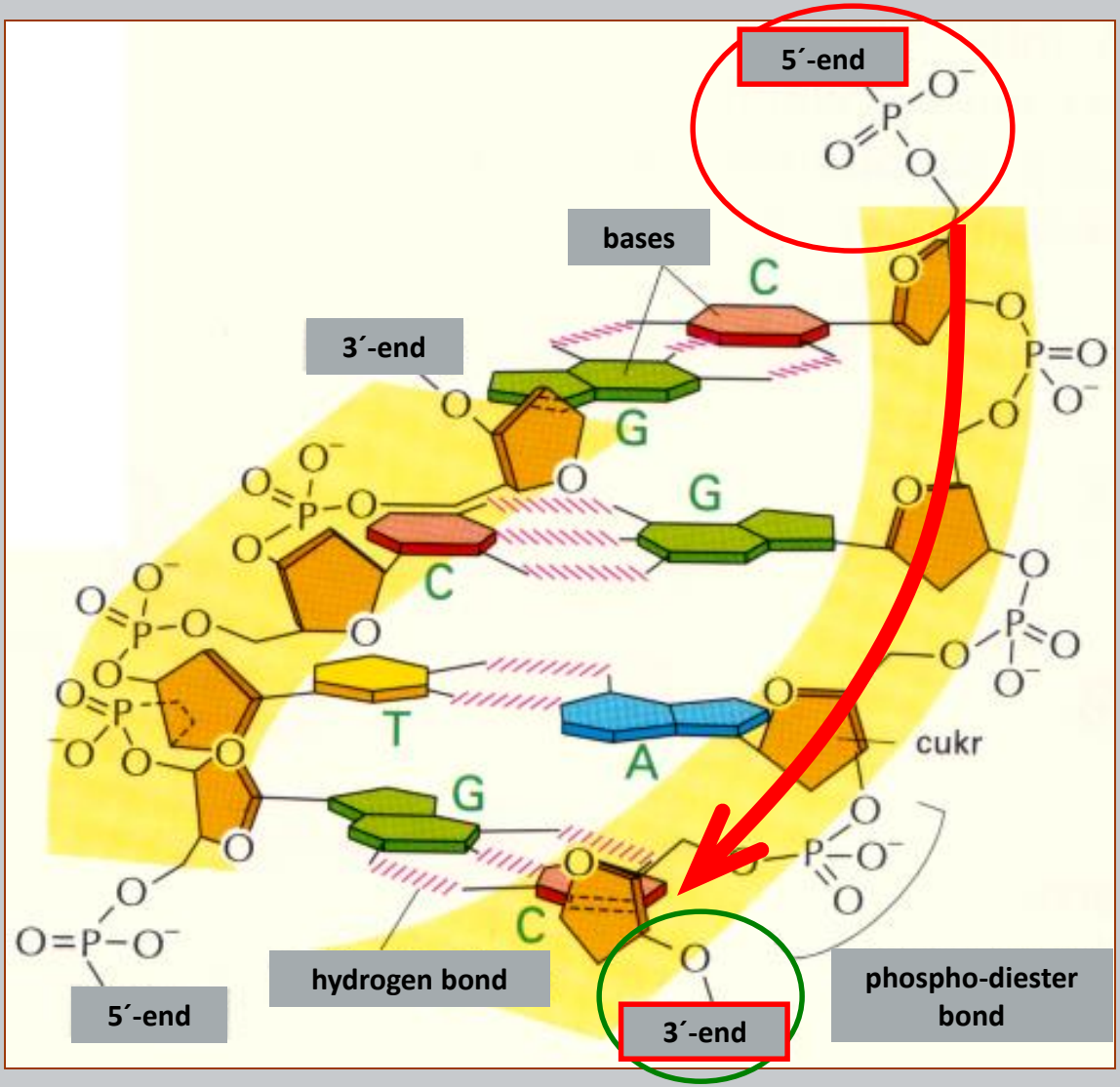
Molecular markers in plant systematics and population biology

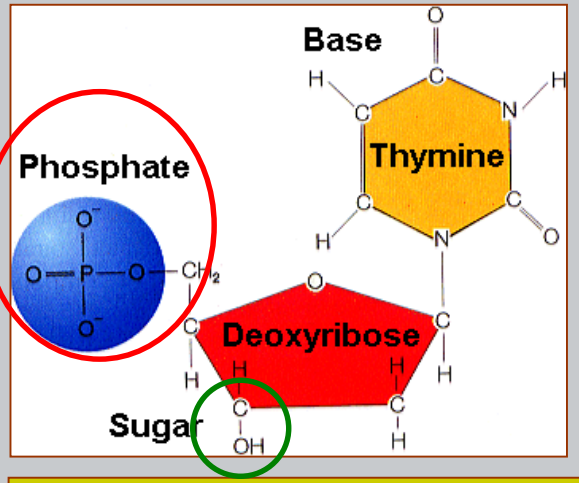
4. DNA, PCR, dominant markers (RAPD, ISSR, AFLP)

Tomáš Fér

tomas.fer@natur.cuni.cz

DNA structure

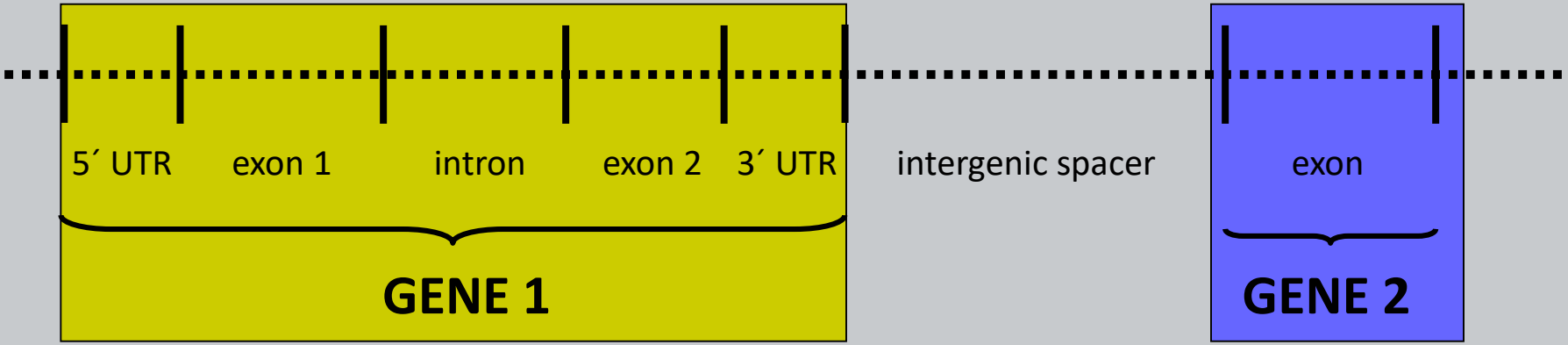




- double helix
- 2 anti-parallel strands
- sugar-phosphate backbone
- bases
  - purines (A, G)
  - pyrimidines (C, T)
- pairing
  - G≡C
  - A=T

DNA structure II. (simplified)

- genetic information – nucleotide sequence (ACGT)



- coding regions – exons – *conserved*
- non-coding regions – introns, spacers – *more variable*

- repetitive DNA
  - *moderately* – tandem genes (rRNA, tRNA), pseudogenes, SINEs, LINEs, LTRs
  - *highly* – minisatellites, microsatellites, telomeres

Plant genome(s)

- nucleus – nDNA
  - chromosomes
  - size – $6.5*10^7$ bp (*Genlisea tuberosa*) – $1.5*10^{11}$ bp (*Paris japonica*)
  - unique information for an individual, recombination

- chloroplast – cpDNA
  - circular molecule
  - 20-240 kbp (*Nicotiana* – 156 kbp)
  - several specific genes (4 – 300 genes), incl. tRNA, rRNA, protein coding

- mitochondrial – mtDNA
  - complex structure, many repetitive sequences
  - 20-2,400 kbp
  - recombinations, insertions/deletions

How to study DNA variation

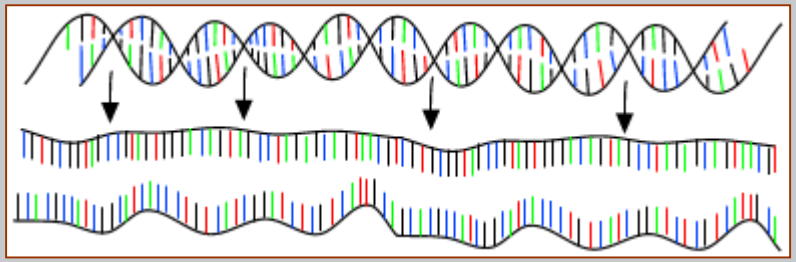
- sequencing (classical, NGS)
  - nucleotide order
- cleavage (restriction)
  - length polymorphism of restriction fragments
  - RFLP, PCR-RFLP
- PCR
  - length polymorphism of amplified fragments
  - RAPD, ISSRs, AFLP, microsatellites, …
- whole genome variability (NGS)
  - de novo sequencing, resequencing
  - target enrichment (Hyb-Seq)
  - SNPs (RADseq …)

Replication *in vitro*
PCR – polymerase chain reaction

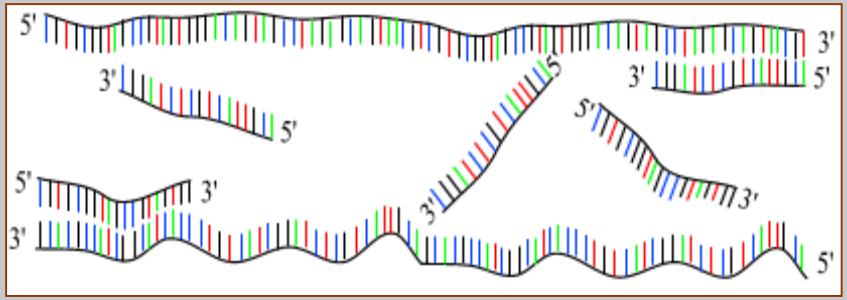
- amplification of particular part of the genome
- we need
  - investigated DNA (template)
  - primers – oligonucleotides (length 10-25 bases)
  - dNTPs – „building stones“ of DNA
  - *Taq* DNA polymerase (thermostable)
  - buffer with $MgCl_2$ – stabilization
- amplification with temperature changes
  - PCR thermocycler

PCR principle
„only“ change of three temperatures

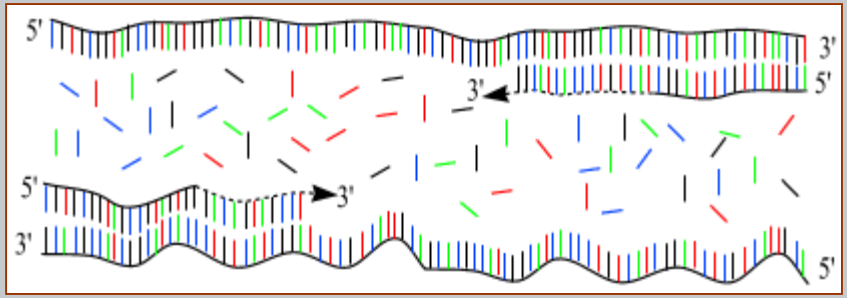
1. 95 ºC – denaturation
   - separation of both DNA strands



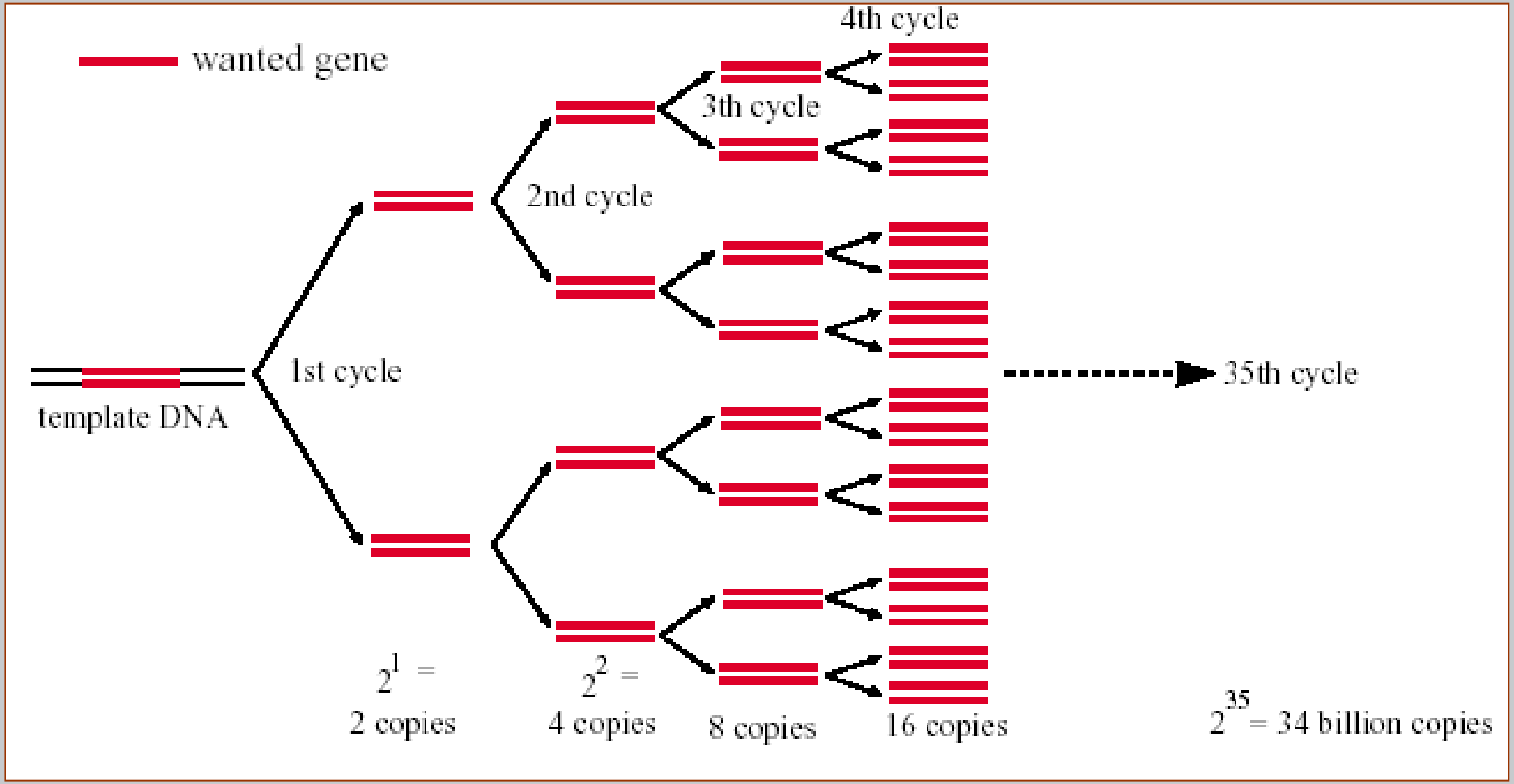
2. 35-65 ºC – annealing
   - specific binding of primers to DNA



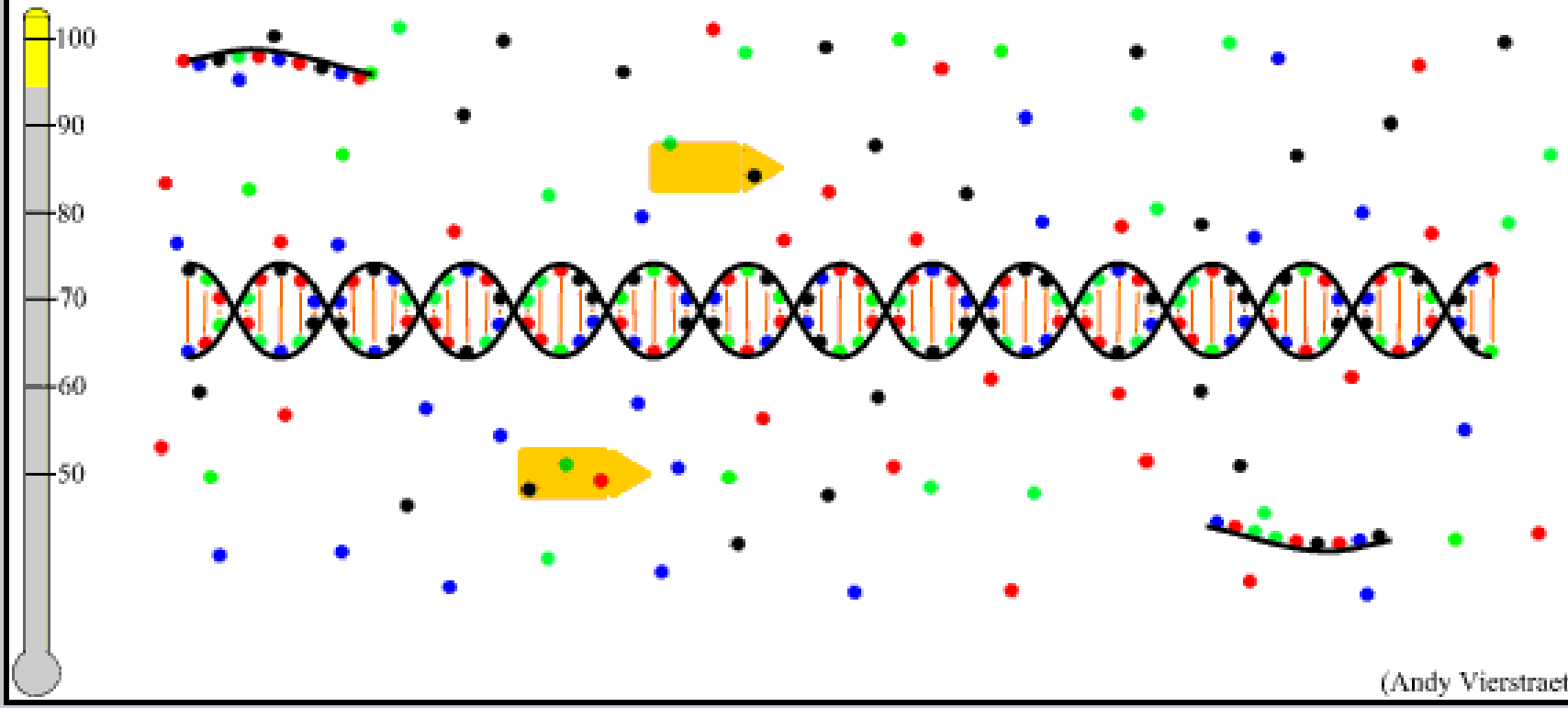
3. 72 ºC – polymeration
   - elongation of DNA strand (5´ → 3´)



Exponencial amplification



PCR : Denaturation 94°C



Overview of PCR techniques

- *RAPD*
  - only one short primer used (length polymorphism)
  - dominant
- *PCR-RFLP*
  - combination of PCR and subsequent fragment restriction with endonuclease (length polymorphism)
  - codominant, often cpDNA (haploid)
- *AFLP*
  - restriction of the whole genome with two endonucleases and subsequent amplification of selected fragments (length polymorphism)
  - dominant
- *microsatellites (SSRs)*
  - amplification of specific repetitive loci (polymorphism is defined by repeat number of 2-6 nucleotide sequence)
  - codominant
- *inter simple sequence repeats (ISSR)*
  - amplification of length polymorphism between SSR loci
  - dominant

Multilocus data

- **dominant** – heterozygotes and homozygotes cannot be distinguished
- **binary** – biallelic data (fragments)
  - presence (dominant allele/heterozygote)
  - absence (recessive allele)
  - i.e., 0-1 scoring
- **anonymous** – unknown genomic origin
- **multilocus** – simultaneous analysis of hundreds of loci, i.e. analysis covers „whole genome“

- RAPD, ISSR, AFLP…

- **codominant**– heterozygotes and homozygotes can be distinguished
- **alellic** – known allelic frequencies in loci, populations…

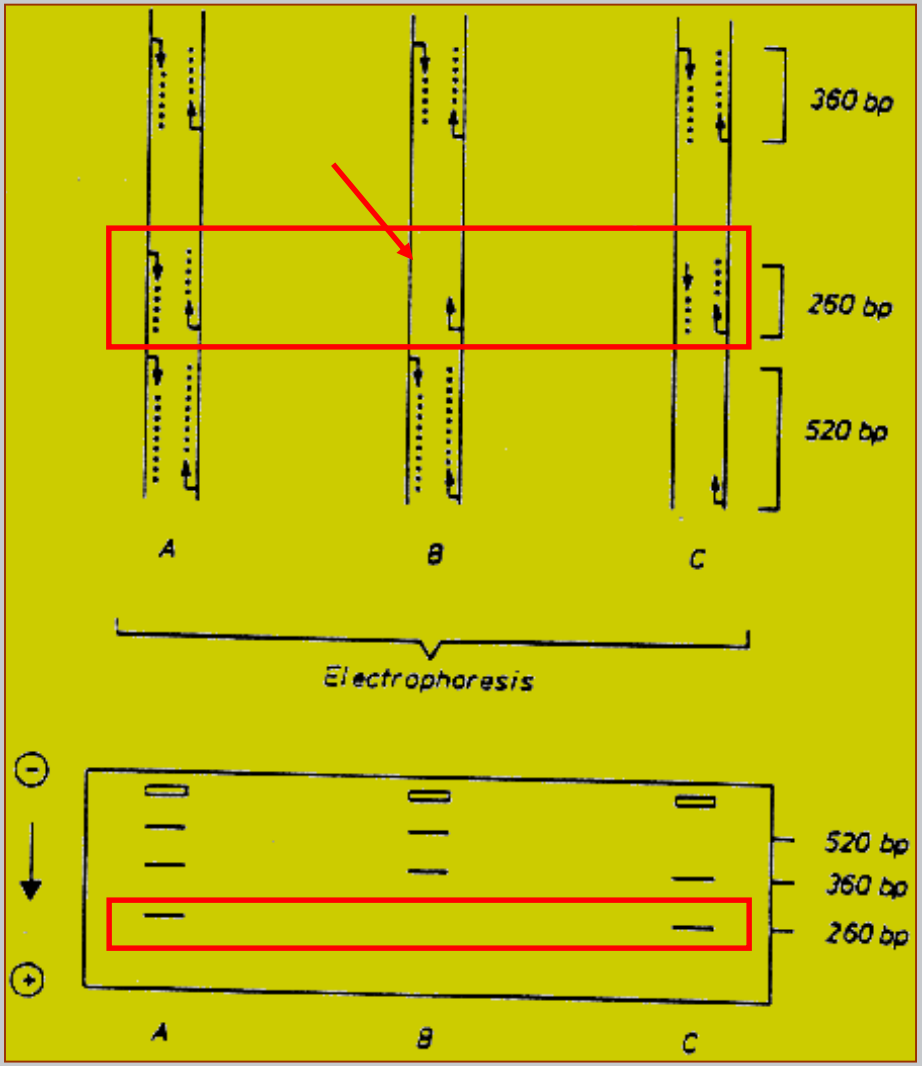
- **anonymous** – unknown genomic origin
- **multilocus** – usually analysis of few loci (5-20)

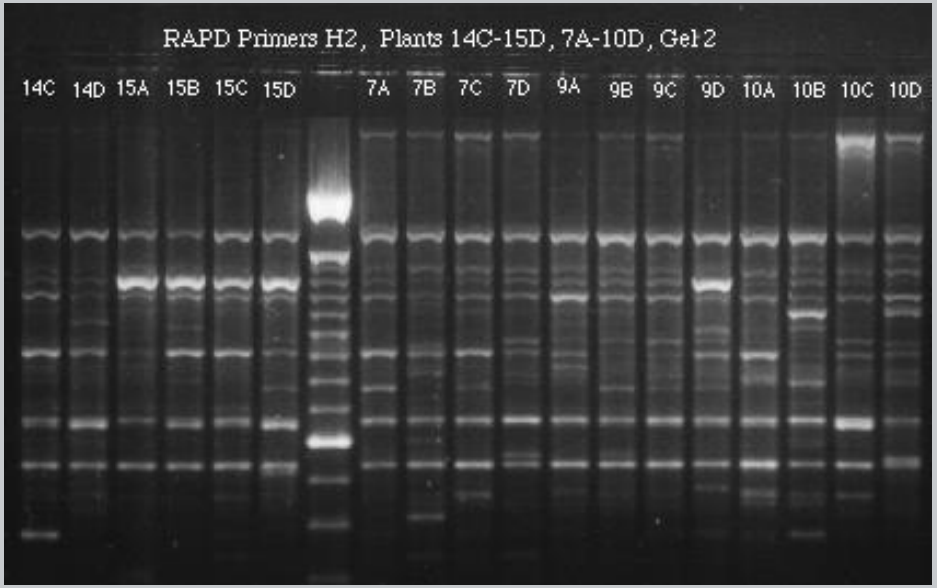
- microsatellites (SSRs), isozymes

Principle of RAPD

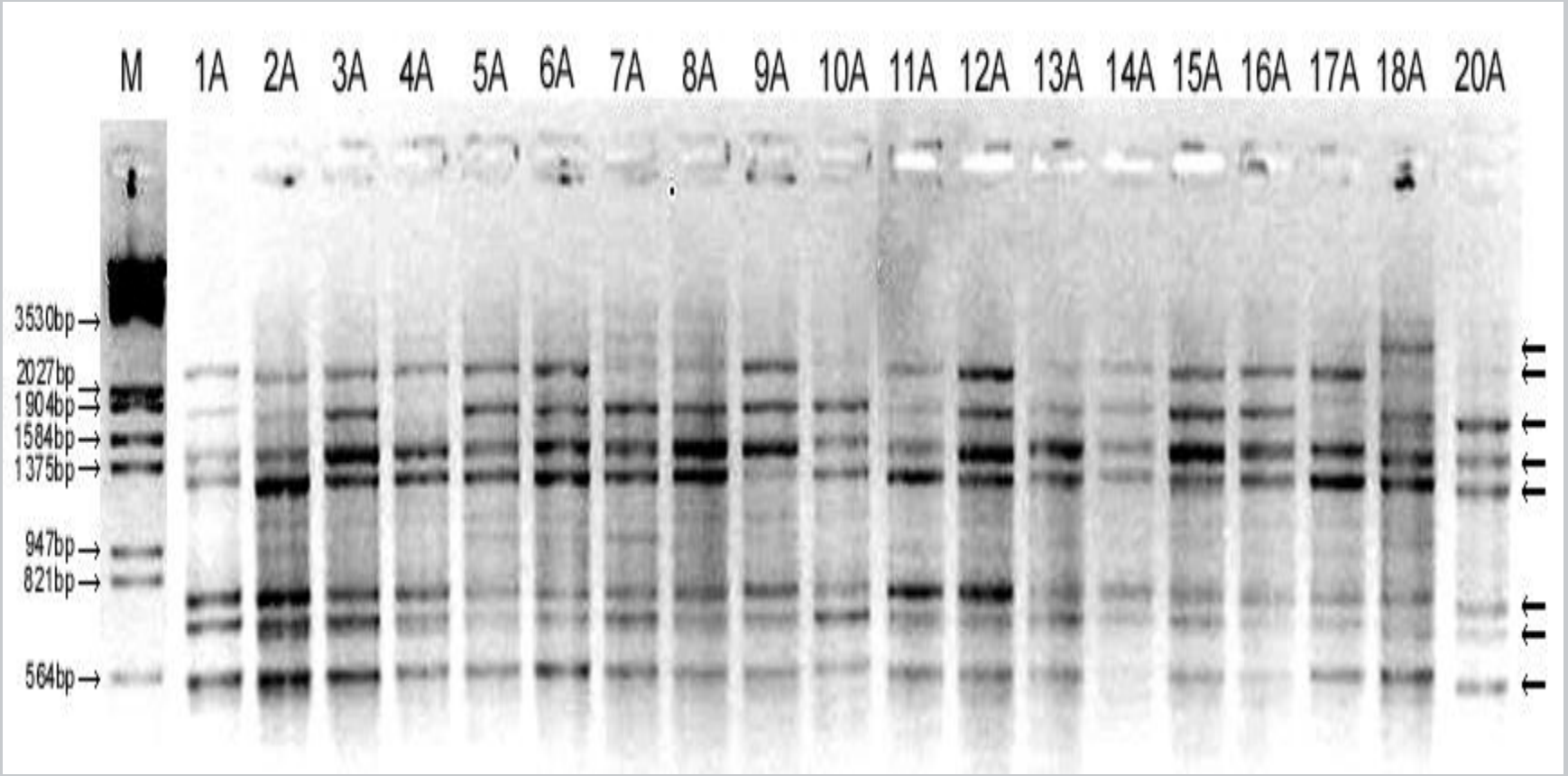
1. PCR with one primer
   - primer – oligonucleotide with arbitrary sequence
   - PCR product is amplified if
     - two identical (or highly similar) priming sites (i.e., sites where primers bind to DNA) exist on antiparalel DNA strands
     - amplifiable distance between these two sites is usually up to ca. 3,000 bp
2. electrophoresis – separation of fragments according to their length
   - horizontal – agarose gel
   - vertical – polyacrylamide gel
3. fragment visualization
   - EtBr (ethidium bromide) – selectively binds to DNA

RAPD


360 bp
260 bp
520 bp
A
B
C
Electrophoresis
520 bp
360 bp
260 bp
A
B
C


RAPD Primers H2, Plants 14C-15D, 7A-10D, Gel 2
14C 14D 15A 15B 15C 15D 7A 7B 7C 7D 9A 9B 9C 9D 10A 10B 10C 10D

RAPD gel example



Gentiana pannonica – Hofhanzlová & Fér (2009)

RAPD

pros

- small amount of DNA is sufficient – ca. 25 ng
- simple and fast method
- highly variable, many markers
- variability throughout whole genome

cons

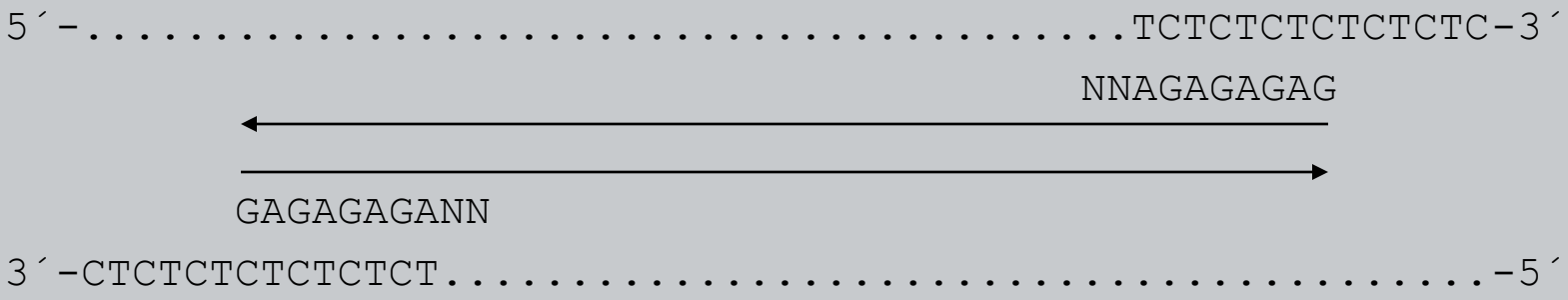
- dominant marker
- less reproducible pattern
- evaluation subjectivity
- standardization necessary

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

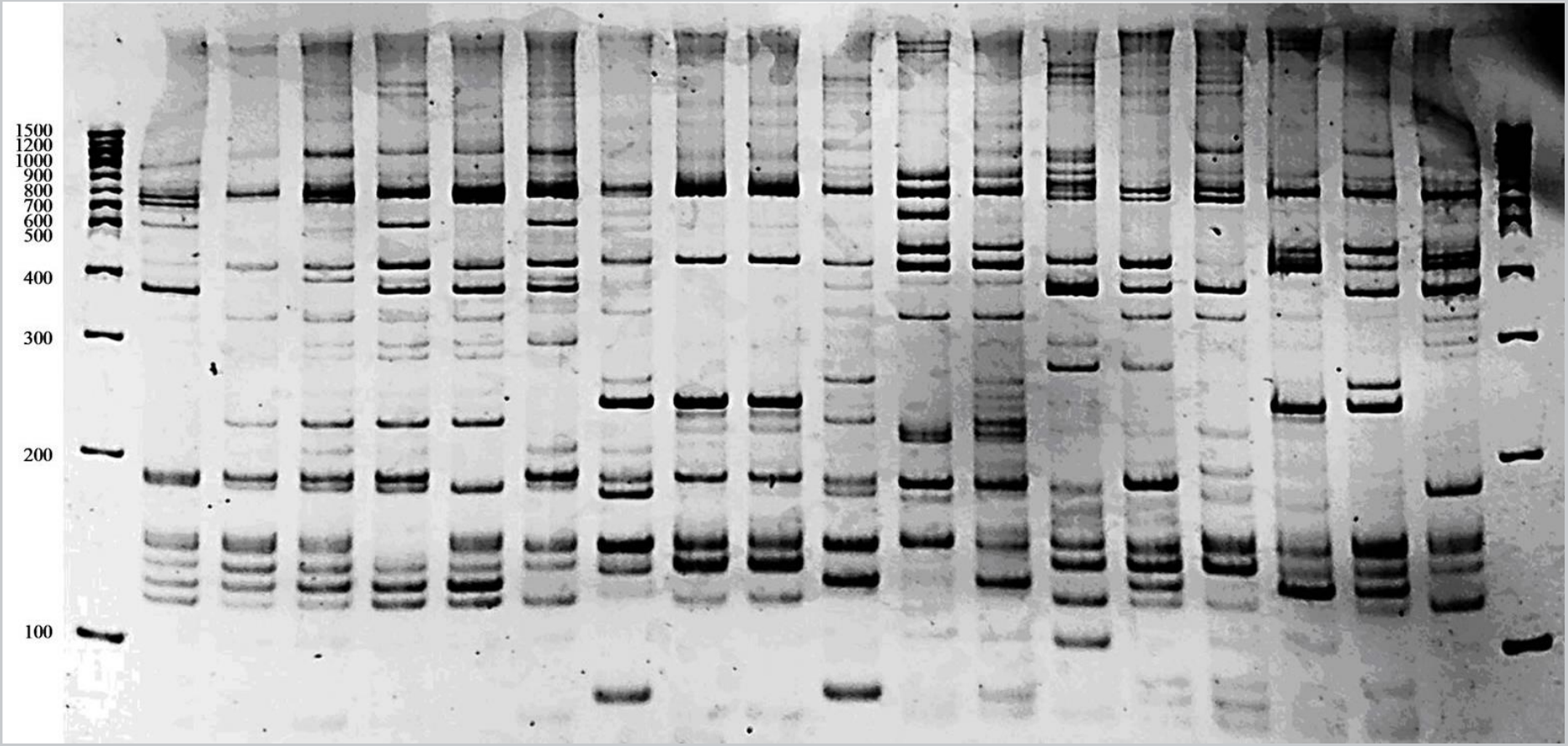
- primer – microsatellite sequence
- amplification of fragments between microsatellite loci
- polymorphism
  - change of number of repeat units in microsatellite
  - indels between microsatellite motifs
- very variable marker
- dominant marker
  - optimization necessary
  - more robust than RAPDs

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

- application of anchored primers



ISSRs gel example



Poa arachnifera, *P. pratensis* – Goldman J. (2013)

ISSRs

pros

- small amount of DNA is sufficient – ca. 25 ng
- simple and fast method
- extremely variable, high number of markers
- more reproducible than RAPD
- variability throughout whole genome

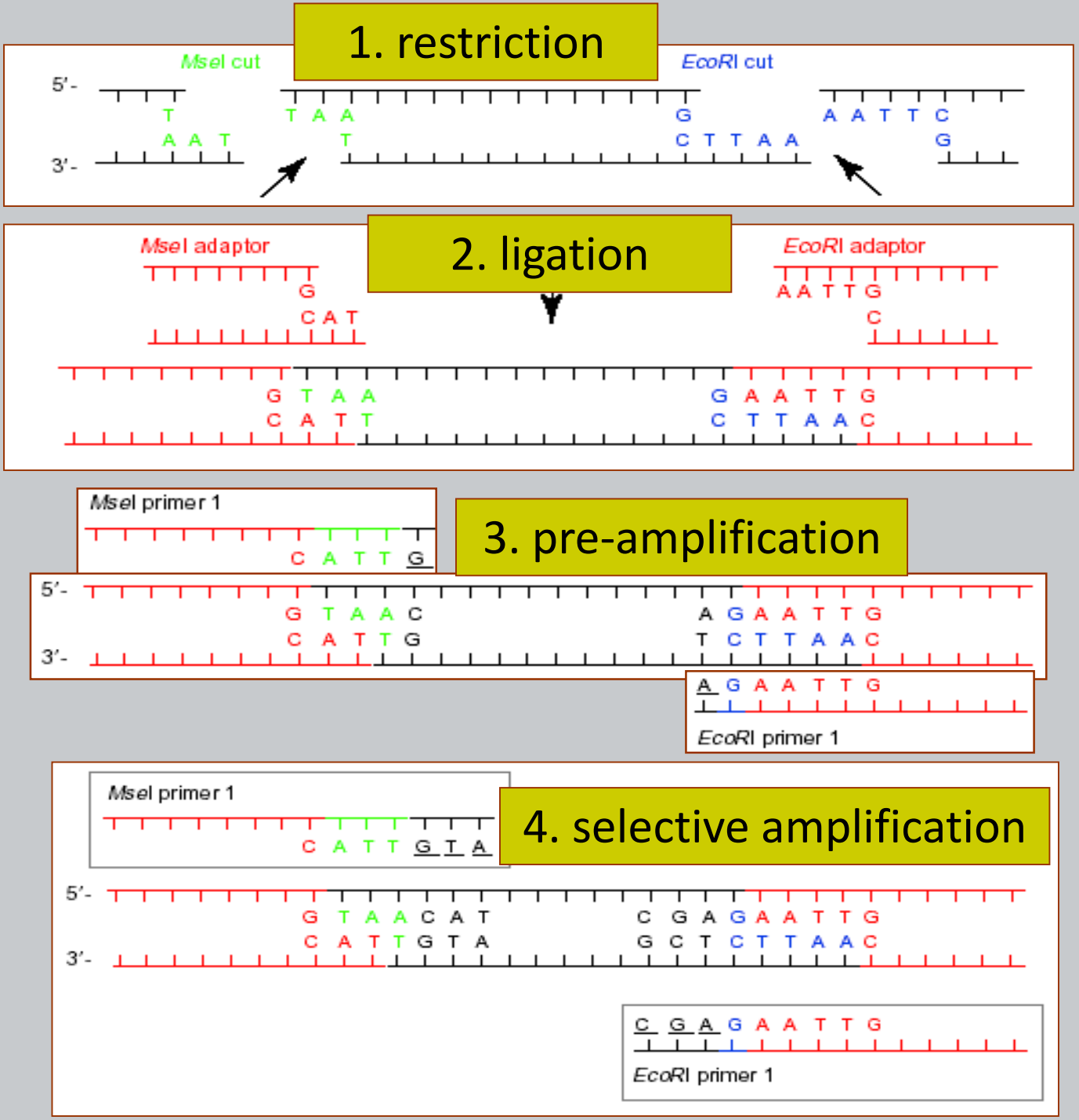
cons

- dominant marker
- sometimes less reproducible
- evaluation subjectivity
- standardization necessary

Amplified **F**ragment **L**ength **P**olymorphism (AFLP)

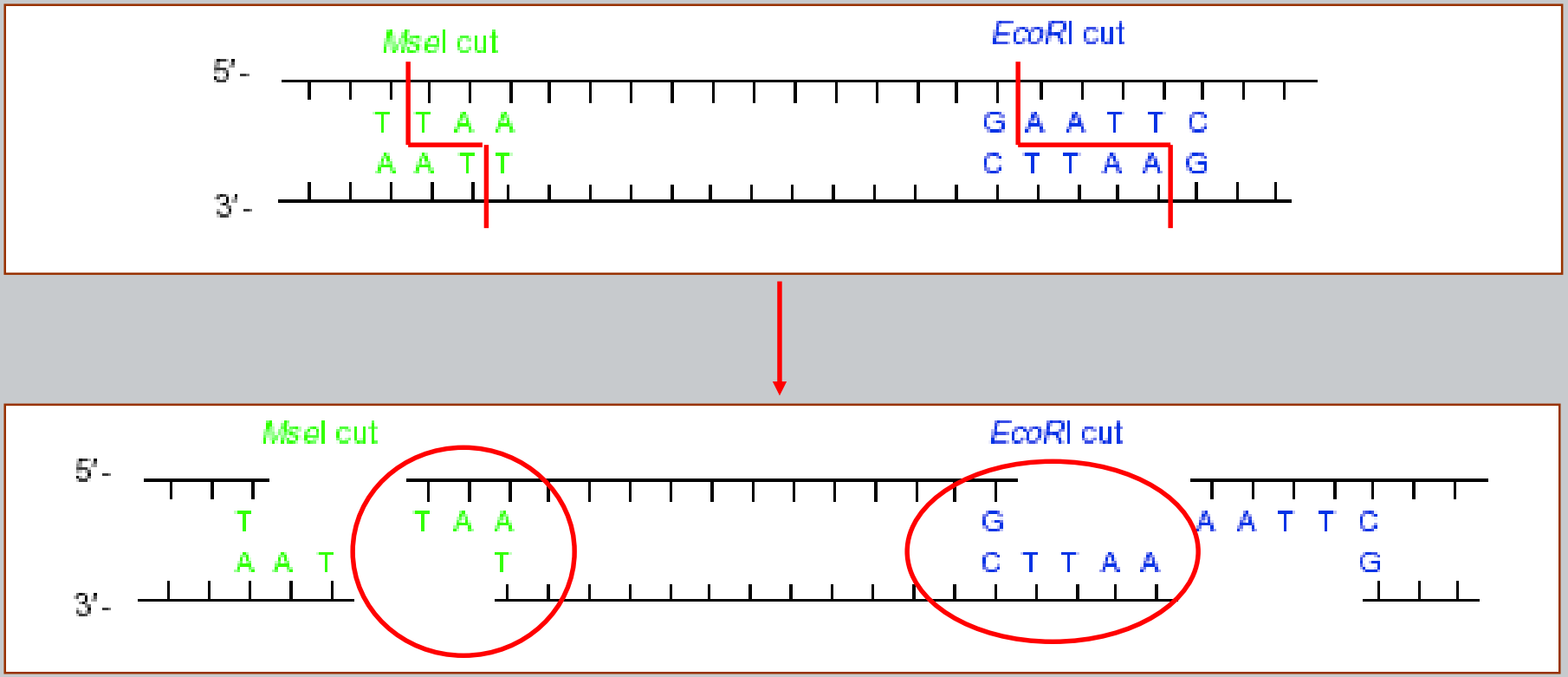
- principle of the method
  - restriction of total DNA
  - selective PCR amplification of selected fragments

total DNA



Mueller & Wolfenbarger (1999), TREE

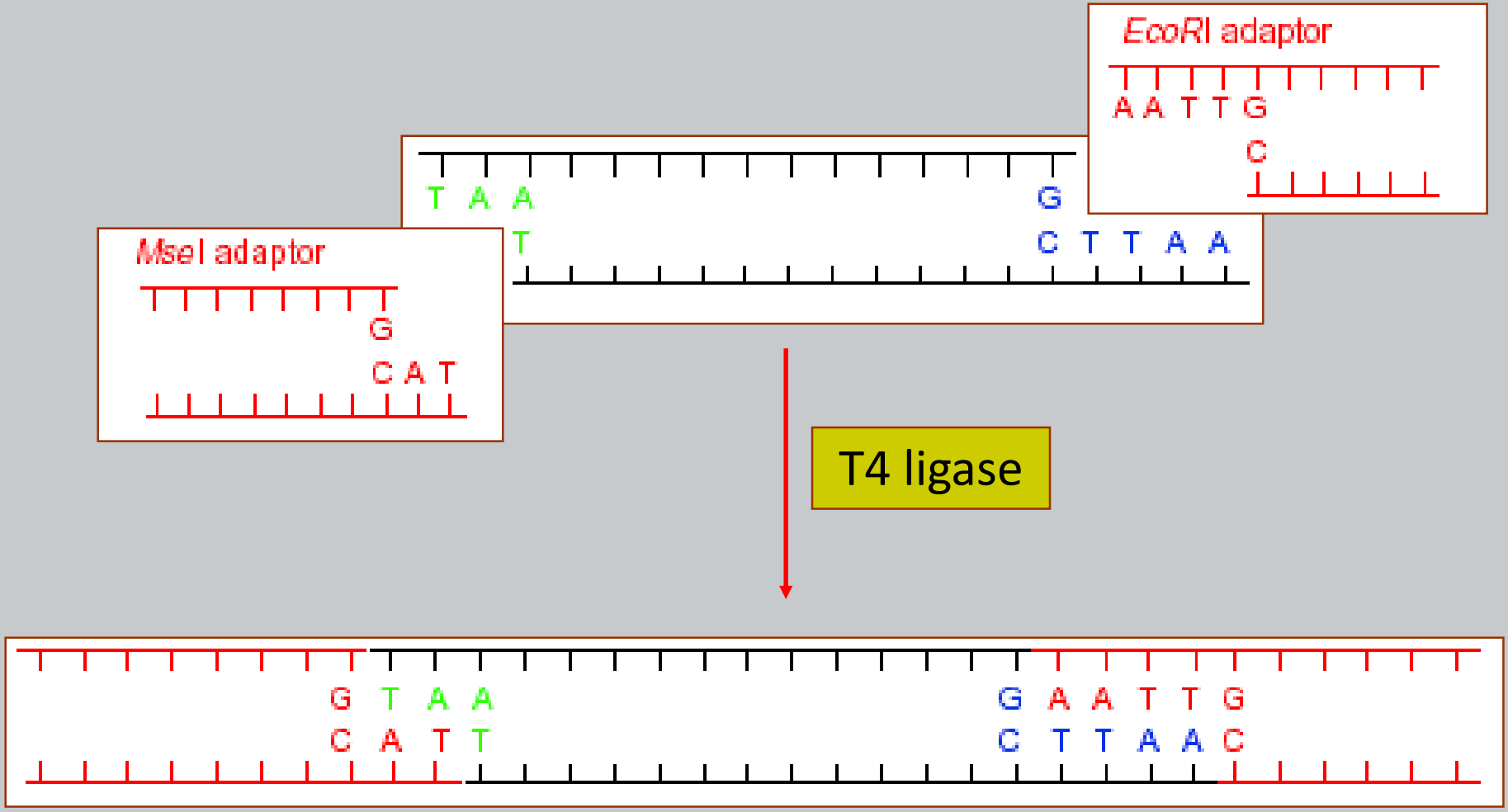
1. Restriction



***Mse*I**	4 bp	„frequent cutter“
***EcoR*I**	6 bp	„rare cutter“

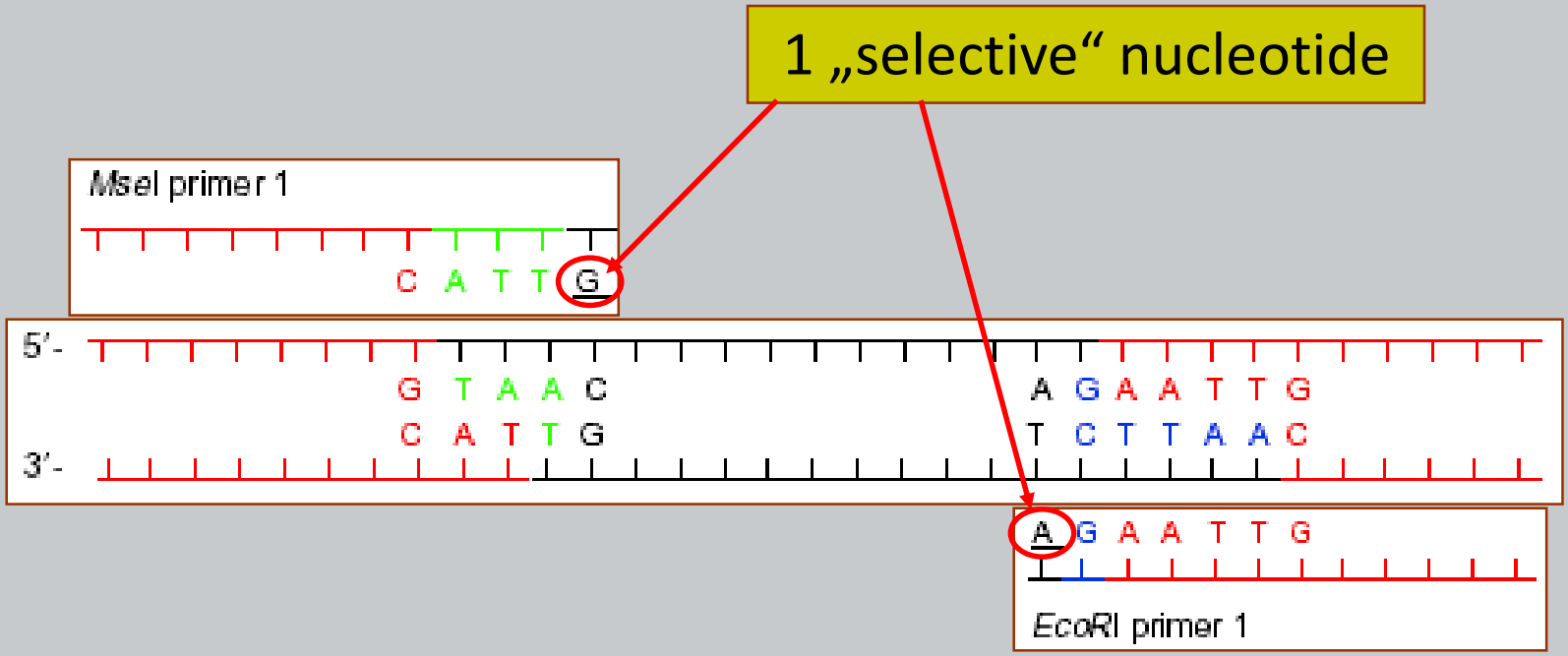
Mueller & Wolfenbarger (1999), TREE

2. Ligation



Mueller & Wolfenbarger (1999), TREE

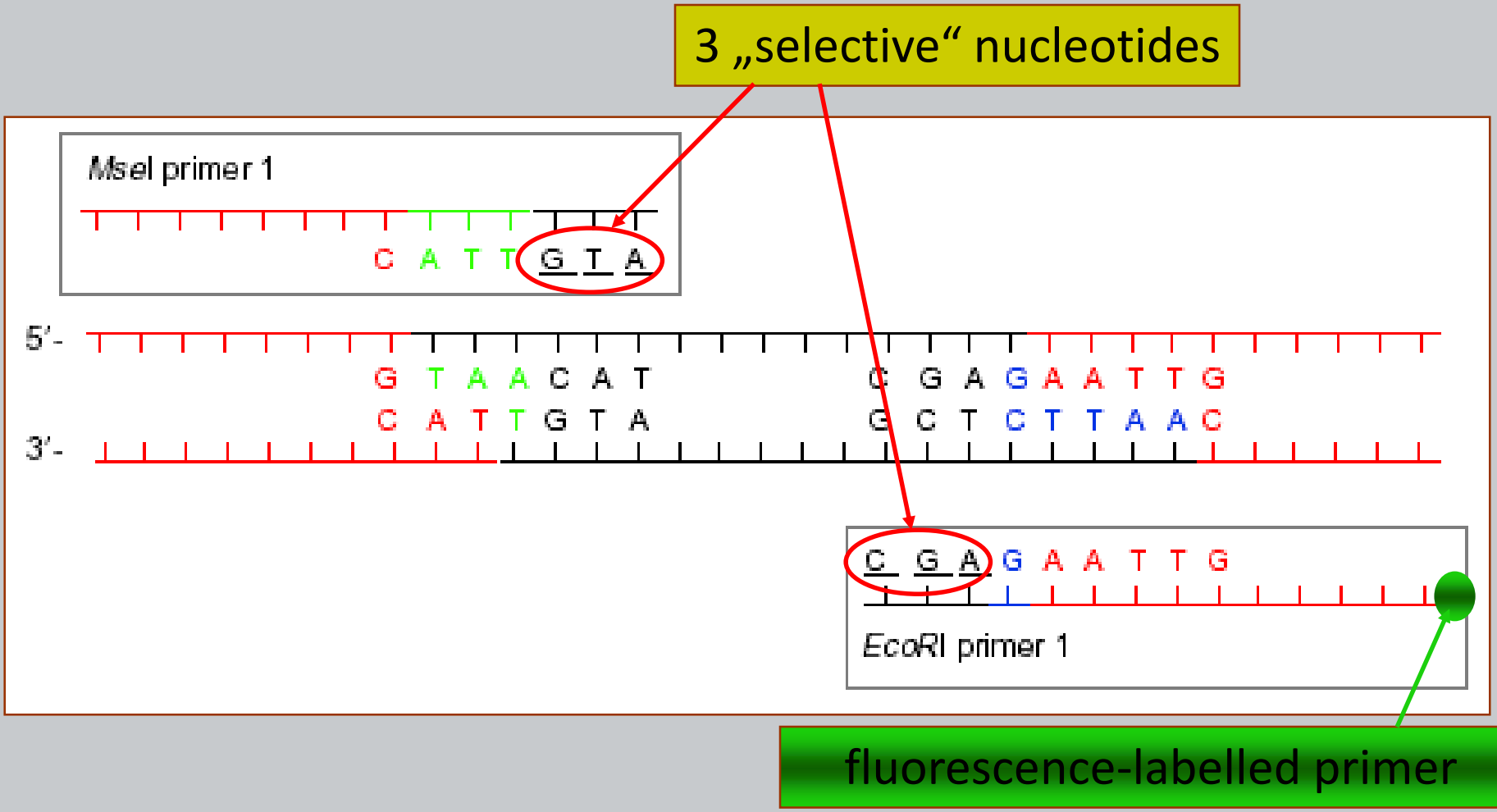
3. Pre-amplification



PCR amplification of 1/16th of all fragments

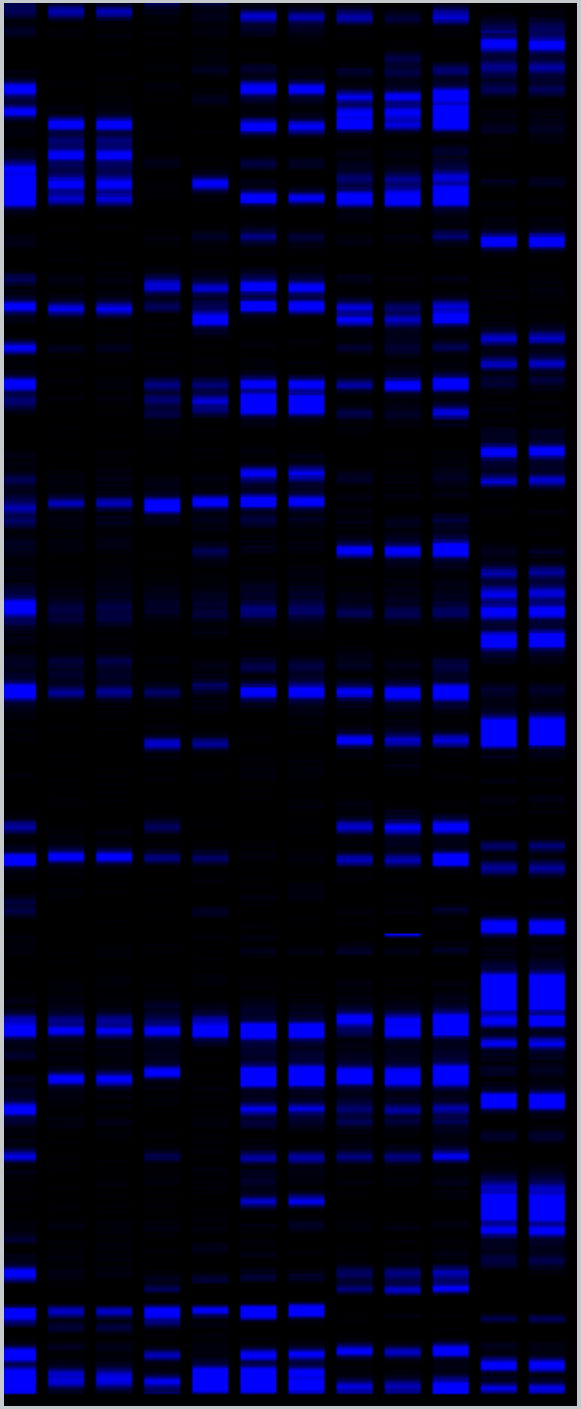
Mueller & Wolfenbarger (1999), TREE

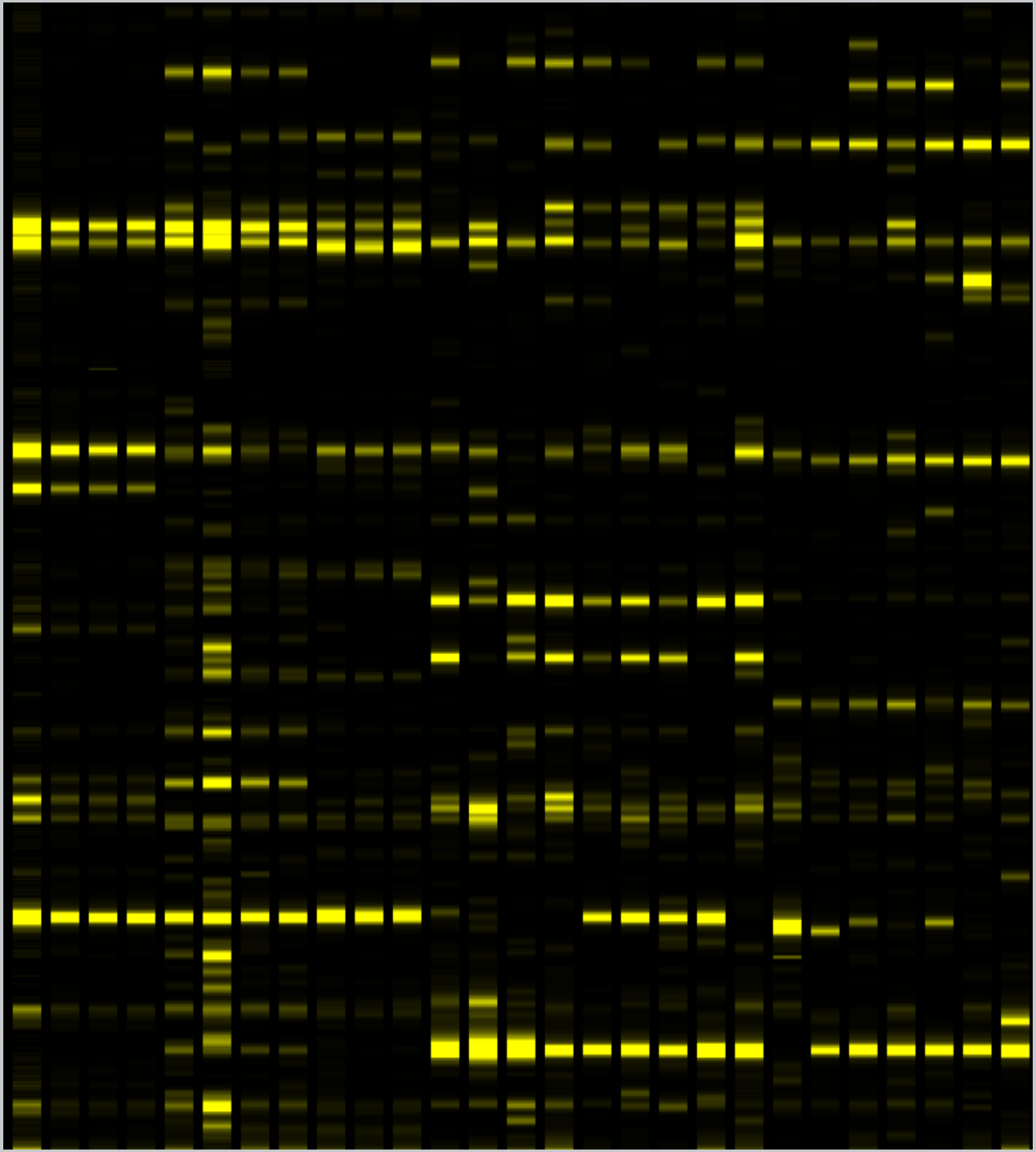
4. Selective amplification

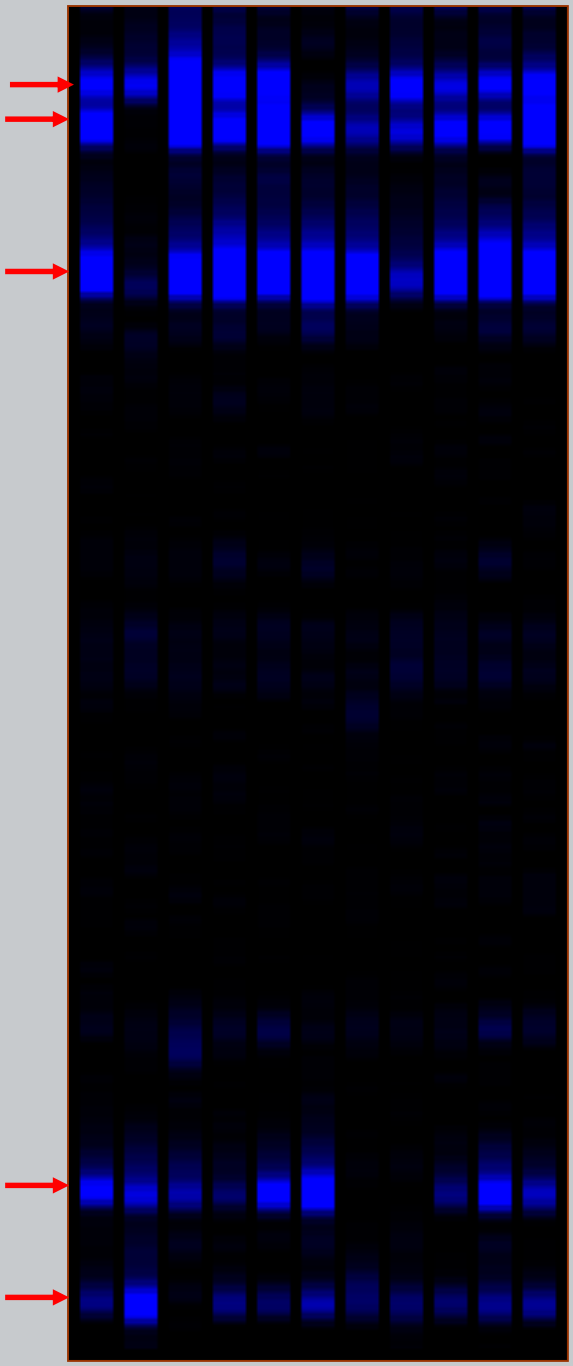


Mueller & Wolfenbarger (1999), TREE

PCR amplification of 1/256th of all fragments







Visualization

- fluorescence-labelled fragments
- use of automated sequencer
- evaluation of fragment pattern
  ← GENOGRAPHER, Gene Marker …
- bands ca. 50-500 bp

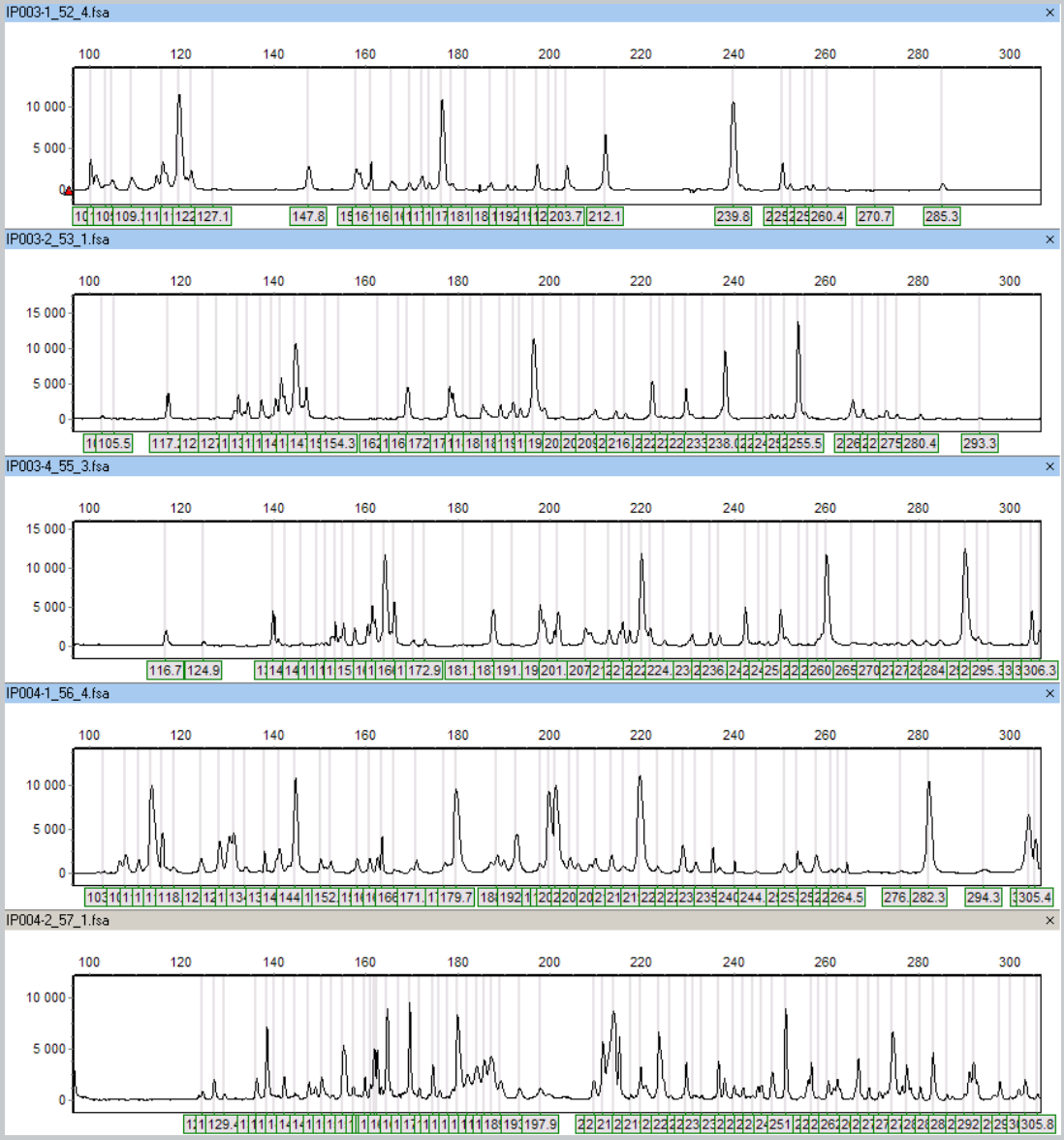
Variability due to

- mutation in restriction site
- insertion/deletion between restriction sites

Assumptions for interpretation

- homology of comigrated fragments
- independence of fragments

AFLP analysis – GeneMarker



		169.6	170.3	170.9	171.7	172.3	172.8	173.8	174.4	175.3	176.1	176.6	177.2	177.7	178.2	178.8
1	IP001-1_49_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	IP002-1_51_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	IP003-1_52_4	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1
4	IP003-2_53_1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
5	IP003-4_55_3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
6	IP004-1_56_4	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
7	IP004-2_57_1	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0
8	IP004-4_59_3	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0
9	IP004-5_60_4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	IP005-1_61_1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
11	IP005-4_64_4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
12	IP006-1_54_2	1	0	0	0	1	0	1	0	0	0	1	0	1	1	0
13	IP008-1_68_4	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
14	IP008-2_69_1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15	IP009-1_72_4	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
16	IP009-2_27_3	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
17	IP010-1_28_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	IP011-2_32_4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
19	IP013-1_36_4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
20	IP013-3_38_2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
21	IP014-1_40_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	IP014-5_44_4	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0
23	IP016-1_48_4	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0

AFLP

pros

- high degree of polymorphism – more than 100 fragments / 1 primer combination
- no previous molecular knowledge about studying organism necessary
- highly reproducible pattern
- variability throughout whole genome

cons

- dominant marker
  - unknown fragment origin
  - homology?
  - intensity?
- optimization necessary (to find best primer combinations for the target species)
- relatively complicated (and expensive) method

Dominant markers – variability assumptions

- *priming sites* (sites where RAPD primers bind to DNA) or AFLP *restriction sites*
  - randomly distributed across the genome
  - i.e., in coding as well as non-coding regions
- variability (presence/absence of fragment with particular length) is expected to be caused by
  - mutation in priming site
  - insertion/deletion in amplified region
- comigrating bands homologous

Evaluation assumptions/problems

- independence of fragments ?
- homology of fragments (comigration of non-homologous fragments ?)
  - different intensity
  - small differences in mobility
- asymetry in probability of loss and gain of fragments (criticisms for application of maximum parsimony method)
- dominance (what is a locus?) – impossible to distinguish homozygotes and heterozygotes

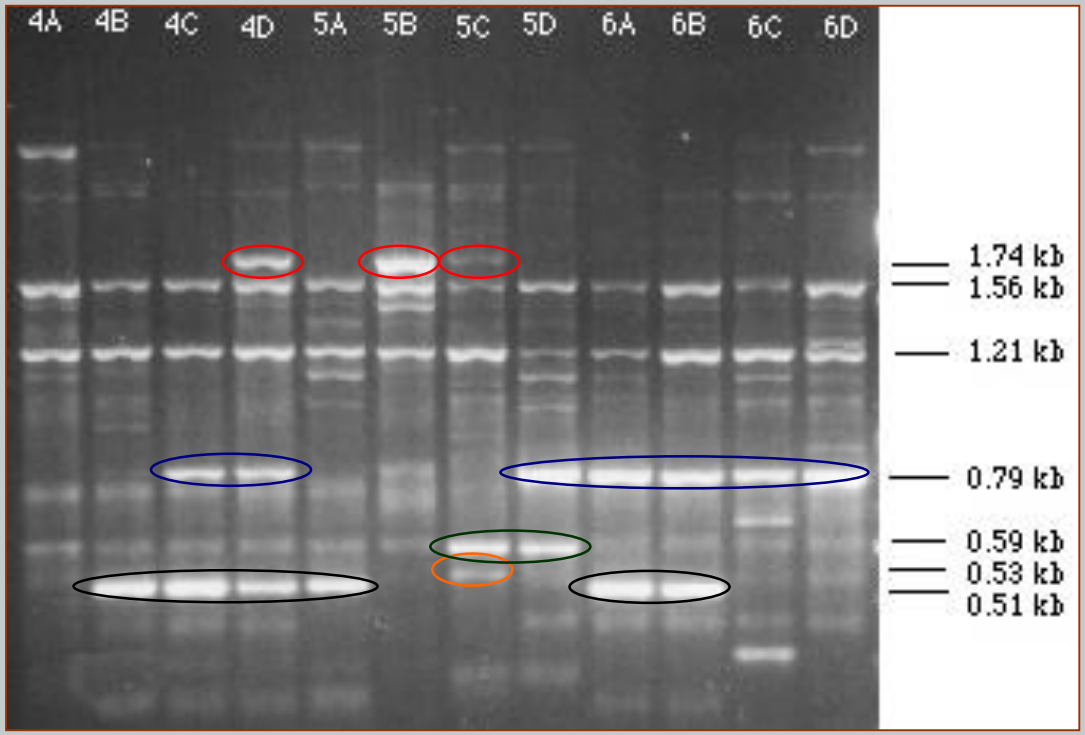
Variability in bands intensity

- different degree of *primer mismatch* – enabled by relatively low annealing temperature
  - confirmed by sequencing of fragments
- heterozygosity
- weaker bands are generally less robust
- intensity is not considered, fragments are evaluated only as present/absent (0/1)

Gel pattern evaluation

- absence/presence of bands – 0-1 matrix
- which bands should be evaluated and which not?
  - in a limited range
  - only bands with higher intensity – higher reproducibility
  - lesser number of reliable fragments increases credibility of the method
  - comparison of repeated reactions for the same sample (from DNA extraction to electrophoresis)
    - only repeated bands should be considered
    - *error rate* calculation

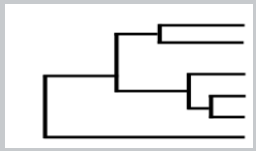
Gel evaluation

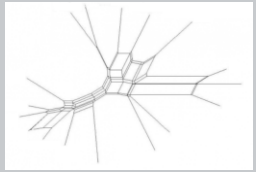


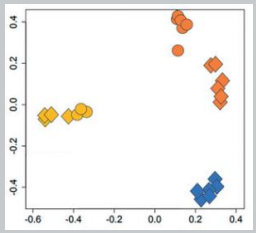
Band (kb)	4-A	4-B	4-C	4-D	5-A	5-B	5-C	5-D	6-A	6-B	6-C	6-D
1.74	0	0	0	1	0	1	1	0	0	0	0	0
1.56	1	1	1	1	1	1	1	1	1	1	1	1
1.21	1	1	1	1	1	1	1	1	1	1	1	1
0.79	0	0	1	1	0	1	0	1	1	1	1	1
0.59	0	0	0	0	0	0	1	1	0	0	0	0
0.53	0	0	0	0	0	0	1	0	0	0	0	0
0.51	0	1	1	1	1	0	0	0	1	1	0	0

Data evaluation

- presence/absence → 0-1 matrix
- number of fragments, % of polymorphic fragments
- number of unique and rare fragments, DW-index
- coefficients of **similarity** (*Nei & Li, Jaccard, …*)
  - dendrograms (UPGMA, NJ …)
  - networks (neighbour-net)
  - PCoA – *principal coordinate analysis*
- *Bayesian clustering* (BAPS 3, Structure…)
- AMOVA (Arlequin …) – variance partitioning, F_{ST} analogue…
- within-population diversity – Shannon diversity index, *average gene diversity*…
- Kinship coefficient (Hardy 2003)

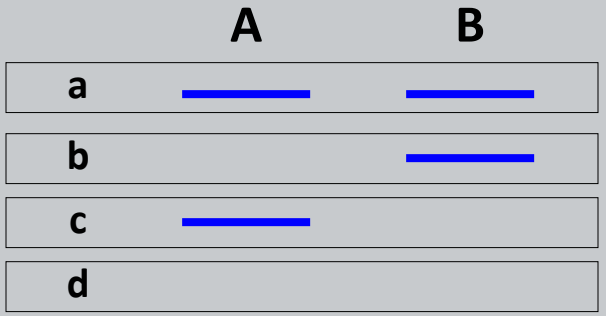






Similarity coefficients

		individual A	
		presence 1	absence 0
individual B	presence 1	a	b
	absence 0	c	d



- Jaccard coefficient (Jaccard 1908)

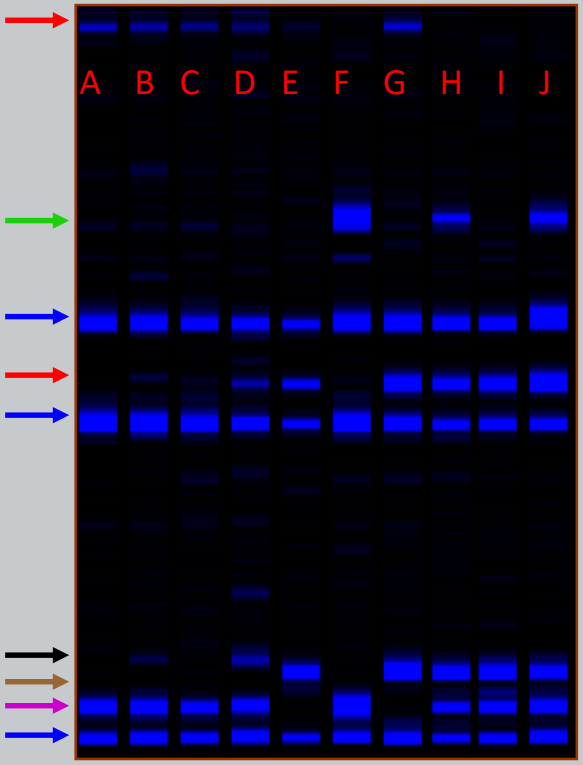
$$\frac{a}{a+b+c}$$

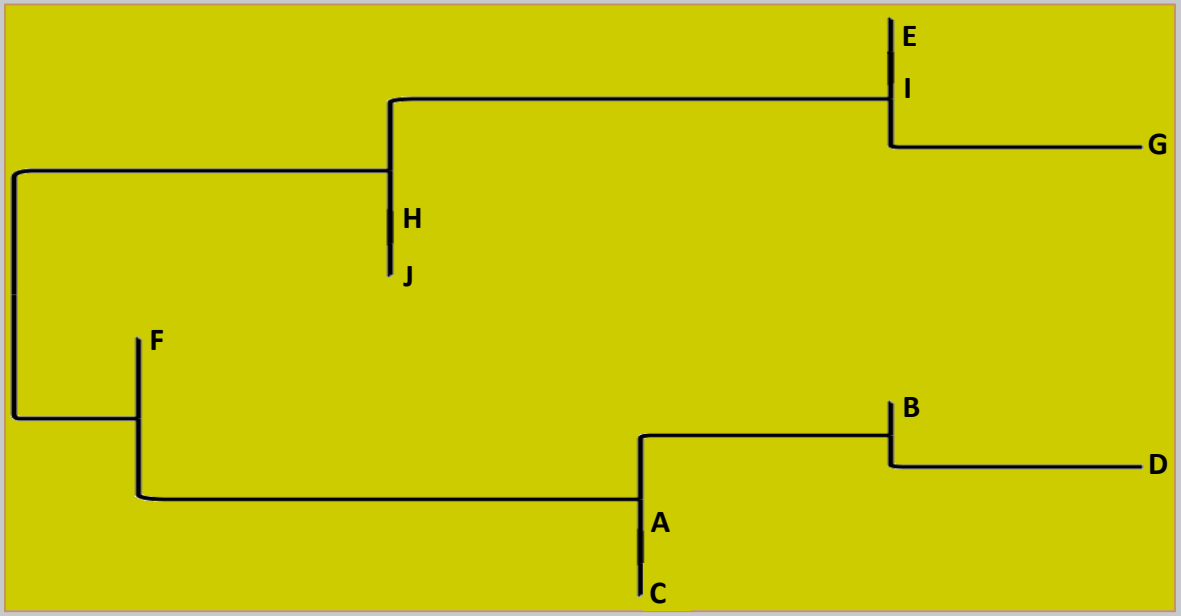
- Dice coefficient (Dice 1945) = Nei & Li 1979, Sørensen 1948

$$\frac{2a}{2a+b+c}$$

- „simple-matching“ coefficient (Sokal & Michener 1958)

$$\frac{a+d}{a+b+c+d}$$



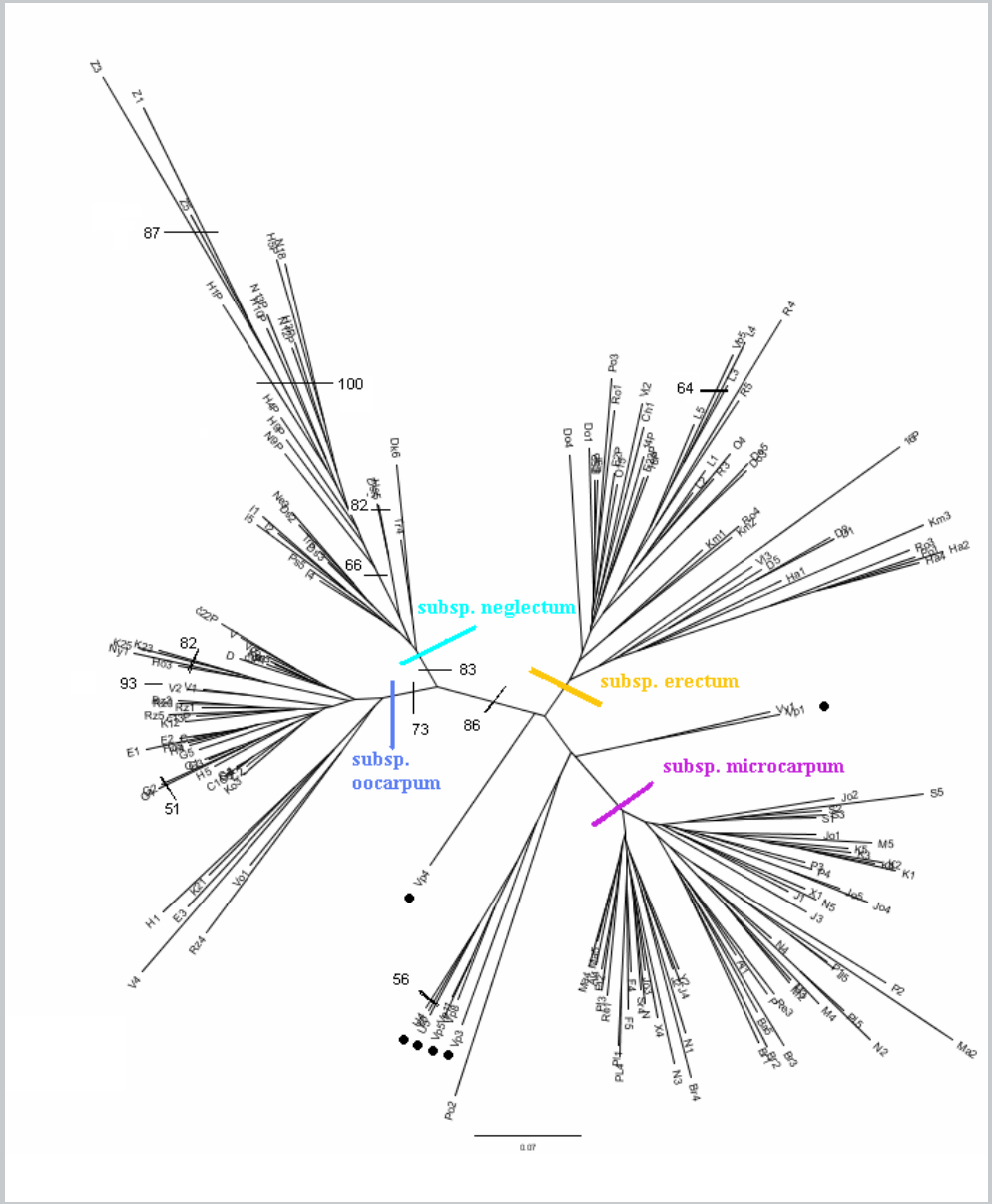


neighbour-joining tree,
midpoint rooting

1	1	1	1			1			
					1		1		1
1	1	1	1	1	1	1	1	1	1
			1	1		1	1	1	1
1	1	1	1	1	1	1	1	1	1
	1		1						
				1		1	1	1	1
1	1	1	1		1		1		1
1	1	1	1	1	1	1	1	1	1

	A	B	C	D	E	F	G	H	I	J
A	0,00000									
B	0,01591	0,00000								
C	0,00000	0,01591	0,00000							
D	0,03046	0,01336	0,03046	0,00000						
E	0,08576	0,10190	0,08576	0,06797	0,00000					
F	0,03731	0,05331	0,03731	0,06797	0,08576	0,00000				
G	0,05331	0,06797	0,05331	0,04389	0,01591	0,10190	0,00000			
H	0,06797	0,08148	0,06797	0,05634	0,03046	0,03046	0,04389	0,00000		
I	0,08576	0,10190	0,08576	0,06797	0,00000	0,08576	0,01591	0,03046	0,00000	
J	0,06797	0,08148	0,06797	0,05634	0,03046	0,03046	0,04389	0,00000	0,03046	0,00000

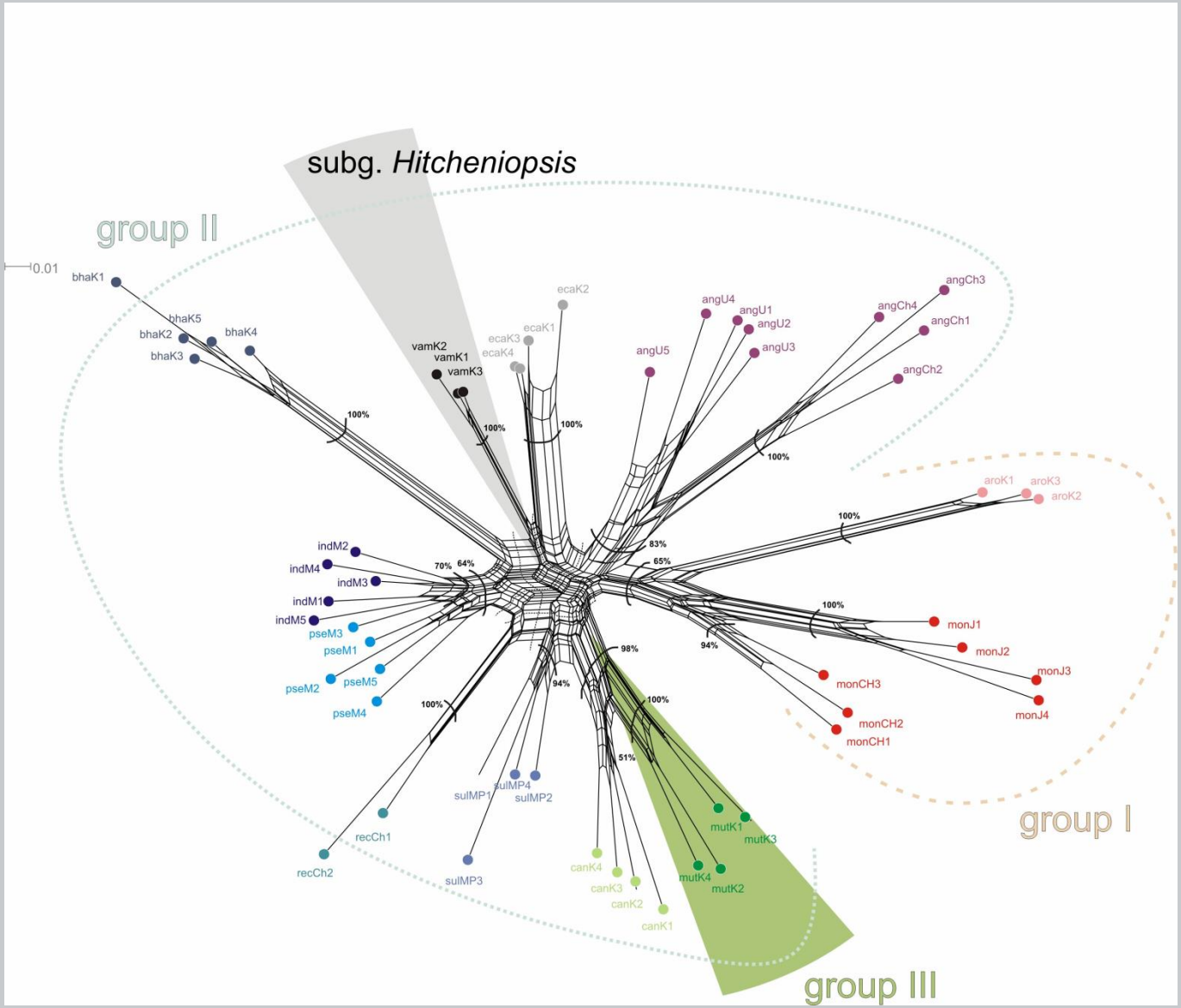
NJ tree – unrooted



Sparganium erectum

Píšová & Fér 2020

Neighbour-network

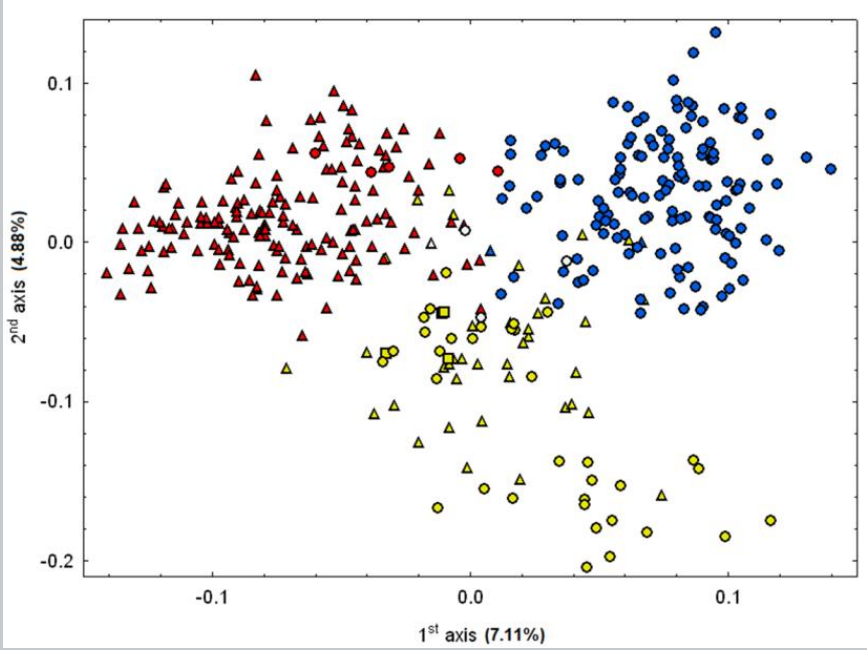


Curcuma

Záveská et al. 2011

Principal **Co**ordinate **A**nalysis
PCoA

- multivariate analysis
- based on matrix of similarity coefficients
- finds directions of highest variability in the dataset
- visualization of variation in the data



Senecio – Dušková et al. 2017

Analysis of **MO**lecular **VA**riance

- population genetic structure testing
- total variance is separated to (co)variance components (groups, populations, individuals)
- which proportion of total variability is
  - among individuals within populations
  - among populations within groups
  - among groups of populations
- calculation of F-statistics analogues (F_{ST})
- software
  - Arlequin v. 3.5 (Excoffier, Laval & Schneider 2005)
  - FAMD etc.

AMOVA design and results :

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	4	33.295	0.05470 Va	15.73
Among populations within groups	5	5.298	0.01250 Vb	3.59
Within populations	662	185.745	0.28058 Vc	80.68
Total	671	224.339	0.34778	

Fixation Indices

FSC : 0.04266

FST : 0.19323

FCT : 0.15728

Population diversity and divergence

- genetic diversity
  - Shannon diversity index
  - average gene diversity

  ➢ reflects reproduction system, recent processes (gene flow, population sizes…)

- *rarity*
  - number (or %) of rare fragments (arbitrary definition of rarity)
  - number (or %) of unique (private) fragments
  - DW-index (*frequency down-weighted marker values*)

  ➢ reflects historical processes (long-term isolation…)

Shannon diversity index

$$H_{Sh} = -\sum_{i=1}^{k} p_i \ln p_i$$

p_i – frequency of *i*-th fragment

Saponaria pumila, AFLP

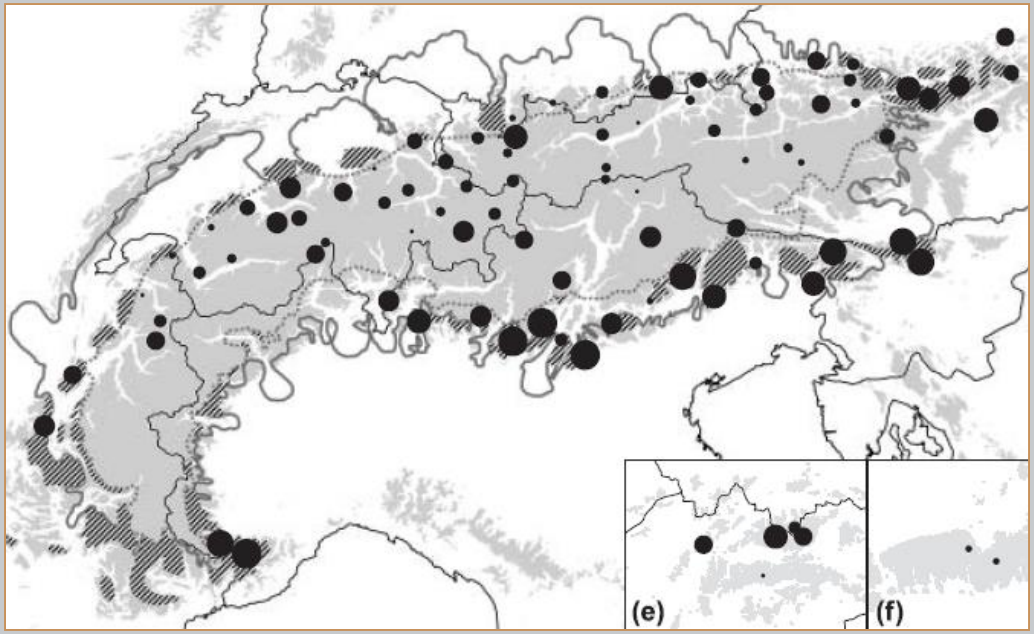
H_{Sh} according to maximal glaciation in the Alps

(Tribsch et al. 2002)

DW-index

frequency down-weighted marker values

- number of occurrences of particular AFLP marker in a population/number of occurrences of the marker in the whole dataset -> sum of the values for all markers
- higher values in long-term isolated populations (accumulation of mutations)
- lower values in newly established populations (recent dispersal)



Ranunculus alpestris, AFLP

distribution of DW-index in the Alps (according to the maximal glaciation)

(Paun et al. 2008)

Mantel test – spatial autocorrelation

- comparison of two matrices → Mantel R_M

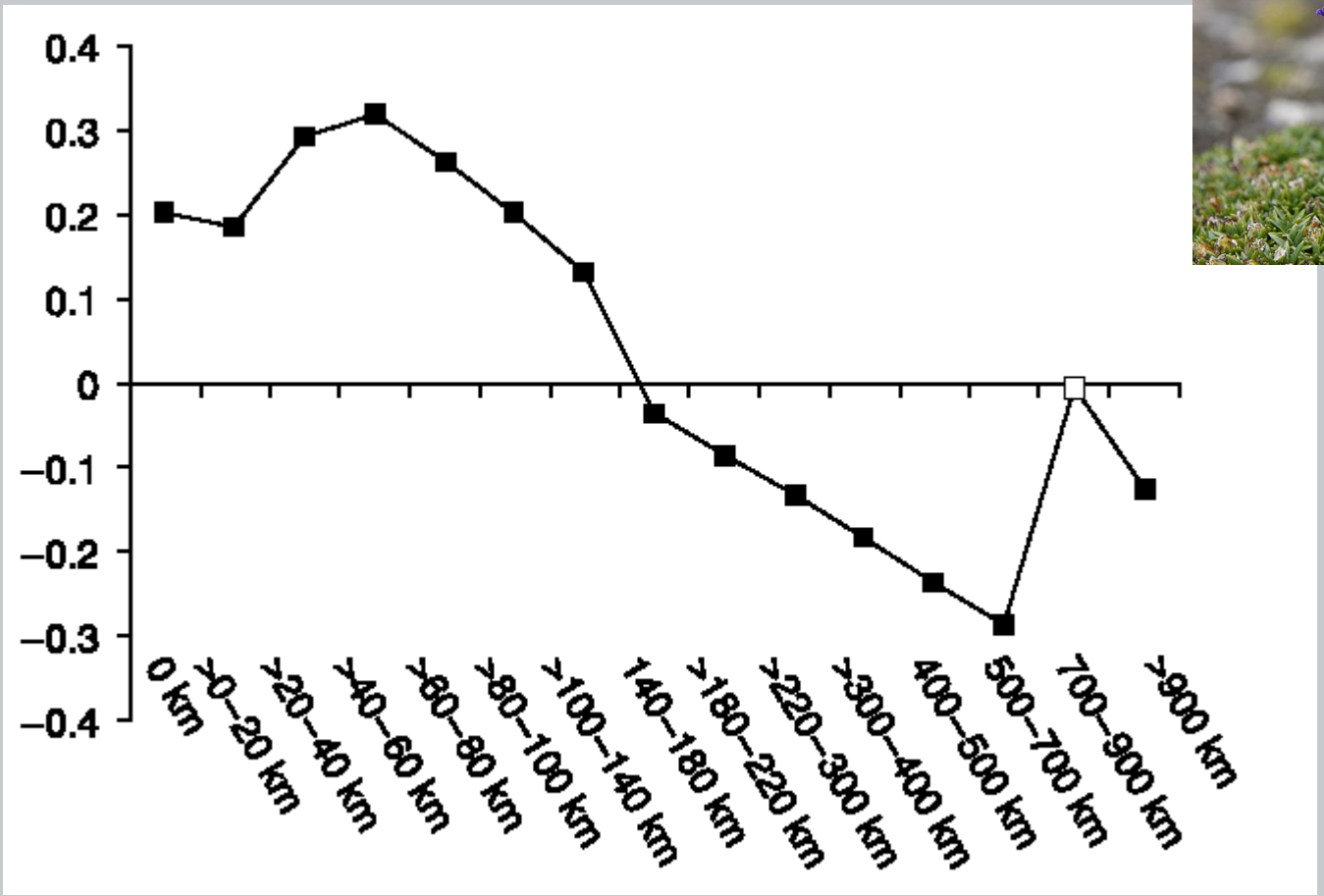
	A	B	C	D	E	F	G	H	I	J
A	0,00000									
B	0,01591	0,00000								
C	0,00000	0,01591	0,00000							
D	0,03046	0,01336	0,03046	0,00000						
E	0,08576	0,10190	0,08576	0,06797	0,00000					
F	0,03731	0,05331	0,03731	0,06797	0,08576	0,00000				
G	0,05331	0,06797	0,05331	0,04389	0,01591	0,10190	0,00000			
H	0,06797	0,08148	0,06797	0,05634	0,03046	0,03046	0,04389	0,00000		
I	0,08576	0,10190	0,08576	0,06797	0,00000	0,08576	0,01591	0,03046	0,00000	
J	0,06797	0,08148	0,06797	0,05634	0,03046	0,03046	0,04389	0,00000	0,03046	0,00000

genetic distances

	A	B	C	D	E	F	G	H	I	J
A	0,0									
B	0,2	0,0								
C	0,2	0,1	0,0							
D	0,1	0,6	0,6	0,0						
E	0,3	0,2	0,6	0,6	0,0					
F	0,3	0,8	0,3	0,2	0,6	0,0				
G	0,3	0,1	0,3	0,8	0,8	0,6	0,0			
H	0,3	0,8	0,6	0,2	0,6	0,2	0,1	0,0		
I	0,8	0,3	0,2	0,6	0,0	0,8	0,2	0,1	0,0	
J	0,8	0,1	0,6	0,1	0,6	0,6	0,3	0,0	0,3	0,0

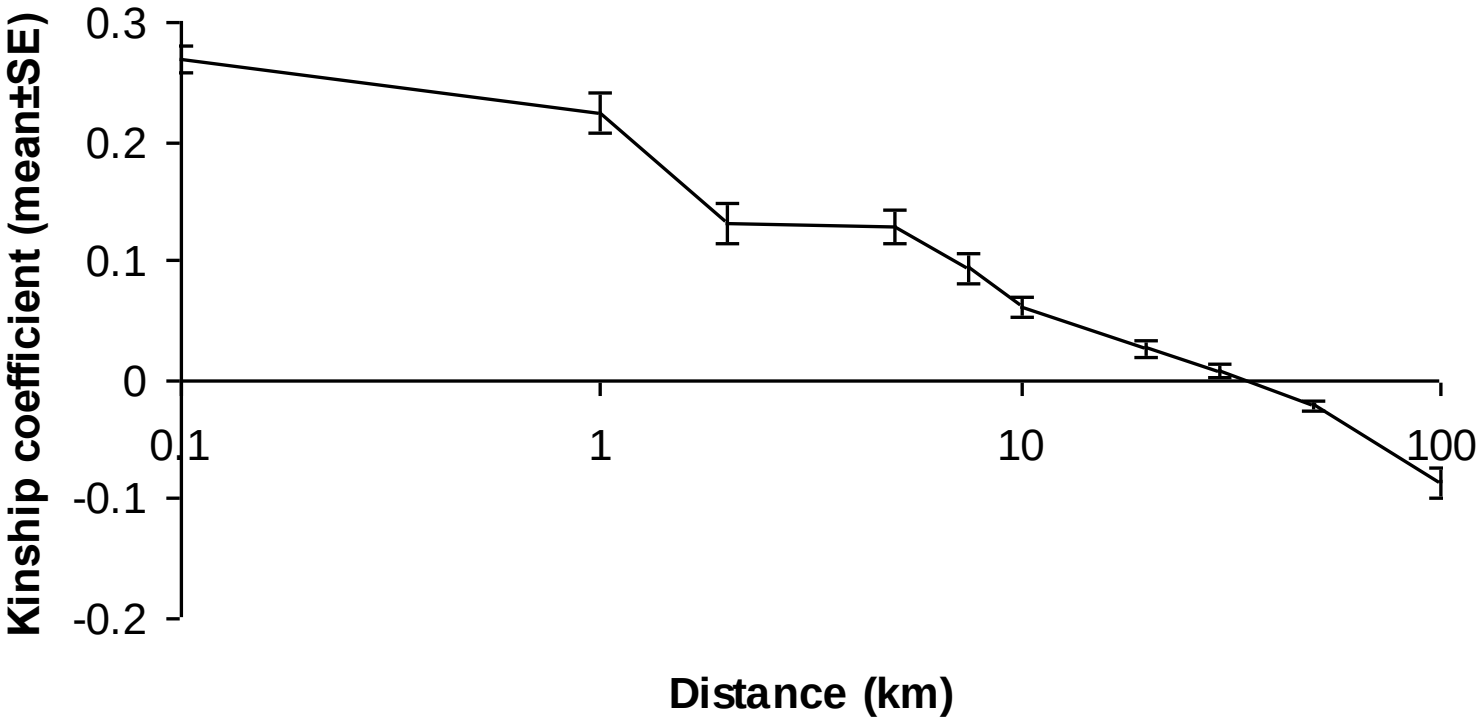
geographic distances

Mantel correlogram



Phyteuma globulariifolium in the Alps (Schönswetter *et al*. 2002)

Spatial autocorrelation



Sparganium erectum in the river Cidlina (Fér & Pfosser unpubl.)

Bayesian clustering

- searching for an optimal partitioning of individuals to *K* clusters, i.e., with maximum negative logarithm of the marginal likelihood
- result is an optimal number of clusters (i.e., „real populations“) and assignment of all individuals to that clusters
- mixture and admixture analysis
- software
  - BAPS – Bayesian analysis of population structure (Corander *et al*. 2003)
  - *Structure* (Pritchard *et al*. 2000)
  - *Geneland* (Guillot *et al.* 2005)

Output from *BAPS v3.2*

```
RESULTS OF INDIVIDUAL LEVEL MIXTURE ANALYSIS:
Data file: S_preprocessed_BAPS.mat
Number of clustered individuals: 258
Number of groups in optimal partition: 8
Log(marginal likelihood) of optimal partition: -8062.66
Best Partition:
Cluster 1: {10, 11, 33, 40, 41, 58, 59, 60, 61, 101, 114, 123,
           124, 126, 131, 132, 137, 159, 161, 162, 164, 166,
           167, 168, 169, 243, 244, 245, 246, 251, 252, 253,
           254}
Cluster 2: {47, 49, 50, 51, 52, 56, 57, 64, 68, 69, 70, 75,
           77, 81, 85, 86, 91, 112, 113, 115, 117, 130, 155,
           173}
Cluster 3: {25, 26, 27, 28, 29, 138, 139, 140, 141, 146, 147,
           148, 149, 150, 151, 156, 160, 163, 165, 184, 185,
           186, 187, 188, 189, 190, 191, 192, 194, 196, 198,
           199, 200, 201, 202, 203, 204, 205, 206, 207}

Changes in log(marginal likelihood) if indvidual i is moved to group j:
  ind      1        2        3        4        5        6        7        8
    1:   -61.0    -20.1    -77.6    -94.0       .0    -67.1   -125.9    -71.3
    2:   -50.4    -18.2    -69.4    -85.4       .0    -60.5   -117.8    -70.9
    3:   -22.5    -60.5    -29.7    -83.0    -64.8       .0    -70.6   -130.4
    4:   -22.9    -58.9    -28.8    -83.2    -61.3       .0    -69.3   -126.6
    5:   -22.3    -53.1    -28.7    -78.5    -61.0       .0    -65.7   -123.4

KL-divergence matrix (Kullback-Leibler):
       1       2       3       4       5       6       7       8
  1
  2   0.415
  3   0.507   1.062
  4   0.407   0.637   1.515
  5   0.782   0.244   1.030   1.433
  6   0.327   0.817   0.269   1.353   0.826
  7   0.583   1.069   1.009   0.780   1.603   0.944
  8   1.050   0.613   1.833   0.910   0.953   1.751   0.776

Probabilities for number of clusters
8    0.9984
9    0.001605
```

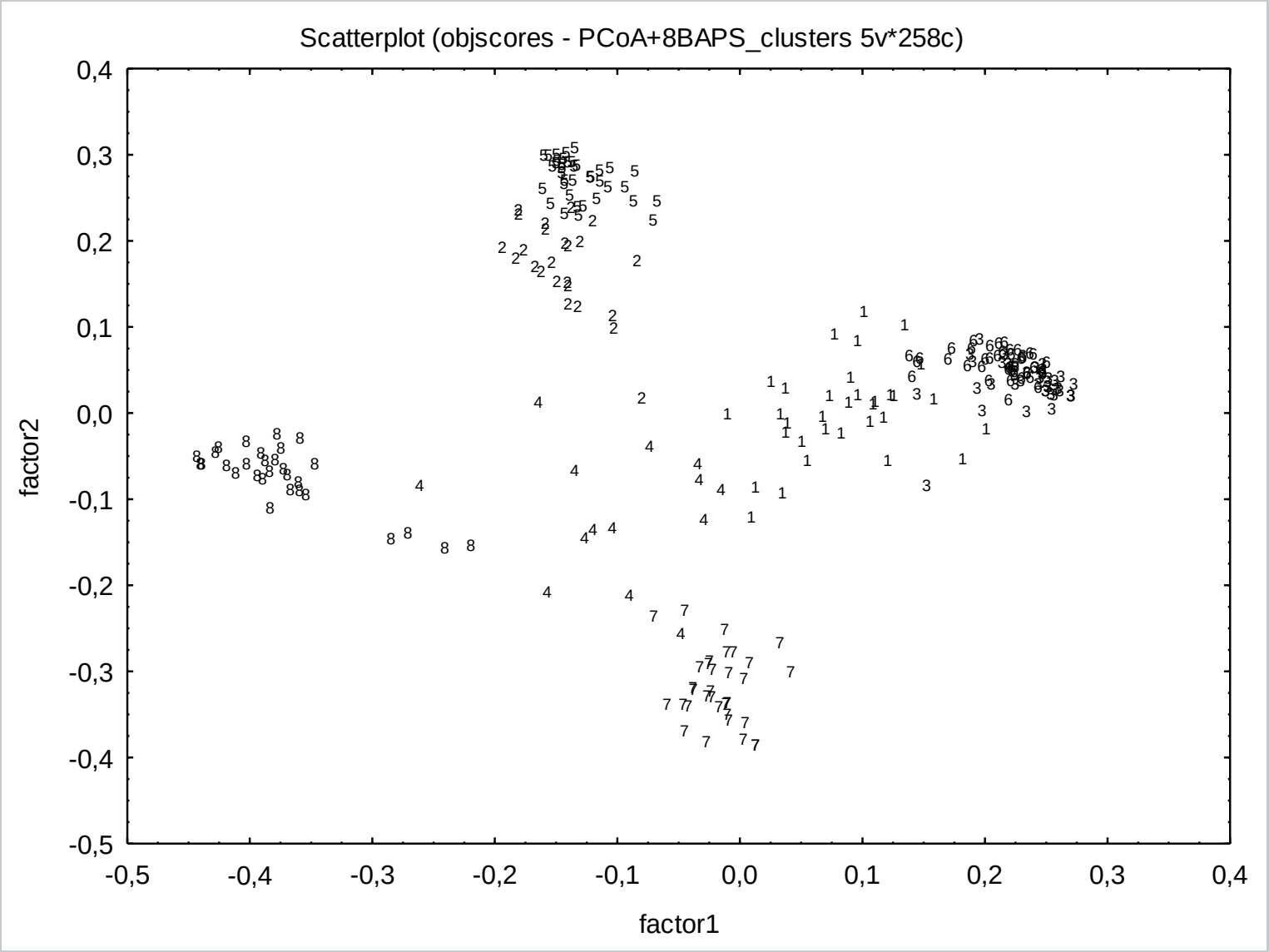
individual assignment to the clusters

change of the model likelihood when an individual is transfered to another cluster

similarity among clusters

model probabilities

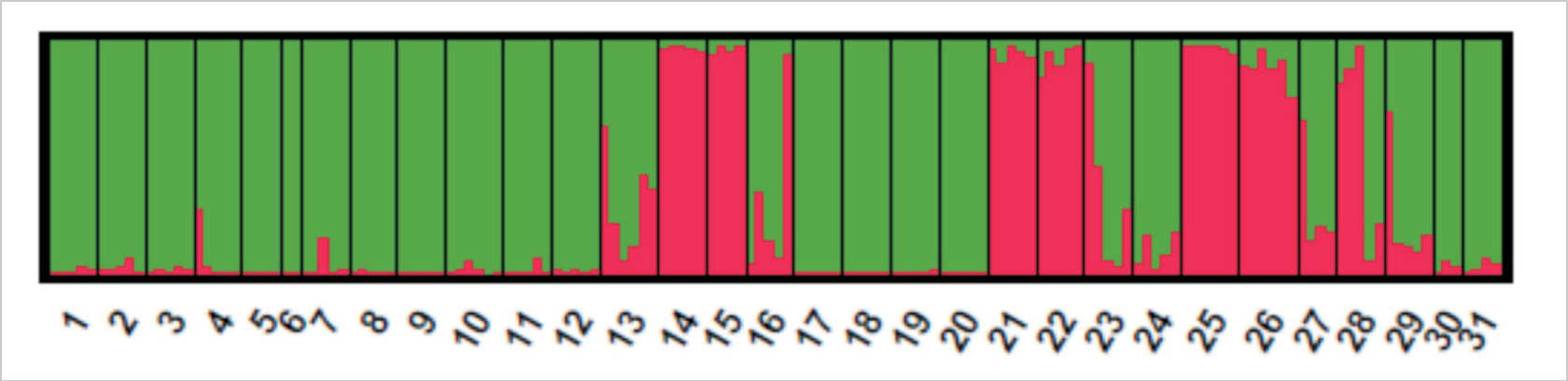
Comparison of PCoA and Bayesian clustering



Results from *Structure*

Structure (Pritchard *et al*. 2000)

- minimalizes deviances from H-W equilibrium
- minimalizes *linkage disequilibrium*
- results subsequently processed with CLUMPP and Distruct



Lonicera nigra – Daneck et al. 2011

Applications of dominant markers

- clone identification
- cultivar identification
- phylogeography
- parentage analysis
  - identification of parents of an individual
- systematics – phylogeny, polyploid speciation…
  - relationships within a species or among closely related species
- hybridization – specific bands for each taxon are shared in hybrid
- …

Clone definition

Escavarage N. et al (1998): Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. Molecular Ecology 7:975-982.

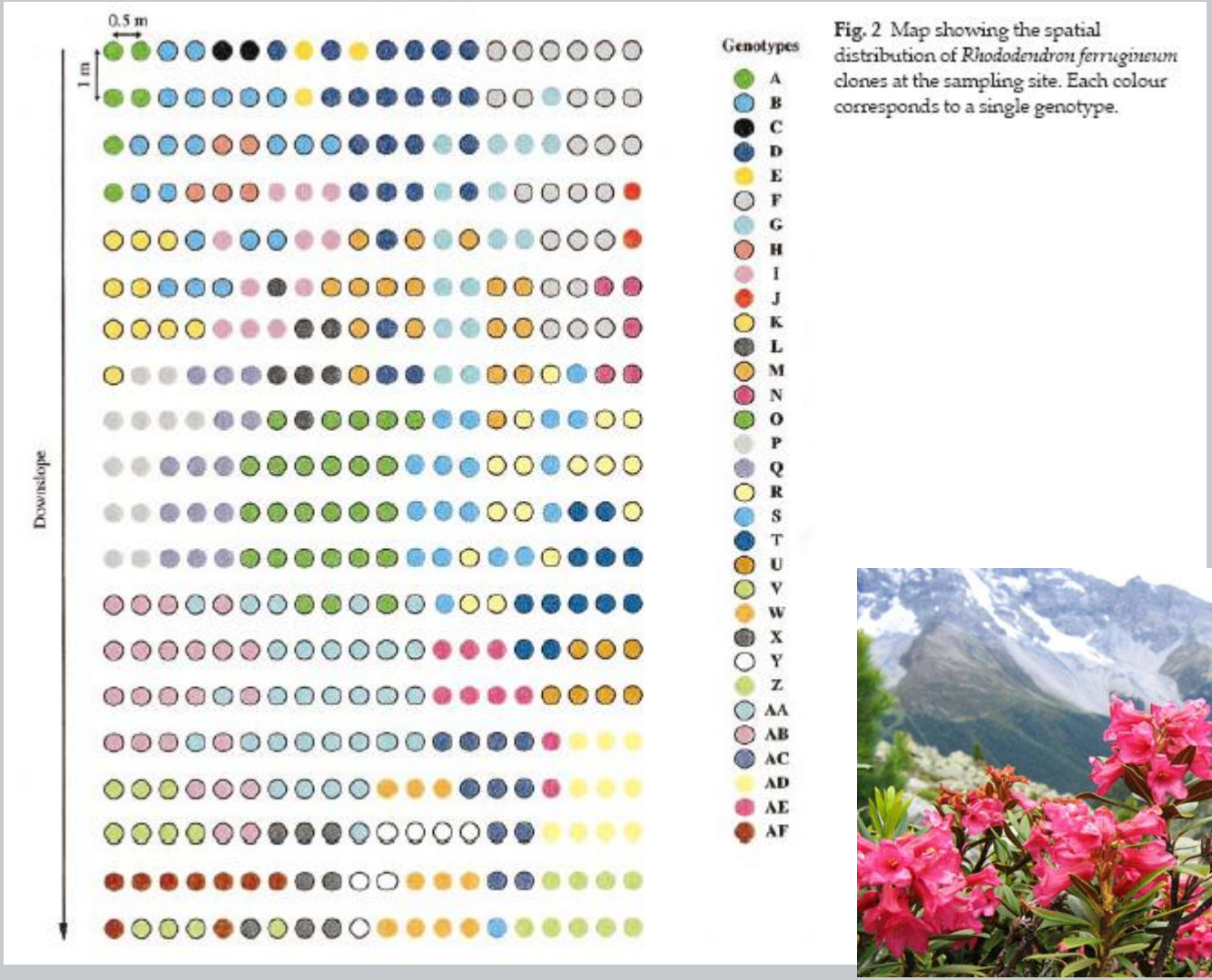
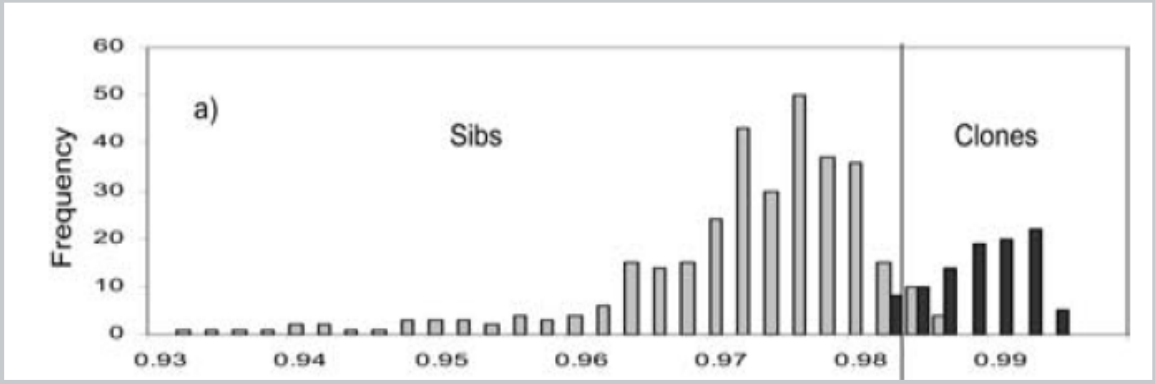


Fig. 2 Map showing the spatial distribution of *Rhododendron ferrugineum* clones at the sampling site. Each colour corresponds to a single genotype.

Clone definition

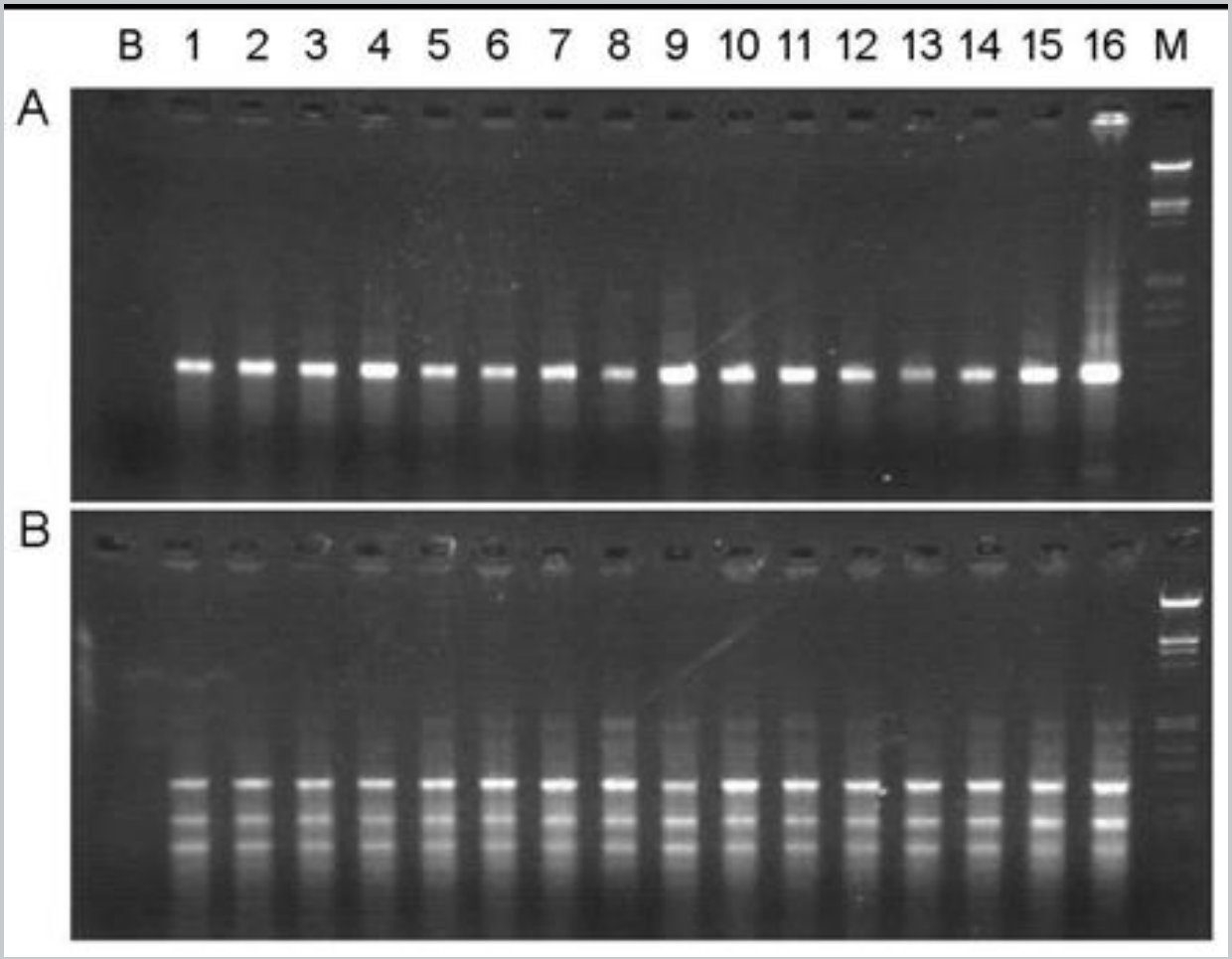
- clone – identical AFLP profile
- *pairwise similarity threshold* – individuals up to certain similarity (e.g., 0.98) are still accepted as clones
- sources of errors
  - contaminations
  - methodological problems (non-specific restriction, PCR inhibition…)
  - scoring errors (+ subjectivity when selecting loci for analysis)
  - somatic mutations



Salix exigua, AFLP – Douhovnikoff & Dodd, 2003

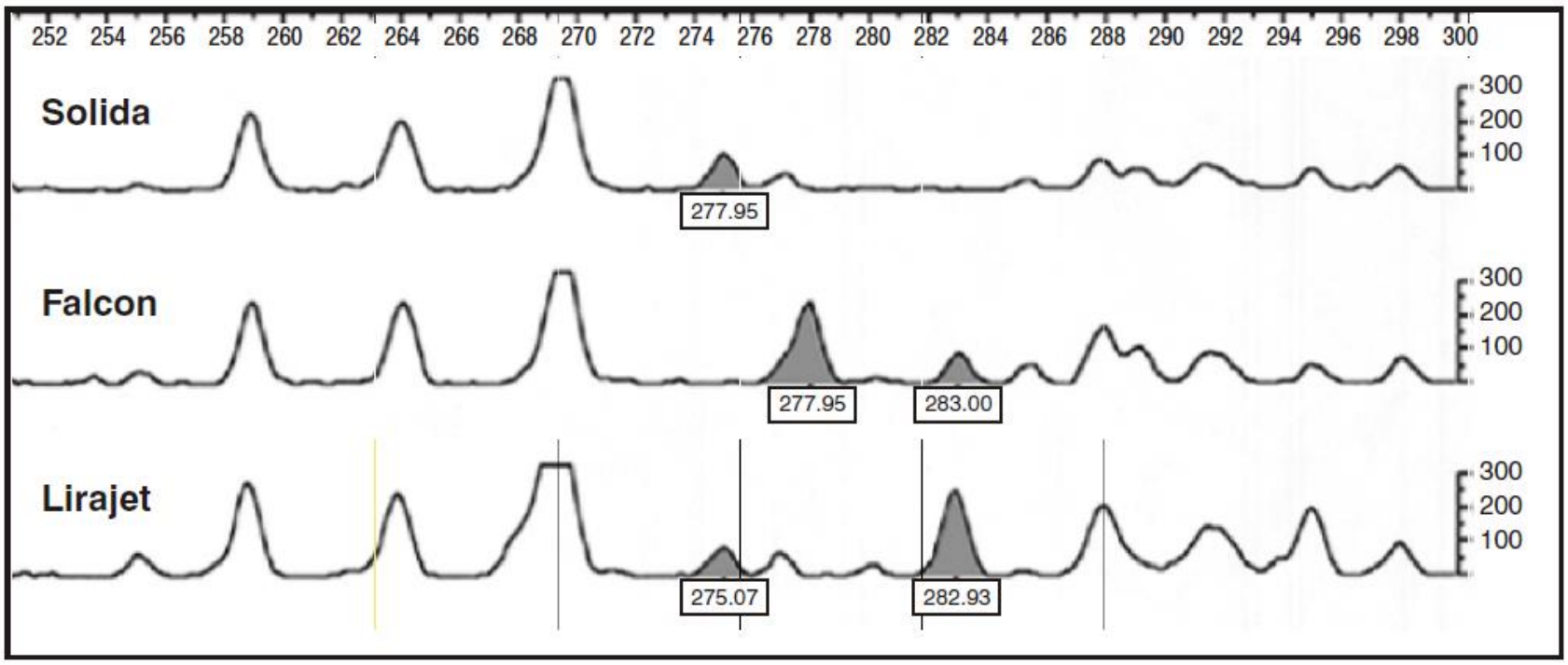
Clone identification

Camellia sinensis, RAPD



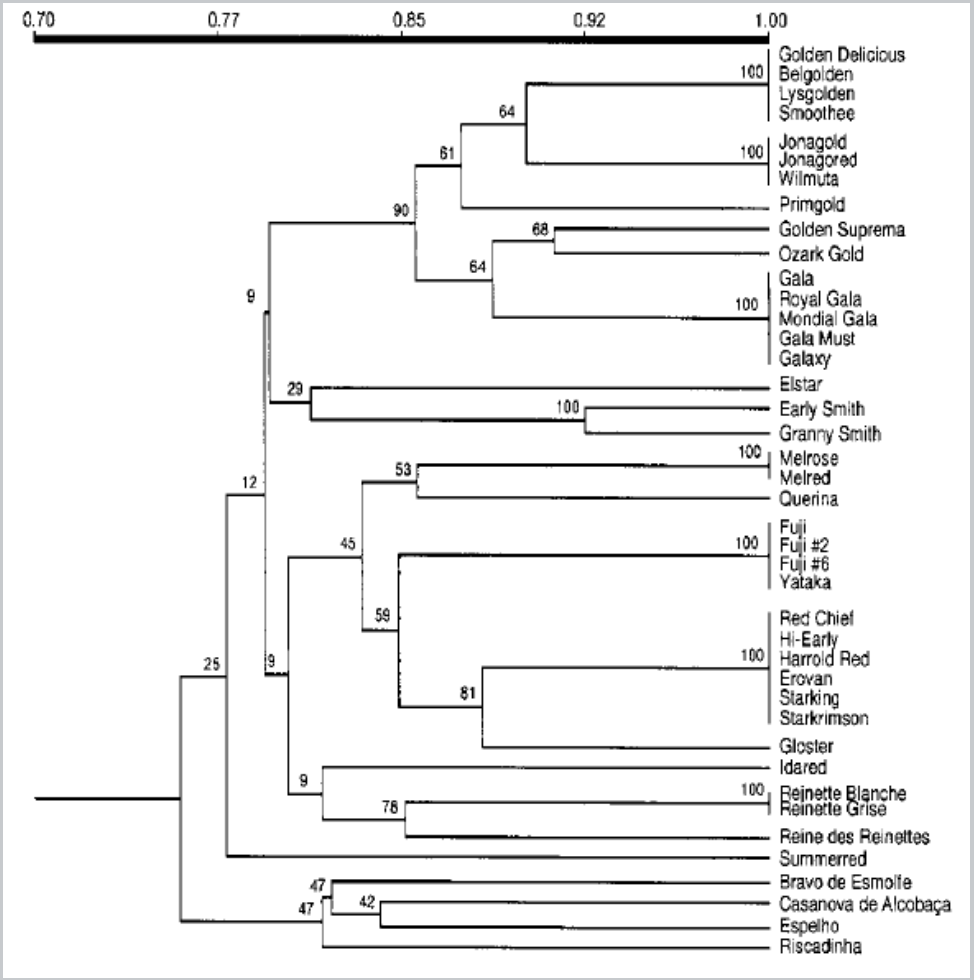
Singh et al. (2004)

Identification of cultivars



Brassica napus, AFLP – Čurn et al., 2006

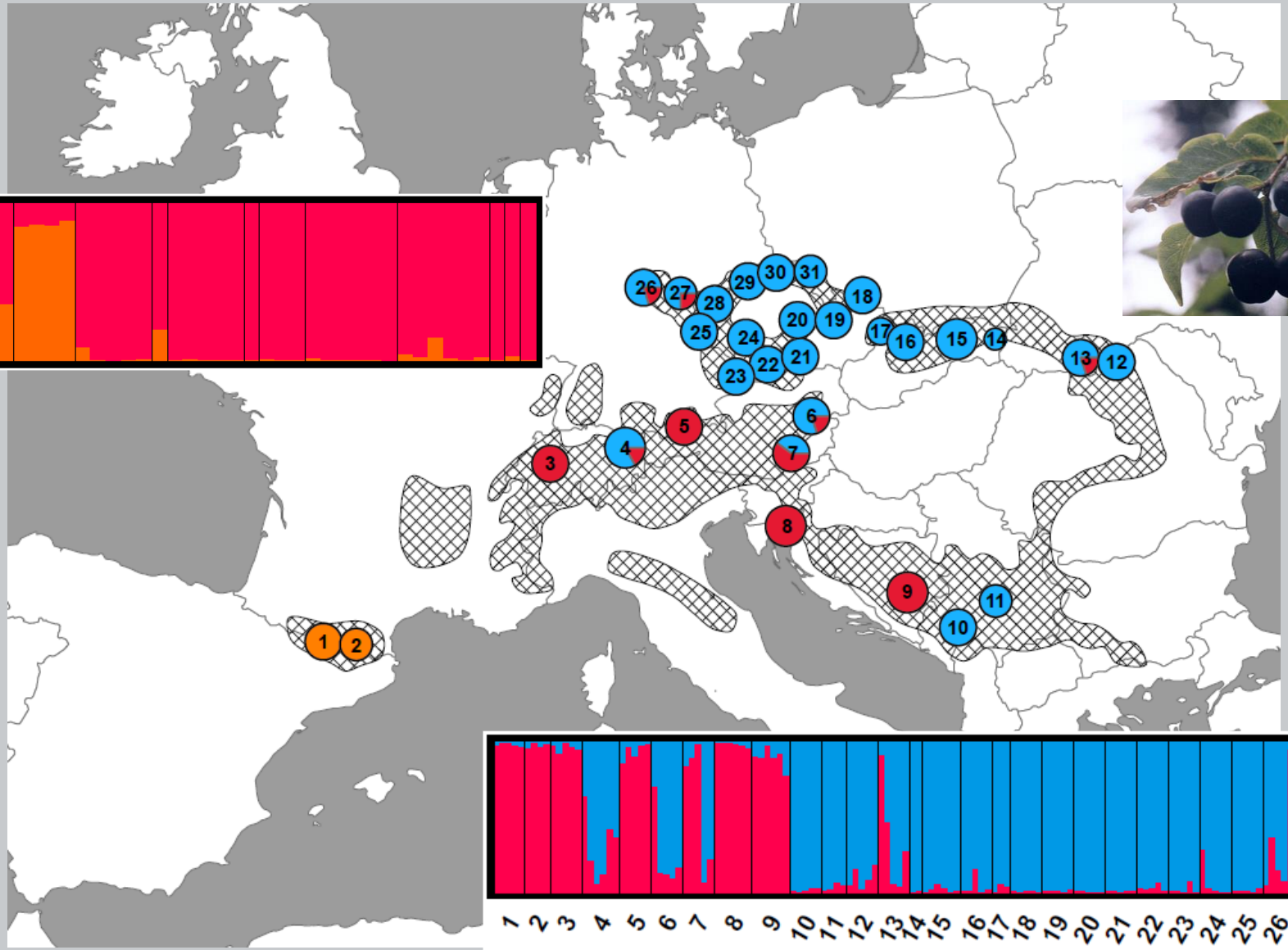
ISSR – apple cultivars



ISSR primer	Number of scored bands	Number of polymorphic bands
(AG)8YT	44	25
(AGC)4YT	40	24
(CA)8R	33	22
(GA)8YG	31	20
DBD(CA)7	32	27
VHV(GT)7	32	26
HVH(TG)7	40	32

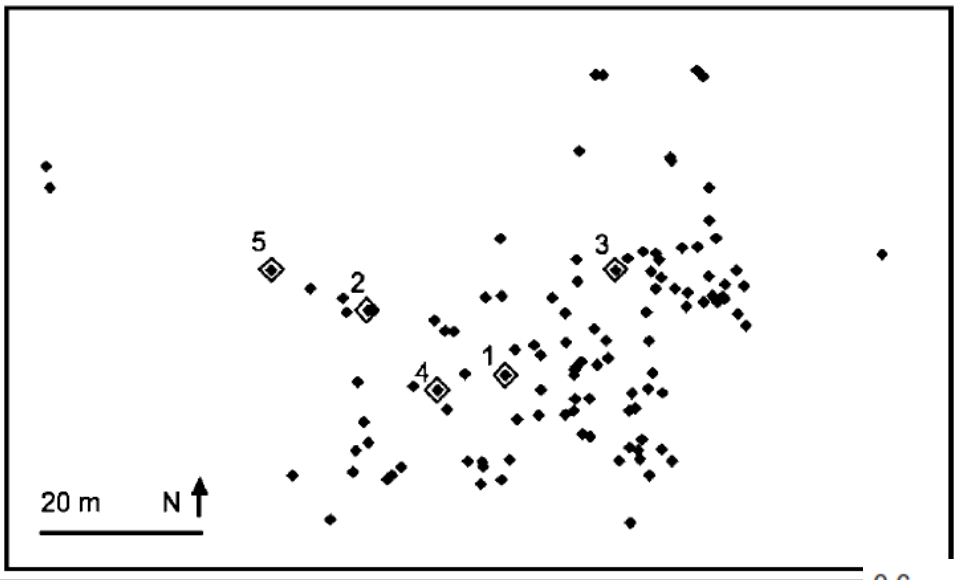
Goulão L. & Oliveira C.M. (2001): Molecular characterisation of cultivars of apple (*Malus* × *domestica* Borkh.) using microsatellite (SSR and ISSR) markers. *Euphytica* 122: 81–89.

Phylogeography

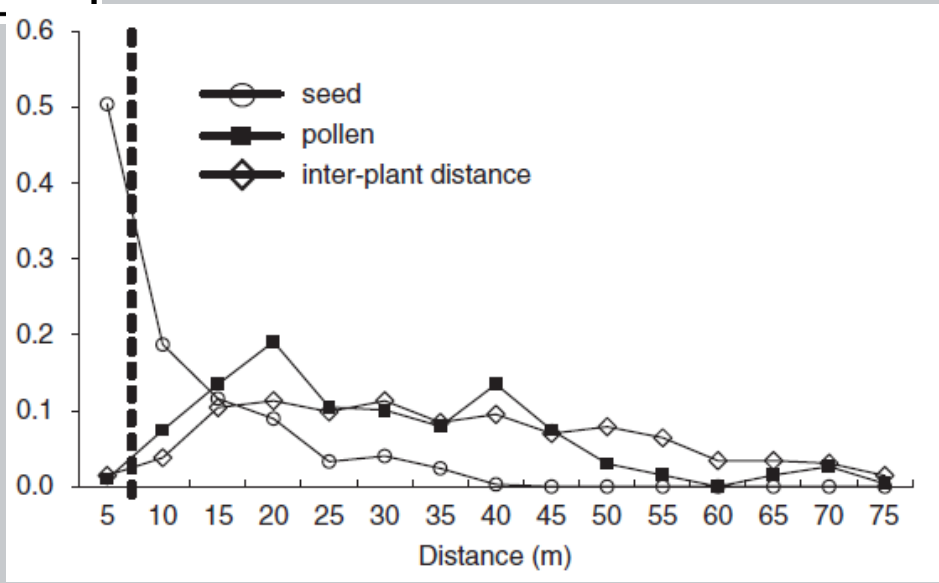


Lonicera nigra – Daneck et al. 2011

Parentage analysis



- assesment of parents of seeds
- analysis of pollen and seed movement



Banksia hookeriana – Krauss et al. 2009

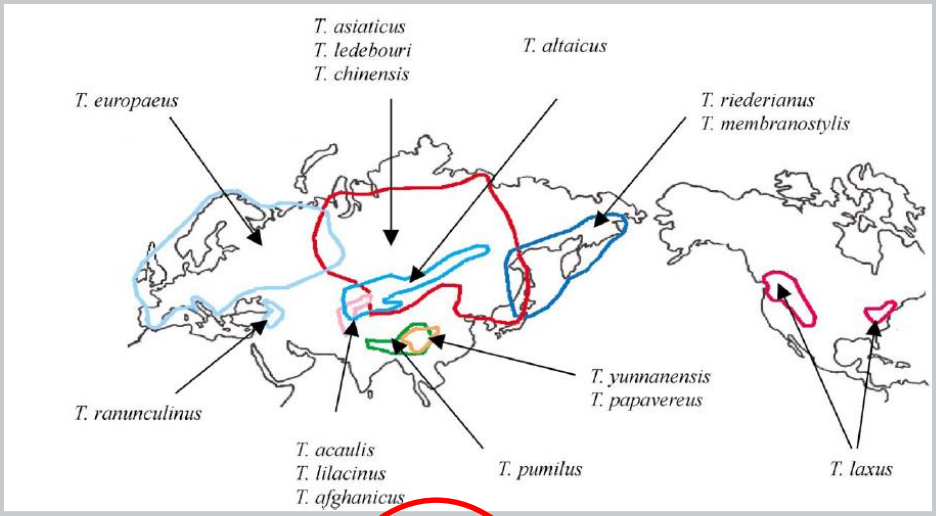
Dominant markers in systematics

- homology – the more similar taxa the higher probability that comigrating fragments are homologous
- useful when differences in ITS sequences are no more than 30 bases (Koopman 2005)
- relationships of species within a genus
- intraspecific level
- evolution of polyploids
- study of hybridization, introgression
- successful when sequences were insufficiently variable

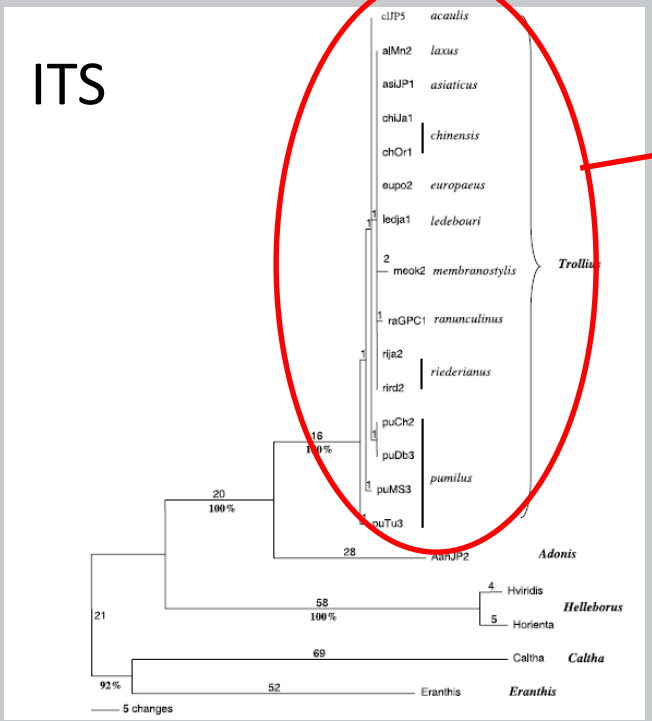
Analyses for systematic purposes

- phenetic methods
  - similarity coefficients – Nei & Li, Jaccard, Dice …
  - cluster analyses – UPGMA, NJ …
- cladistic methods
  - phylogeny reconstruction
  - character changes (0 → 1 → 0)
  - maximum parsimony
  - cladogram

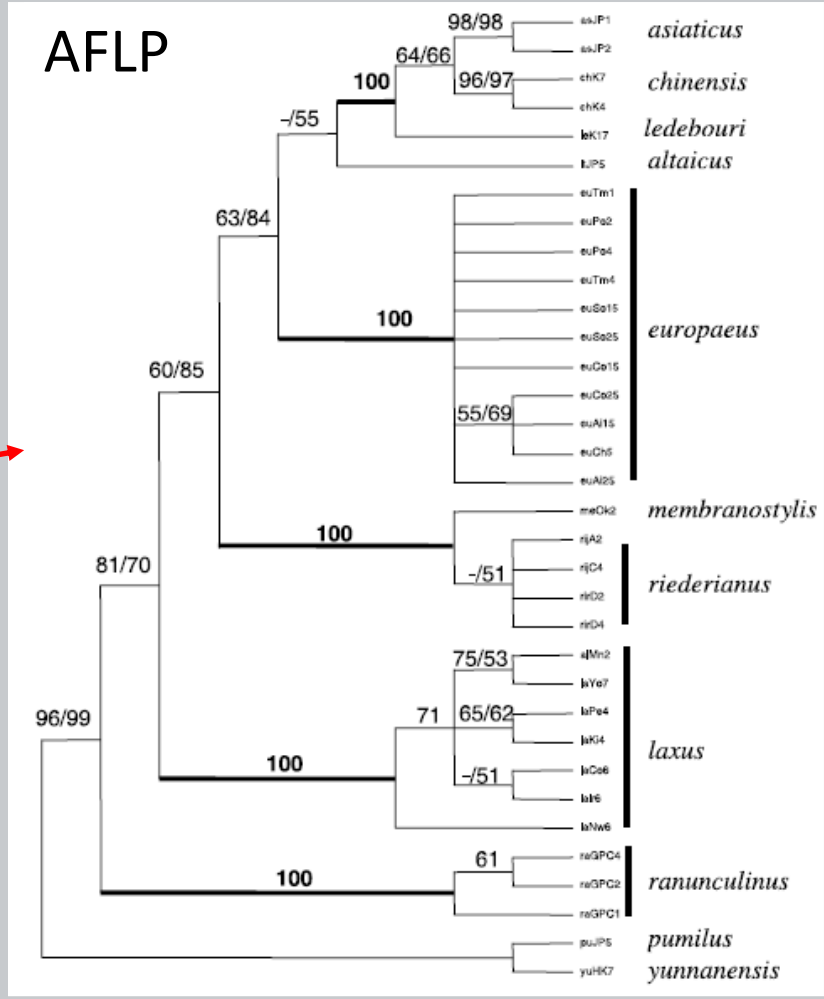
Phylogeny reconstruction



ITS



AFLP



Trollius L. – Després *et al.*, 2003

Study of hybridization and polyploid evolution

Guo Y.-P. et al. (2006): Hybrid origin and differentiation of two tetraploid *Achillea* species in East Asia: molecular, morphological and ecogeographical evidence. Molecular Ecology 15:133-144.

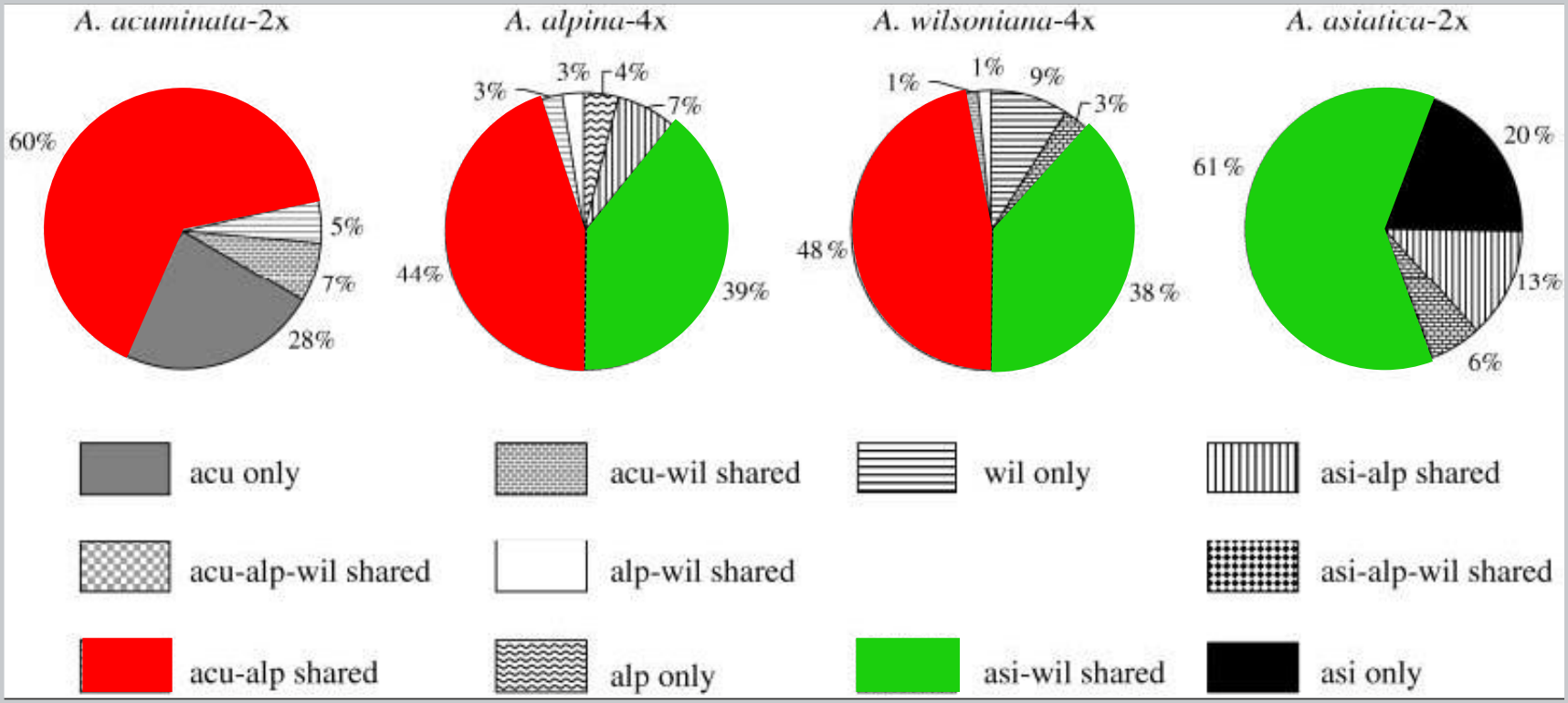


Fig. 2 Frequencies of species-specific and species-shared AFLP bands in populations of *Achillea acuminata*-2x, *Achillea alpina*-4x, *Achillea wilsoniana*-4x, and *Achillea asiatica*-2x in East Asia. The 4x-species share many AFLP bands with the two presumed 2x parental species, but also exhibit a few specific bands.

Polyploid complexes

15x
12x
11x
9x
6x

Y10, Y64 | Y60
K — *C. zedoaria* — As

C. sp "ranchi"

C. oligantha

Y30
C. leucorrhiza

Y69, Y120
C. zanthorrhiza

Y98

G35 | B69
B32, B9 | G58 | Y122
K1 + K2 — *C. aeruginosa* — As

G90 | Y97 | Y48, Y85, Y99
K — *C. longa* — SL

B125, B45, Y101 | B27, B56
Y96, G82, G86 | B21, B116
Ch — *C. montana* — J

Y119, Y117, B76, B39, Y56,
G32, G45, Y9, Y113, B92
C. aromatica

B67 | G11
Y89
U *C. angustifolia* Ch

Y86
B60, B65, B74
C. aurantiaca

Y17, B4, Y22, Y66
B11, B46, B85, G91
C. bhatii

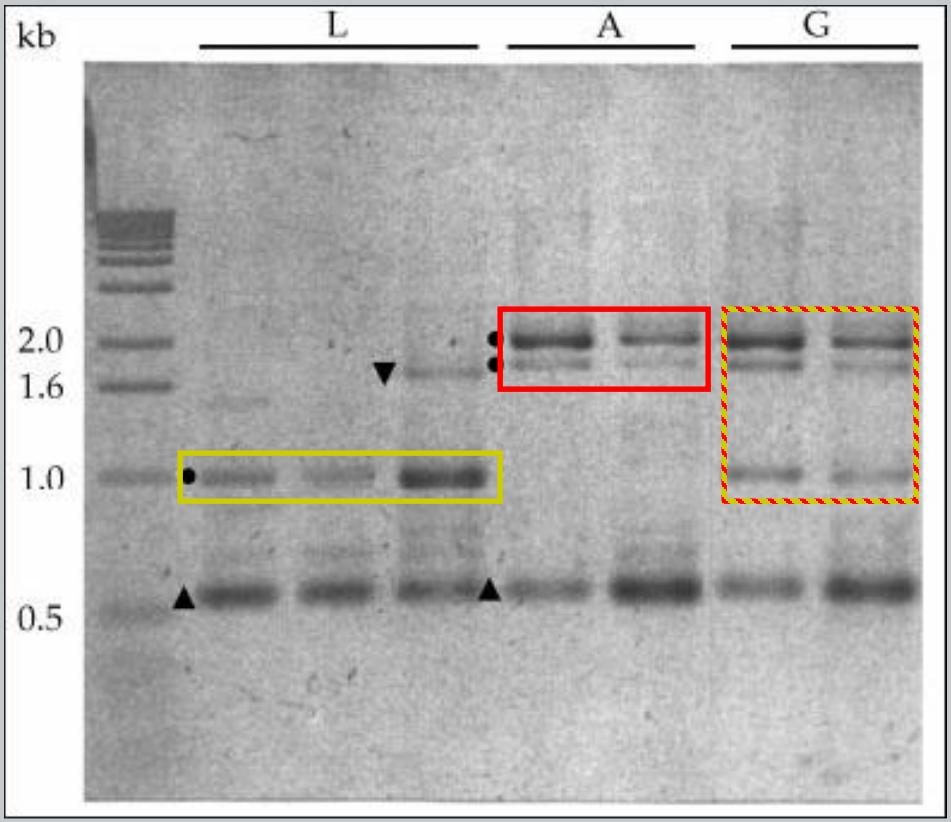
Genome size group I

Genome size group II

Curcuma – Záveská et al. 2011

Hybridization

Typha ×*glauca*

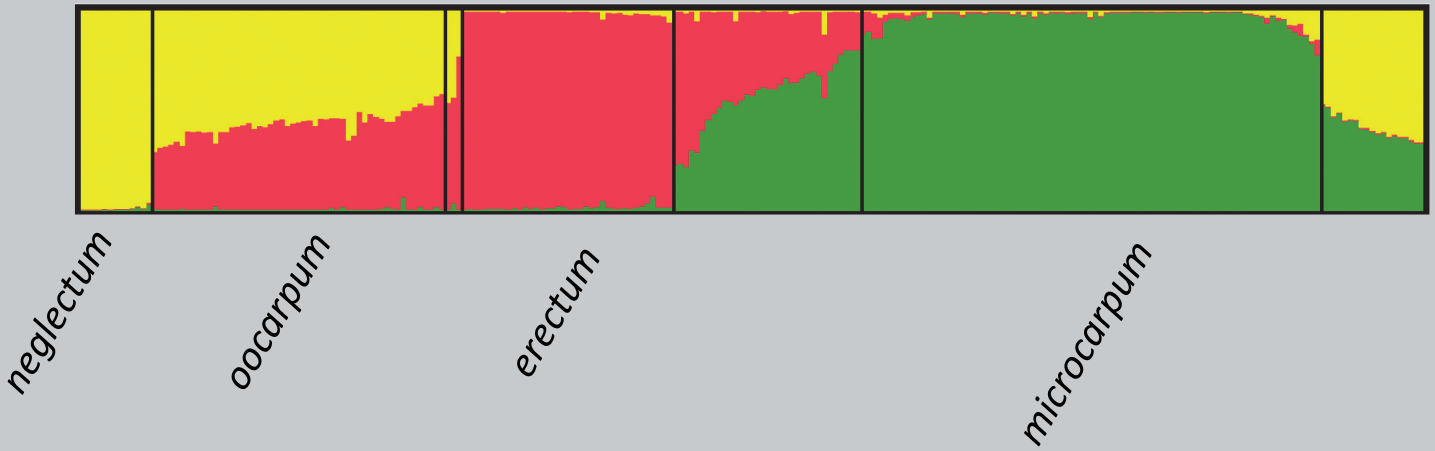


- *T. latifolia*
- *T. angustifolia*

Kuehn et al. (1999), RAPD

Testing a hybrid origin

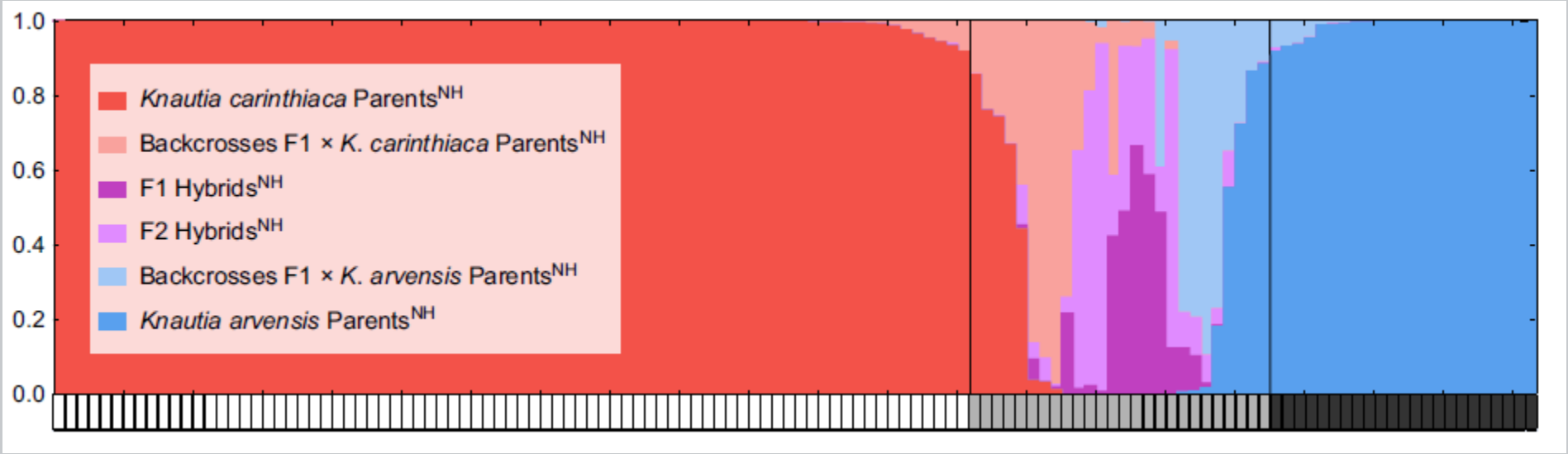
- hybrid shares fragments with both parents
- probabilistic (model) detection
  - Structure (Pritchard et al. 2000; Falush et al. 2003)
  - Newhybrids (Anderson & Thompson 2002)



4 subspecies of *Sparganium erectum,* AFLP – Píšová & Fér (2020)

Testing a hybrid origin

- Newhybrids (Anderson & Thompson 2002)
  - probability of F1, F2 or back-crosses with both parents



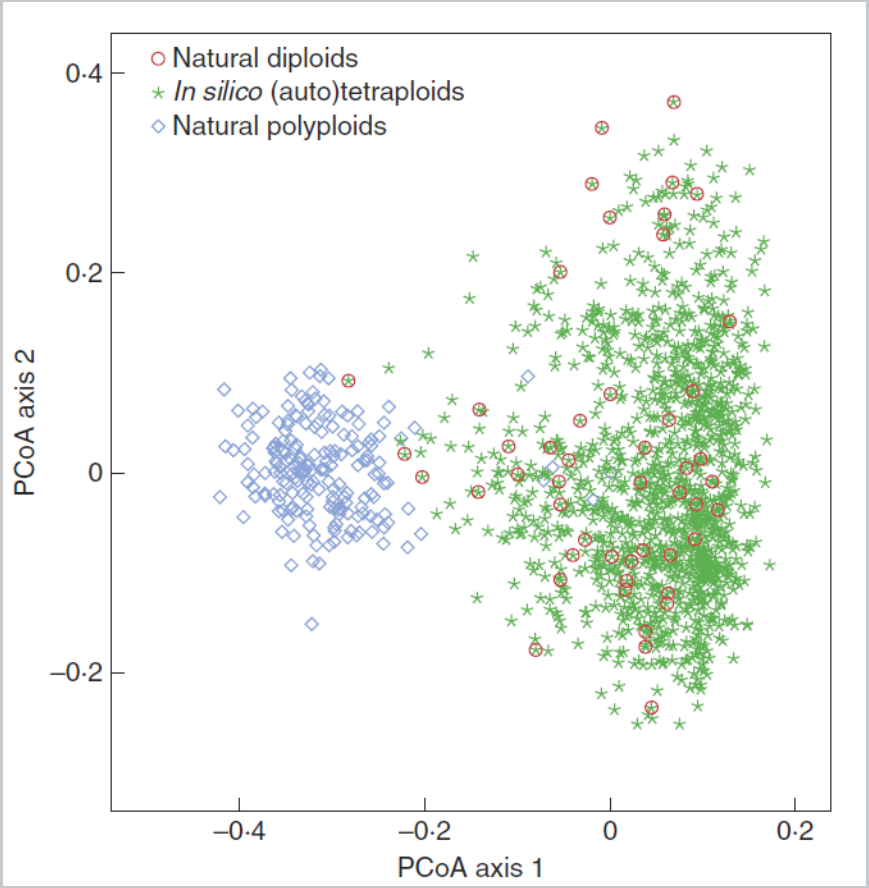
Knautia carinthiaca

Čertner et al. (2015)

Knautia arvensis

Auto- and allopolyploid origin

in-silico hybrids



tetraploids
hexaploids

diploids

Greiner et al. (2013)

Leucanthemum pluriflorum

Population study

Tribsch A., Schönswetter P. & Stuessy T. (2002): *Saponaria pumila* (Caryophyllaceae) and the Ice Age in the European Alps. *American Journal of Botany* 89(12): 2024-2033

Systematic study

Schenk M.F. et al. (2008): Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics & Genomes* 4: 911–924

Literature

Harris S.A. (1999): *RAPDs in systematics – a useful methodology ?* In: Hollingsworth & al. [eds.]: Molecular systematics and plant evolution, pp. 211-228

Wolfe A.D. & Liston A. (1998): *Contributions of PCR-based methods to plant systematics and evolutionary biology*. In: Soltis D.E. & al. [eds.]: Molecular systematics of plants. II. DNA sequencing, pp. 43-86

Simmons M.P., Zhang L.-B., Webb C.T., Müller K. (2007): *A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference*. Molecular Phylogenetics and Evolution 42: 528–542.

Vos P. et al. (1995): *AFLP: a new technique for DNA fingerprinting.* Nucleic Acids Research, 23(21):4407-4414.

Mueller U.G. & Wolfenbarger L.L. (1999): *AFLP genotyping and fingerprinting*. Trends in Ecology & Evolution 14:389-394.

Meudt H.M. & Clarke A.C. (2007): *Almost forgotten or latest practice? AFLP applications, analyses and advances*. Trends in Plant Science:106-117.

Bonin A. et al. (2004): *How to track and assess genotyping errors in population genetics studies*. Molecular Ecology 13:3261-3273.

Bonin A. et al. (2007): *Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists*. Molecular Ecology 16:3737-3758.

Robinson J.P. & Harris S.A. (1999): *Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective*. In: Gillet E.M.[ed.]: Which DNA marker for which purpose? http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Koopman W.M.J. (2005): *Phylogenetic signal in AFLP data sets*. Systematic Biology 54:197-217.