought to arise in the replication process.
- New experimental procedure uses plasmids to artificially produce microsatellites that can be introduced into cells and easily monitored.
Plasmid data supports several aspects of germline transmission studies

- Mismatch repair system is critical in maintaining microsatellite stability
- Mutation rate positively correlates with sequence length
- A repeat sequence position during transcription (whether it's on the coding or complimentary strand) does not affect mutation rate
- There have been varying results as to whether the actual microsatellite sequence affects mutation rate
- Insertion mutations outnumber deletions in eukaryotes
Can we use an approximated rate of mutation and a poorly understood mutation model to study population divergence or the amount of gene flow between populations?