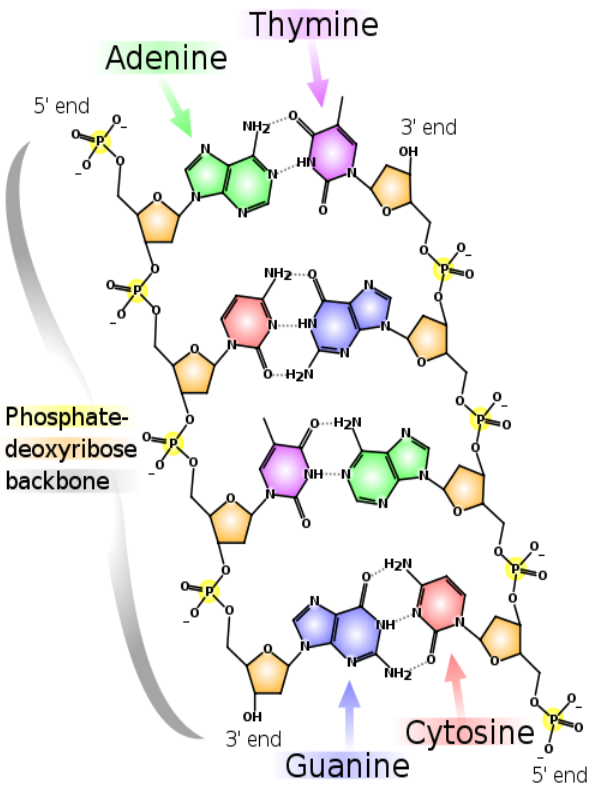
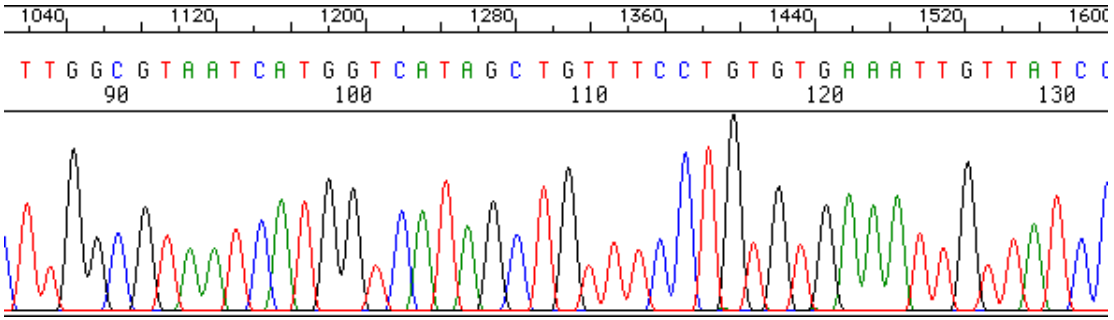
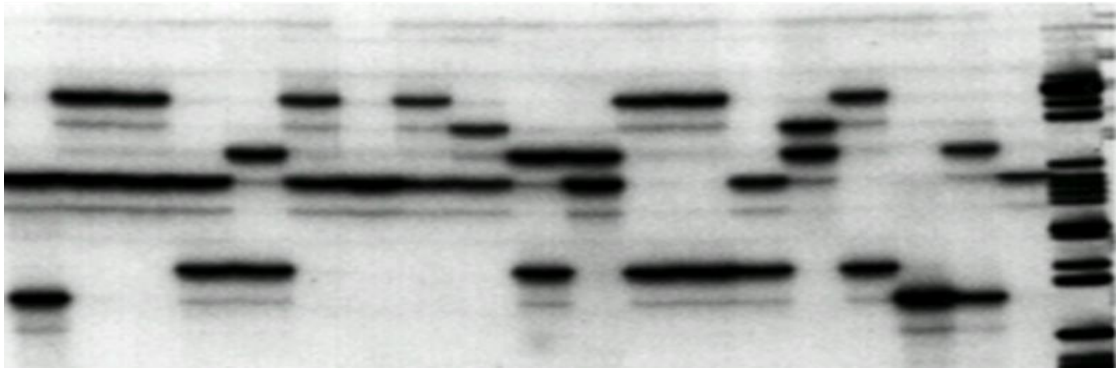
Molecular Markers in Conservation







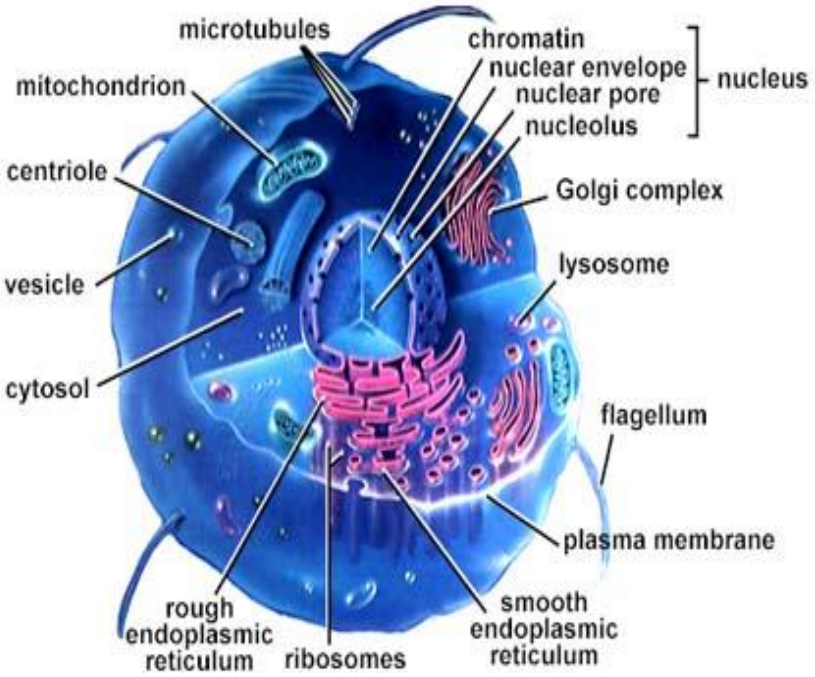
Molecular Markers

- A **molecular marker** is segment of DNA whose characteristics can be measured and make inference to the ecology and evolution of individuals, populations, and species
- A wide range of genetic markers are available - choice of molecular marker depends on the specific conservation genetic application
- Inappropriate marker selection can seriously compromise the ability to address research objectives
- Constant and rapid development of new and sophisticated laboratory methods requires keeping up with technology.

Molecular Markers

- The following are commonly used, or have been commonly used in conservation genetics
  - DNA sequences
  - Microsatellites
  - Minisatellites
  - SNPs (Single Nucleotide Polymorphisms)
  - RAPDs (Random Amplification of Polymorphic DNA)
  - AFLPs (Amplified Fragment Length Polymorphisms)
  - RFLPs (Restriction Fragment Length Polymorphisms)
  - SSCPs (Single Strand Conformational Polymorphisms)

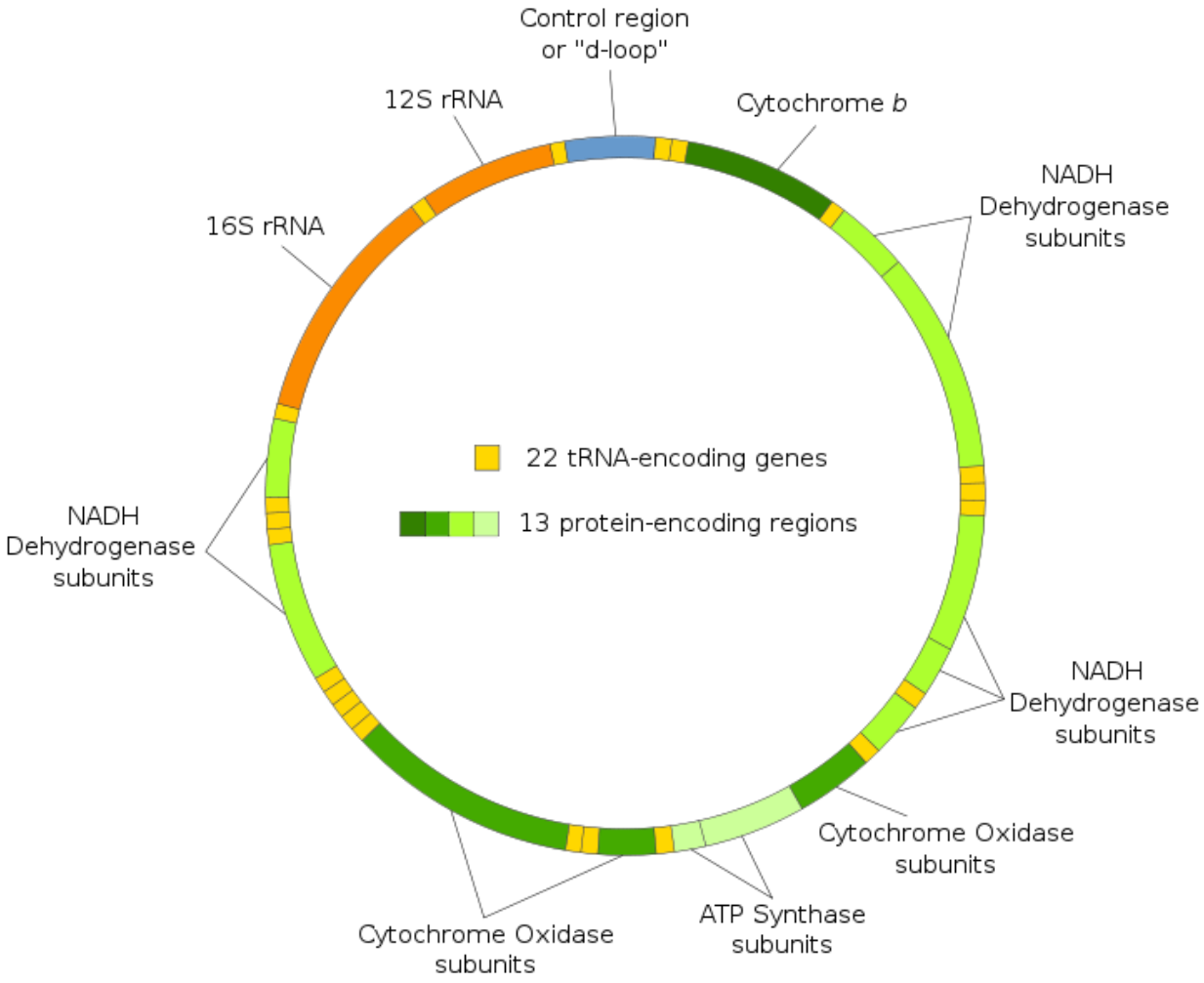
Same Cell – Different Genomes



- Nuclear DNA
- Mitochondrial DNA

Mitochondrial DNA

- Circular molecule that codes for 35 proteins and enzymes
- Maternally inherited – traces matrilines
- Some segments are highly conserved
- No recombination

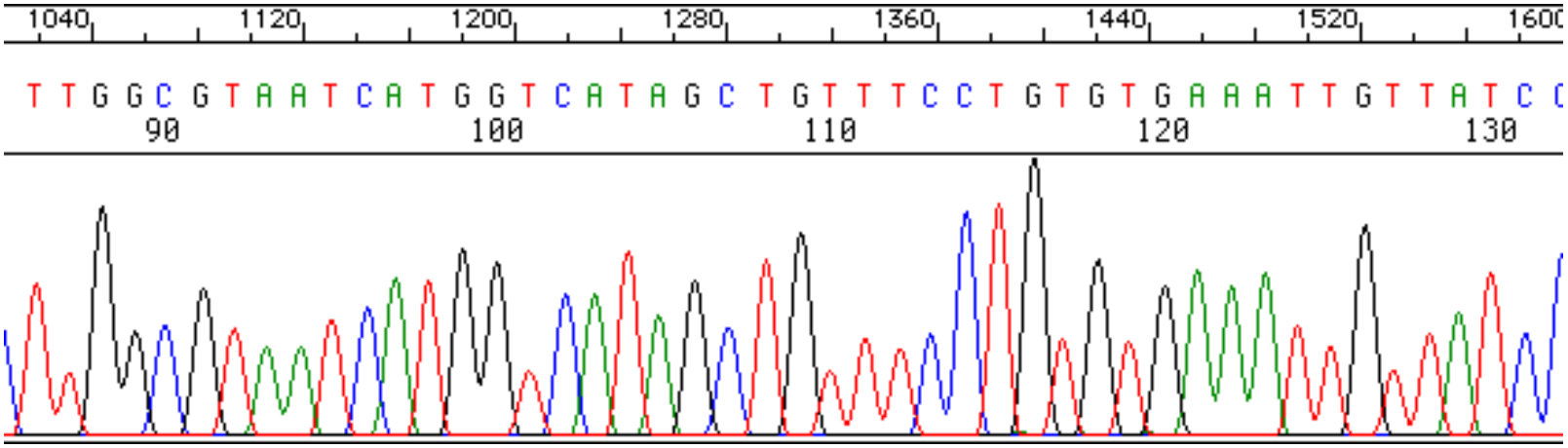


Nuclear vs Mitochondrial DNA

Characteristic	**Mitochondrial**	**Nuclear**
Number of Copies	Hundreds	Two
Ploidy	Haploid	Diploid
Inheritance	Maternal	Bi-parental
Mutation rate	High	Variable
Recombination	None	Yes
Rate of genetic drift	$1/N_e$	$1/2N_e$
Number of loci characterized	1	Usually many
Typical applications in conservation genetics	Phylogeography and phylogenetics	All, depending on marker type

DNA Sequences

- Can be obtained from both mitochondrial and nuclear DNA
- Requires prior amplification of targeted sequence
- Expensive and laborious compared to other markers, but provides ultimate characterization of genetic diversity .



Restriction Fragment Length Polymorphisms (typically with mtDNA)

- Use restriction enzymes from bacteria such as *E. coli* that cleave DNA at specific sequences

- Eg. *E. coli*RI cleaves GAATTC (and reverse complement)

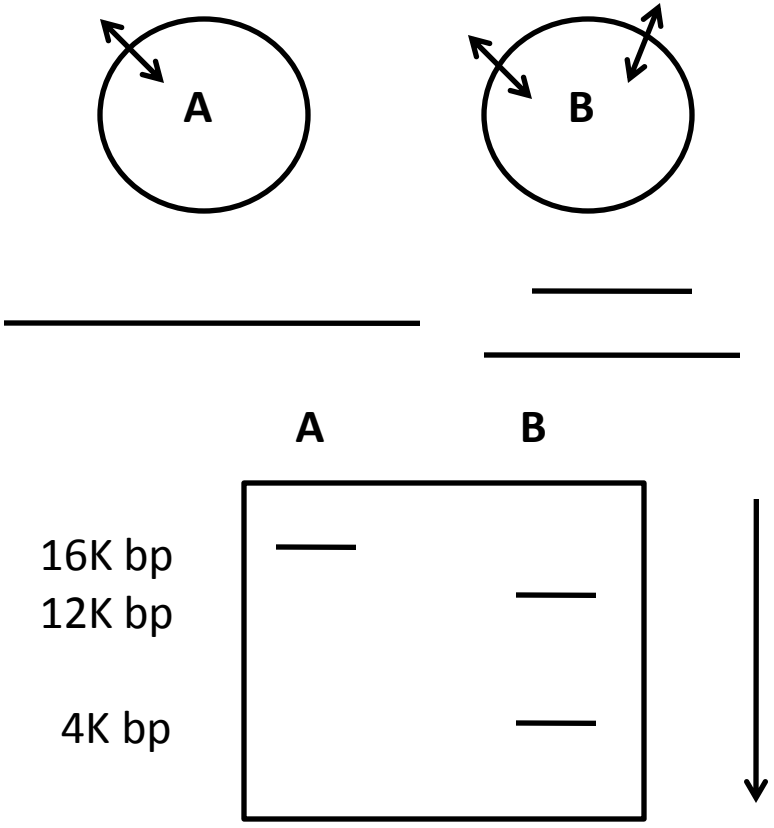
5’-XXXX**G**˅**AATTC**XXXX-3’

3’-XXXX**CTTAAG**˄XXXX-5’

5’-XXXX**G** **AATTC**XXXX-3’

3’-XXXX**CTTAA** **G**XXXX-5’

•E.g. sequence differences in a 16K base pair mtDNA revealed by restriction enzyme analysis



Microsatellites
(nuclear DNA)

- Microsatellites have become the most frequently used DNA marker in conservation genetics.

- Consist of short tandem repeats in a DNA sequence, e.g., cgtcgtcgtcgtcgt or $(cgt)^5$, usually between 75 and 300 bp long.

- Not expressed in that they do not code for proteins and are considered 'selectively neutral'.

- Highly polymorphic (many alleles) even in rare species such as polar bears (due to high mutation rates), ~10^{-5} to 10^{-3} mutations per locus per generation.

Microsatellite Applications

- Detecting population structure within species
- Detecting differences between closely related species
- Bottleneck testing and effective population size estimation
- Assigning individuals to populations
- Estimating migration and gene flow
- Each individual's genotypes also serves as a "genetic tag"

  -Individual identification, parentage assignment, kinship analyses

Advantages of Microsatellites

- **Locus-specific** - the identity of loci are known, in contrast to multi-locus markers such as minisatellites, AFLPs, RAPDs.

- **Codominant** - heterozygotes can be distinguished from homozygotes, in contrast to RAPDs and AFLPs (dominant markers).

- **PCR-based** in that only tiny amounts of tissue and can work on non-invasively sampled or degraded DNA

- Highly **polymorphic** ("hypervariable") due to high mutation rates (10^{-5} to 10^{-3} mutations per locus per generation)

- Useful at a **range of scales** from individual ID to fine-scale phylogenies

Limitations of Microsatellites

•Microsatellite DNA is rarely useful for higher-level systematics because mutation rates are too high.

-Primer sites may not be conserved - primers used for Species A may not even amplify in Species B.

-**Homoplasy** becomes much more likely in distantly related species and we can no longer safely assume that two alleles **identical in state** are **identical by descent** (from a common ancestor).

• Cost and time limits number of microsatellite loci that can be employed (usually 10-20).

Limitations of Microsatellites

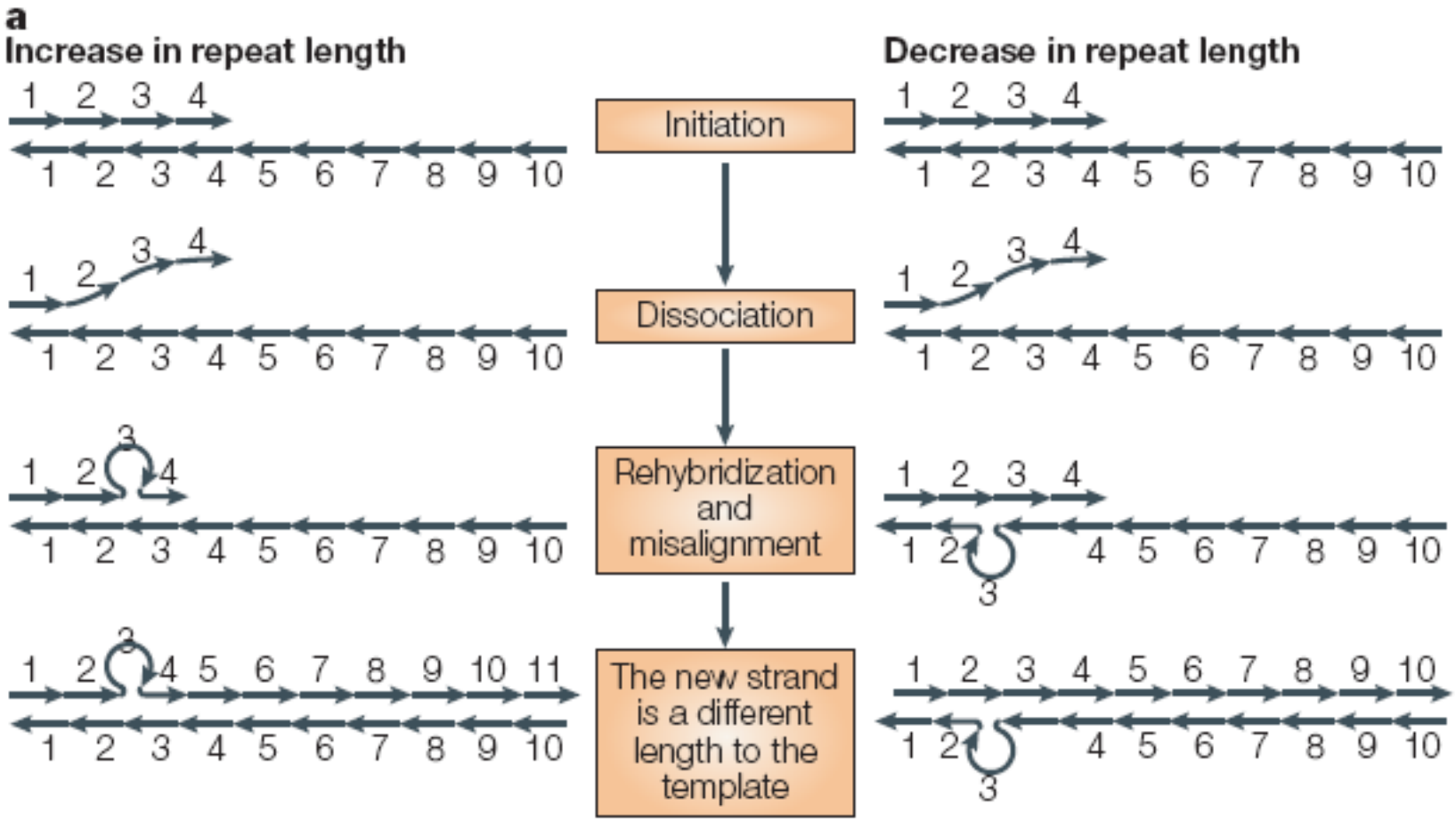
- **Null alleles -** one of the two alleles in a heterozygote does not amplify due to a mutation in a primer binding site, giving the appearance of a homozygote.

- **Allelic dropout** - one of the two alleles (usually of the largest allele) in a heterozygote does not amplify, giving the appearance of a homozygote.

- Results include the mis-identification of individuals, biased allele frequencies, etc…

Mutation Process for Microsatellites

•Mutations in microsatellites are due to ‘strand slippage’ during DNA replication and results in the ‘offspring allele’ having greater or less than the number of repeats of the ‘parent allele’



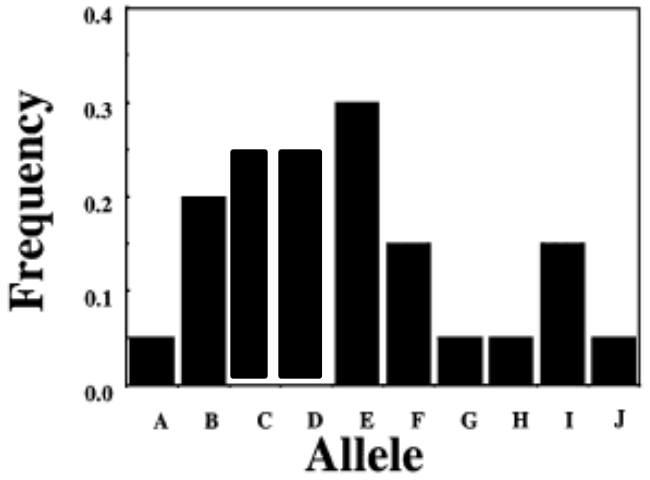
Microsatellite Mutation Models

- Number of repeats added or lost as a result of mutations
  - -Stepwise mutation model (SMM) – each mutation results in the loss or gain of a single repeat
  - -Infinite allele model (IAM) – each mutation results in a new and unique allele such that there are an infinite number of possible alleles
  - -Two-phase mutation model (TPM) – a combination of single- and multi-step mutations. But at what frequency do multistep mutations occur and how large are such mutations
- Different motifs (e.g. di- vs tetra-nucleotides) can have different mutation processes
- Possible limits to the number of repeats in a microsatellite
- Possible biases towards additions or deletions of repeats
- Presence of imperfect repeats (e.g., GATAGATA**C**GATAGATA)

Why are Mutation Models Important to Consider?

Understanding demographic history depends on it…

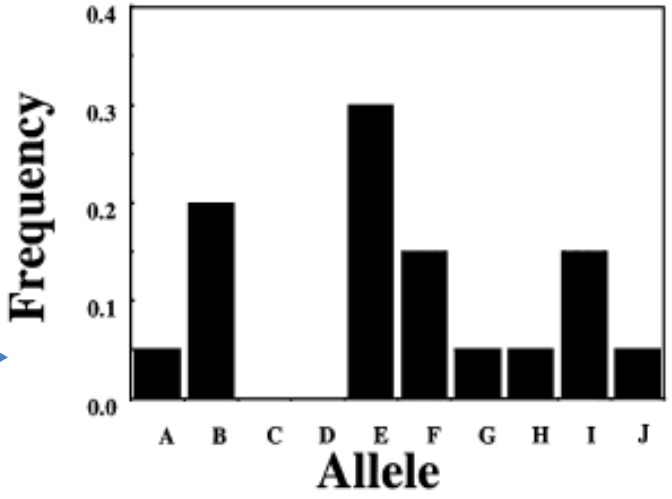
Non-bottlenecked population with microsatellite evolving according to a SMM



Bottlenecked population with microsatellite evolving according to a SMM

OR

Non-bottlenecked population with microsatellite evolving according to a TPM

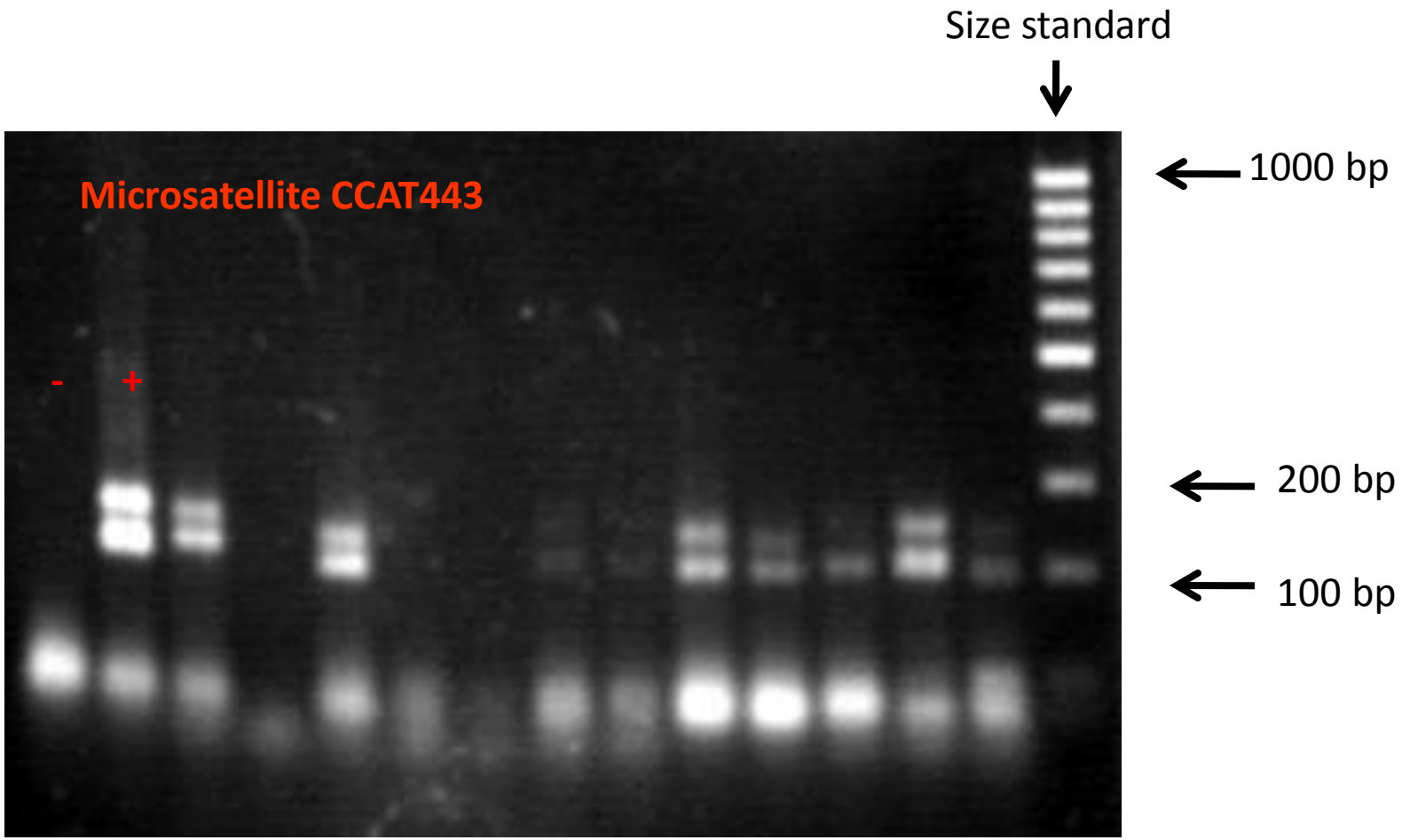


Flow Chart for Working with Microsatellites

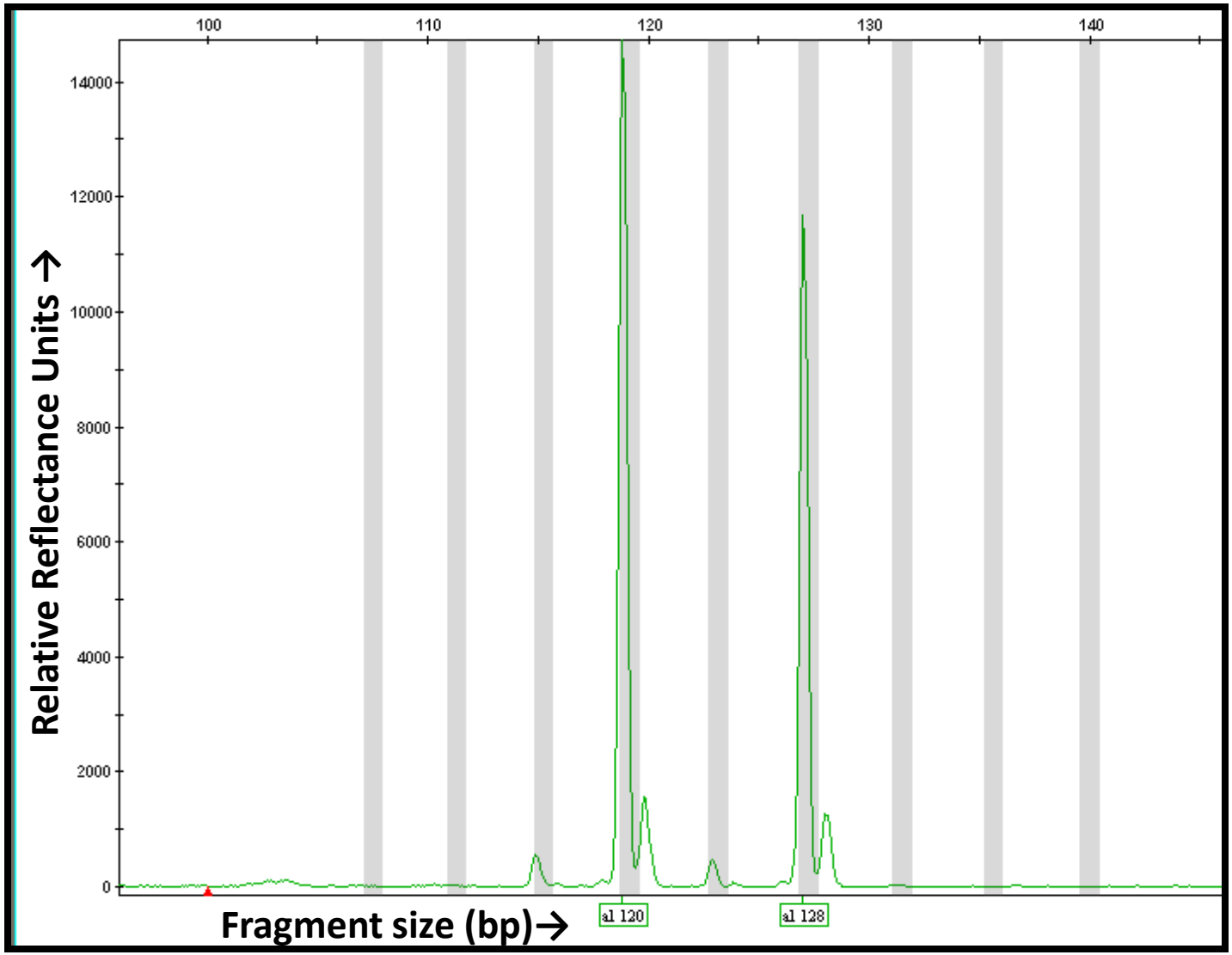
- Development, selection, and optimization of primers/loci
  - Use primers developed for other species, or
  - Develop species specific primers
- Collected tissue
- Extract DNA from tissue
- Amplify extracted DNA using selected primers in PCR reactions
  - Attach fluorescent label to primers for subsequent allele sizing (also known as calling or scoring)
  - Loci can be amplified individually or “multiplexed”
- Load labeled PCR product (and size standard) onto a sequencer
  - Analyze sequencer output and score alleles

Microsatellites

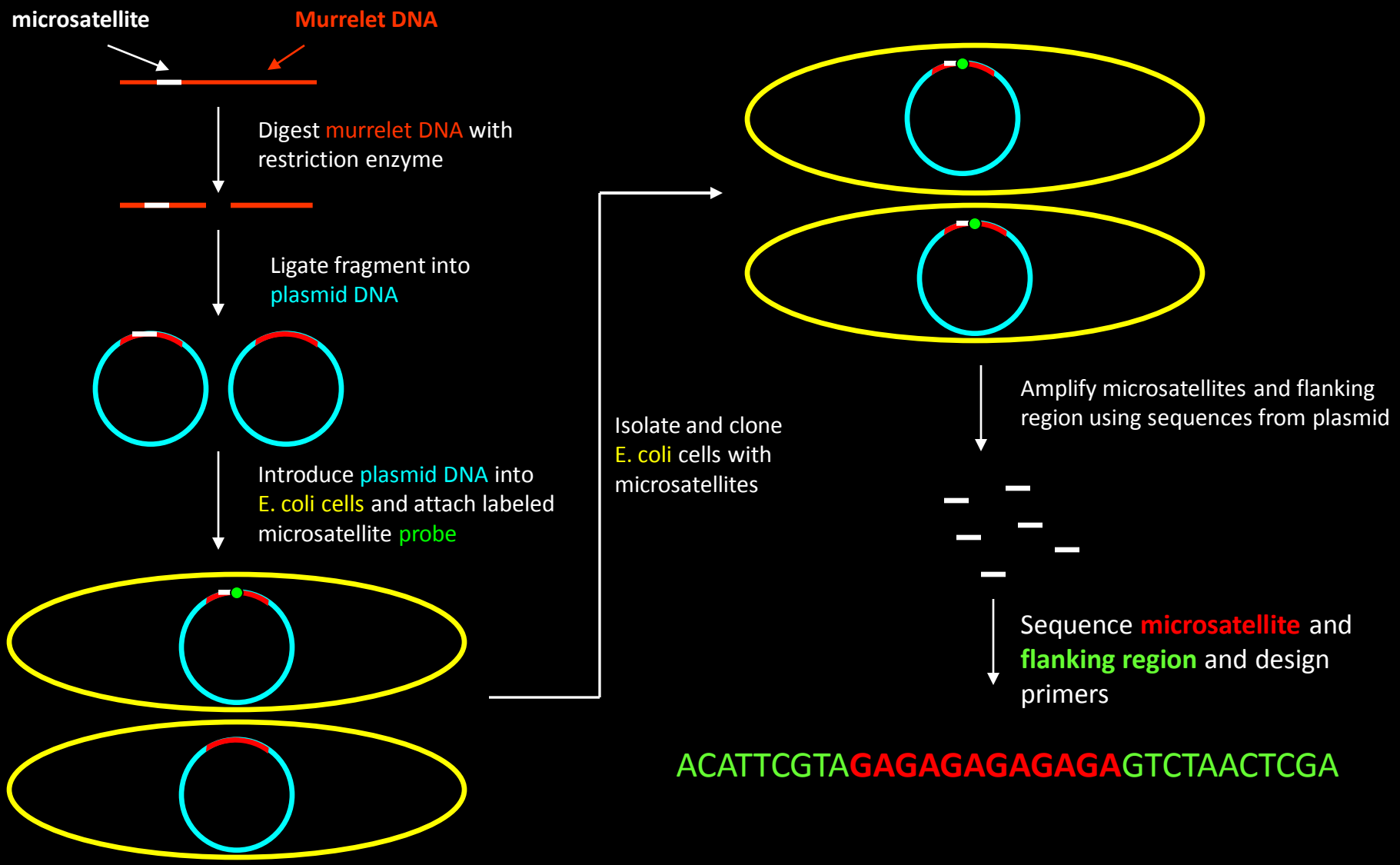
Amplified microsatellite viewed on an agarose gel



A Microsatellite Electropherogram (from an automatic sequencers)

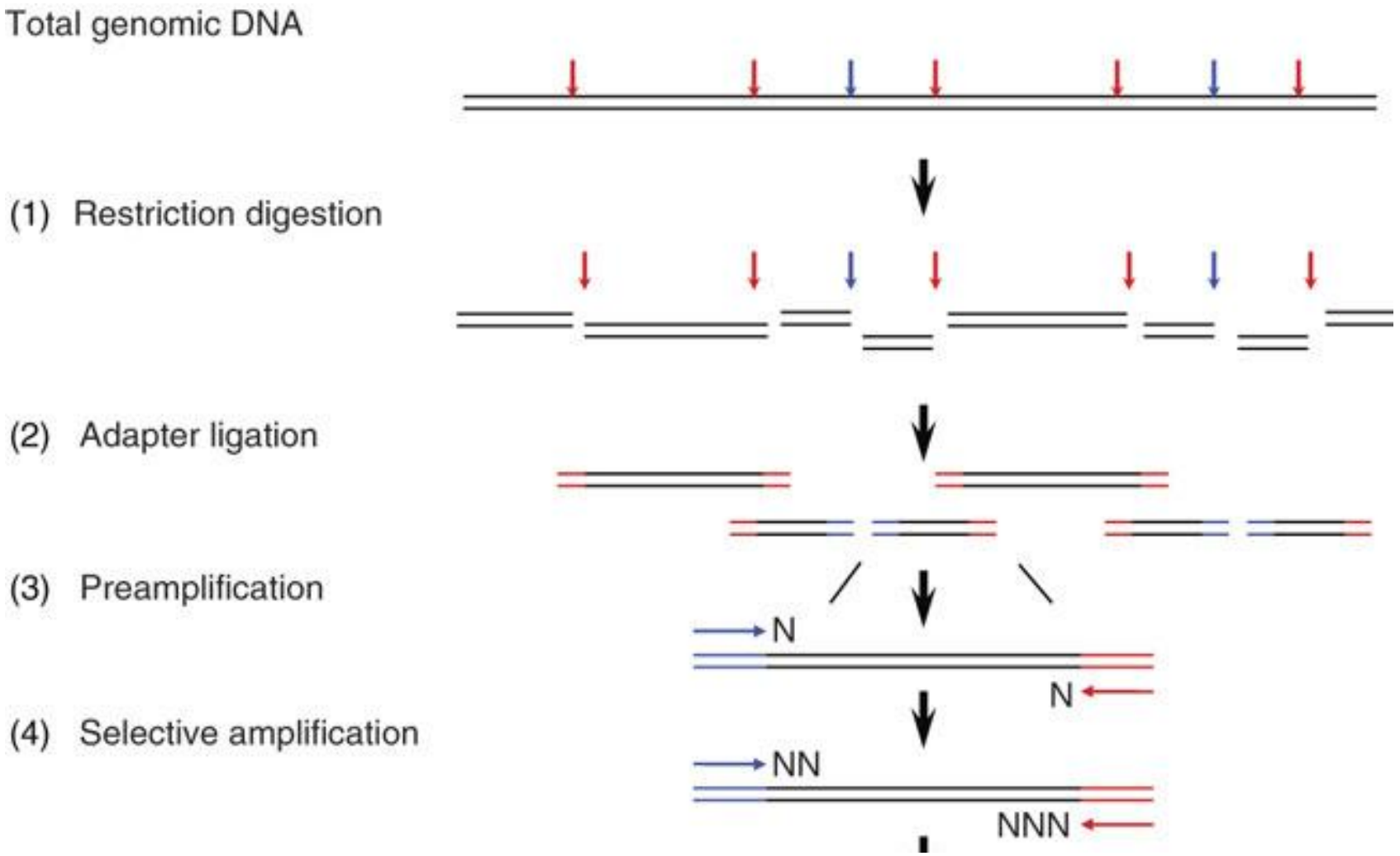


Cloning a Microsatellite Library



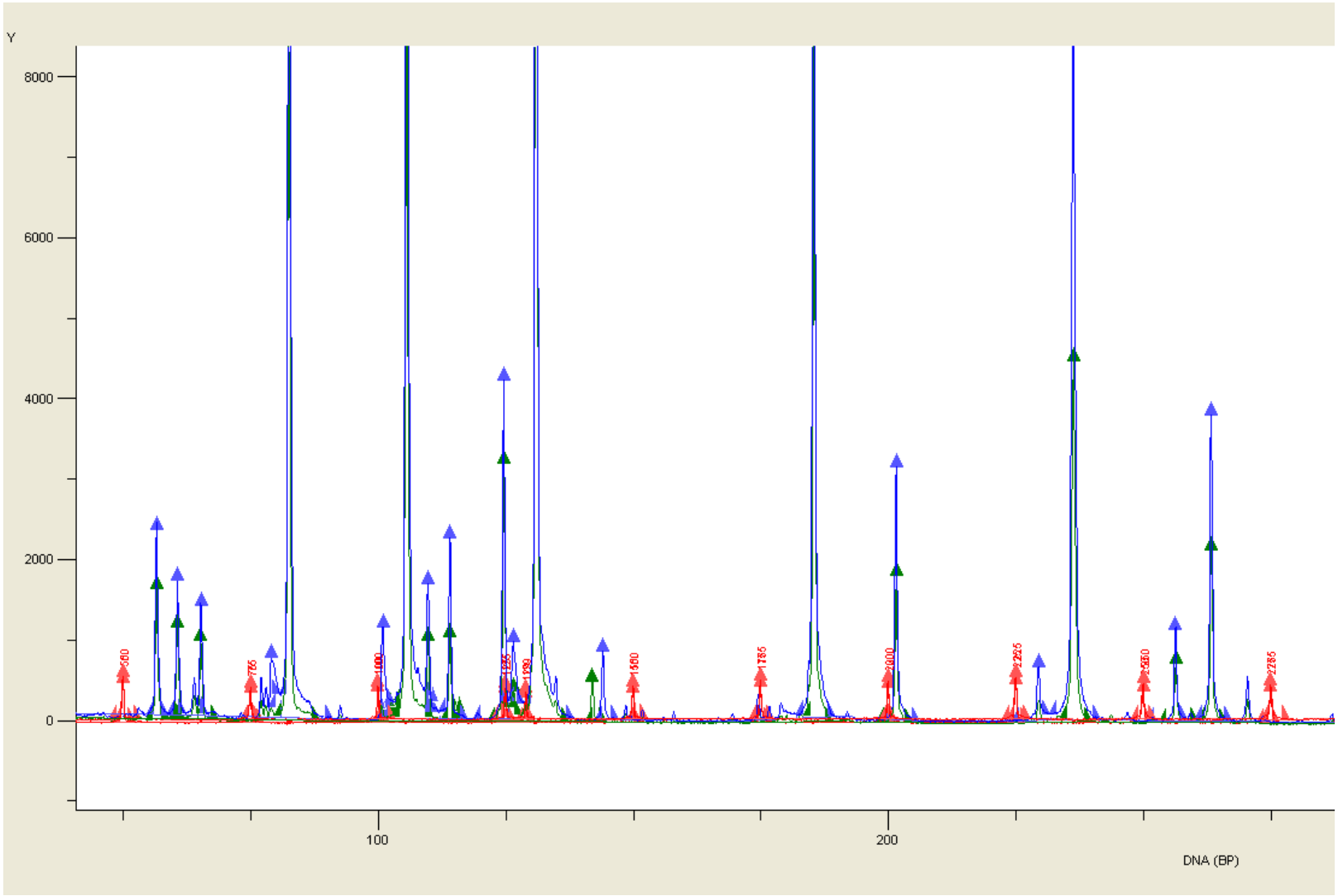
Locus	Primer sequences	T_a	Cycles	N	No. of alleles	Size range (bp)	H_O	H_E	P value	I	GenBank Accession n
BmaAGGT503	F: CTCAGCAAAACCAGGAAAATA	59/53	22/8	15	4	218–260	0.47	0.55	< 0.2141	0.25	DQ173166
	R: TTTAAGTCTAATATTGGTCTCTCAGC										
BmaATAC370	F: CCTGATGACCTTTGATGGCTCT	55/53	24/8	15	6	186–204	0.87	0.74	< 0.9127	0.11	DQ173167
	R: ACCTGTGCCTGCGTTGGT										
BmaATTT351	F: TGGGAATATCTTTTGGTTTGG	59/53	22/8	15	4	165–207	0.53	0.73	< 0.0450	0.12	DQ173168
	R: TCCAGCCTTTCCTTGTCTCTA										
BmaCCAT301	F: AGATCTATCCCTTGGCTGGA	59/53	22/8	15	6	152–172	0.87	0.78	< 0.8229	0.079	DQ173169
	R: TATCTGCCAAAATCTGCTGAA										
BmaCCAT443	F: TGCCAGGCCATCTACTTTAAT	59/53	22/8	15	9	178–214	0.93	0.85	< 0.8630	0.037	DQ173170
	R: GCTTATCTTTCCCTCCATCCT										
BmaGACA340	F: GGCCATCTGAGTTGGATAAAA	59/53	22/8	15	2	136–140	0.40	0.32	< 0.9999	0.51	DQ173171
	R: GTTGGGTGGATCATGGTTTAG										
BmaGACA456	F: ACTGGTCTCTTTGCTTGATGG	59/53	23/12	14	4	395–407	0.64	0.68	< 0.3909	0.16	DQ173172
	R: GGAAGAGCACACCTTTACCAG										
BmaGATA365	F: GCTTTATCTGTGGCAACACTG	59/53	22/8	15	7	225–253	0.80	0.73	< 0.7765	0.10	DQ173173
	R: GCTGTAGGGAGGATATGATGC										
BmaGATA439	F: GAGGGGAGGGTGTATCTTTTC	59/53	22/8	15	9	315–351	0.80	0.78	< 0.5944	0.068	DQ173174
	R: ATGTCACTCTGGTGGAGAACC										
BmaGATA464	F: GCACCATGCTCAGATCACTAA	59/53	23/12	15	6	414–438	0.47	0.66	< 0.0272	0.14	DQ173175
	R: ATCTGTGCTTGAGGGAGAGAA										
BmaGATA465	F: TCAGAGGGGGAAACAACATAG	59/53	22/8	15	12	245–303	0.47	0.88	**< 0.0001**	0.028	DQ173176
	R: GGGAATTTGCATTCAGTCTGT										
BmaGATA553	F: TTGTGAGAGGGTCACTTATCAAAT	59/53	22/8	15	8	136–165	0.73	0.78	< 0.2850	0.069	DQ173177
	R: CATCTCTCTTTCAGAAGAGCAGTC										
BmaGGAT313	F: CTCTAAAGGTCCCTTCCAACC	59/53	22/8	15	5	235–251	0.73	0.77	< 0.3592	0.088	DQ173178
	R: TGACTTCACAGTTCCTCATGC										
BmaGGAT368	F: AATCACCAAGGATAAAGGATGATA	59/53	22/8	15	11	212–293	0.93	0.87	< 0.7955	0.029	DQ173179
	R: AGGGGACCTGCCCATATATTA										
BmaGTTT332	F: TCTCCAAATCCAGAAAAATGG	52/53	22/8	15	4	171–197	0.27	0.60	**< 0.0011**	0.22	DQ173180
	R: ATAATCCTGTGAGGGGTTTCC										
BmaGTTT428	F: GCATGTAACAAGTCCATTTGC	52/53	22/8	15	2	143–147	0.13	0.39	< 0.0188	0.45	DQ173181
	R: CAGGGGCAGCTTAAGTAAAGT										

AFLP procedure



http://www.nature.com/nprot/journal/v2/n6/fig_tab/nprot.2007.175_F1.html

AFLP trace

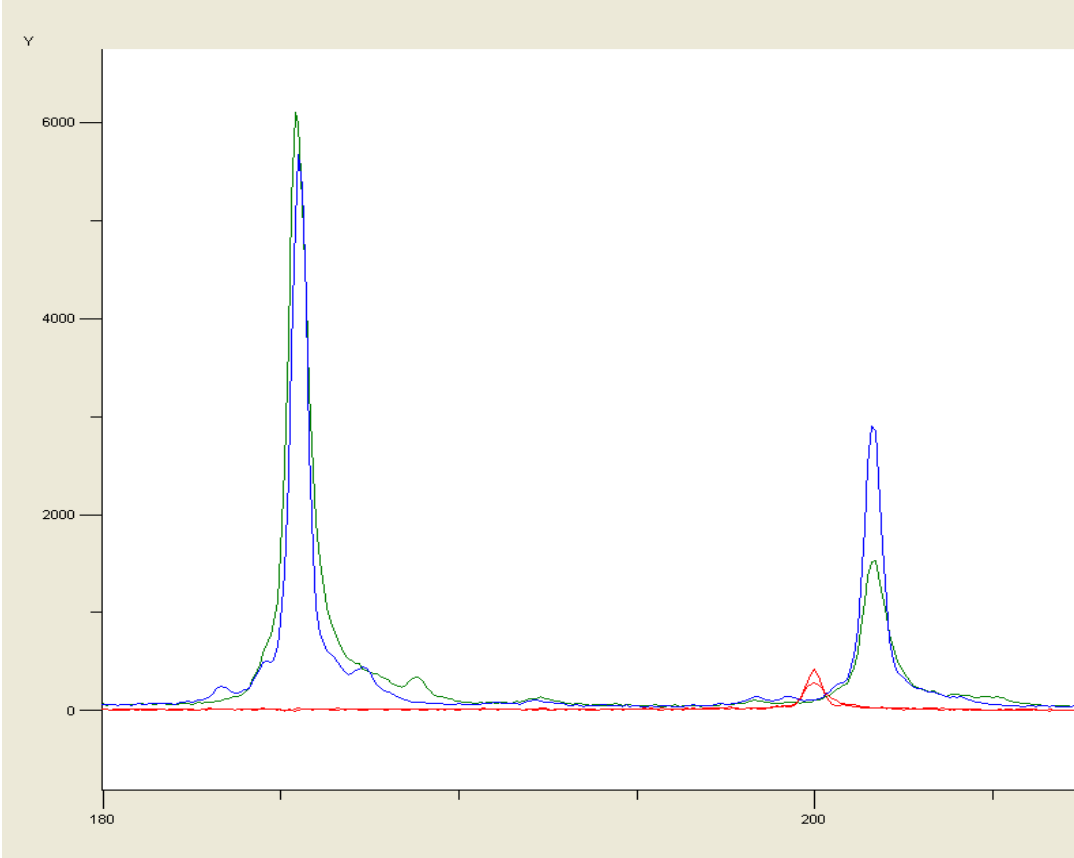


Advantages of AFLPs

- Quickly scan entire genome for polymorphisms
- Large number of loci sample
- Replicability
- High resolution
- No prior knowledge of genetic makeup required
  - Across taxa
- Inexpensive

Limits of AFLPs

- Dominant marker
  - However, can be genotyped
    - By assessing the relative amount of DNA within AFLP markers
    - Technically demanding
- Less information per loci than microsats
- Difficulty in amplifying from ancient DNA



Single Nucleotide Polymorphisms (SNPs)

- SNPs - nucleotide site in a DNA sequence where more than one nucleotide (G, A, T or C) is present in the population.

- Any polymorphic nucleotide is a locus – lots of loci possible (26,000 described for humans by Akey et al. 2002), but only 2 alleles per locus so more loci will be needed

- Using many (~hundreds) loci means that some loci will be physically linked and not statistically independent samples of the demographic and evolutionary history of a population or species.

- SNP discovery in non-model organisms still challenging.

Marker	Genome	Cost	Develop Time	Inheritance	Typical Applications
Microsatellites	Nuclear	Med	High	Co-dominant	Individual identification, kinship, population structure, demographic history
DNA fingerprints	Nuclear	Med	Low	Dominant	Individual identification
RAPDS	Nuclear	Low	Low	Dominant	population structure, demographic history
AFLPS	Nuclear	Med	Low	Dominant	population structure, demographic history
RFLPs	mtDNA	Med	Low	Co-dominant	Population structure, phylogeography
DNA sequences	Nuclear and mtDNA	High	None	Co-dominant	Species ID, population structure, phylogeography, phylogenetics
SNPs	Nuclear	Med	High	Co-dominant	Same as msats