Molecular markers in plant systematics and population biology

1. Introduction – overview of the methods

Tomáš Fér

tomas.fer@natur.cuni.cz

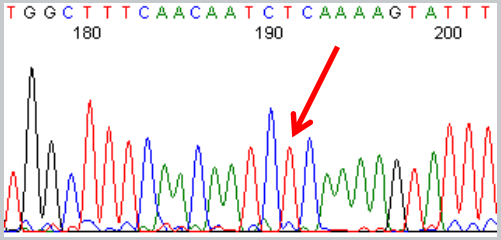
What are molecular markers ?

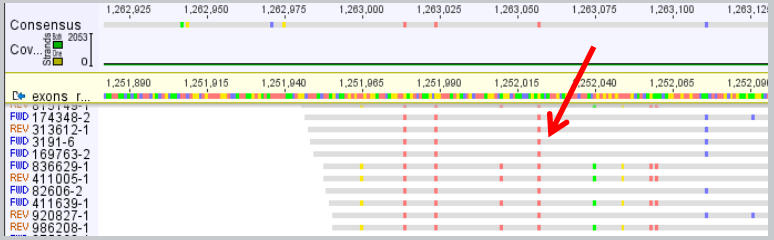
- information about an organism obtained from analysis of its molecules – proteins, DNA, RNA
- marker – information unit – targeted or randomly chosen part of the total information
- markers tell about genetic similarity (kinship) of individuals, populations, species or higher taxa

Marker examples

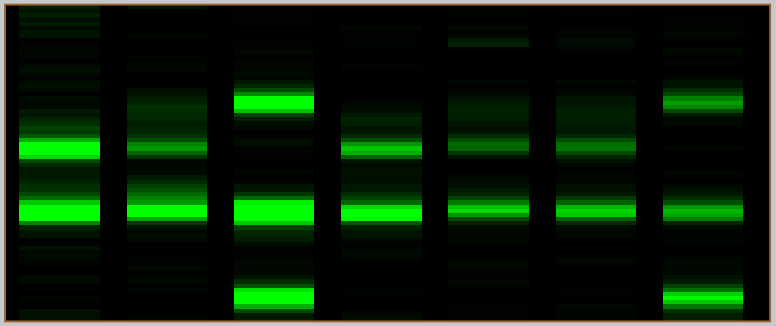
- information about enzyme molecule (e.g., its charge and mobility)

- sequence of nucleotides in DNA chain or specific (variable) position (SNP – single nucleotide polymorphism)





- length of DNA fragment



Aspects of molecular data

- give an information about individual genotype/haplotype
- information not dependent on environment conditions (no plasticity)
- neutral vs. non-neutral markers
  - selective neutrality – i.e., no influence on individual fitness
    - non-coding DNA regions or synonymous substitutions
    - identical functionality of all isozymes
  - non-neutrality – under selection
- qualitative information – fragment presence, allele, nucleotide
- unique information about organism – clone identification
- subject to modification during generative reproduction – recombination

Deposition of genetic information

nucleus

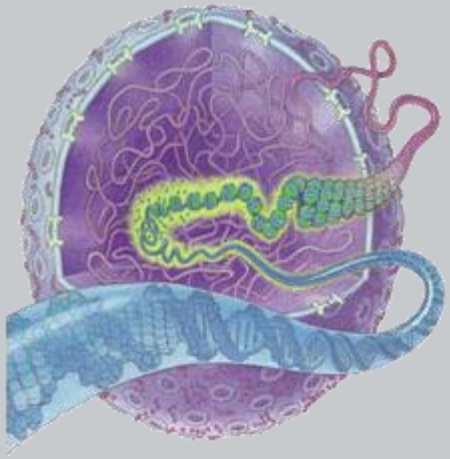
- different ploidy level
- recombination
- biparental transfer

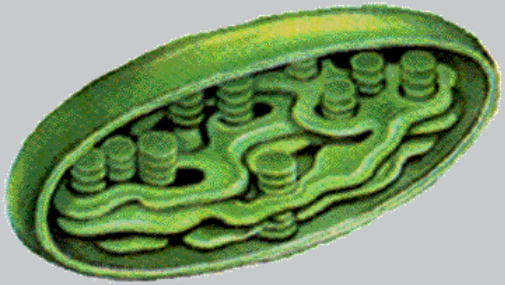
plastids

- 1 circular molecule
- without recombination
- uniparental transfer

mitochondrion

- structurally complicated
- common restructuring
- uniparental transfer





Genome characteristics

	nDNA animals	plants	cpDNA	mtDNA animals	plants
heritability	biparental	biparental	angiosperms maternal, conifers paternal	maternal	maternal
structure	linear	linear	circular	circular	circular, complex
size (kb)	$4.9 \times 10^4 – 7.0 \times 10^8$	$5.0 \times 10^4 – 3.0 \times 10^8$	71 – 214	15 – 20	200 – 2400
substitution rate	3.5×10^{-9}	$4.1 – 5.7 \times 10^{-9}$	$0.86 – 1.20 \times 10^{-9}$	56×10^{-9}	$0.36 – 0.50 \times 10^{-9}$
substitution rate relative to plant mtDNA	8.1	11.4	2.4	130.2	1.0
foreign sequences	common	common	rare	rare	common
structural mutations	common	common	rare	rare	common
recombination	yes	yes	intramolecular	no	inter- and intramolecular

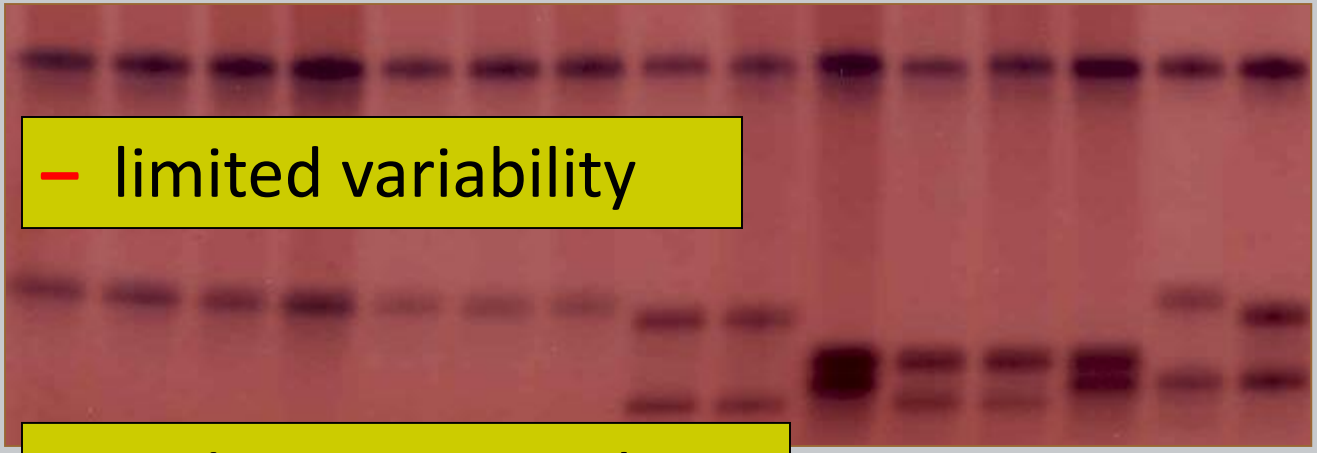
Lowe et al. 2004

Molecular markers overview

1. proteins – isozymes
2. DNA markers
   - RFLP (**R**estriction **F**ragment **L**ength **P**olymorphism)
   - PCR based – analysis of DNA fragments
     - order of nucleotides – DNA sequences
     - „whole genome“ analysis – fragment length polymorphism
       - RAPD (**R**andom **A**mplified **P**olymorphic **D**NA)
       - AFLP (**A**mplified **F**ragment **L**ength **P**olymorphism)
       - ISSRs (**I**nter **S**imple **S**equence **R**epeats)
     - information from specific genome regions
       - PCR-RFLP (**P**olymerase **C**hain **R**eaction – RFLP)
       - microsatellites (**S**imple **S**equence **R**epeats – SSRs)
       - SSCP (**S**ingle **S**train **C**onformation **P**olymorphism)…
   - whole genome markers – SNP, whole genome sequencing
     - RADseq
     - Hyb-Seq (target enrichment)
     - de novo sequencing, re-sequencing
     - RNA-seq (transcriptome)

Isoenzymes (isozymes, allozymes)

- proteins catalysing basic biochemical reactions
- extracted from living tissues – preferably leaves
- individual molecules (=alleles) electrophoretically separated according to differences in their mobility (electric charge)
- visualisation by „color reactions“



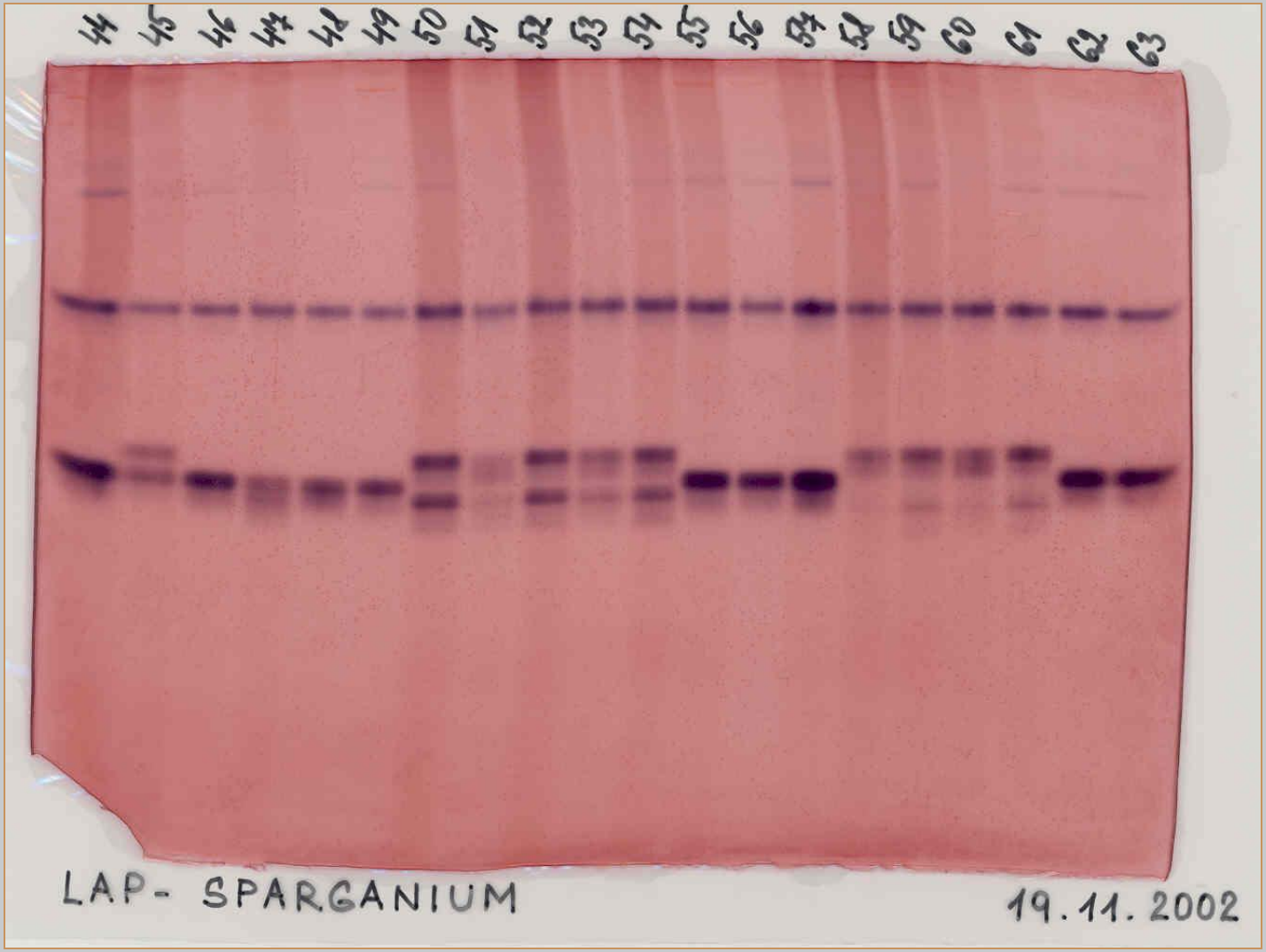
– limited variability

+ codominant marker

1st locus
1 allele

2nd locus
5 alleles

Isoenzyme gel example



RFLP

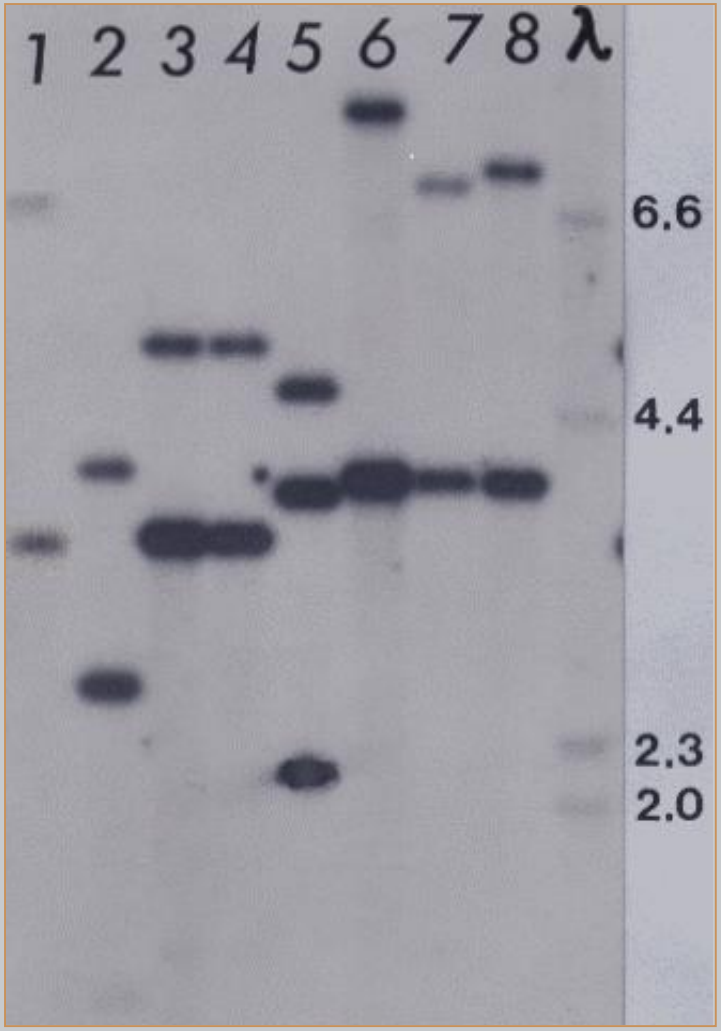
Restriction **F**ragment **L**ength **P**olymorphism

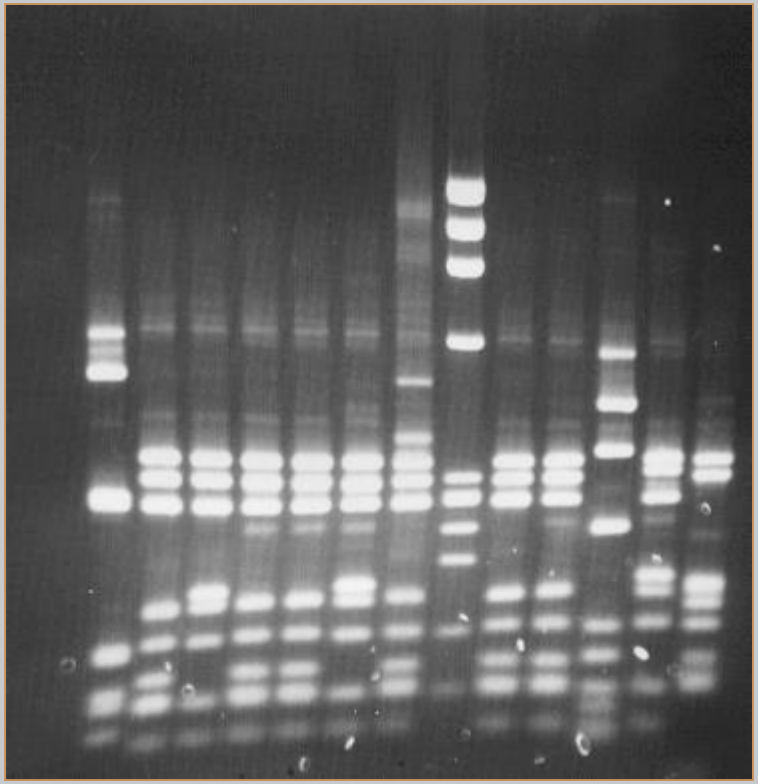
- DNA is specifically cleaved to fragments by restriction endonuclease
- electrophoresis – length separation
- large amount of fragments – specific part is visualized by hybridization with labelled probe (*Southern blotting*) – e.g., visualization of cpDNA only
- variability – insertion/deletion or mutation in restriction site

\+ highly reproducible pattern

– labelled probes required

RFLP gels examples

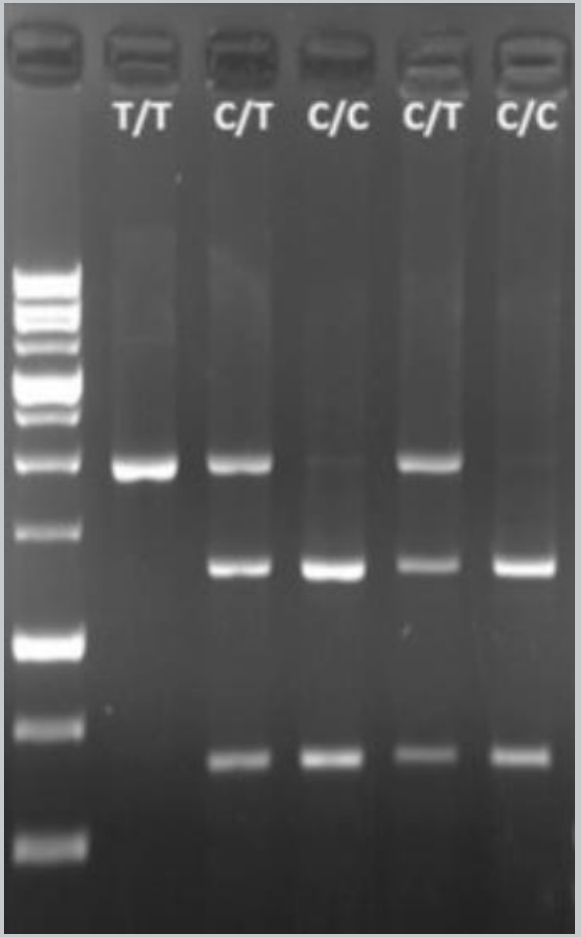




http://www.ufpe.br/biolmol/Tec-mol-biol/RFLP-real.JPG

PCR-RFLP

- amplification of specific DNA region using PCR with two specific primers
- use of consensual primers (e.g., cpDNA)
  - applicable to almost all plant species
- restriction of amplified region by different restriction enyzmes
- electrophoresis of fragments



Lye et al., 2021

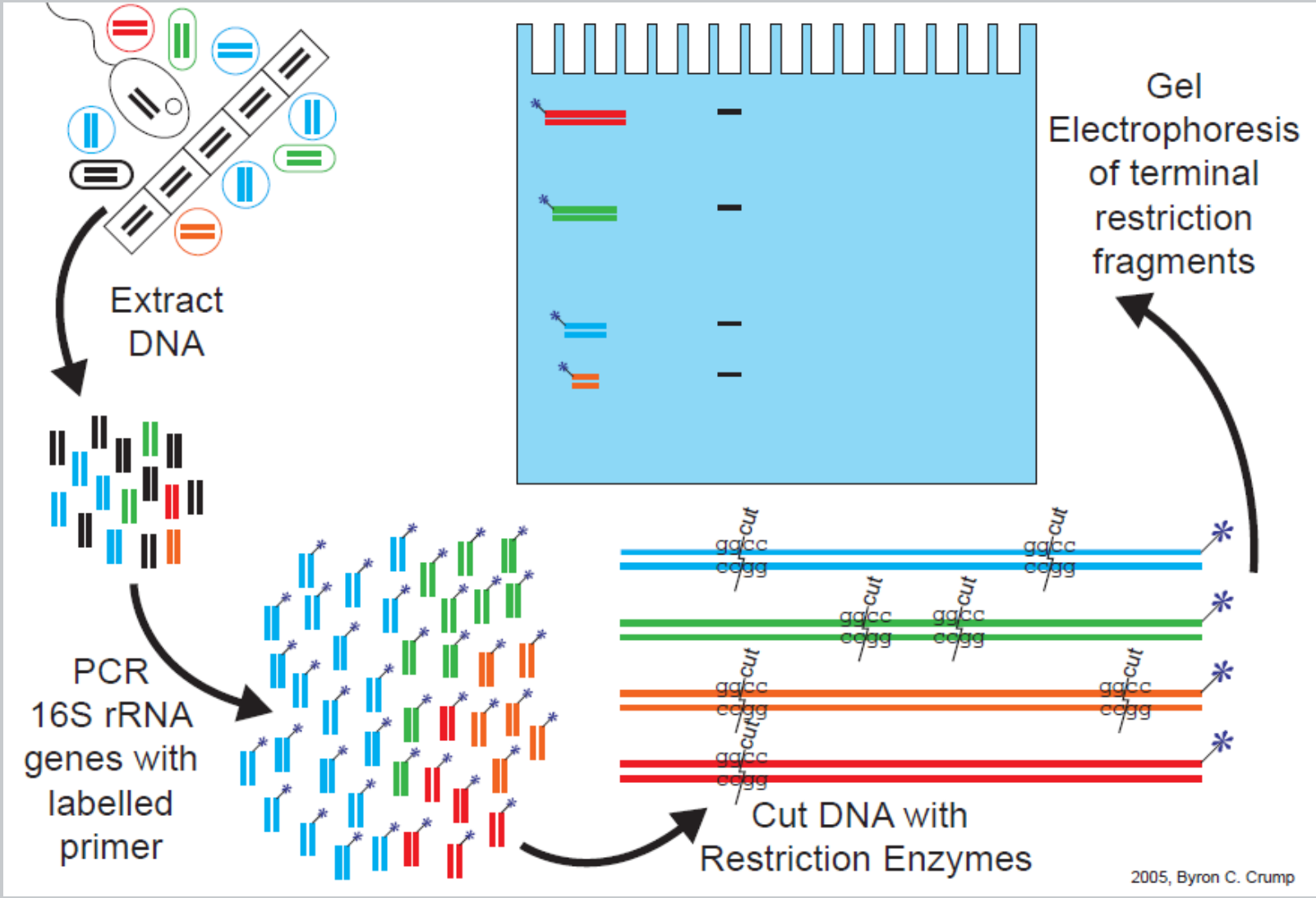
\+ universal method

\+ codominat marker when ITS is used

– lower variability when cpDNA is used

tRFLP

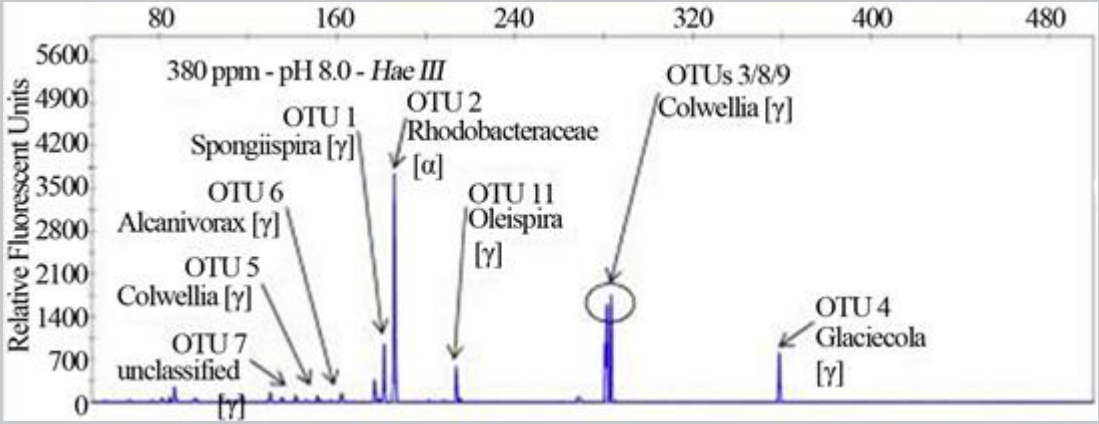
terminal - **R**estriction **F**ragment **L**ength **P**olymorphism



http://hpl.umces.edu/faculty/bcrump/TRFLP.pdf

tRFLP

terminal - **R**estriction **F**ragment **L**ength **P**olymorphism

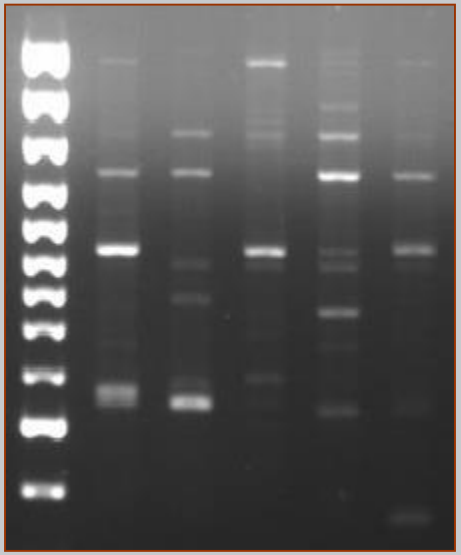


Siu N. et al. (2014). Open J. Ecol. 4: 434-455

RAPD

Random **A**mplified **P**olymorphic **D**NA

- fragments are generated using PCR with one arbitrary primer (decanucleotide)
- electrophoretic separation of fragments according to their length
- polymorphism is caused by
  - mutation in the place where primers anneal (*priming site*)
  - insertion/deletion in the amplified DNA

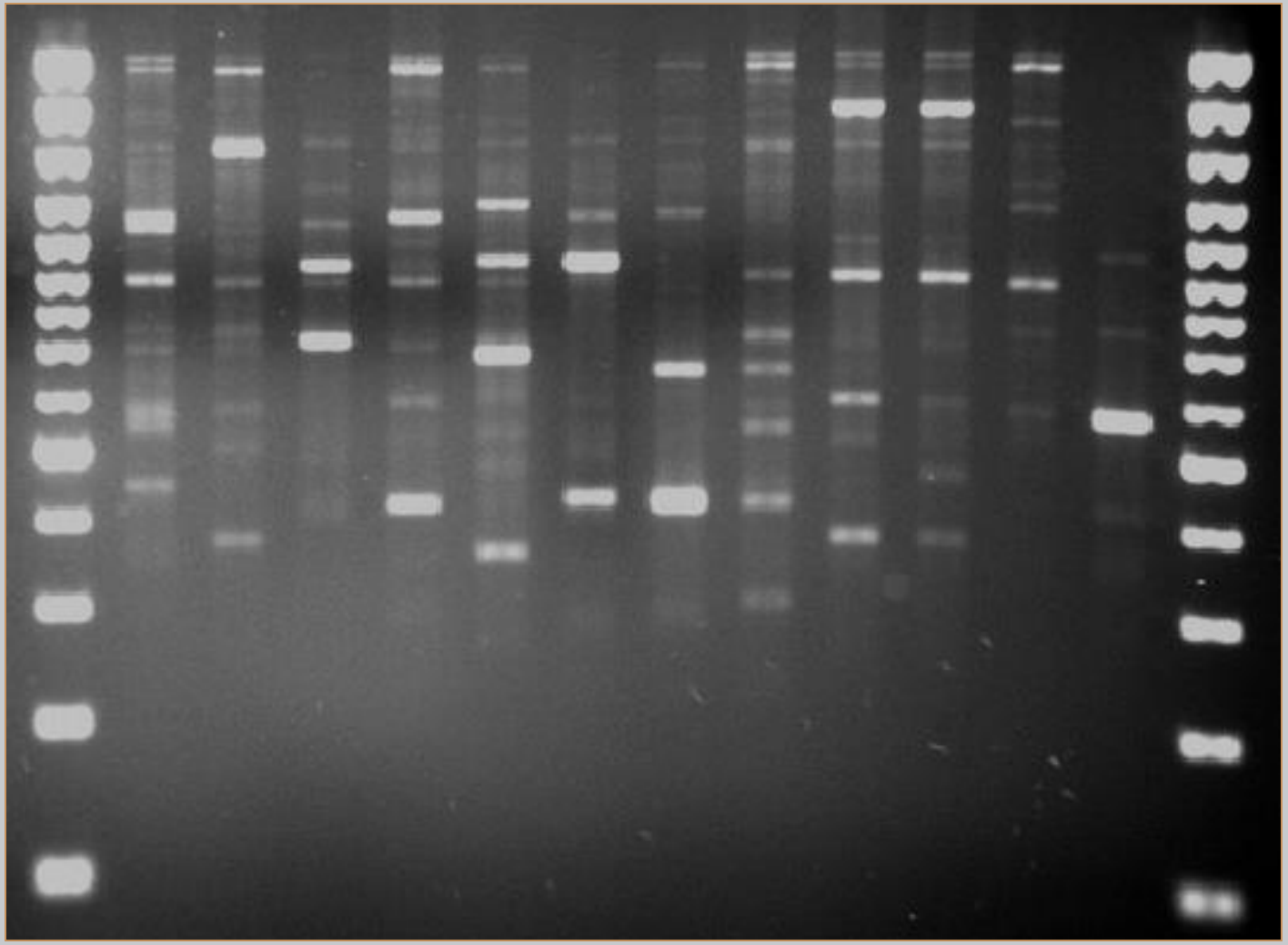


\+ simple method

– dominant marker

– results difficult to reproduce

RAPD gel example



AFLP

Amplified **F**ragment **L**ength **P**olymorphism

- combination of RFLP and subsequent PCR of selected fragments
- DNA restriction by two different enzymes
- selective amplification of a subset of fragments
- fluorescence visualization of fragments on the gel (using automated sequencer)

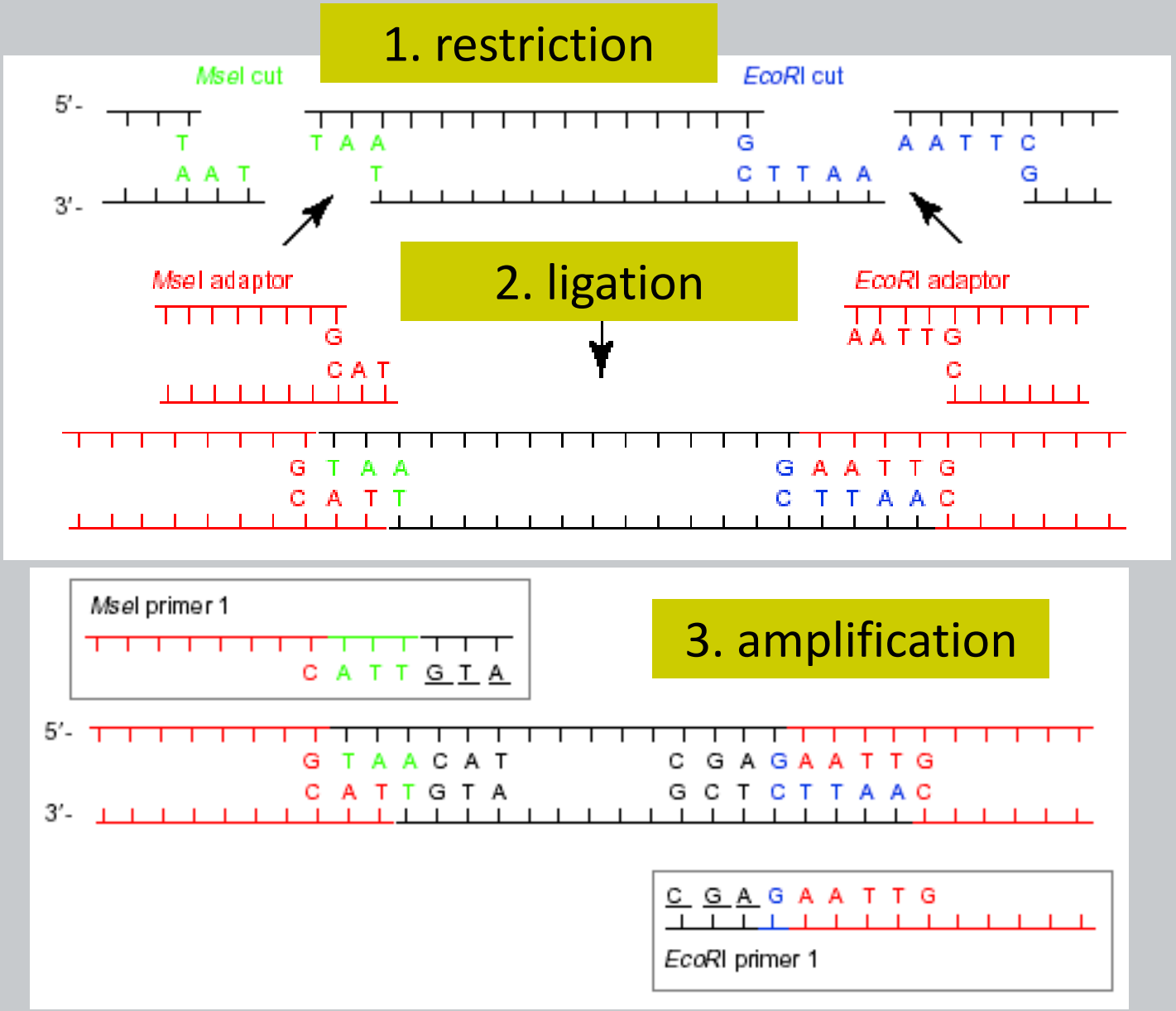
\+ highly polymorphic, reflects variability of „whole“ genome

\+ high reproducibility, reliability

– dominant marker – homo- a heterozygotes cannot be distinguished

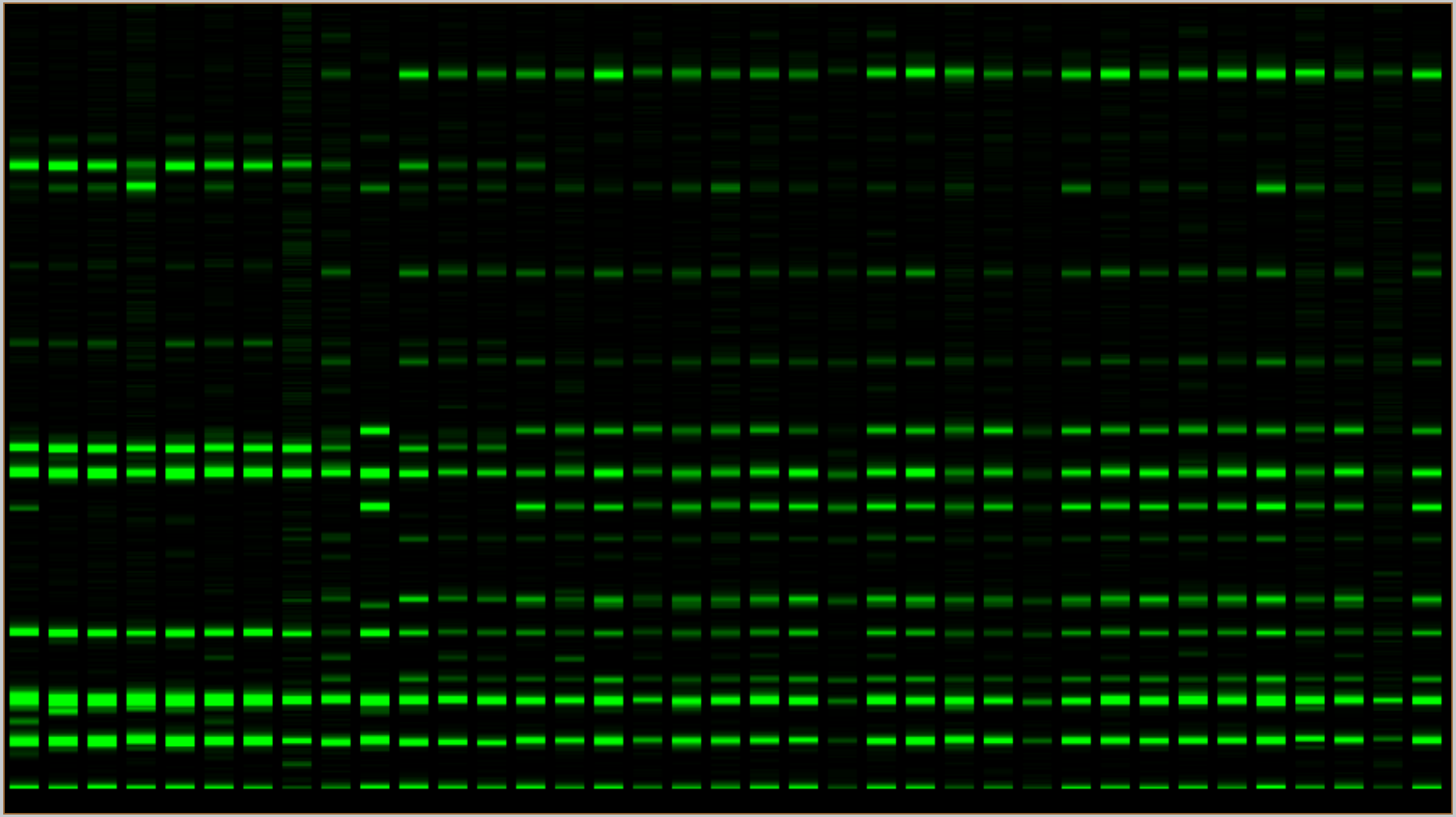
AFLP – Amplified Fragment Length Polymorphism

total DNA



Mueller & Wolfenbarger (1999), TREE

AFLP gel example



Microsatellites

SSRs – **s**imple **s**equence **r**epeats

- tandem repeats of several nucleotides

```
AGGCTATATATATAGGCA  1

AGGCTATATATA--GGCA  2
```

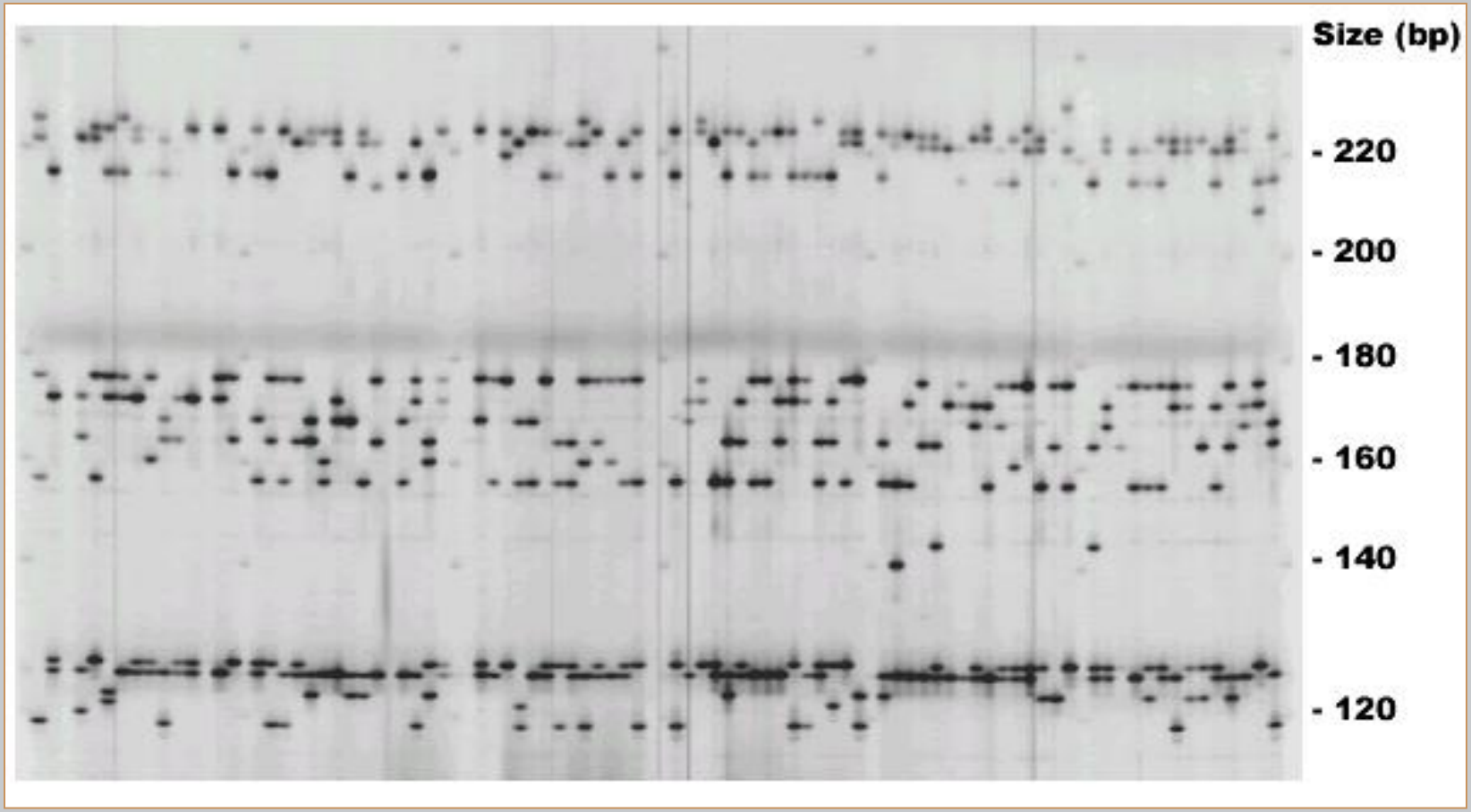
- alelles – differ by number of repeats

\+ codominant marker, highly variable

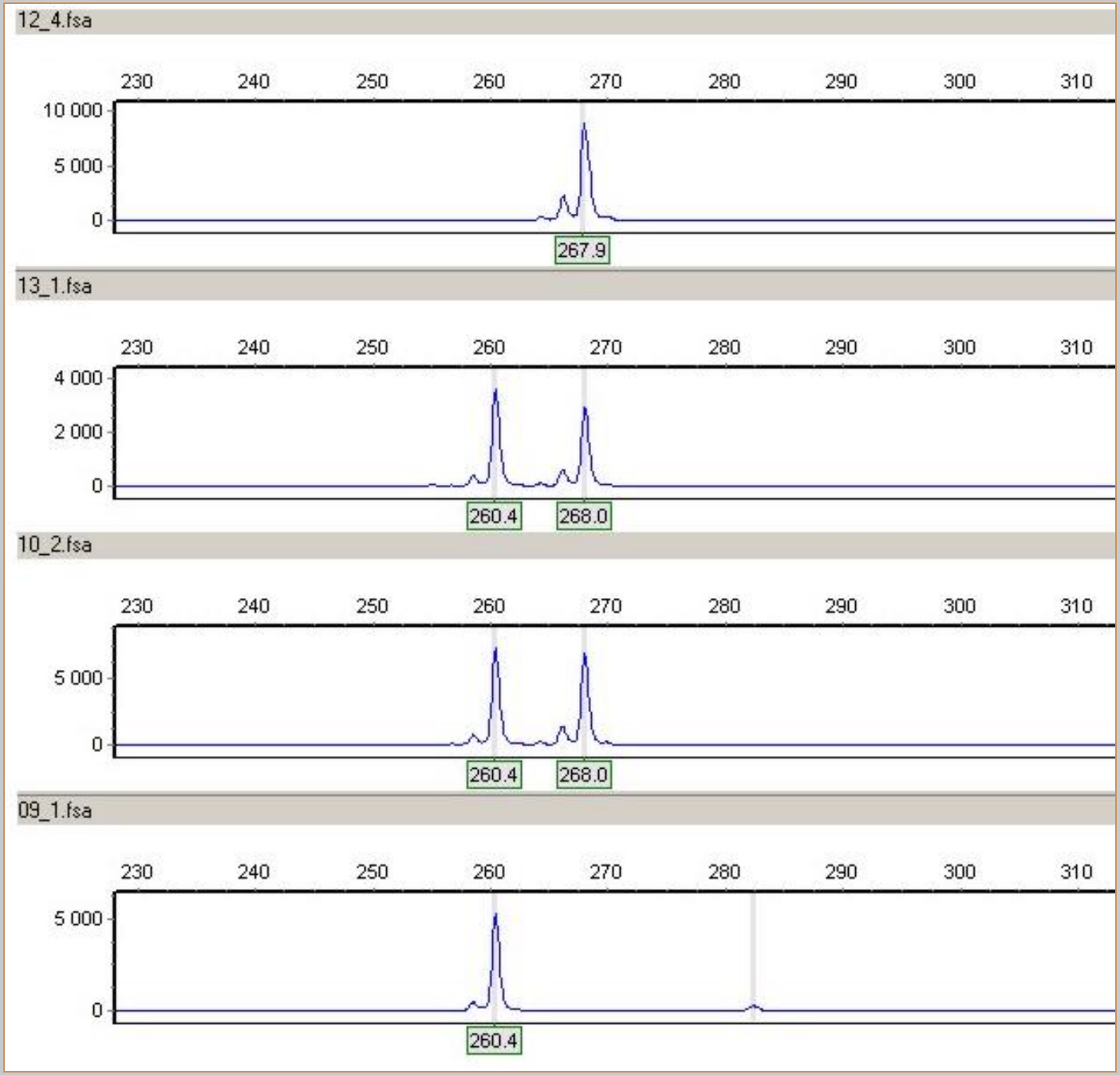
\+ relationships among alleles can be assessed

– necessery to develop primers for the study species

Example of microsatellite analysis



http://www.qub.ac.uk/bb-old/prodohl/gel/Images/multiplex_set_1b.gif



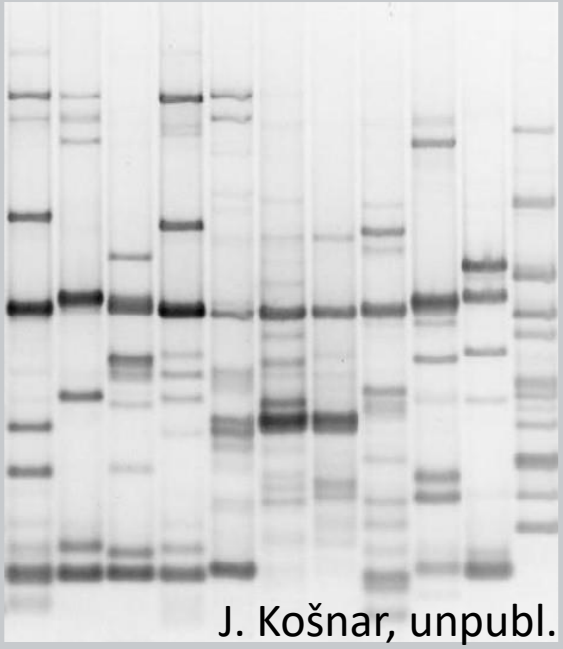
Microsatellites - homozygotes and heterozygotes (*Nuphar lutea*)

ISSRs – **I**nter **S**imple **S**equence **R**epeat**s**

- length variation of regions between microsatellite loci
- primer – microsatellite sequence

\+ variable marker, simple

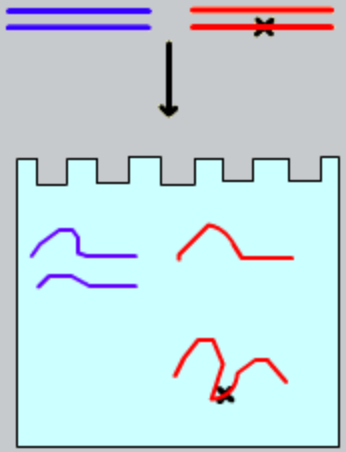
– dominant marker



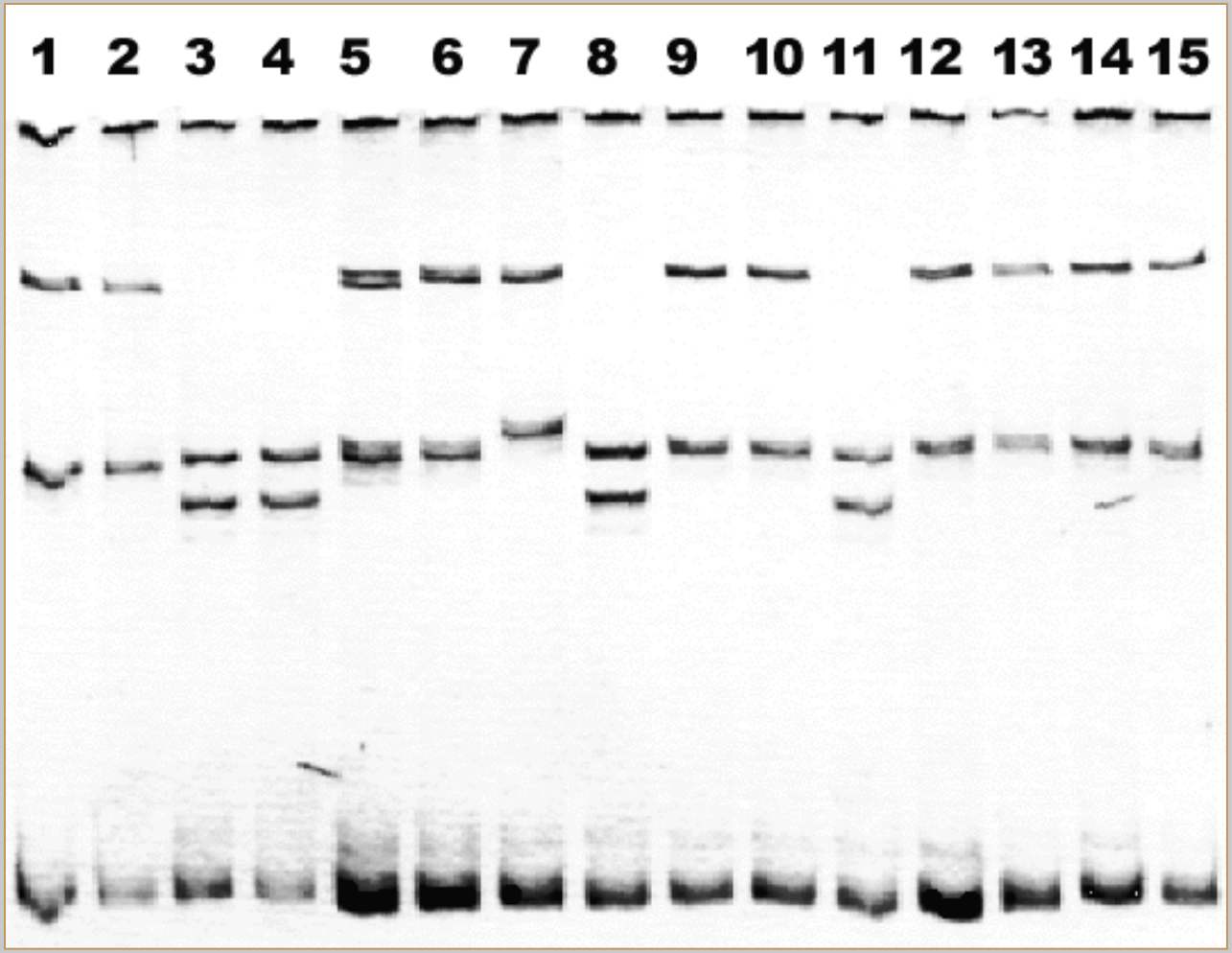
J. Košnar, unpubl.

SSCP – **S**ingle **S**trand **C**onformation **P**olymorphism

- method for finding unknown point mutation
- PCR amplification of target region
- denaturation – electrophoresis of ssDNA
- mutation changes tertiary structure (conformation) of the chain and thus its mobility in gel

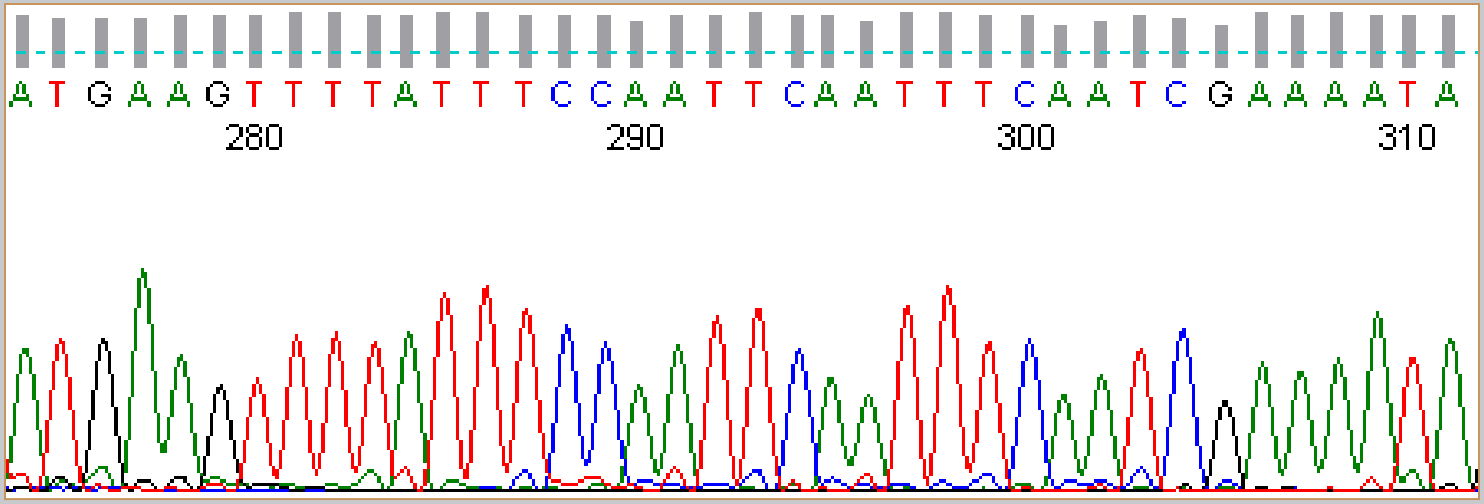


SSCP gel example



DNA sequencing

- determination of the sequence of nucleotides in DNA chain
- Sanger sequencing – use of automated sequencers – *fluorescence base labelling*
- specific primers for PCR amplification of the target region



Sequencing

1. coding genes – *conservative*
   - systematics at the level of families, genera (*rbc*L*)*
2. spacers, introns – *variable regions*
   - systematics at the level of genera, species and below (*trn*L-*trn*F, *atp*B-*rbc*L, ITS)

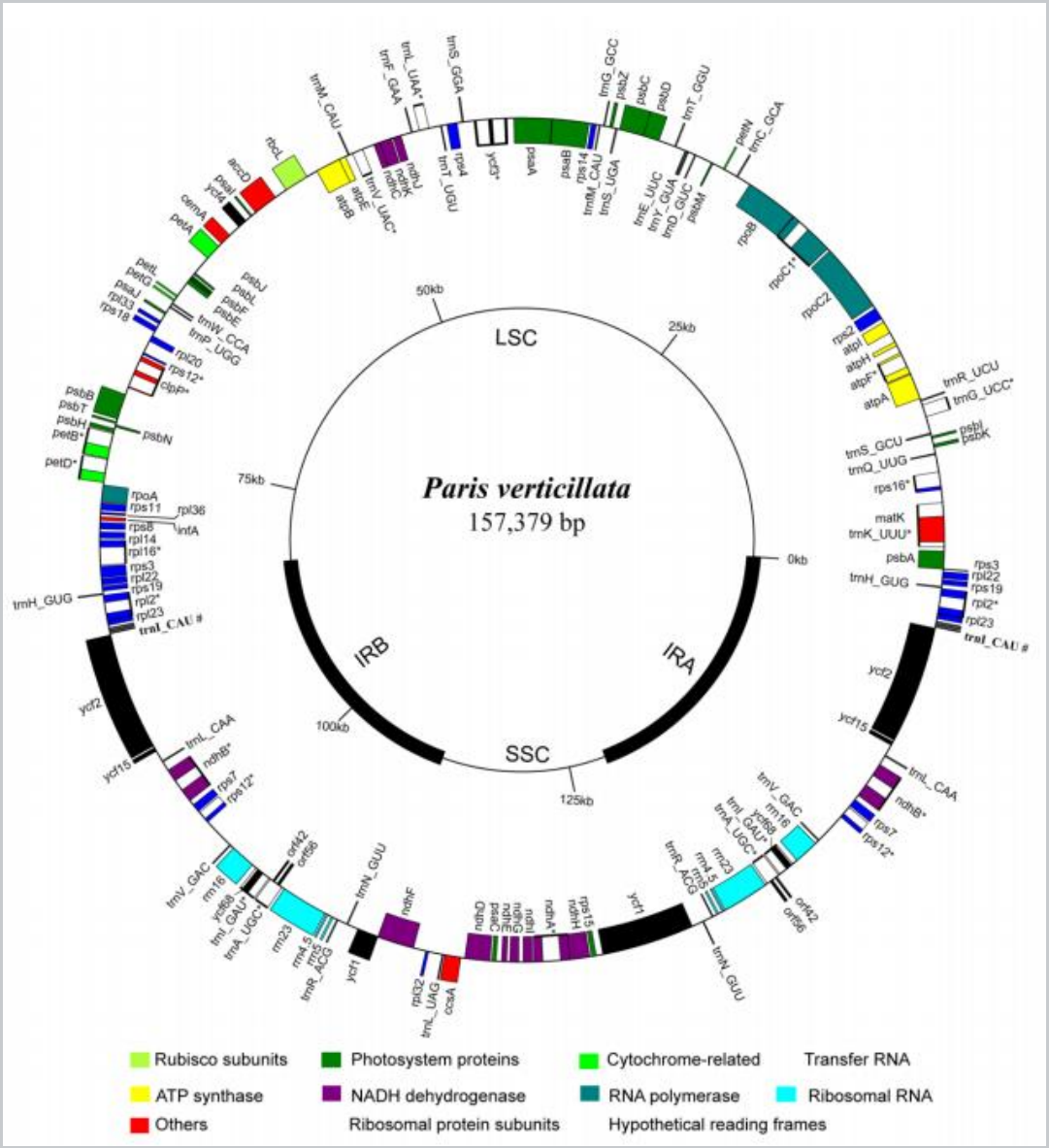
chloroplast genes

- *rbc*L
- *atp*B
- *mat*K …

nuclear genes

- ITS
- 18S rDNA
- 26S rDNA …

Chloroplast DNA



Khoa et al., 2014

Next generation sequencing – NGS

masivelly parallel sequencing, high-throughput sequencing

- parallel sequencing of millions of fragments
- bioinformatics to deal with huge amount of information

several platforms

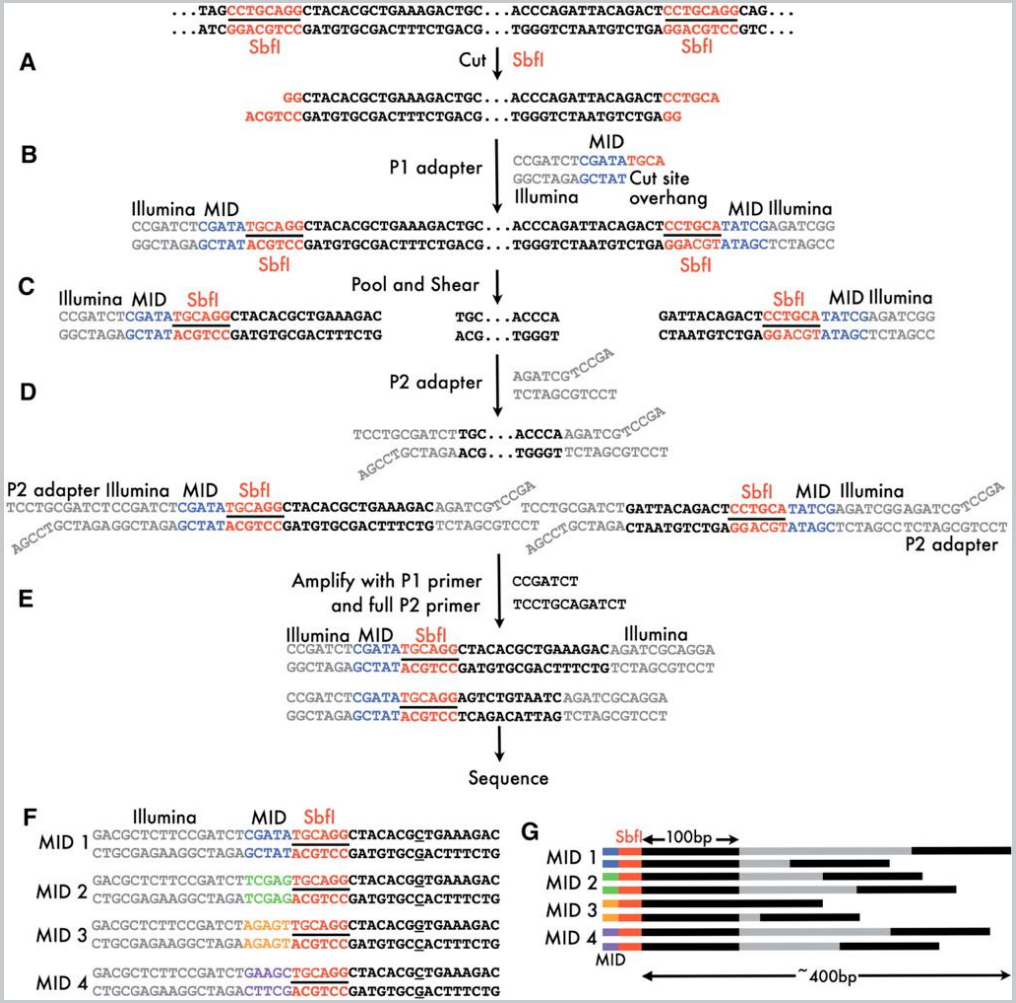
- Roche (454) – pyrosequencing
- Illumina (Solexa)
- SOLiD (ABI)
- Pacific Biosciences
- Ion Torrent
- Oxford Nanopore…

diverse approaches/applications

- shotgun (re)sequencing (genome, transcriptome – RNAseq)
- sequence capture (target enrichment) – Hyb-Seq
- restriction-based reduction – RADseq
- amplicon sequencing
- metasequencing …

RAD-sequencing

Restriction-site-associated DNA sequencing



Davey J.W. & Blaxter M.L. (2011): *RADSeq: next-generation population genetics*. Briefings in Functional Genomics 9: 416-423.

Davey J.W. et al. (2011): *Genome-wide genetic marker discovery and genotyping using next-generation sequencing*. Nature Reviews 12: 499-510.

SNP

single-nucleotide polymorphism

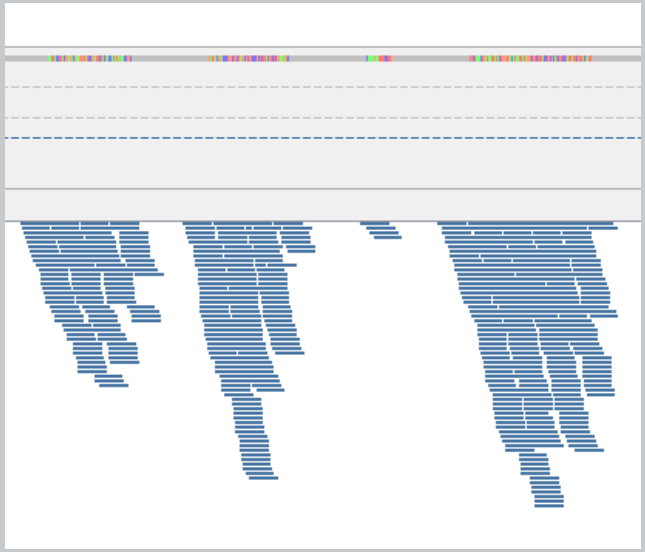
- variability in DNA at one particular site (base mutation)

…ACTGGAGTCGACTG…
…ACTGGAGACGACTG…

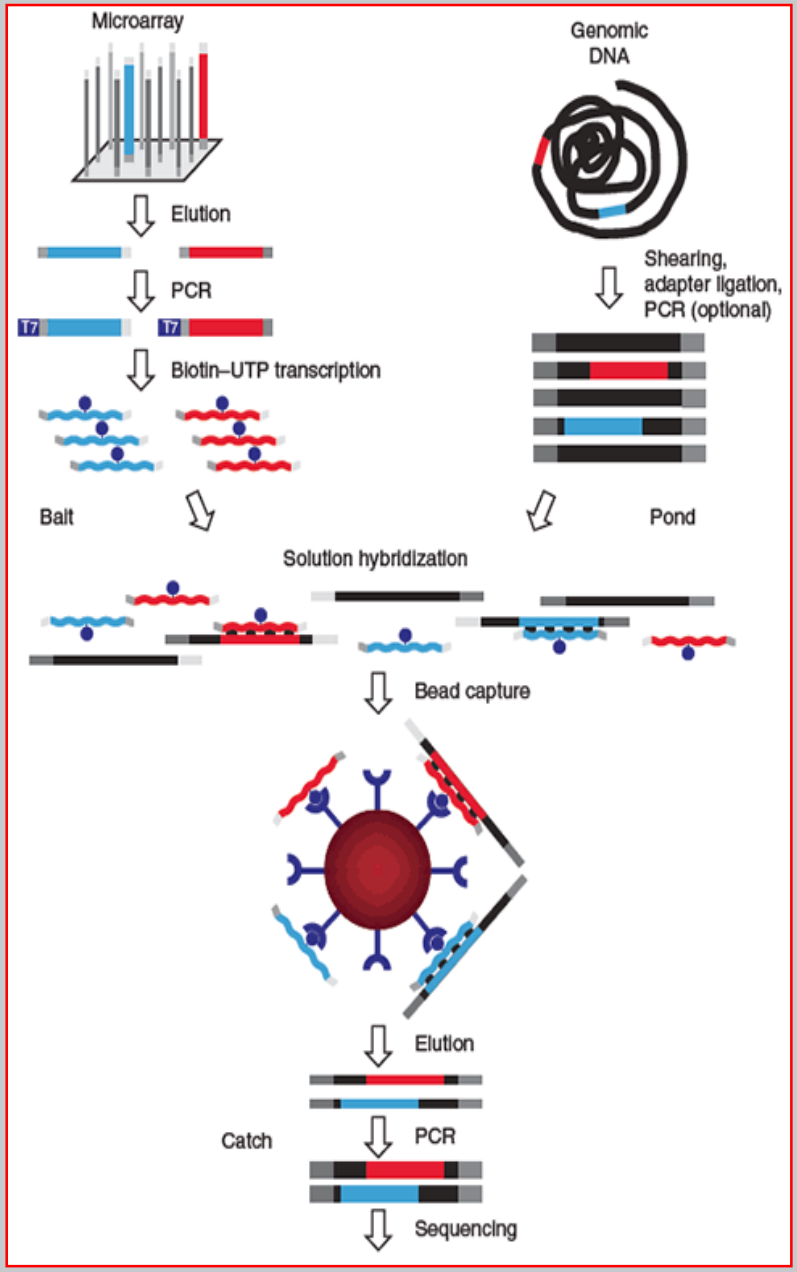
- mostly biallelic – two variants (e.g., A/T)
- codominant marker
- known SNPs × new variants detection
- detection (see http://en.wikipedia.org/wiki/SNP_genotyping)
  - sequencing
  - PCR-RFLP
  - SSCP, TGGE, DGGE (conformation based)
  - hybridization (allele-specific probes)
  - MALDI-TOF mass spectrophotometry (Sequenom MassARRAY…)
  - microarrays
  - **NGS (e.g., resequencing, RADseq)**

Hyb-Seq

- solution phase hybridization
- enrichment of target sequences



Weitemier et al. (2014) Appl Plant Sci. 2: apps.1400042
Cronn et al. (2012) Amer. J. Bot 99: 291-311
Lemmon et al. (2012) Syst. Biol.
McCormack et al. (2012) Syst. Biol.
Bi et al. (2012) BMC Genomics



Arbor Biosciences (Mycroarray)

Types of molecular markers
important differences

- variability high × low

high	**low**
microsatellites	allozymes
AFLP	chloroplast markers
SNPs	

Types of molecular markers
important differences

- variability — high × low
- heritability — dominant × codominant

dominant	**codominant**
AFLP	allozymes
RAPD	microsatellites
RFLP	(PCR-RFLP of, e.g., ITS)
	SNPs

Types of molecular markers
important differences

- variability — high × low
- heritability — dominant × codominant
- recombination — yes × no

yes	no
nuclear markers (diploid, polyploid)	organelar (cp, mt) (haploid)

Types of molecular markers
important differences

- variability — high × low
- heritability — dominant × codominant
- recombination — yes × no
- transfer to next gener. — biparental × uniparental

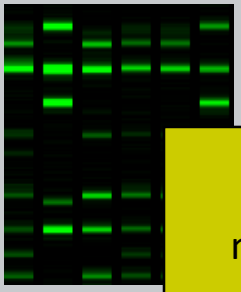
biparental	**uniparental**
nuclear markers	angiosperms – cp and mt DNA maternally (seeds)
(seeds and pollen)	gymnosperms – cp DNA paternally (pollen), mtDNA maternally

Types of molecular markers
important differences

- variability — high × low
- heritability — dominant × codominant
- recombination — yes × no
- transfer to next gener. — biparental × uniparental
- mutation rate — high × low

high	low
microsatellites	allozymes
introns, spacers	exons (esp. 1st and 2nd positions)

Which factors influence markers?



mutations

selection

individuals

DNA markers
non-coding or synonymous mutations

allozyme markers
DNA non-synonymous

gene flow

genetic drift

population level

variability increase

variability decrease

Utility of markers in different types of studies

		Fragment-based			Sequencing				NGS		
	Allo-zymes	RAPD	AFLP	SSR	nDNA	cpDNA	mtDNA (plant)	mtDNA (animal)	Hyb-Seq	RADseq	resequencing
Genetic diversity	++	++	++	++	+++	++	+	++	+++	+++	+++
Population differentiation	+++	++	++	++	+++	++	++	+++	++?	+++	+++
Gene flow	++	(+)	(+)	+++	+++	++	(+)	++	?	+++	+++
Polyploidy	+++	-	(+)	+	++	++	-	-	+++	++	+++
Hybridization	++	++	++	+	++	++	+	+	+++	+++	+++
Phylogeny	(+)	-	++	(+)	+++	+++	(+)	+++	+++	++	+++
Individual genotyping	(+)	+++	+++	+++	+++	-	-	-	?		+++
Phylogeography	(+)	-	++	-	(+)	+++	(+)	+++	(+)	+++	+++
Selection	(+)	(+)	(+)	+	++	-	-	-	++	++	+++
Diversification	?	?	(+)	-	++	++	?	++	+++	+++	+++

+++	excellent	(+)	has been used
++	good	-	unlikely to be usefull or useless
+	OK	?	uncertain or not used

partially based on Lowe et al. 2004

Literature

Besse P. (2014): Molecular Plant Taxonomy. Methods and protocols.

Lemmon E.M. & Lemmon A.R. (2013): High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst*, 44, 99–121.

Henry R.J. (2012): Molecular Markers in Plants.

Höglund J. (2009): Evolutionary Conservation Genetics.

Weising K. et al. (2005): DNA fingerprinting in plants. Principles, methods, and applications.

Lowe A., Harris S. & Ashton P. (2004): Ecological Genetics: Design, Analysis, and Application.

Avise J.C. (2004): Molecular markers, natural history and evolution.

Baker A.J. (2000): Molecular methods in ecology.

Beebee T. & Rowe G. (2004): An introduction to molecular ecology.

Henry R.J. (2001): Plant genotyping. The DNA fingerprinting of plants.

Karp A. et al. (1998): Molecular tools for screening biodiversity.

Jennings W.B. (2017): Phylogenomic data acquisition. Principles and practice.

Molecular markers in botany

1.	(4.10.)	**molecular markers** – characteristics, differences, technique overview
2.	(11.10.)	**molecular markers** – overview of applications and questions
3.	(18.10.)	**isozymes** – electrophoresis, evaluation of codominant data, population genetics
4.	(25.10.)	**DNA** – structure, PCR techniques, applications, dominant markers (AFLP, RAPD, ISSRs…), data evaluation
5.	(1.11.)	**restriction techniques** (RFLP, PCR-RFLP), **cpDNA**, phylogeography
6.	(8.11.)	**microsatellites** – nuclear, chloroplast, isolation, data evaluation, applications
7.	(15.11.)	Sanger **sequencing** – cpDNA, genes and non-coding regions
8.	(22.11.)	**sequencing II** – nuclear DNA, nrDNA, ITS, low-copy genes
9.	(29.11.)	**NGS** (next-generation sequencing) – principles and applications...
10.	(6.12.)	**RADseq, resequencing** – SNP analysis, population genomics
11.	(13.12.)	**HybSeq** – target enrichment methods, phylogenomics
12.	(3.1.2024)	student presentations, exam…

http://botany.natur.cuni.cz/fer/markers/indexE.htm