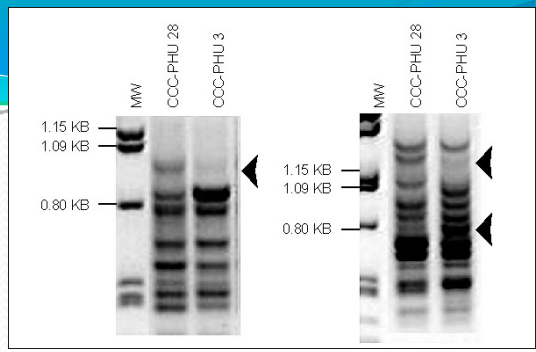
# B3 Olympic High School Science Summer Camp

Dr. Jennifer Weller

Summer 2011

# Genetics and Markers

- Measure aspects of an organisms structure and function
  - How do genes and environment influence individual differences and fitness?
- What aspects should we measure?
  - Morphological (measurements of the body of the organism)
  - Behavioral
  - Molecular




# Genetic markers

- Phenotype is the result of genotype
  - How much of a trait depends only on the genes?
  - How much do differences in genes drive differences in appearance?
  - How different are organisms in the same species from each other?
  - How different are organisms in difference genus' from each other?

```
Human    MPPGTARPGSRGCPIGTGGVLSSQIKVAHRP----EKINQFIEEIRQLDMDLEEHQGEMNQKYKE
Chimp    MPPGTARPGSRGCPIGTGGVLSSQIKVAHRP----EKKNQFIEEIRQLDMDLEEHQGEMNQKYKE
Dog      LPPGTARPGSRGGPIGTGGVLSSQIKVADRP----EKINQFSEEIRQLDMDLEEHQGEMNQKYKE
Mouse    MPPTTARPGSRGGPLGTGGVLSSQIKVADRP----EKINQFSEEIRQLDMDLEEHQGEMNQKYKE
Rat      MPPATARPGSRGGPLGTGGVLSSQIKVADRP----EKINQFSEEIRQLDMDLEEHQGEMNQKYKE
```

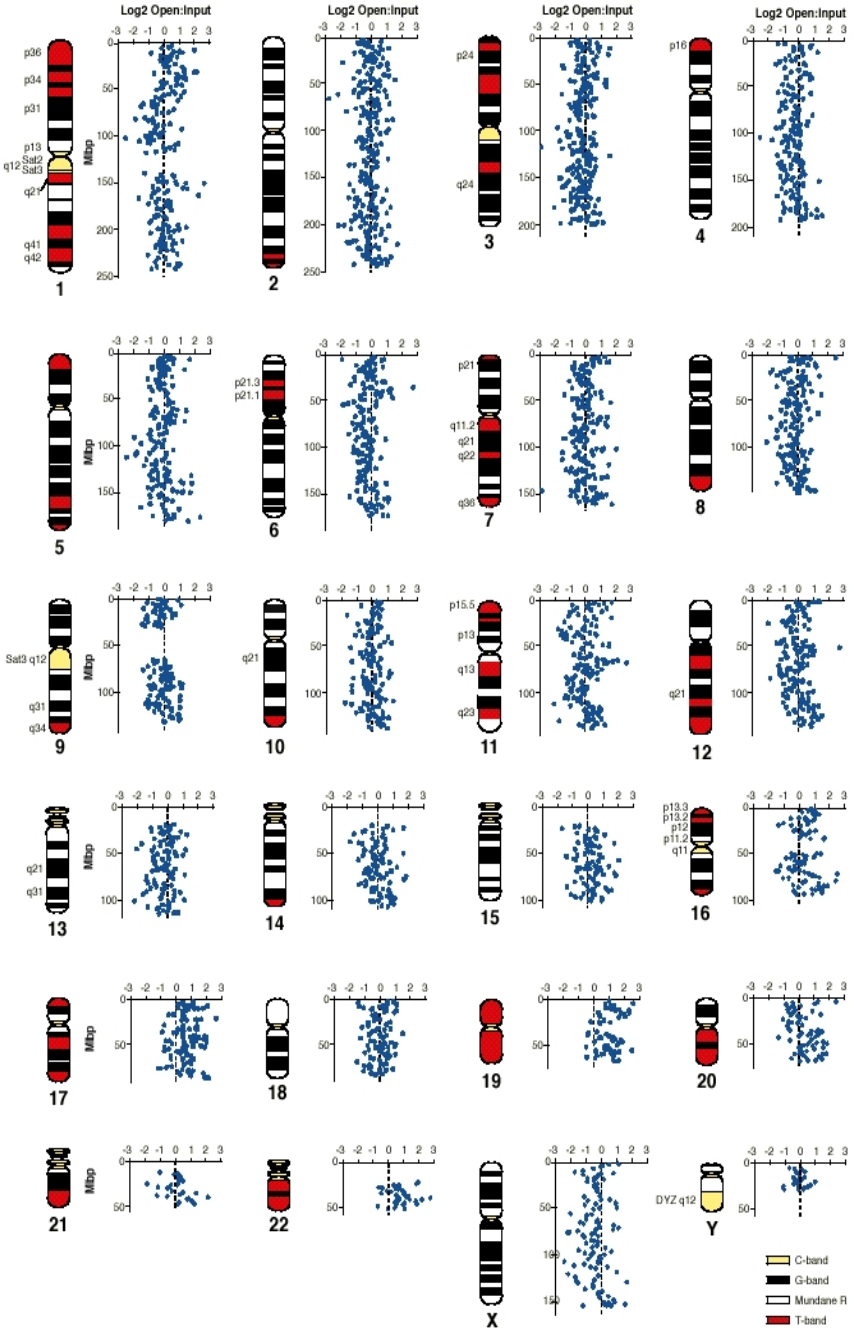
# Relatedness

- Whales to Hippos
- Even-toed, hoofed mammals are called artiodactyls. This is further broken down to
  - Camels + llamas
  - Cattle + deer
  - Pigs + peccaries
  - Hippopotamuses
- Cetaceans don't have toes – ancestors had an even number of appendages.

# What is a molecular marker?

- A piece of DNA on a chromosome
  - May be part of or closely linked to a gene that makes a protein that affects cell survival
  - May be part of controlling elements
  - May be in the larger area of ‘non-coding’ DNA
- Markers have a known location
  - What is being marked?

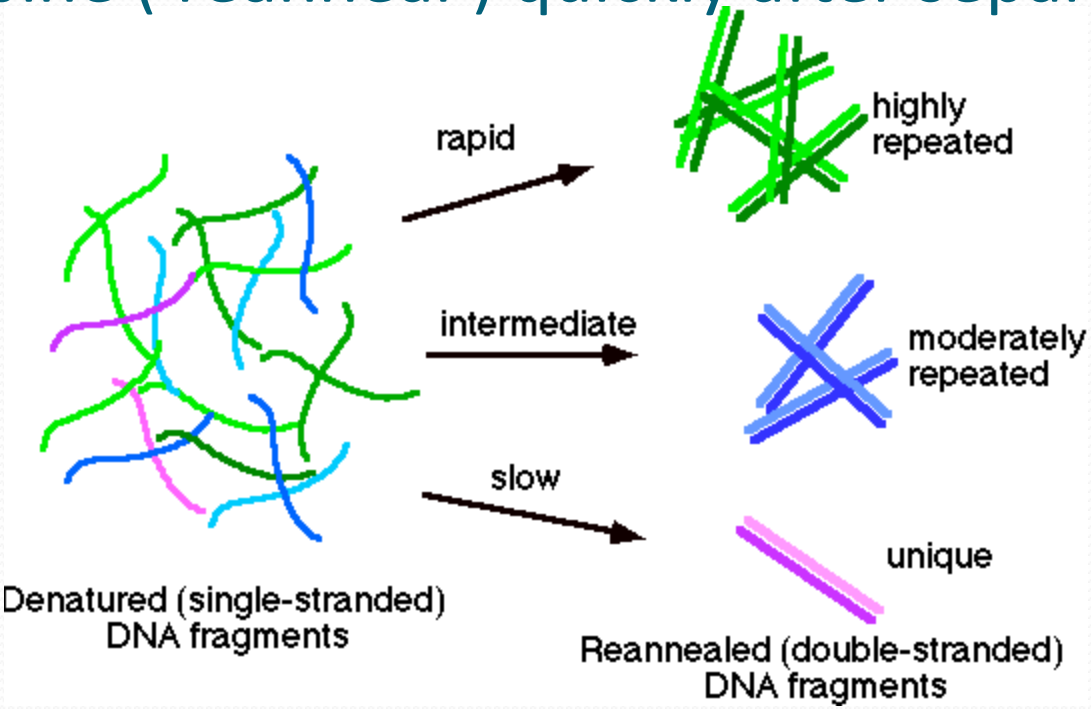
Human genome
23 pairs of chromosomes
3 X $10^9$ base pairs of DNA
Different families of repeated sequences
Sex chromosomes (X and Y)



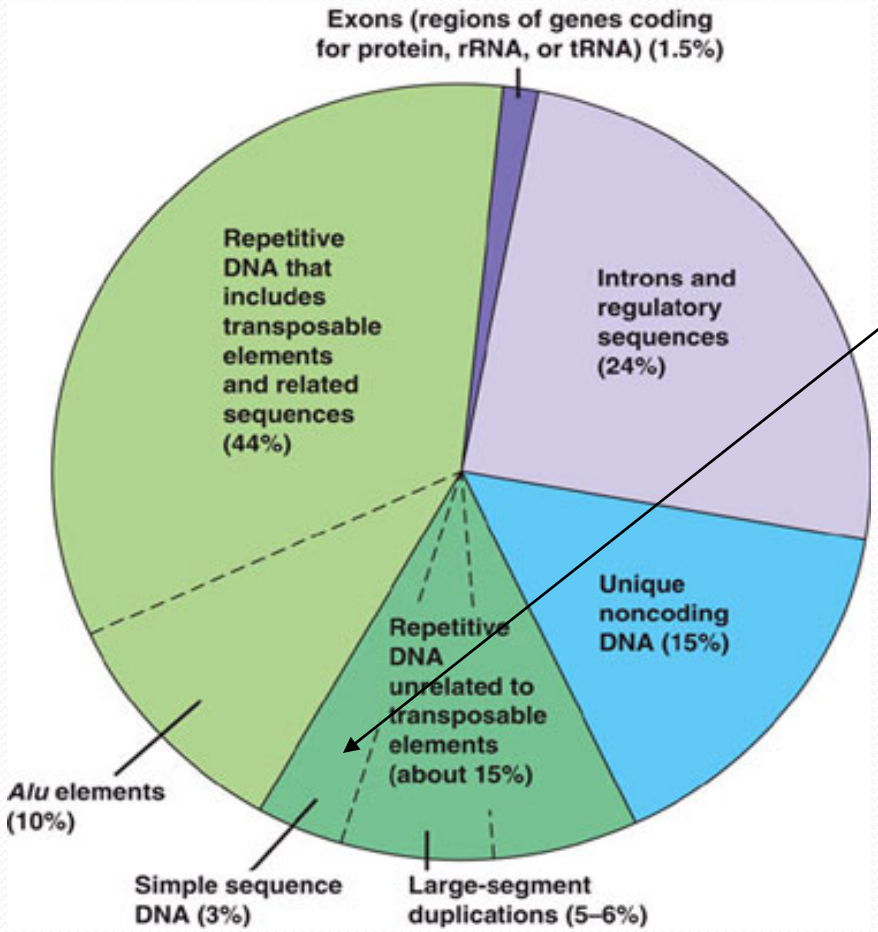

Chromosome banding occurs at reproducible positions, so the bands are another level of marker.

Studies reveal that gene-rich and gene-poor regions correspond to banding patterns within the genome.

DNA can have very simple sequences, like CACACACACA, or sequences that are unique across a long distance. The simple ones are present at much higher concentration than the unique ones, so the recombine ('reanneal') quickly after separation.



# Composition of the human genome




Simple sequence repeats can expand quickly if in non-coding regions, and are a rich source of variation for individual identification purposes.

SSRs are common in the human genome
Estimated 96,000 4 bp repeats in the human genome

**Table 14 SSR content of the human genome**

| Length of repeat unit | Average bases per Mb | Average number of SSR elements per Mb |
|---|---|---|
| 1 | 1,660 | 36.7 |
| 2 | 5,046 | 43.1 |
| 3 | 1,013 | 11.8 |
| 4 | 3,383 | 32.5 |
| 5 | 2,686 | 17.6 |
| 6 | 1,376 | 15.2 |
| 7 | 906 | 8.4 |
| 8 | 1,139 | 11.1 |
| 9 | 900 | 8.6 |
| 10 | 1,576 | 8.6 |
| 11 | 770 | 8.7 |

SSRs were identified by using the computer program Tandem Repeat Finder with the following parameters: match score 2, mismatch score 3, indel 5, minimum alignment 50, maximum repeat length 500, minimum repeat length 1.

# Interspersed repeats in the human genome

Classes of interspersed repeat in the human genome

| | | | Length | Copy number | Fraction of genome |
|---|---|---|---|---|---|
| LINEs | Autonomous |   | 6–8 kb | 850,000 | 21% |
| SINEs | Non-autonomous |   | 100–300 bp | 1,500,000 | 13% |
| Retrovirus-like elements | Autonomous |   | 6–11 kb | 450,000 | 8% |
| | Non-autonomous |   | 1.5–3 kb | | |
| DNA transposon fossils | Autonomous |   | 2–3 kb | 300,000 | 3% |
| | Non-autonomous |  | 80–3,000 bp | | |

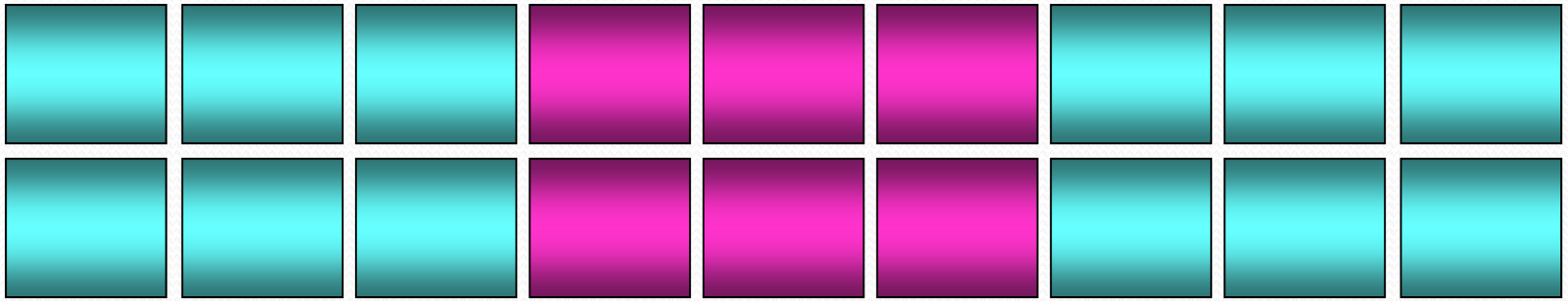
# SSRs and Length Polymorphisms

Simple sequence repeats have a certain number of units of the repeat: $(CA)_{13}$ has 13 of the CA units in a row.

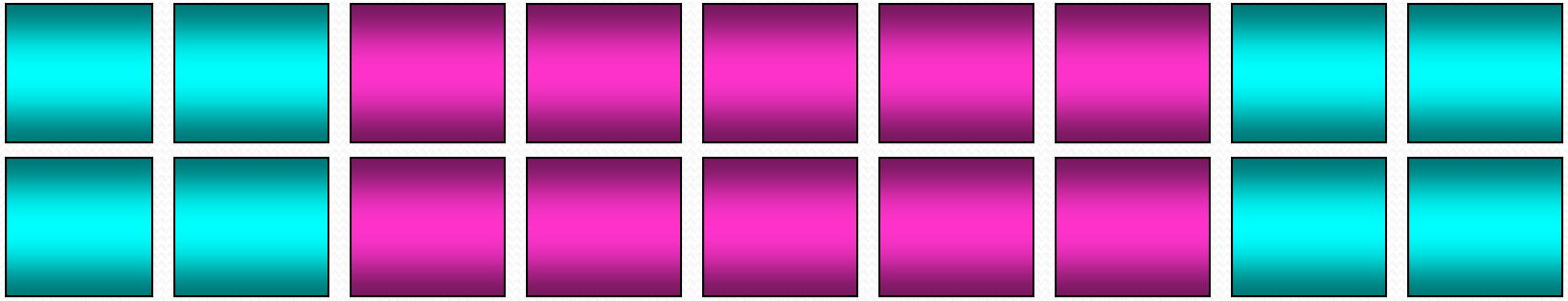
CA

The number of units can vary if the polymerase stutters (loses its place).

Two individuals differ by the number of repeats:



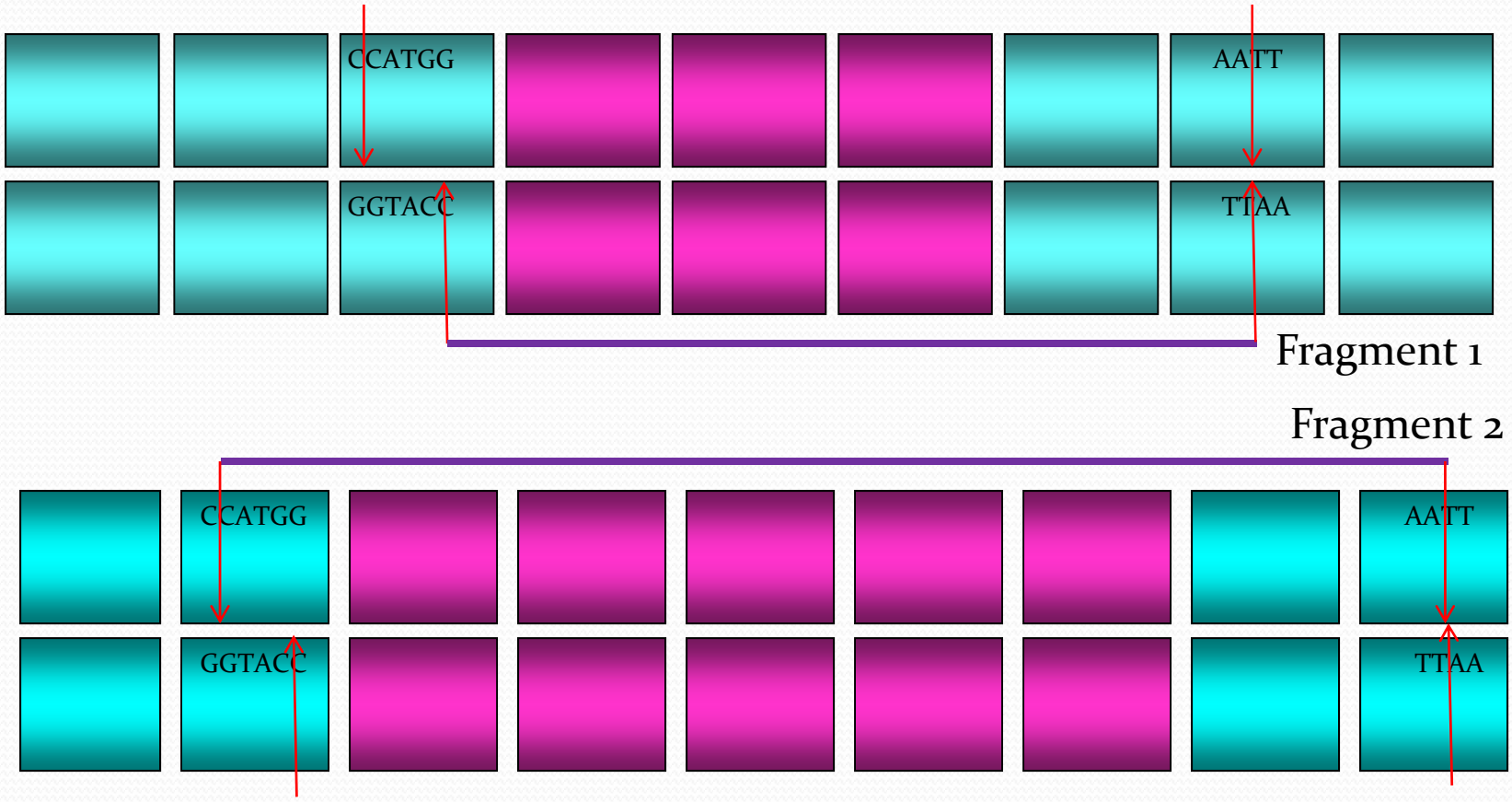
3 repeats vs 5 repeats → produces <u>length differences.</u>



# Turning SSRs into markers

You can use a restriction enzyme that recognizes a sequence just outside the boundaries.

But it takes a lot of DNA and you will have a lot of other fragments cluttering up the gel



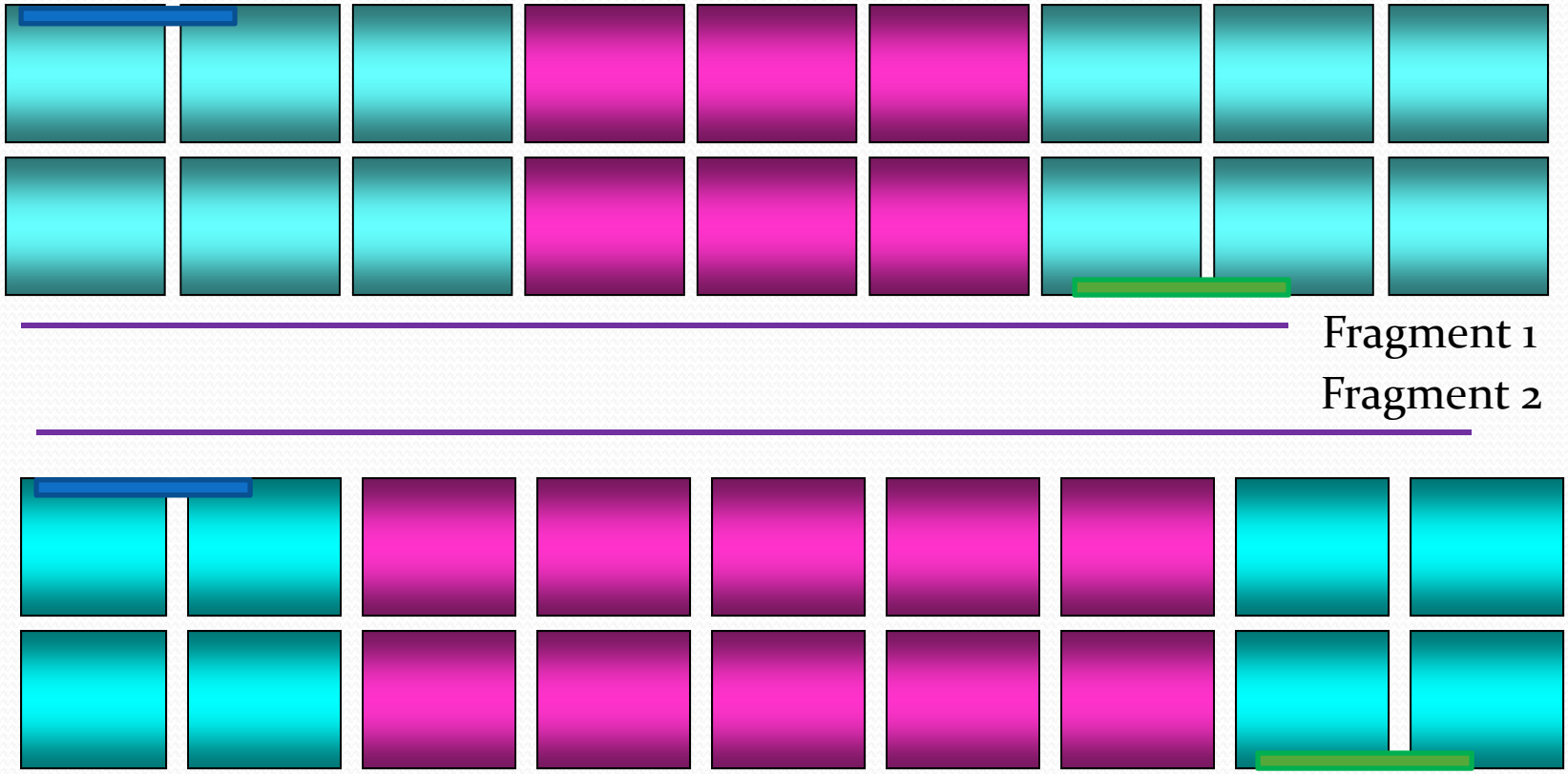
# Turning SSRs into markers

You can use PCR with primer sequences that lie just outside the boundaries.

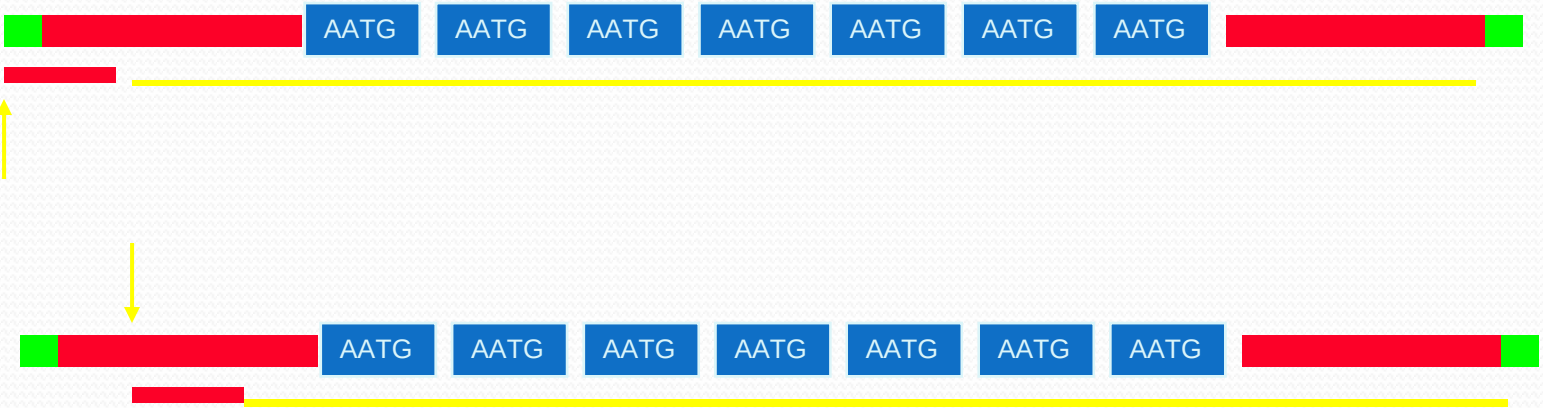
Primer 1

Primer 2

Primers match conserved sequence, lie just outside the repeated region and amplify across is, so most of the PCR product is the repeated region



# How Do Primers Influence Amplified Fragment (Amplicon) Length?



# Fragment Size Estimation