

Molecular markers in plant systematics and population biology

5. RFLP, PCR-RFLP, cpDNA, mtDNA

Tomáš Fér

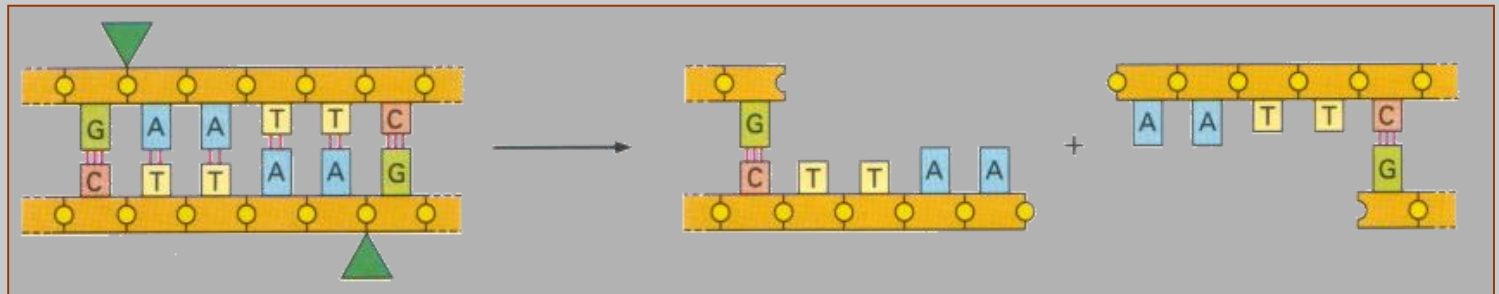
tomas.fer@natur.cuni.cz

Restriction Fragment Length Polymorphism

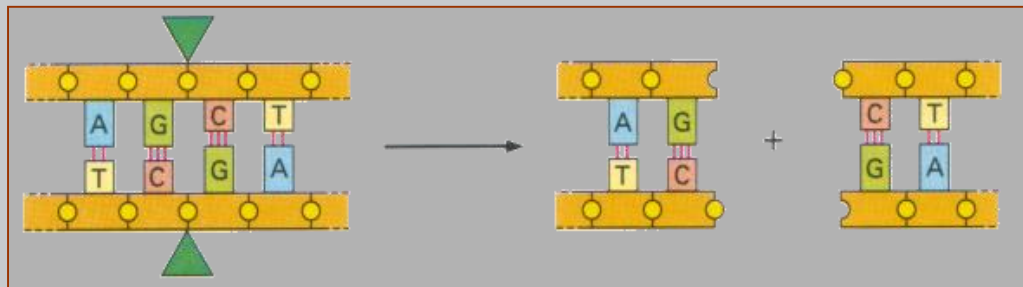
- polymorphism in length of the fragments produced by restriction of DNA
- basic principle of the method
 - cleavage of total DNA to fragments with the use of restriction endonucleases
 - electrophoretic separation of fragments according to their length
 - transfer of fragments to the membrane
 - hybridization with labelled probe

Restriction endonucleases

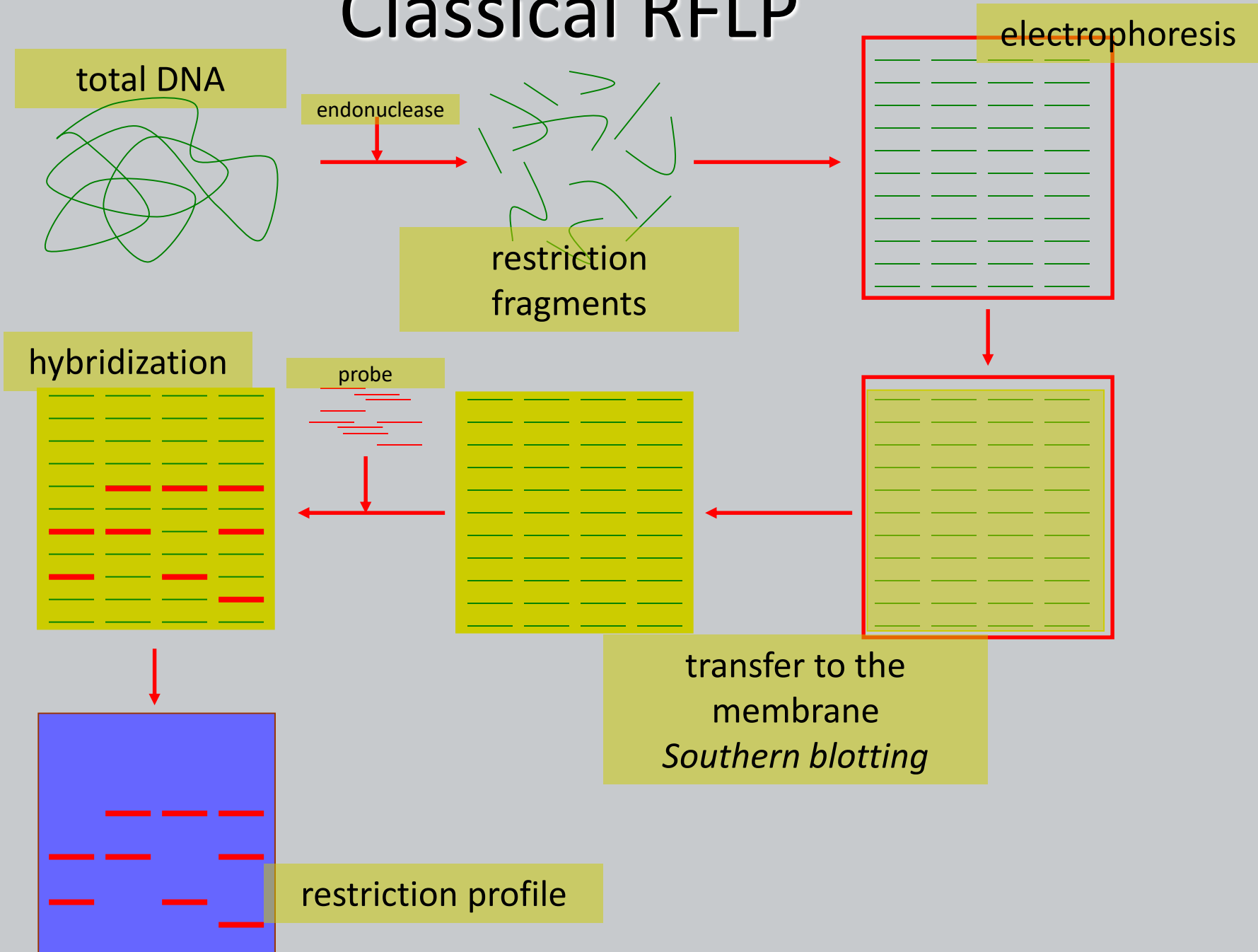
- enzymes isolated from bacteria
 - *EcoRI* – *Escherichia coli*, *AluI* - *Arthrobacter luteus* ...
- specific cleavage of dsDNA (*double stranded*)
- recognise particular sequence
 - palindrom – symmetry along the central point
 - mostly 4 bp or 6 bp
 - asymmetric cleavage (*sticky ends*) – e.g., *EcoRI*



- symmetric cleavage (*blunt ends*) – e.g., *AluI*



Classical RFLP

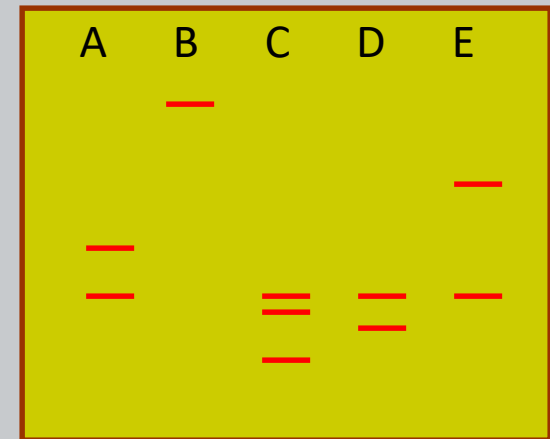
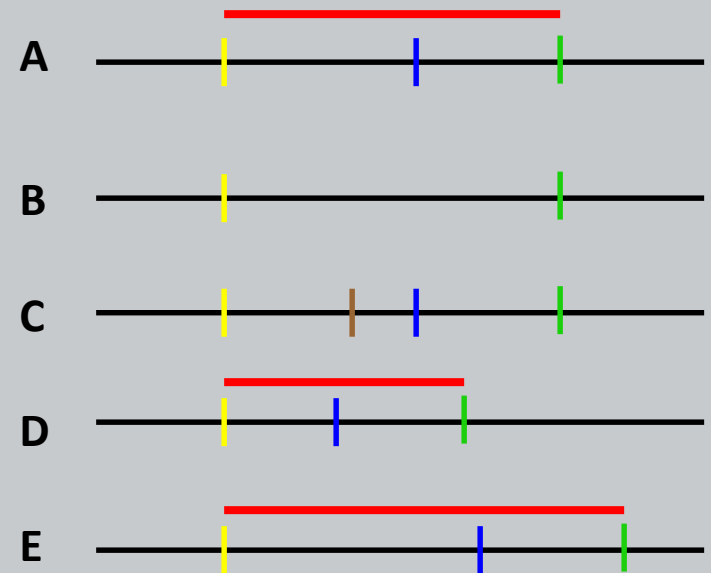


Probe preparation

- bacterial cloning of particular part of DNA (gene, operon)
- (non)radioactive product labelling – ^{32}P , ^{35}S , fluorescence etc.
- using probe from the studied (or similar) taxa
- i.e., rDNA probe – visualization of DNA fragments originating from this region
- frequent use of cpDNA probes

Polymorphism

- *mutation* in the restriction site
 - loss (B) is more probable than gain of the new site (one change on whichever position is sufficient)
 - new site (C)
 - specific change
- *insertion* (E) or *deletion* (D) between two restriction sites



— hybridizing part of DNA
 —+— restriction site on template DNA

RFLP

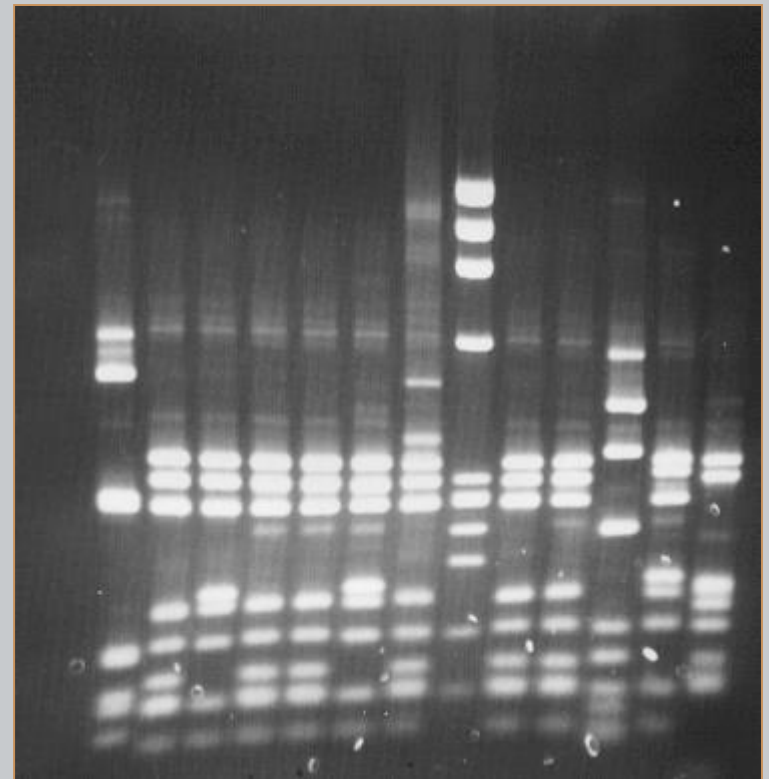
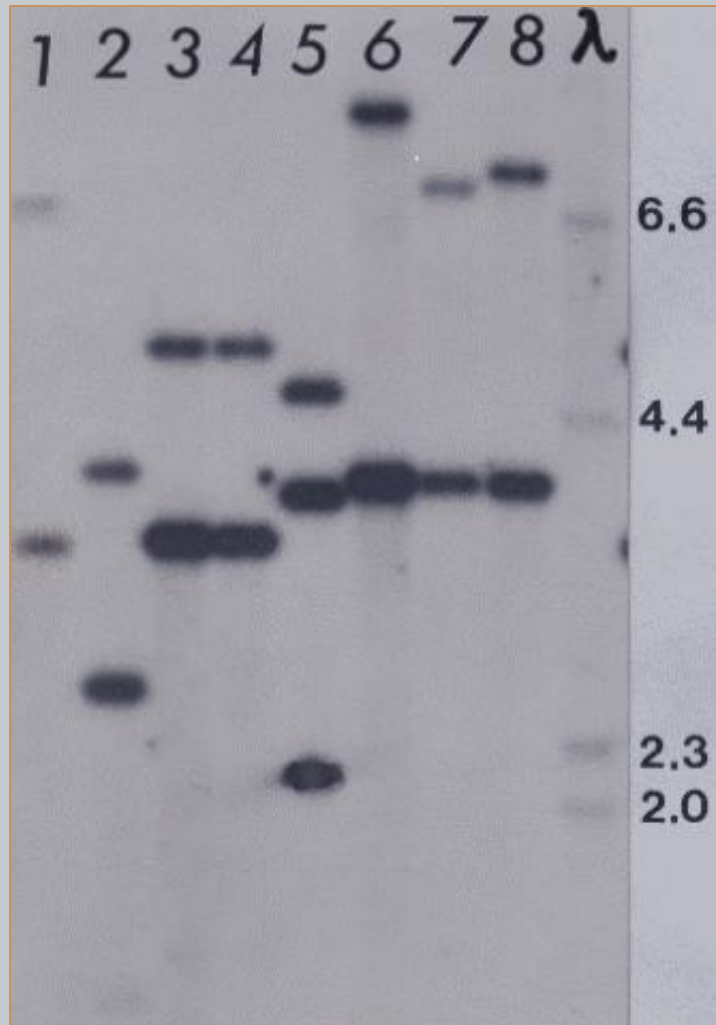
pros

- highly reproducible pattern
- variability in particular part of DNA (e.g., cpDNA, rDNA ...)

cons

- large amount of DNA necessary
- need for an amount of labelled probe
- expensive and complicated
 - blotting equipment
 - work with labelled materials (detection etc.)

RFLP gels examples



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

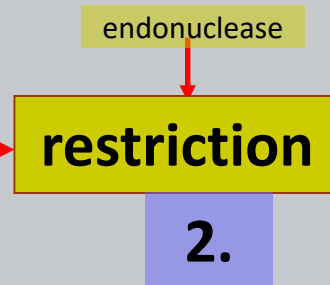
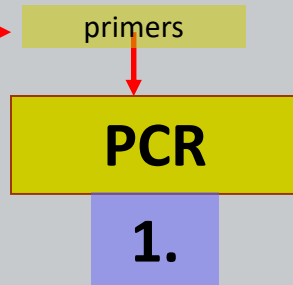
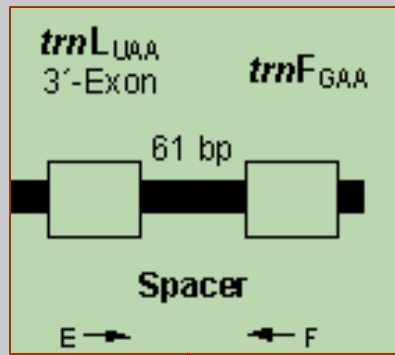
PCR-RFLP

- or – CAPS (**C**leaved **A**mplified **P**olymorphic **S**equence)
- principle of the method
 - amplification of target DNA using a pair of primers
 - restriction of PCR product with endonuclease
 - electrophoresis
 - fragment visualization using, e.g., ethidium bromide

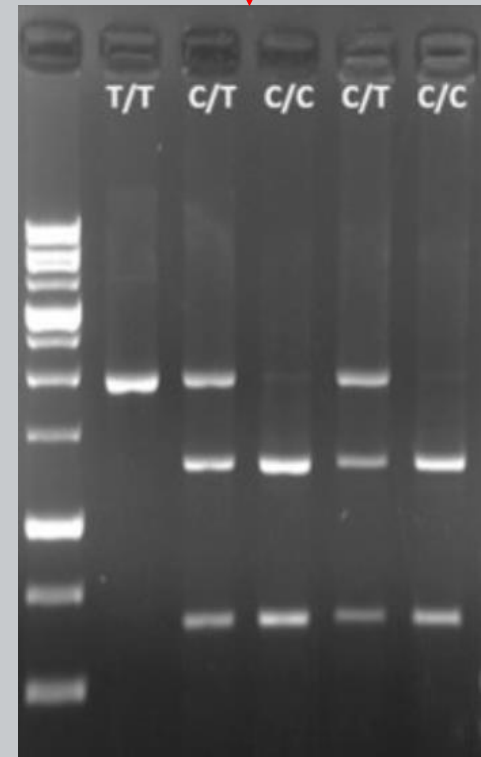
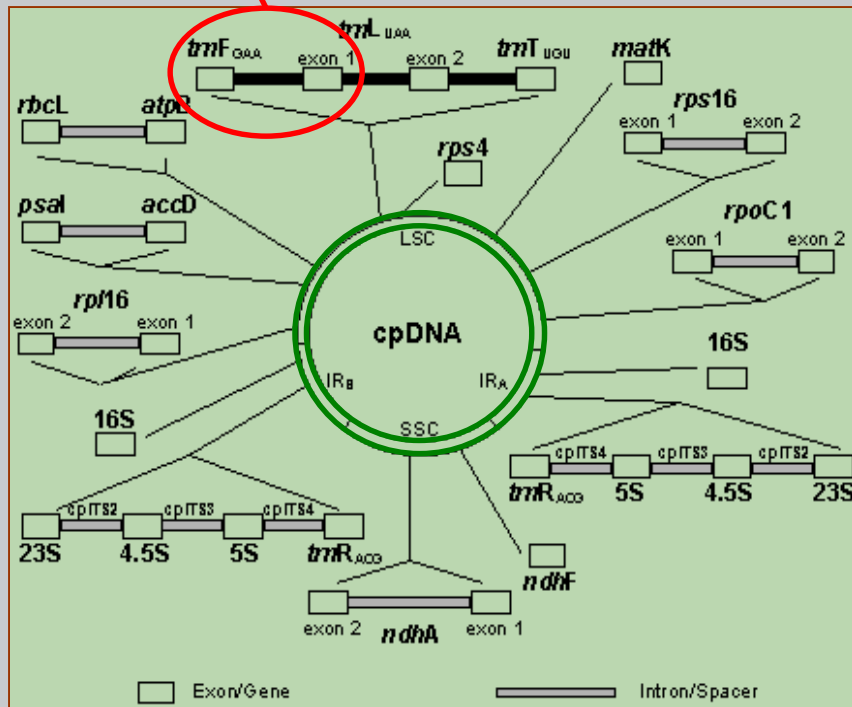
PCR-RFLP

- advantages
 - minimum amount of DNA necessary
 - no need for blotting and for radioactivity-labelled material – simple method
- common use – cpDNA and rDNA (i.e., ITS)
 - PCR amplification of non-coding regions
 - use of universal primers

PCR-RFLP cpDNA



electrophoresis

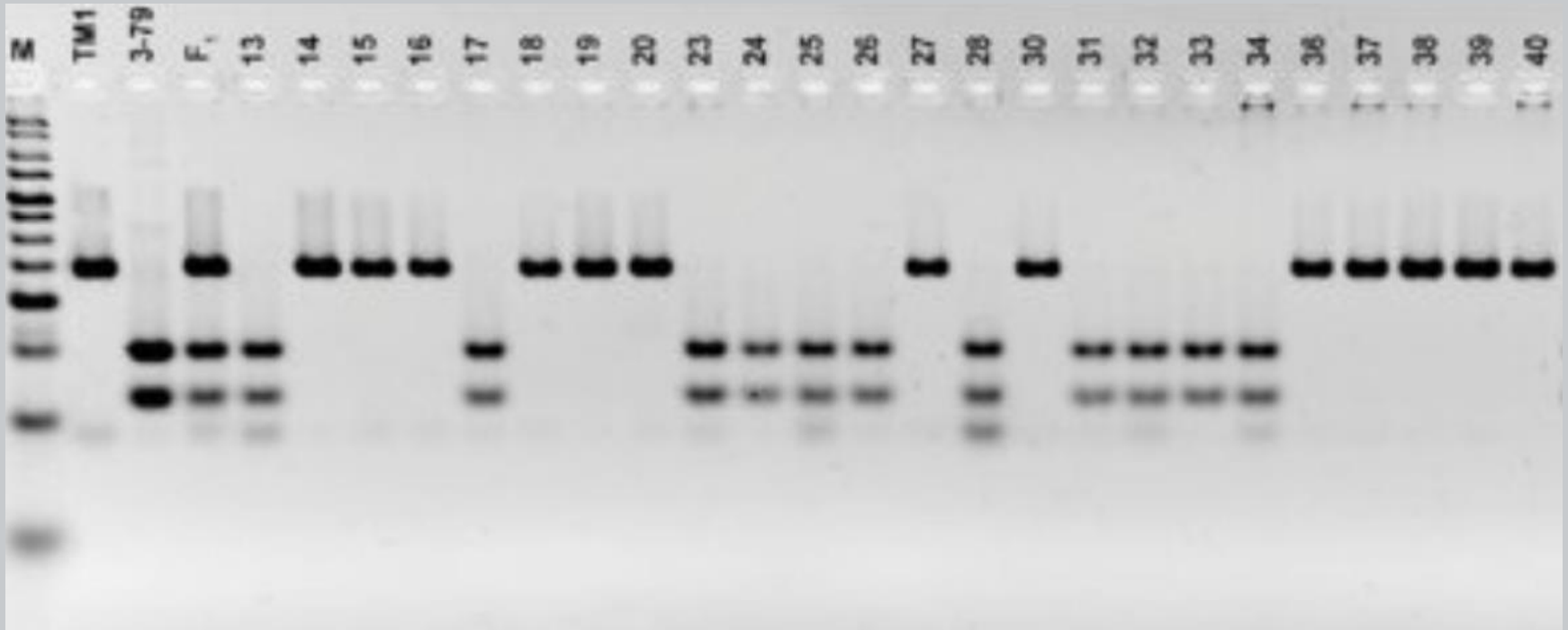


dCAPS — derived Cleaved Amplified Polymorphic Sequence

- Neff et al. (1998), *Plant Journal* 14: 387–392
- when the SNP of interest does not alter the restriction site of an available RE
- mismatches in the dCAPS primer create a restriction site
- dCAPS Finder 2.0 – <http://helix.wustl.edu/dcaps/>

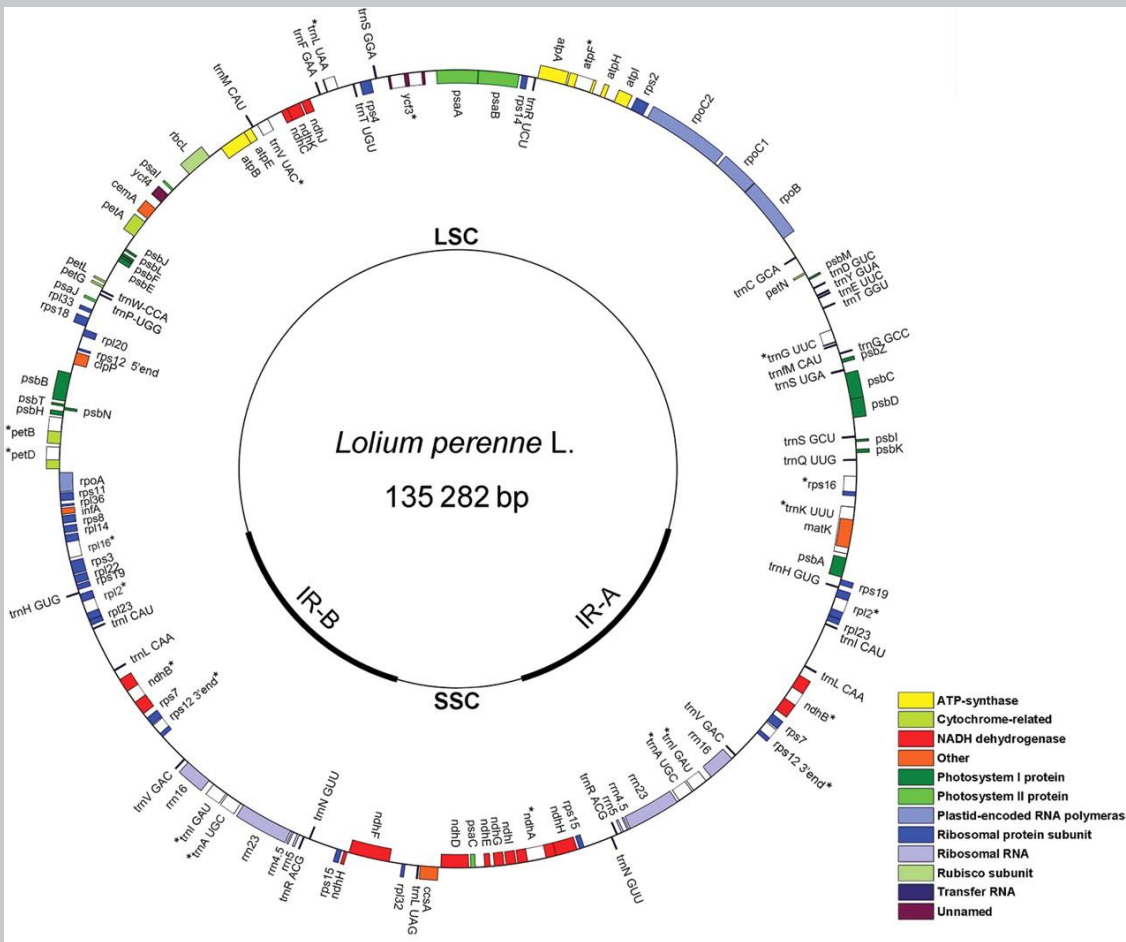
```
type1          GTGGAAGAAGCTCGATGAGGCTTTGGGG
type2          GTGGAAGAAGCTCGATGAGGCTTTGAGG
BsII           CCNNNNN|NNGG
dCAP primer    GTGGAAGAAGCTCGACCAGGCTTTG
BsII digestion
type1          GTGGAAGAAGCTCGACCAGGCT / TTGGGG
type2          GTGGAAGAAGCTCGACCAGGCTTTGAGG
```

dCAPS gel example



Kushanov et al., 2016

Chloroplast genome (cpDNA)



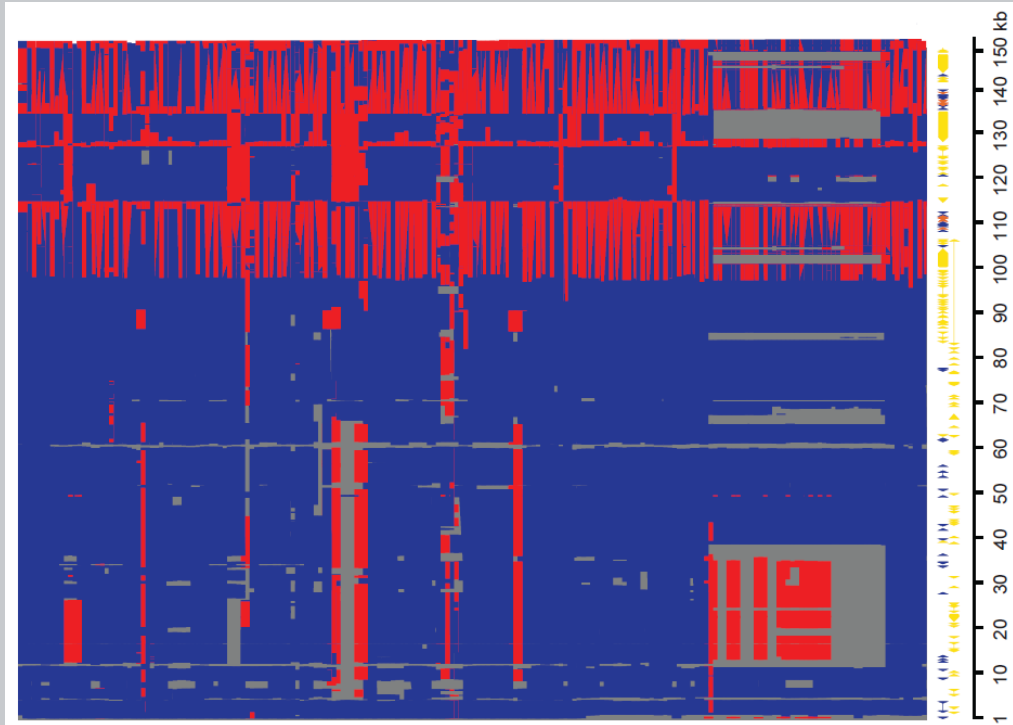
- circular molecule
- 45-220 kb
- 30-100 in plastide
- homoplasmy
- **2 subunits** – LSC, SSC
- IR – inverted repeats (10-30 kb)
- genes for
 - proteins (70-88)
 - photosystem – *ps*
 - ribosomal protein – *rp*
 - RUBISCO – *rbc*
 - tRNA – *trn* (30-35)
 - rRNA (2x4)
 - ...

Specificity of cpDNA

- relatively *conserved* – low mutation rate
 - extraordinary degree of functional and structural conservation
 - constantly high selective pressure on genes for the photometabolic pathways
 - genes in IR evolve about three times more slowly than those in the SC regions
- but – existence of intraspecific *variation*
- mutation types
 - point
 - insertions/deletions – frequent in non-coding regions
- *non-recombinant* unit
- *haploid* – haplotypes
- uniparental inheritance
 - *maternal* transfer in angiosperms – via seeds
 - *paternal* transfer in gymnosperms – via pollen

Plastome structure

- conservative among land plants



Wicke & Schneeweiss 2014

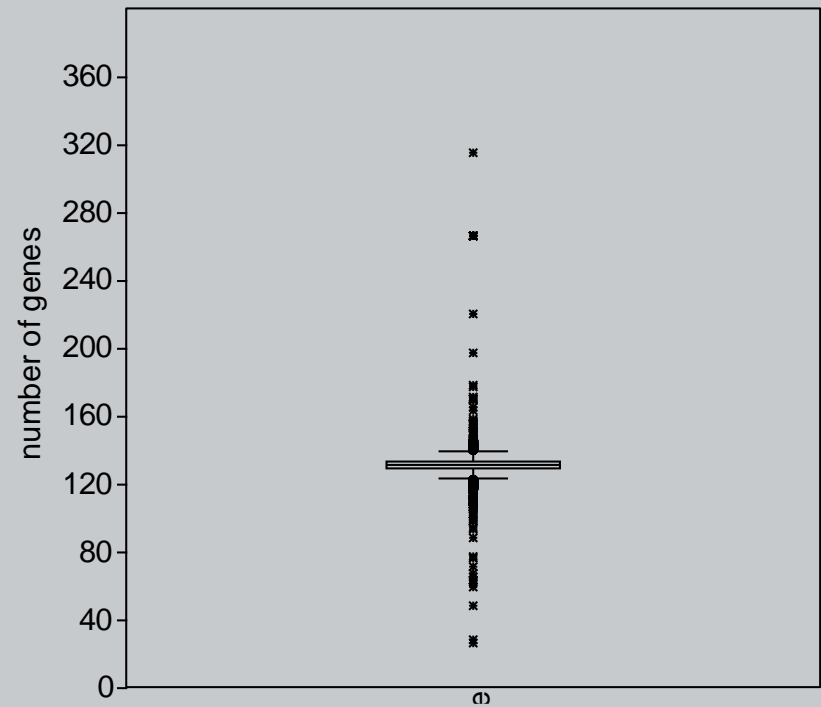
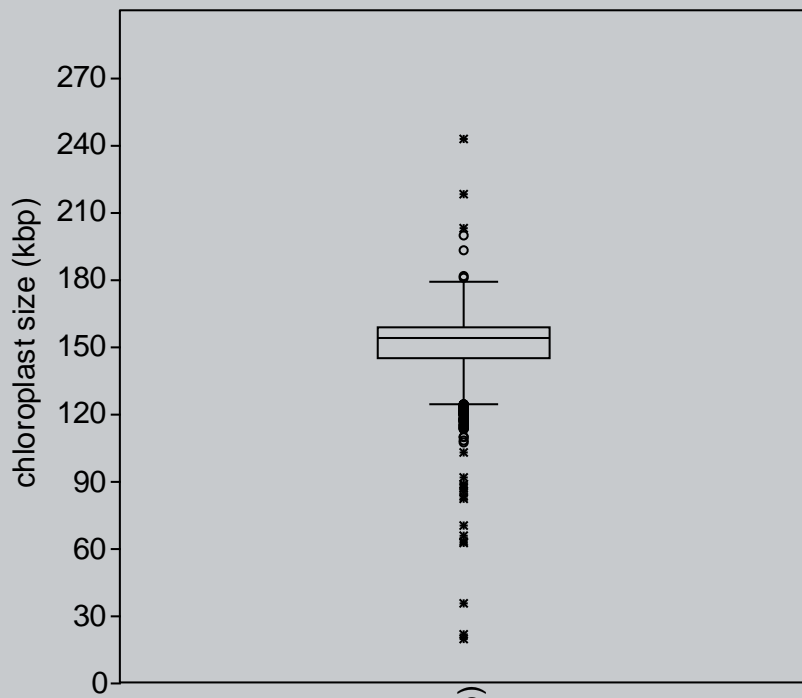
- reconfigured in non-photosynthetic heterotrophic plants
 - loss of many genes
 - structural reconfigurations (inversions, IR reduction or loss)
 - overall relaxed selection on plastome architecture

Plastome structure

- gene gain is extremely rare
 - three gained before the transition to land (*matK*, *ycf1*, *ycf2*)
 - more than a dozen were lost/transferred to the nuclear genome
- inserts to nuclear DNA
 - NUPTs – nuclear plastid DNAs
 - NUMTs – nuclear mitochondrial DNA
- inserts to mitochondrial genome
 - MIPTs – mitochondrial inserts of plastid DNAs
- NUPTs, NUMTs and MIPTs are paralogs of the genuine organelle DNA copies (promiscuous DNA)
 - large inserts in pericentromeric regions
 - later fragmented by TE insertions and reshuffled away from centromere
 - distribution is species-specific

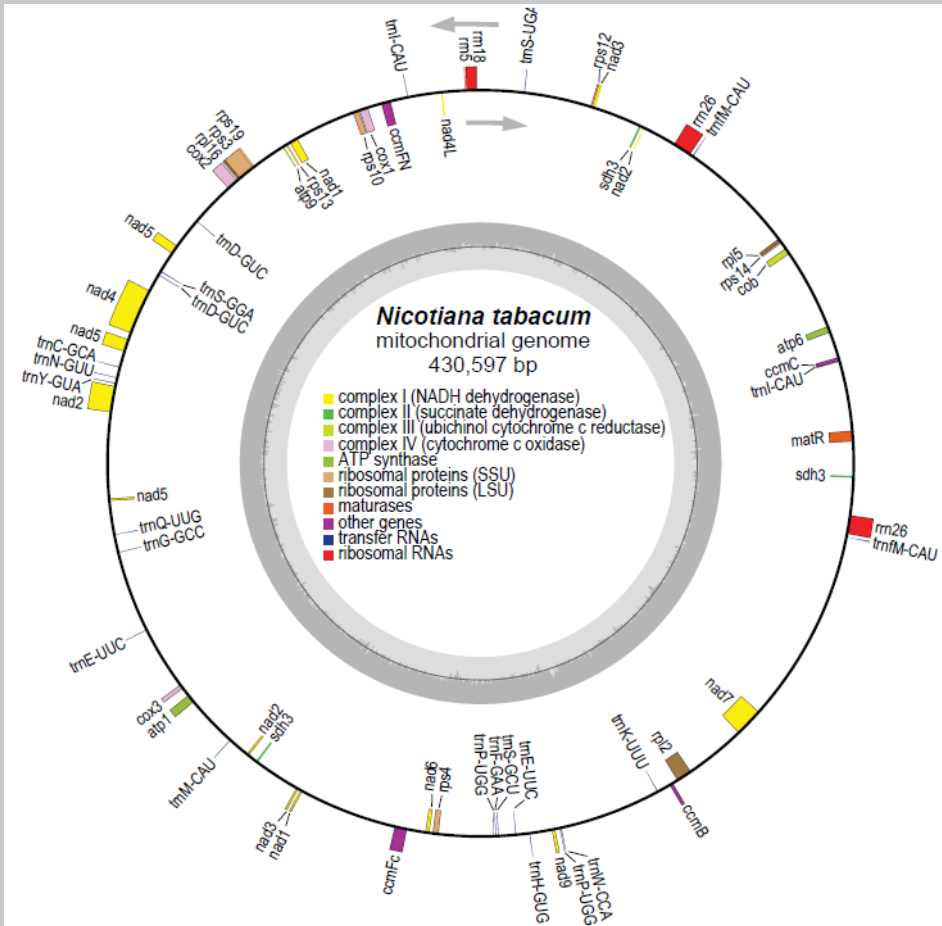
Chloroplast size, nr. of genes

- cca 1,700 sequenced plastomes (land plants)
- size 150 kbp (19 – 243)
- 131 (26 – 315) genes: 84 proteins, 8 rRNA, 37 tRNA



<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/>

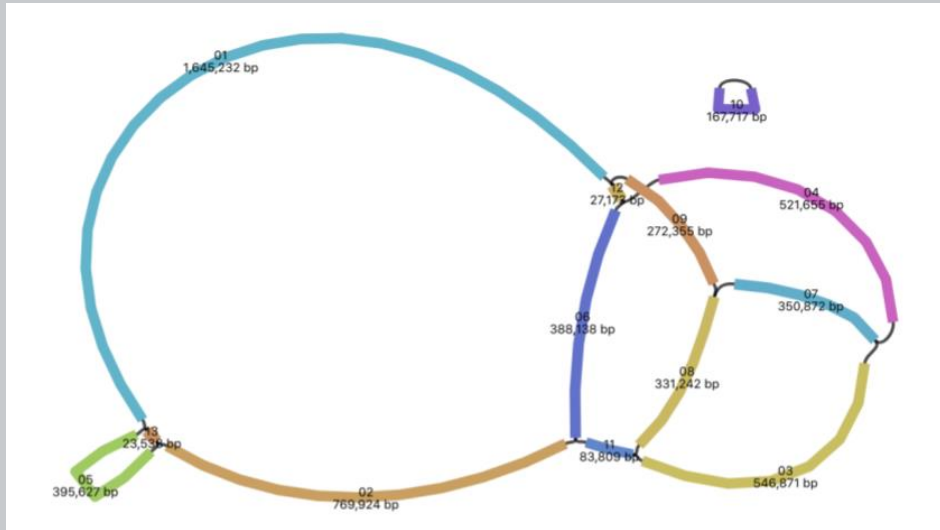
Mitochondrial DNA – chondrome



- often inflated in size
- less than 20% coding regions (19-64 genes), max. 40 kb (excl. introns)
 - respiratory chain complexes (nad, sdh, cob, cox, atp)
 - cytochrome c maturation (ccm)
 - ribosomal proteins for the small and large subunit (rpl, rps)
 - translocase (mttB/tatC) ...
 - 4 ribosomal genes
- substitution rate is low
- foreign DNA (other cellular genomes or other organisms)
- often inflated in size
 - from 101 kb (non-vascular plants)
 - to 11.3 Mbp in *Silene conica*

- rarely used for phylogenetic and phylogenomic purposes
- lower resolution, lower support and/or incongruent with plastome-derived phylogenies

Mitochondrial DNA – complicated structure



- several segments (loops)
- hard to assemble

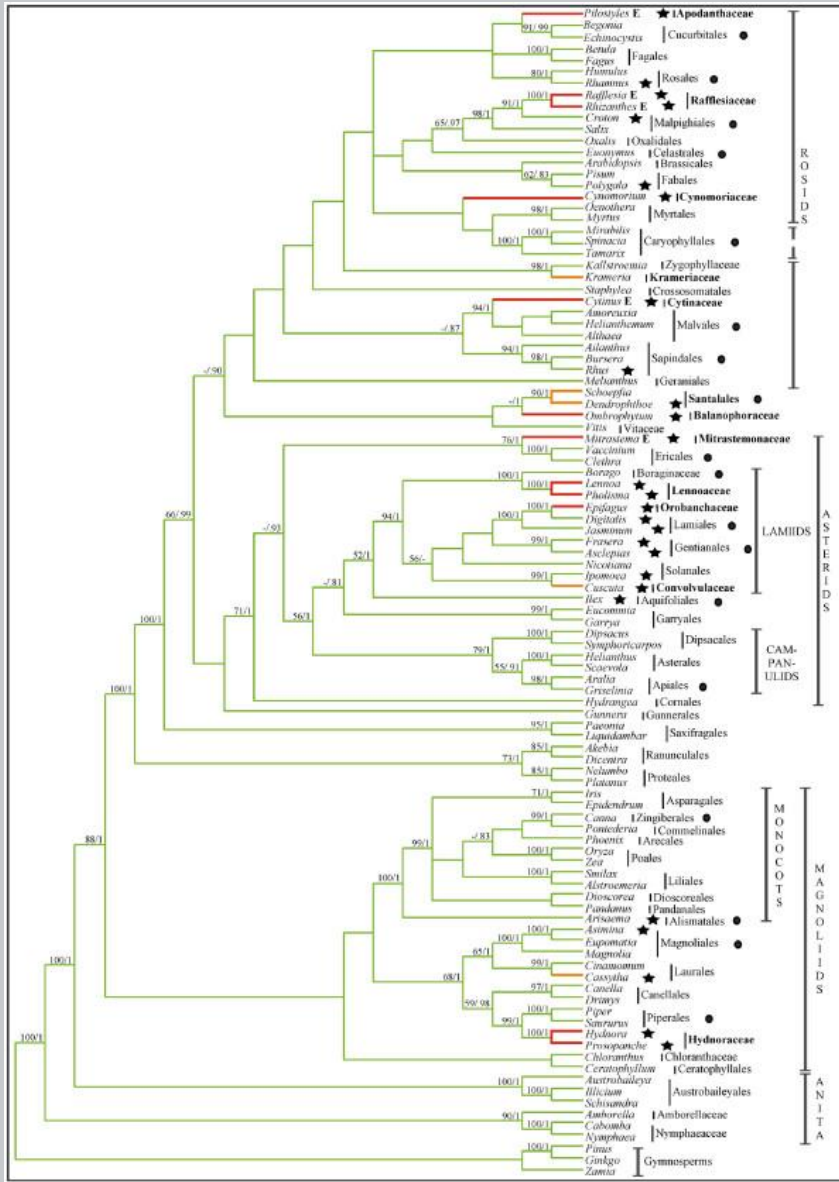
Picea sitchensis

5.52 Mbp, 13 segments

Jackman et al. 2019



Mitochondrial DNA – phylogeny



3 mtDNA genes

- *atp1*
- *cox1*
- *matR*
- angiosperms
- 11 independent origins of parasitism

Applications

- phylogeography
 - geography of gene lineages (haplotypes)
 - reconstruction of postglacial recolonization
- study of gene flow by seeds
 - what is the influence of cpDNA to the total genetic differentiation of populations
- systematics
 - phylogeny reconstruction
- study of hybridization
 - identification of maternal taxon (individual)

Phylogeography

influence of historical factors (typically glaciation) to the na geographical distribution of gene lineages

- maximum glaciation – 20,000-18,000 BP
- maximum (concentrated) variability in the Mediterranean
- 3 major refugia – Iberian Peninsula, the Appeninnes, the Balkans
- only small part of the variability migrated back to the Central and Northern Europe
- recolonization started circa 13,000 BP
- we can trace individual lineages (*cpDNA haplotypes*) and correlate them with their geographical distribution

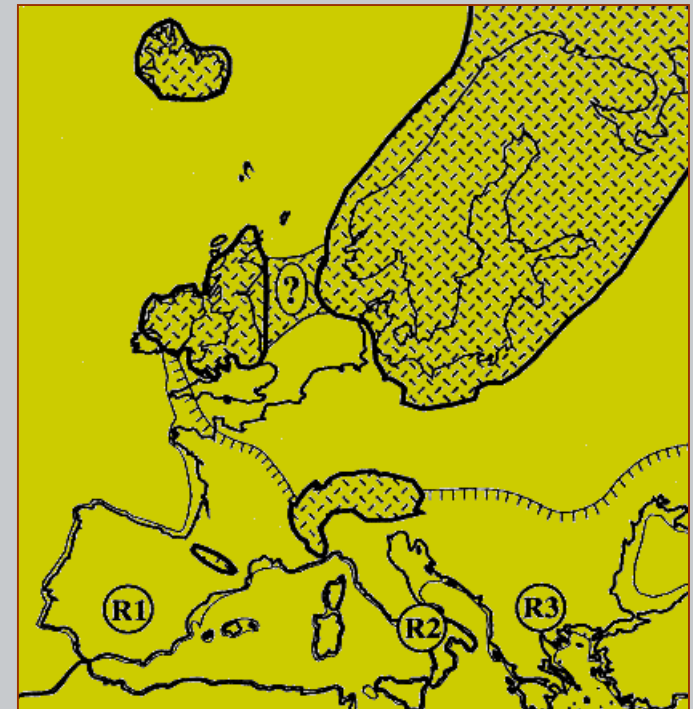


maximum extent of glaciation during last ice age



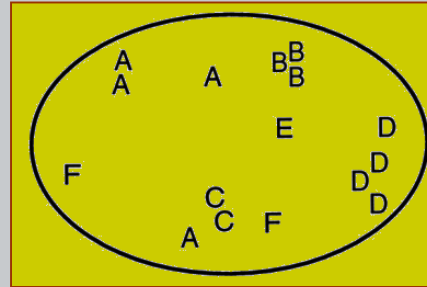
permafrost

R1, R2, R3 – major refugia

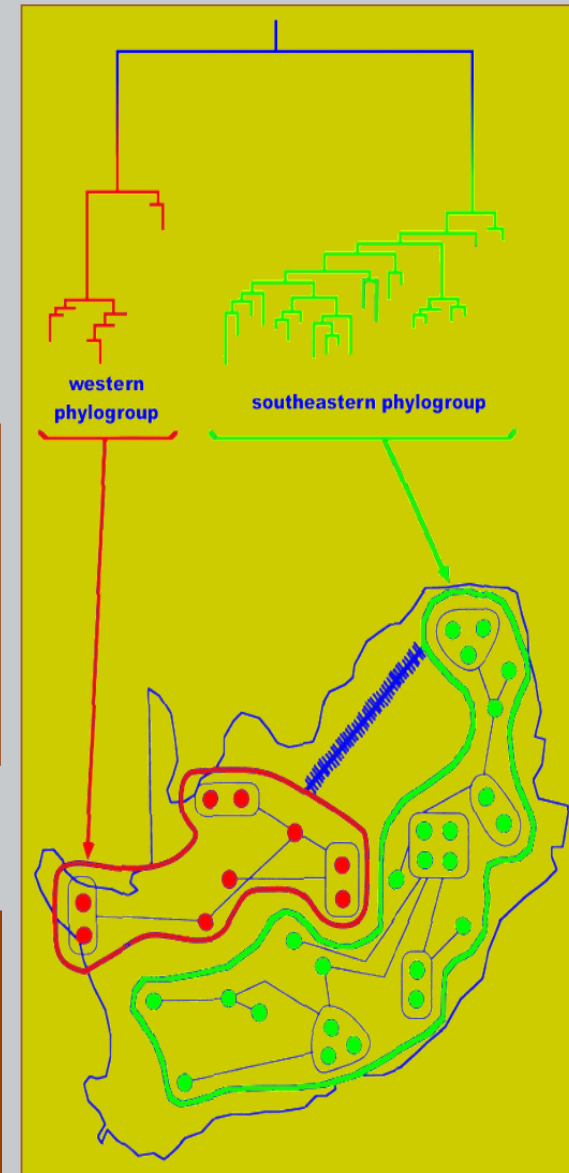
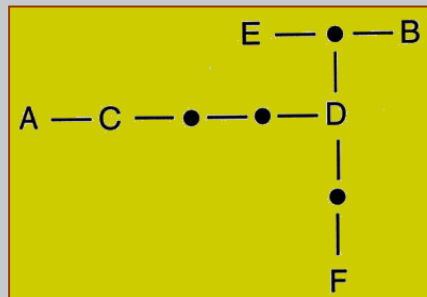
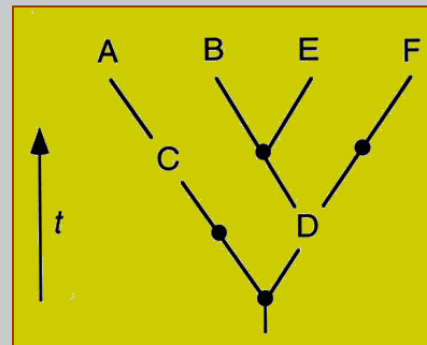


cpDNA haplotypes at the population level

- geographical distribution of alleles
- new alleles originate due to mutations
- all existing alleles are derived from a single ancestor allele
→ existed sometime in the past

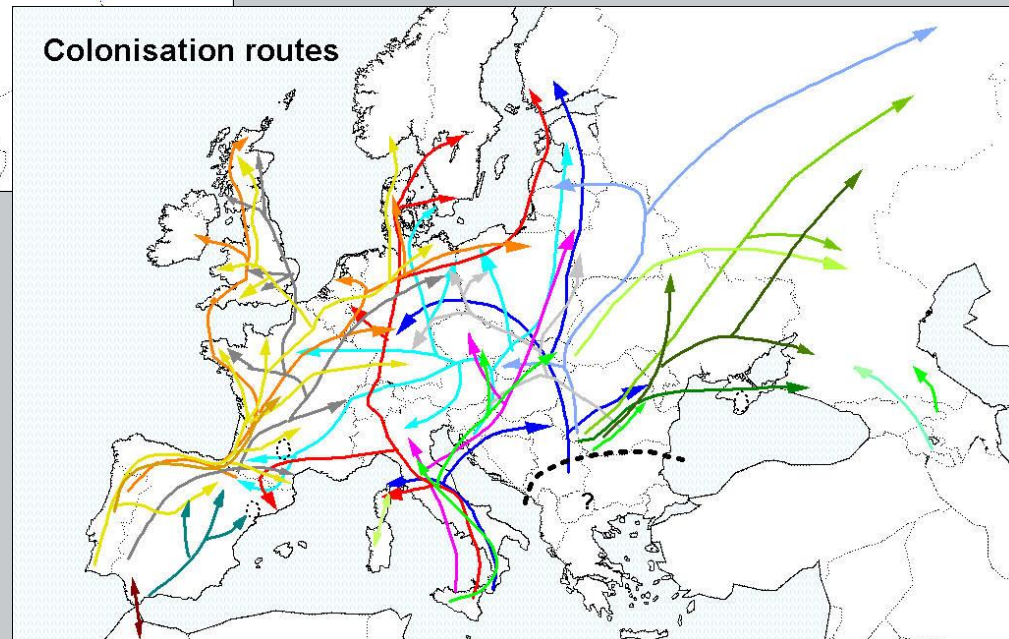
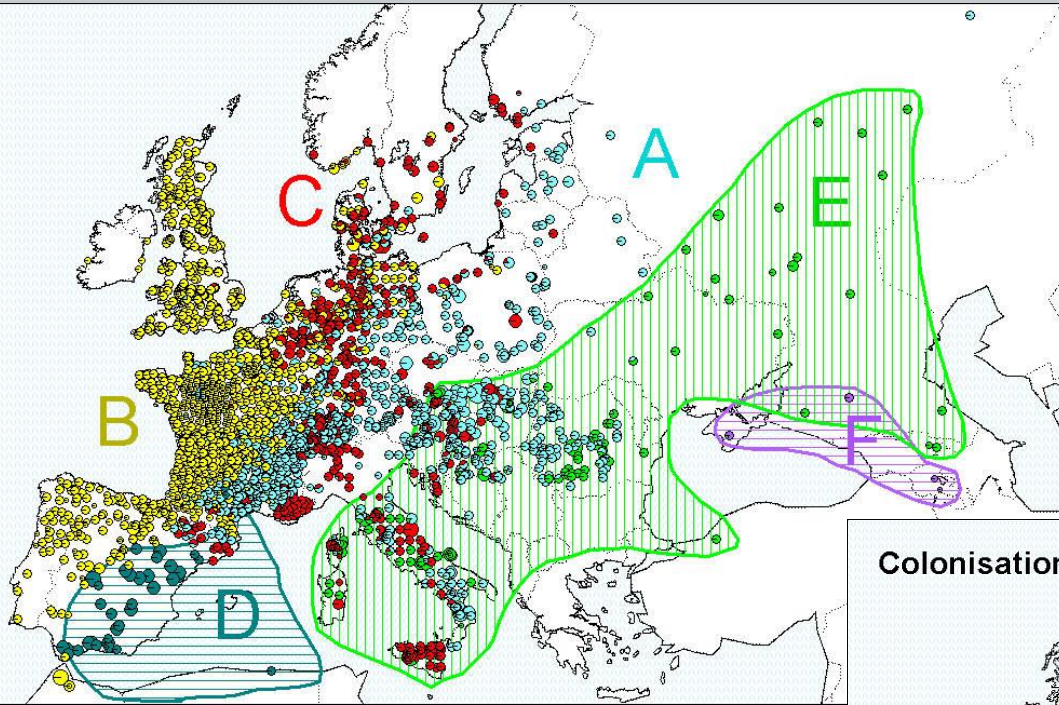


- problems with polarity – use of *minimum spanning tree*
 - minimalizes number of mutations



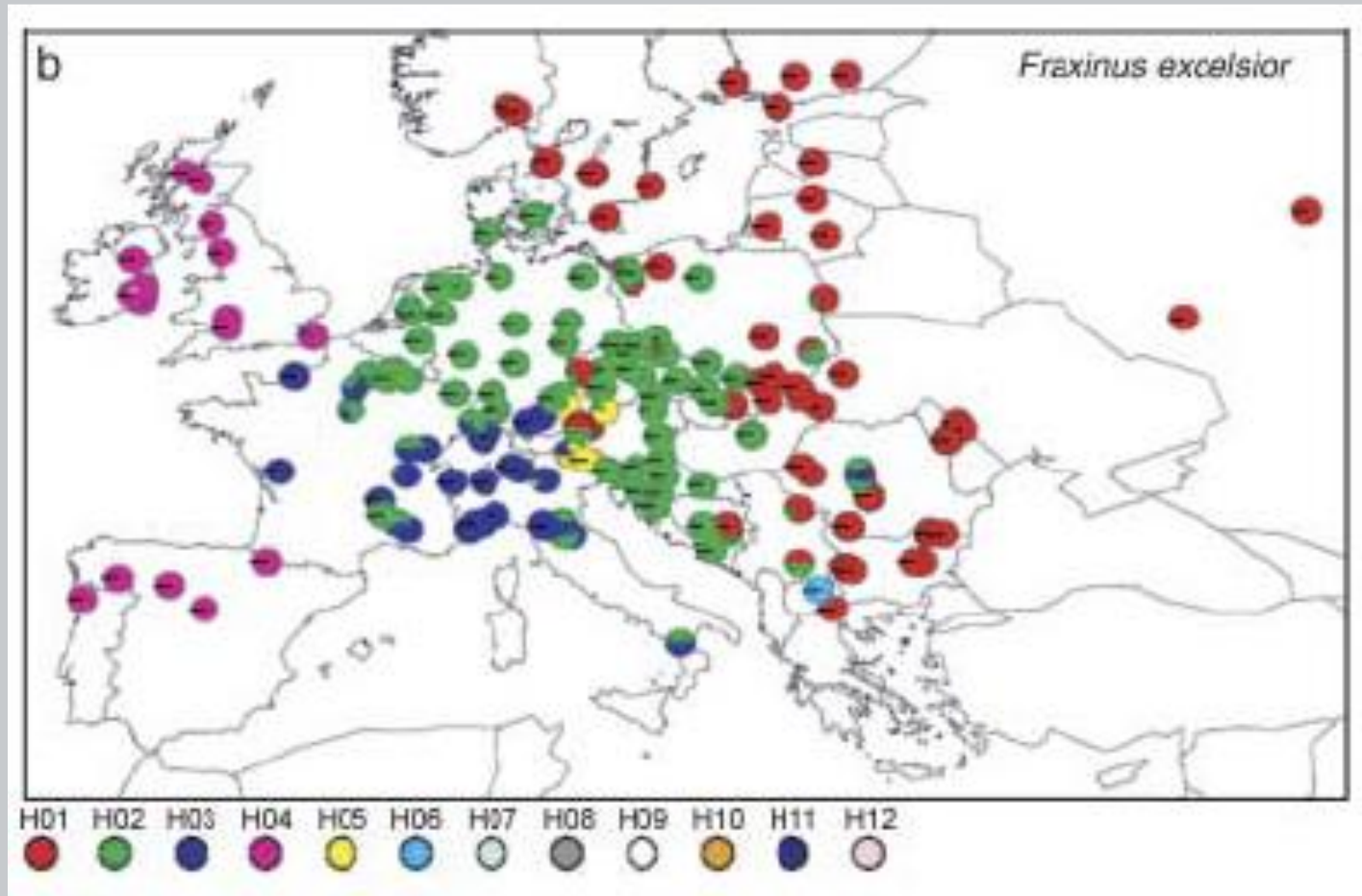
Postglacial recolonisation of Europe

Quercus sp.



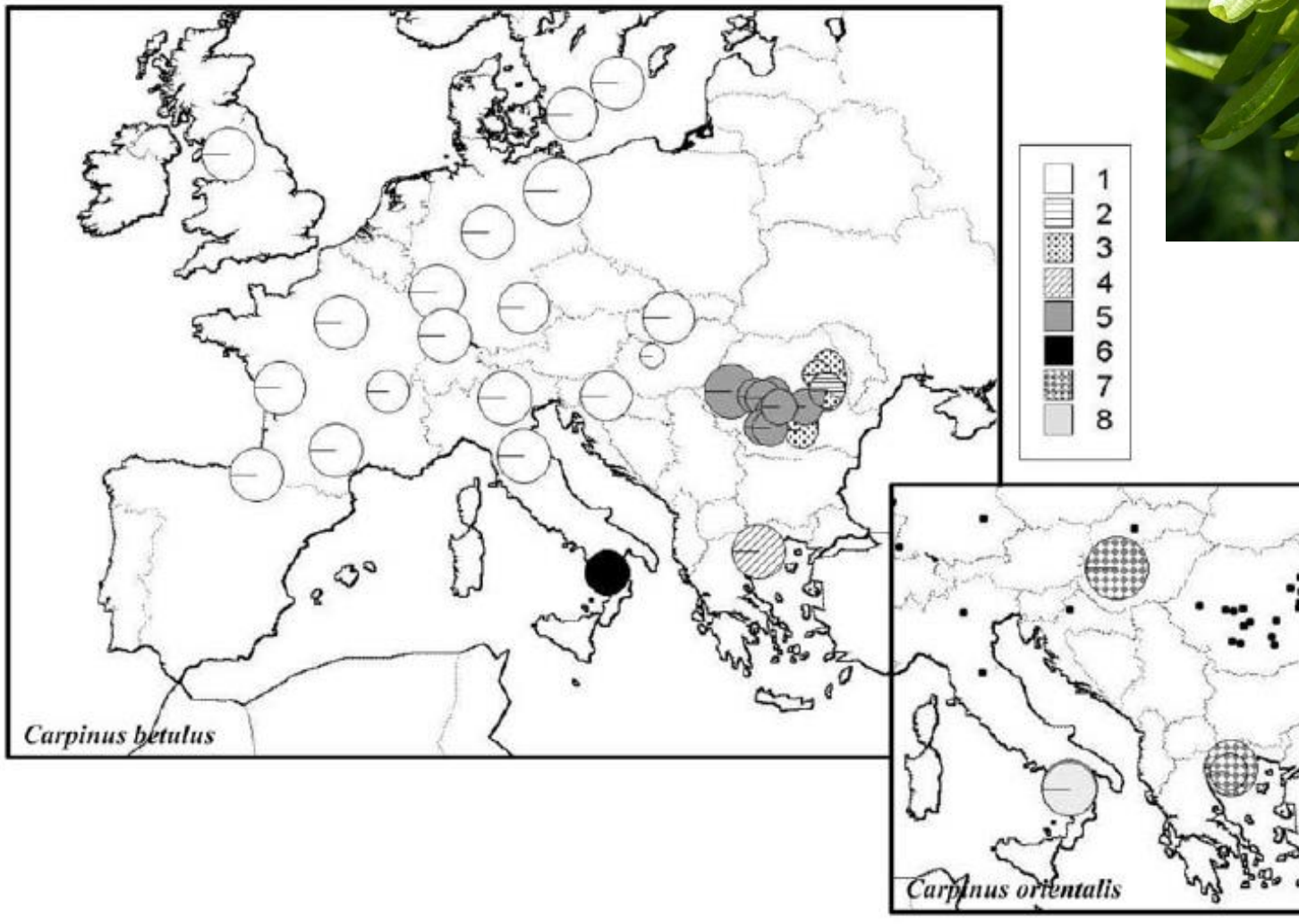
Fraxinus excelsior

Heuertz et al. 2006

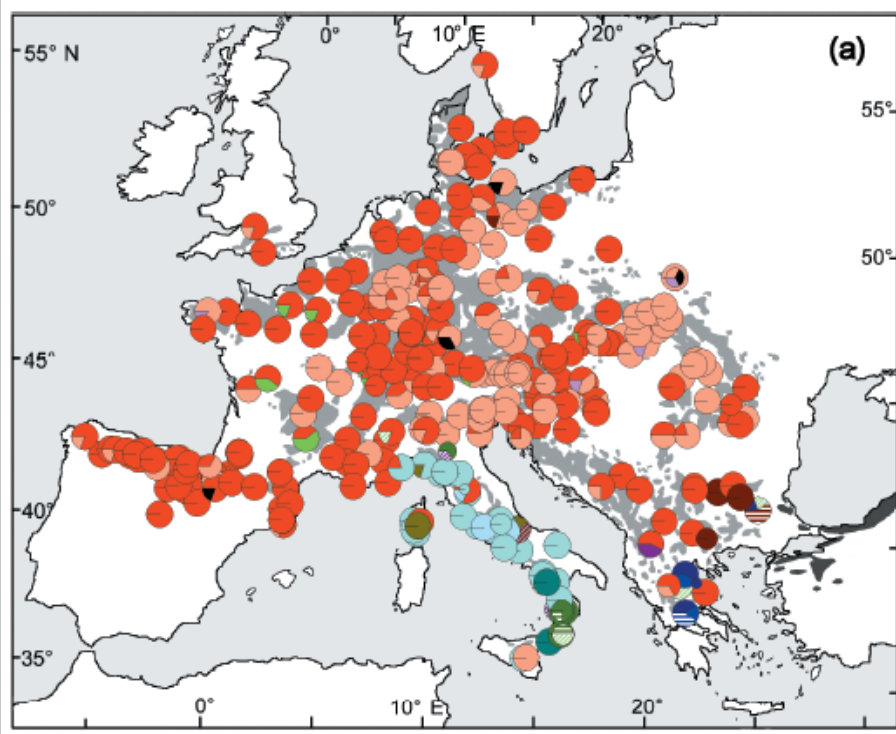


Carpinus betulus

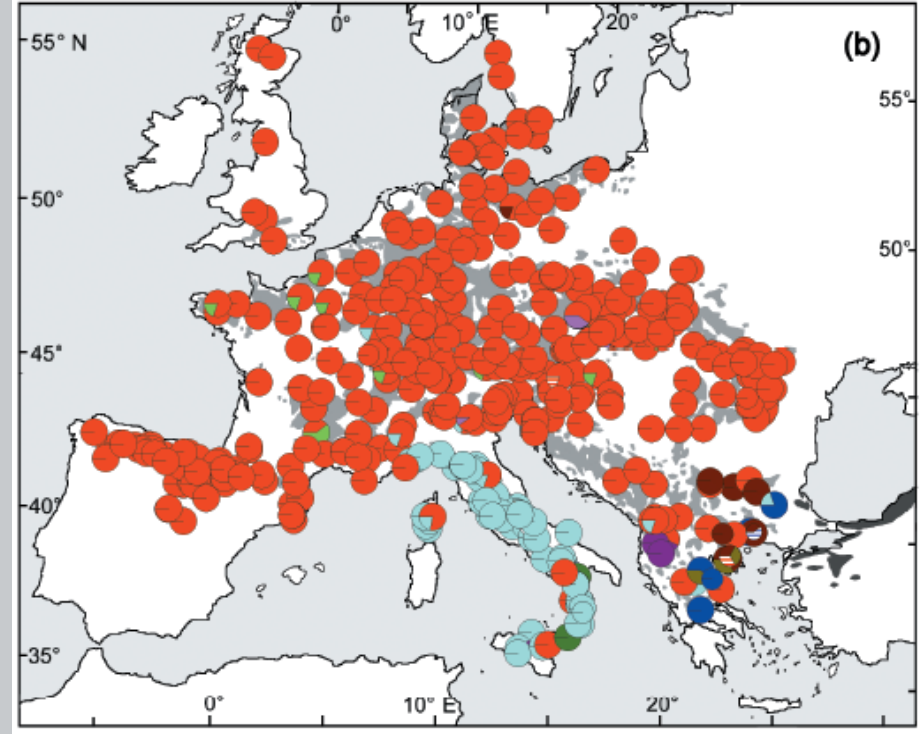
Grivet & Petit 2003



Reconstruction of beech migration



cpDNA microsatellites

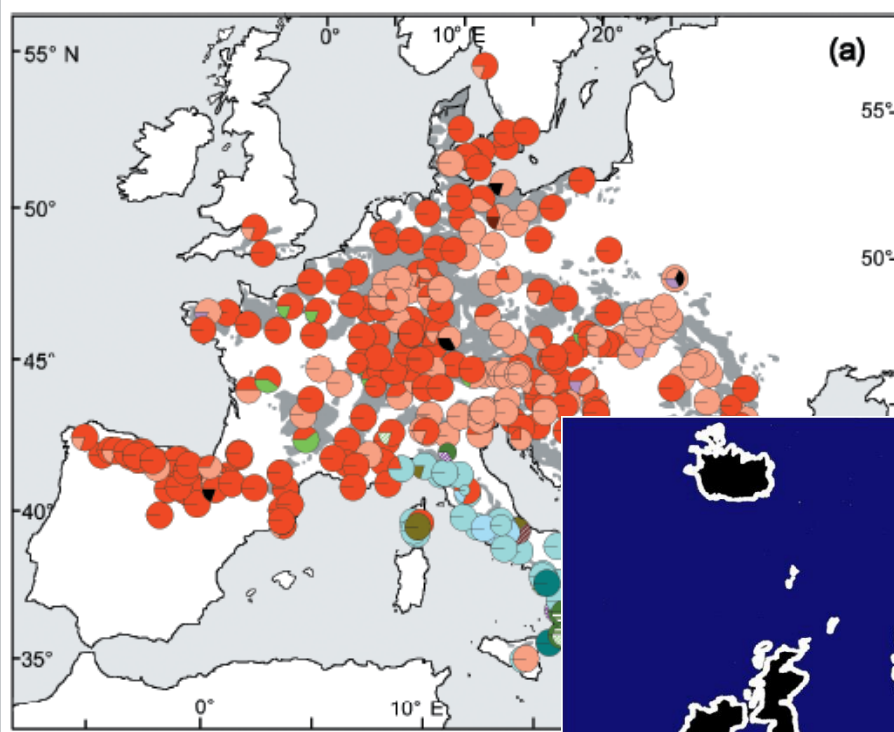


PCR-RFLP cpDNA

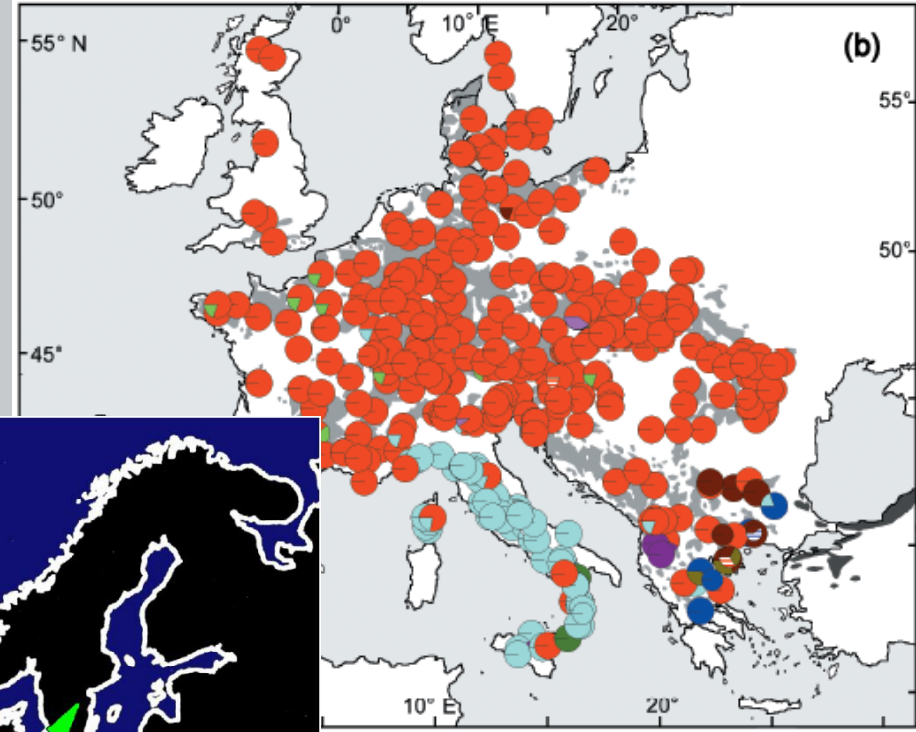


Magri et al. 2006

Reconstruction of beech migration



cpDNA microsat

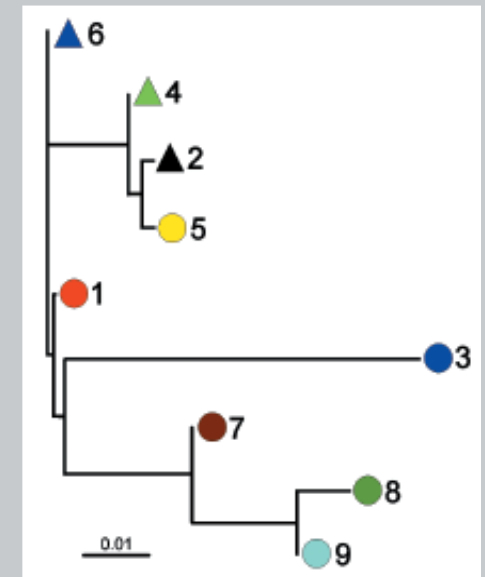
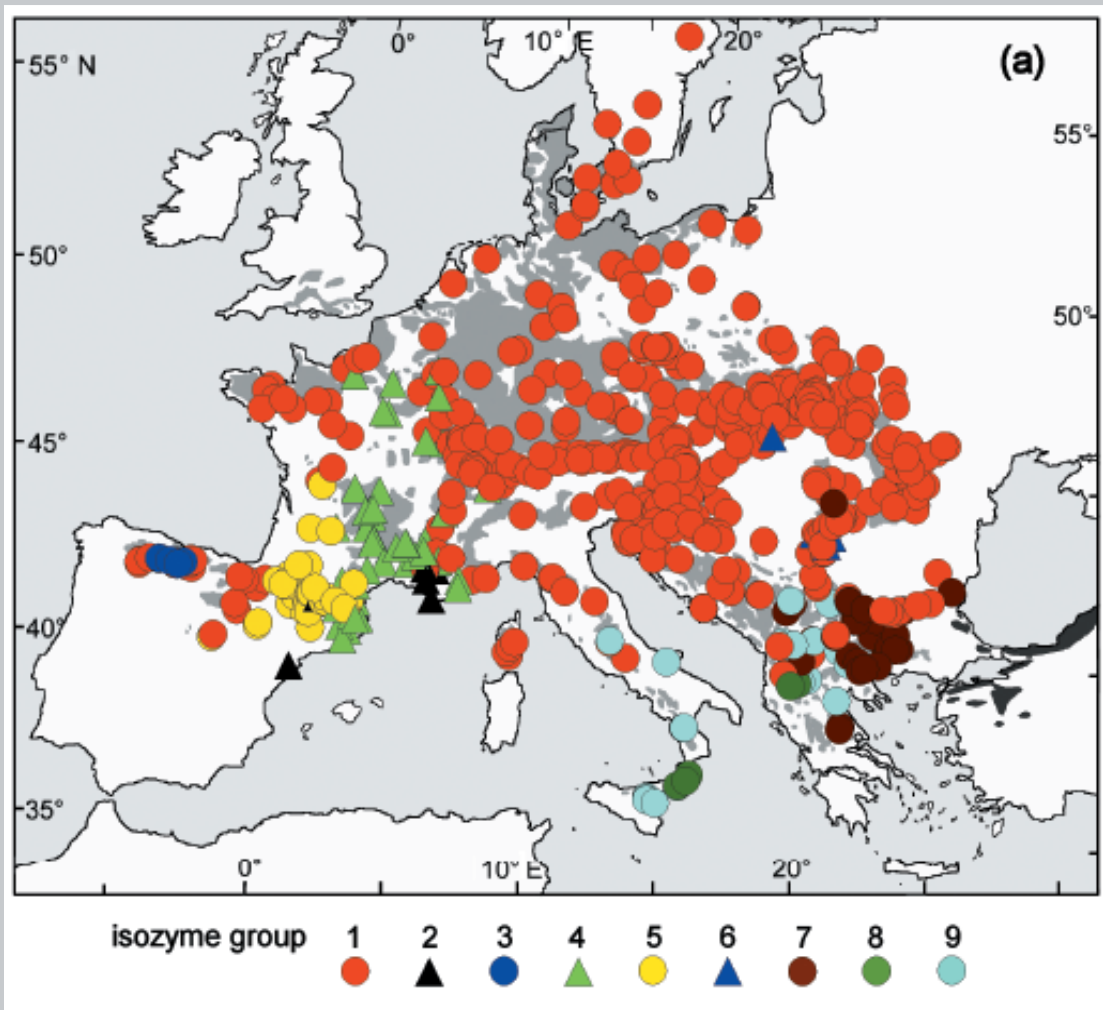


RFLP cpDNA



Magri et al. 2006

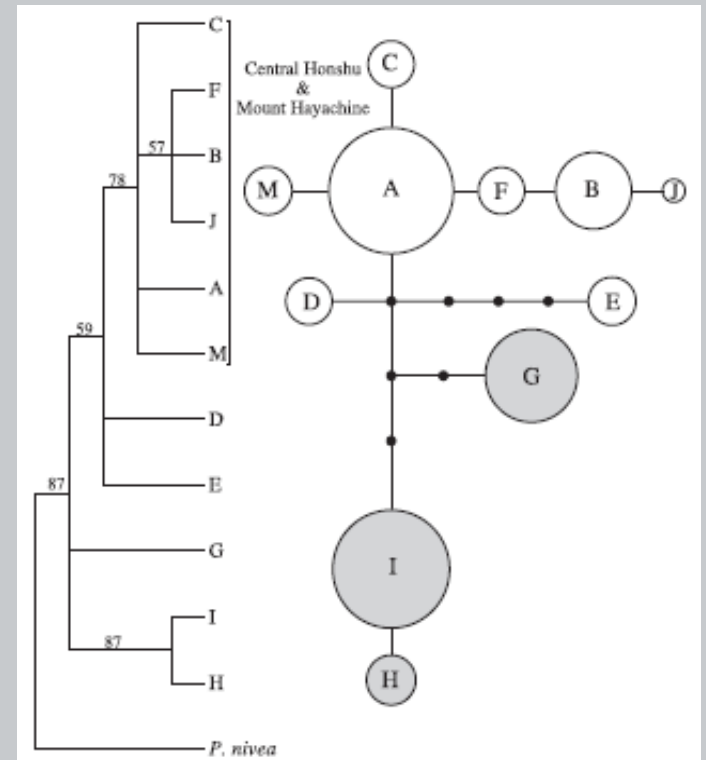
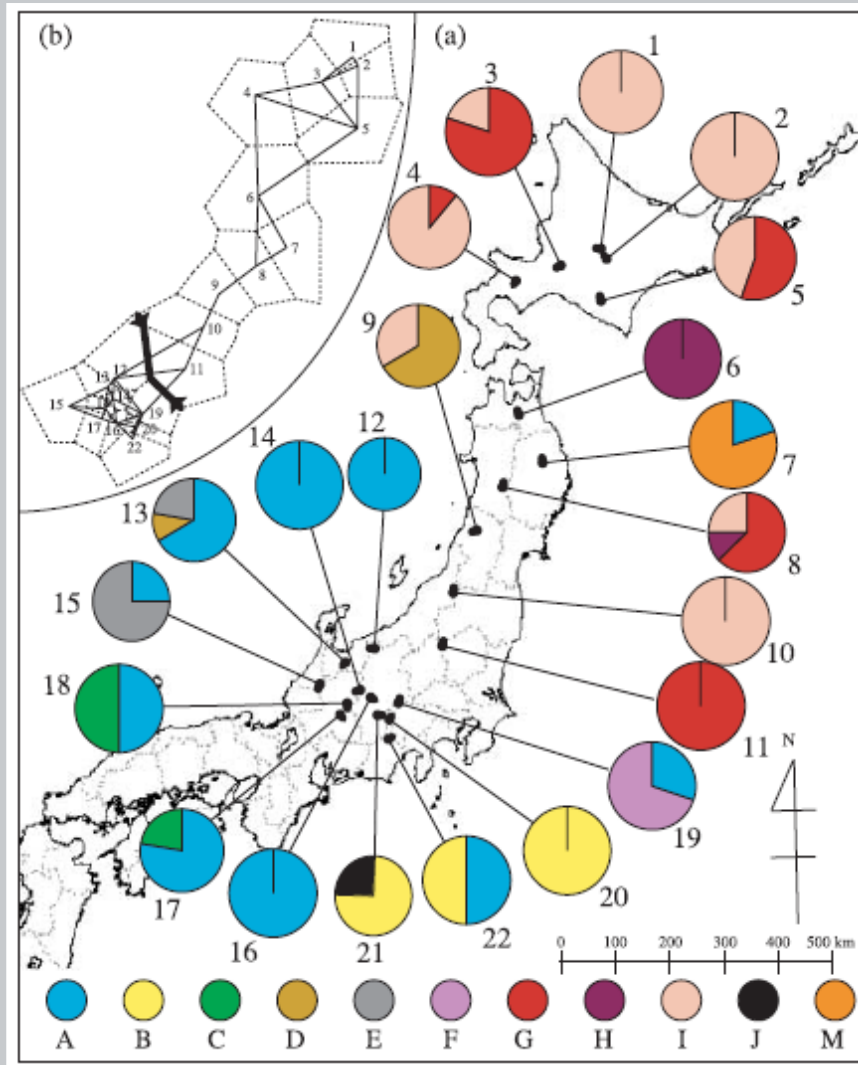
Fagus sylvatica – isozyme analysis



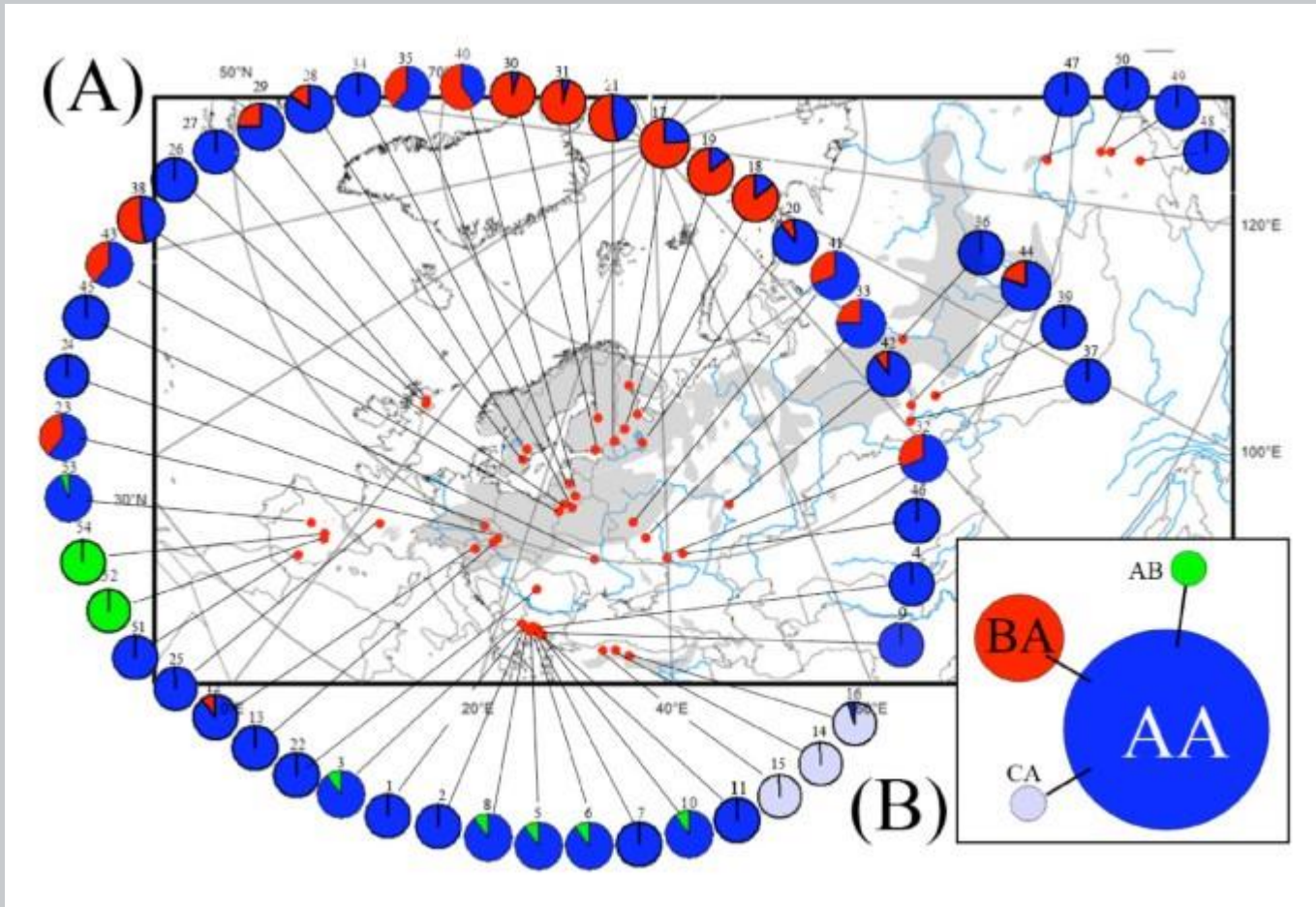
Magri et al. 2006

Phylogeography of alpine plants in Japan

Potentilla matsumurae, Ikeda et al. 2006



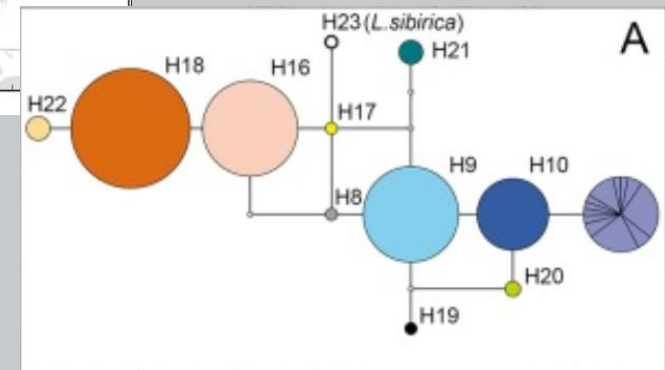
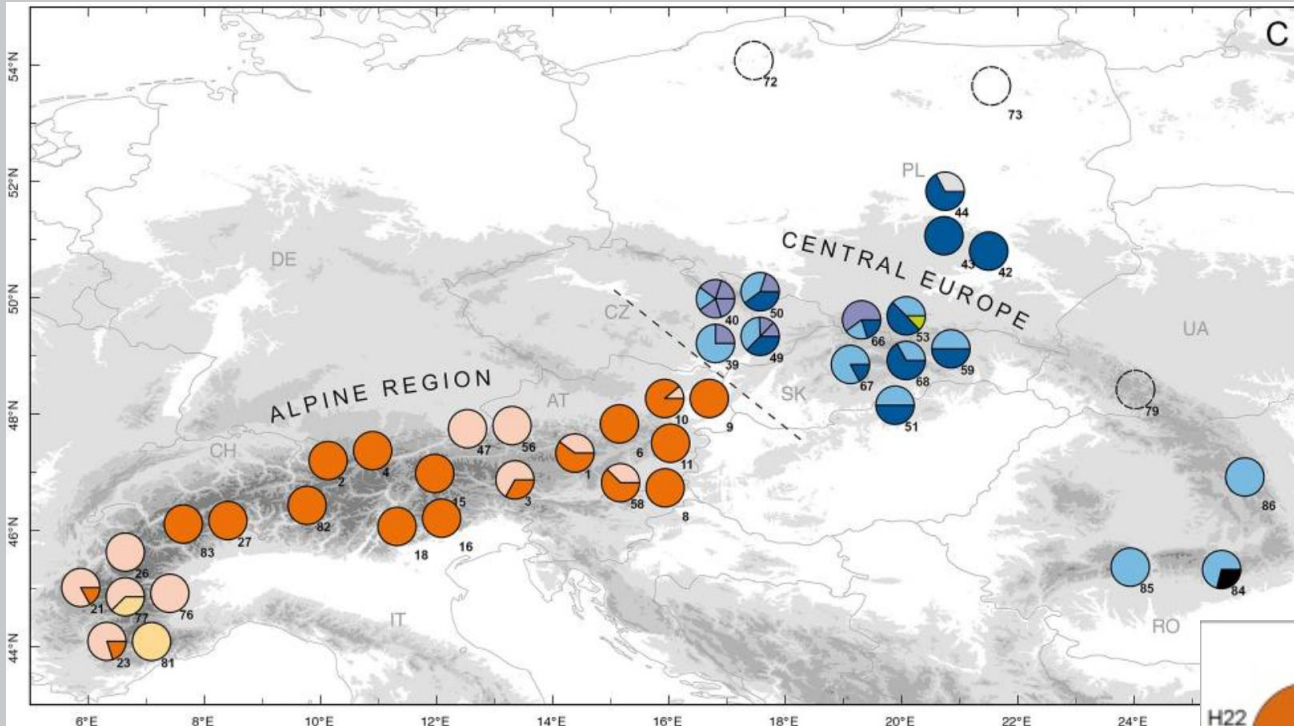
Phylogeography based on mtDNA Gymnosperms



Pinus sylvestris – intron 1 of *nad7*, 4 haplotypes
Naydenov et al. 2007

Phylogeography based on mtDNA

Larix decidua

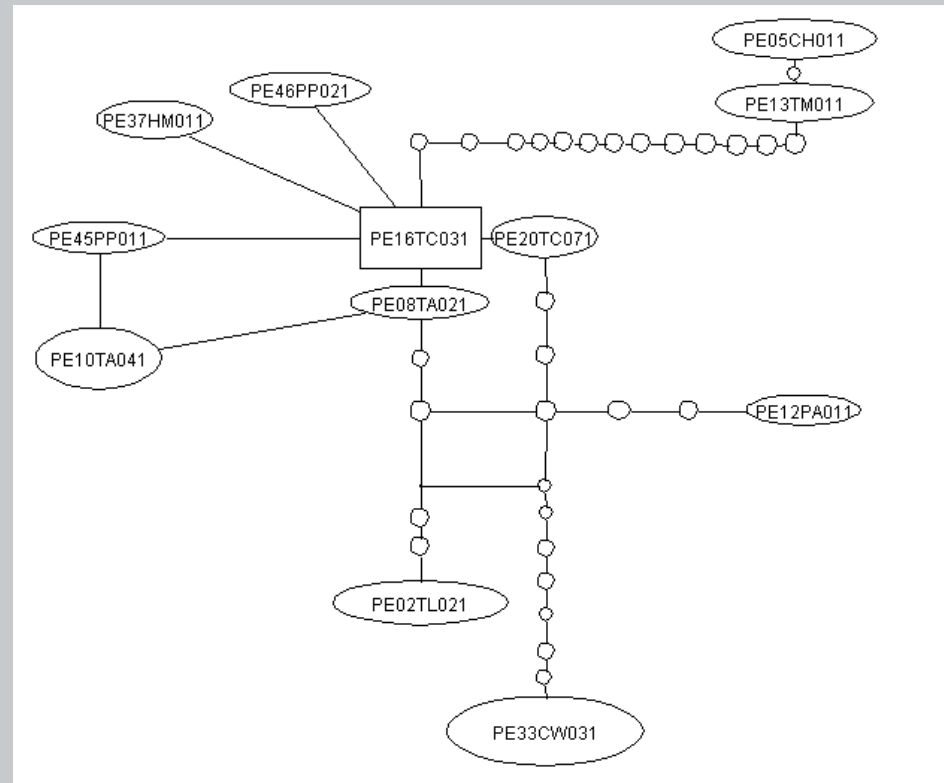


2 regions (UBC460 and *atpA*) – 2 haplotype groups

Wagner et al. 2005

Analysis of phylogeographical data – *statistical parsimony network of haplotypes*

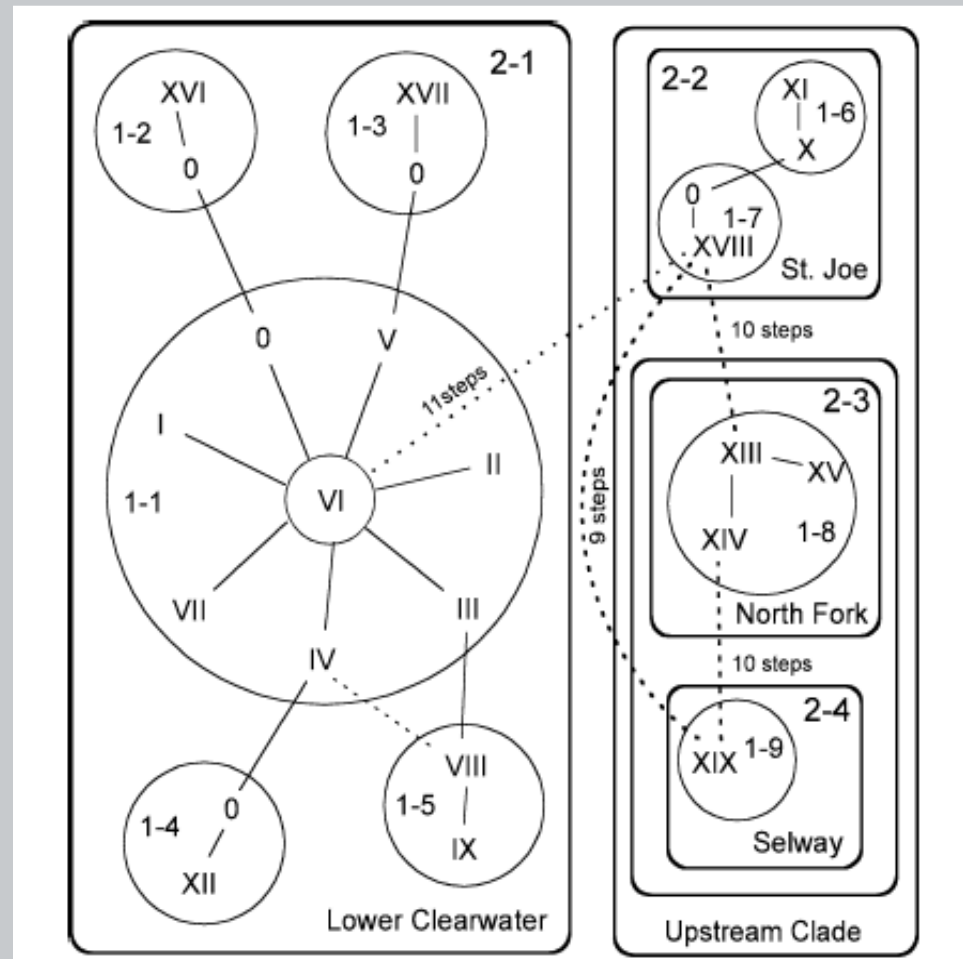
- *software – TCS 1.21*



- Posada D & Crandall KA (2001): *Intraspecific gene genealogies: trees grafting into networks*. Trends Ecol. Evol. 16(1): 37-45.
- Clement M, Posada D & Crandall KA (2000): *TCS: a computer program to estimate gene genealogies*. Molecular Ecology 9 (10): 1657-1660.
- Templeton, A. R., K. A. Crandall & C. F. Sing 1992. *A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation*. Genetics 132: 619-633.

Analysis of phylogeographical data – *nested clade analysis*









- software – GeoDis
- Templeton AR (1998): *Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history*. *Molecular Ecology* 7: 381-397.
- Posada D, Crandall KA and Templeton AR (2000): *GeoDis: A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes*. *Molecular Ecology* 9(4): 487-488.
- Posada D, Crandall KA, Templeton AR (2006): *Nested clade analysis statistics*. *Molecular Ecology Notes* 6: 590-593.



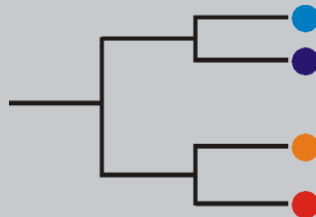
Analysis of phylogeographical data – comparison of F_{ST} and N_{ST} – unordered vs. ordered alleles









genetic distances among alleles (haplotypes)

- without phylogeny (F_{ST} – unordered alleles – frequencies)

				
	0	1	1	1
	1	0	1	1
	1	1	0	1
	1	1	1	0

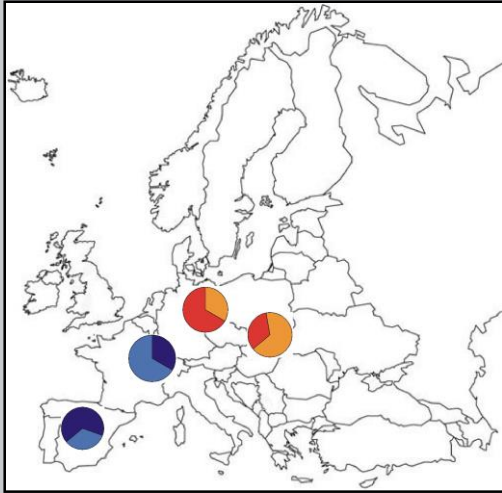
- with phylogeny (N_{ST} – ordered alleles – distances among haplotypes are taken into account)



				
	0	1	3	3
	1	0	2	2
	3	2	0	2
	3	2	2	0

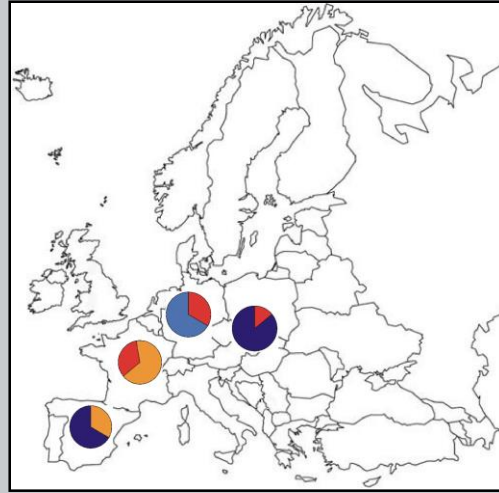
- testing – F_{ST} vs. N_{ST} vs. 0
- software – *PERMUT*, *SPAGeDi*
- Pons O. & Petit RJ (1996): *Measuring and testing genetic differentiation with ordered versus unordered alleles*. *Genetics* 144: 1237-1245.

Analysis of phylogeographical data – comparison of F_{ST} and N_{ST} – unordered vs. ordered alleles



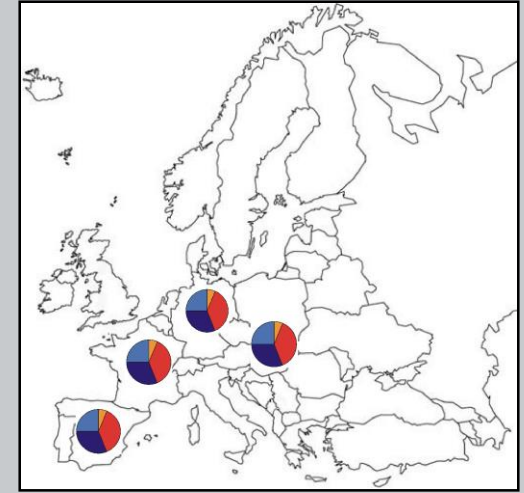
phylogeographical pattern

$$N_{ST} > F_{ST} > 0$$



genetic structure without
phylogeographical pattern

$$N_{ST} = F_{ST} > 0$$



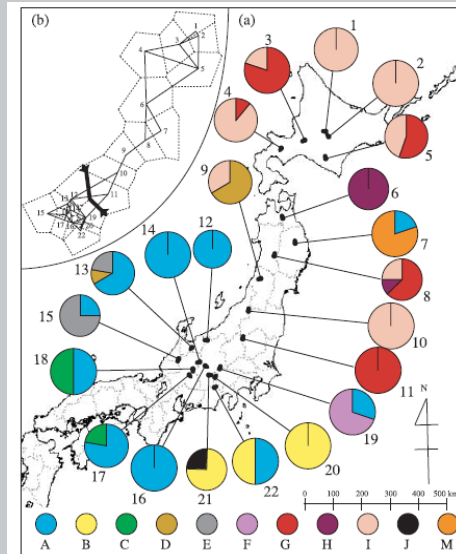
no genetic structure

$$N_{ST} = F_{ST} = 0$$

phylogeographical pattern – if mutation rate > migration

Analysis of phylogeographical data – *identification of sharp changes in genetic similarity*

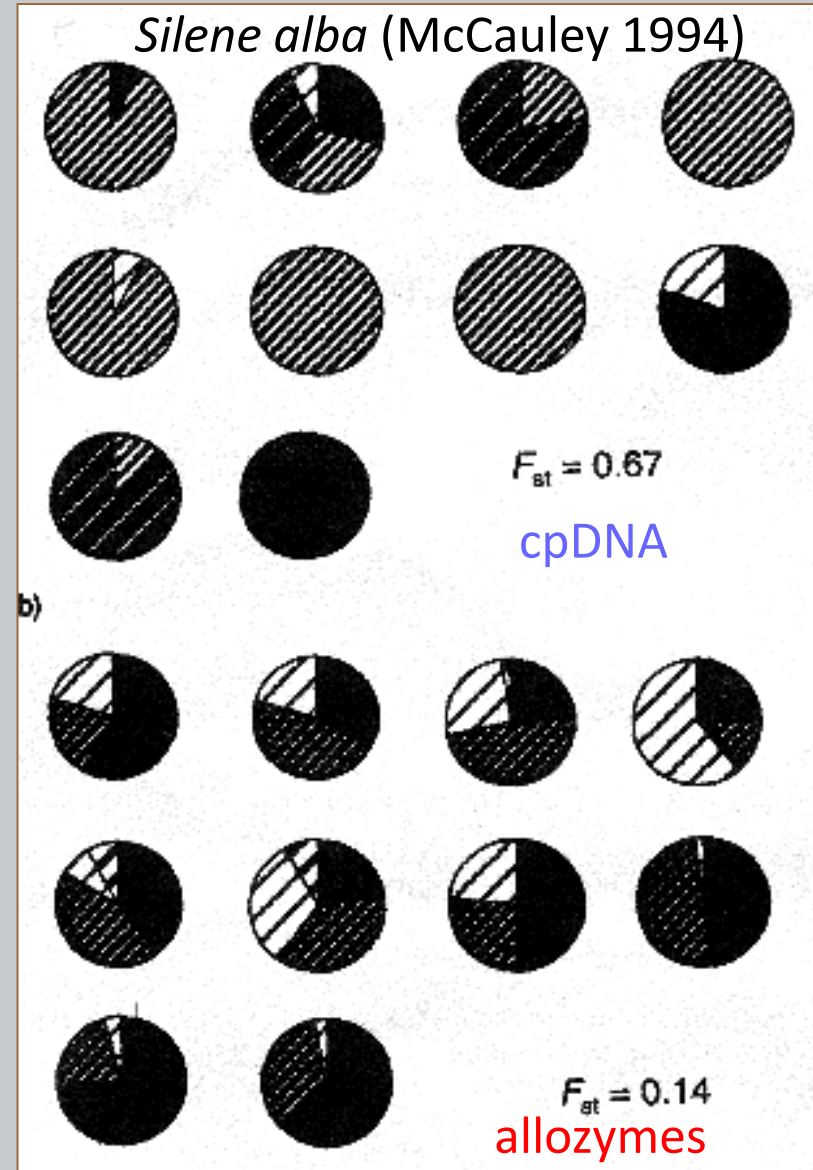
- software – *Barrier*
- *Mani et al. 2004*
- *Mani & Guérard 2004*



Proportion of pollen and seed transfer to the gene flow

- **pollen** - haploid **nuclear DNA**
- **seeds** - diploid **nuclear DNA**
- **cpDNA**

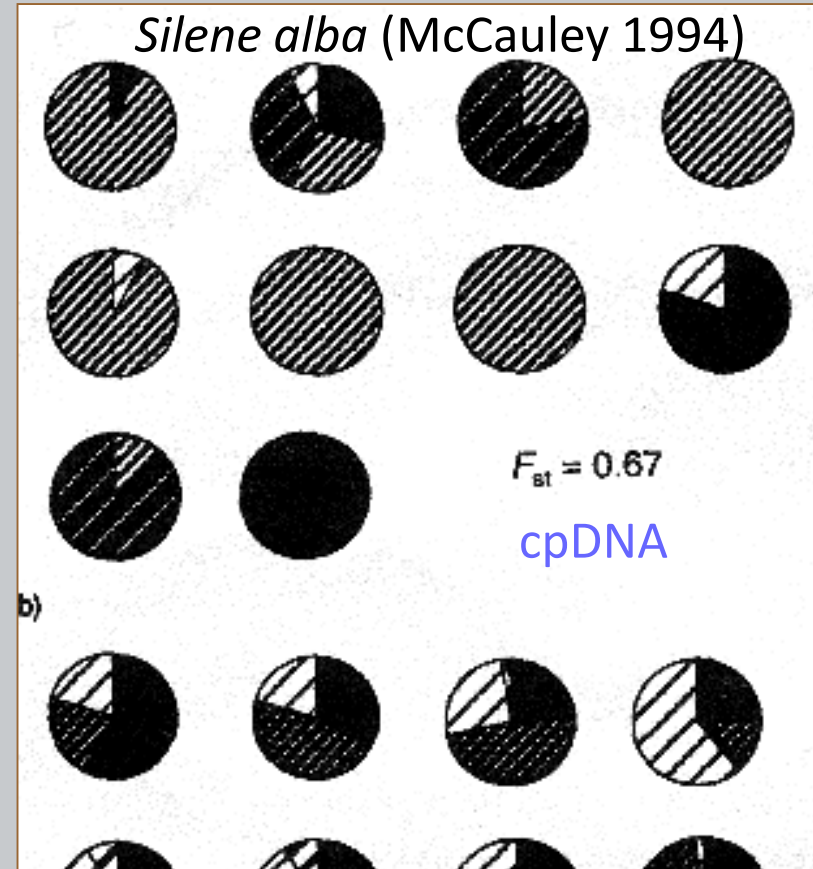
$$\frac{\text{pollen migration}}{\text{seed migration}} \approx \frac{\left(\frac{1}{F_{STb}} - 1\right) - 2\left(\frac{1}{F_{STm}} - 1\right)}{\left(\frac{1}{F_{STm}} - 1\right)}$$



Proportion of pollen and seed transfer to the gene flow

- **pollen** - haploid **nuclear DNA**
- **seeds** - diploid **nuclear DNA**
- **cpDNA**

$$\frac{\text{pollen migration}}{\text{seed migration}} \approx \frac{\left(\frac{1}{F_{STb}} - 1\right) - 2\left(\frac{1}{F_{STm}} - 1\right)}{\left(\frac{1}{F_{STm}} - 1\right)}$$

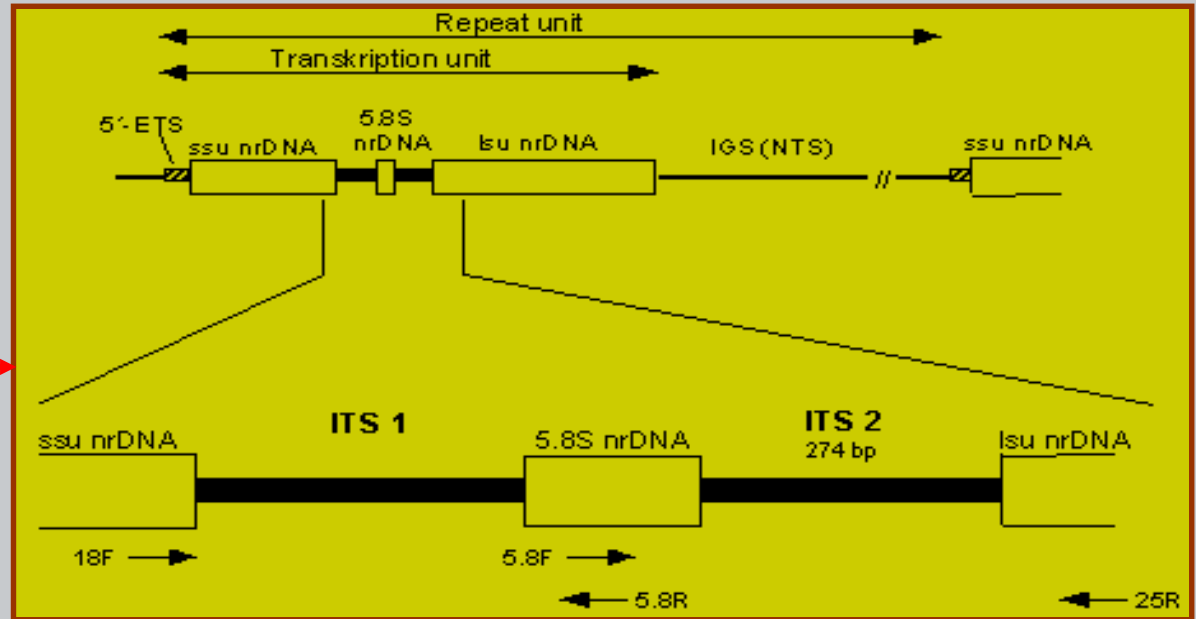


<i>Plant species</i>	<i>Pollen dispersal</i>	<i>Seed dispersal</i>	<i>Pollen/seed migration rate</i>	<i>Reference</i>
<i>Quercus</i> sp.	Wind	Bird	196	Kremer et al. (1991)
<i>Pinus contorta</i>	Wind	Wind	28	Dong and Wagner (1993)
<i>Argania spinosa</i>	Insect	Ruminant	2.5	El Mousadik and Petit (1996)
<i>Pinus sylvestris</i> (Scotland)	Wind	Wind	18	Sinclair et al. (1998)
<i>Pinus sylvestris</i> (Spain)	Wind	Wind	105	Sinclair et al. (1999)

RFLP in systematics

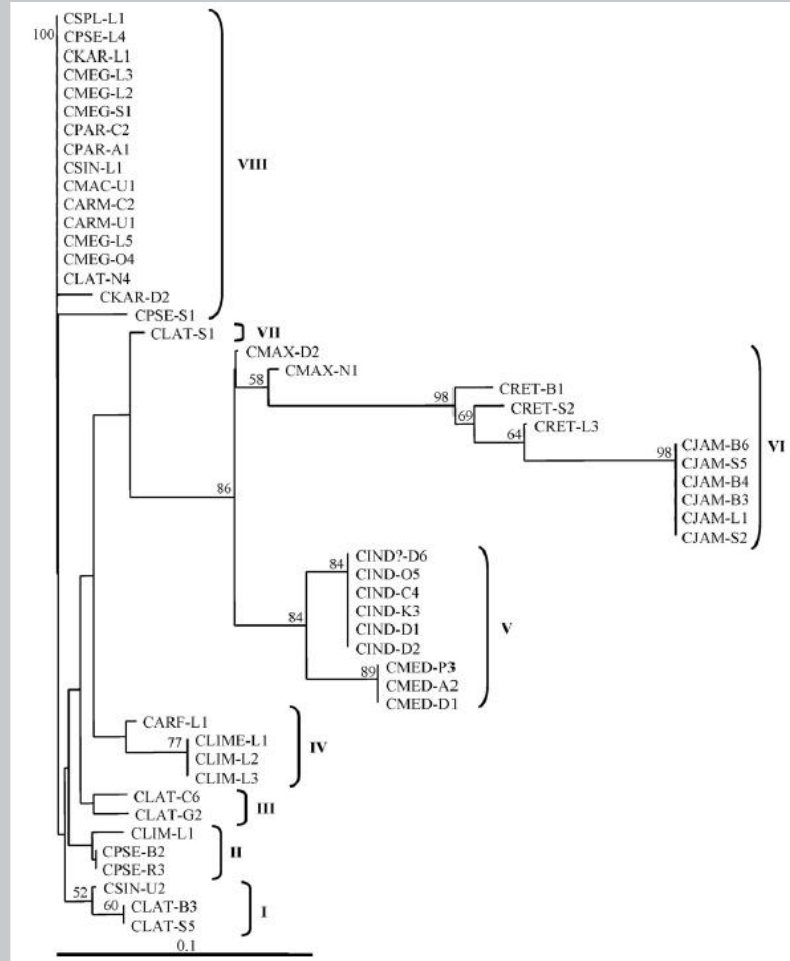
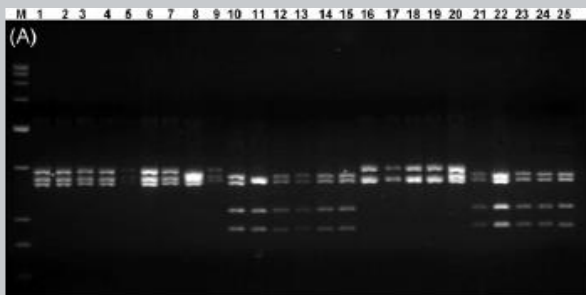
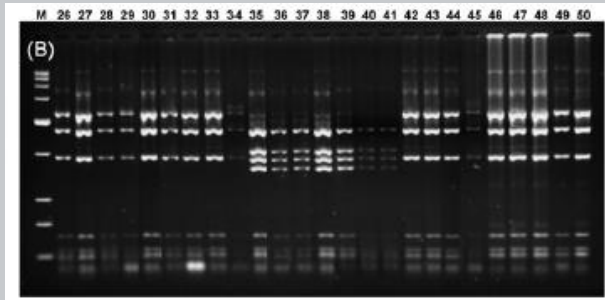
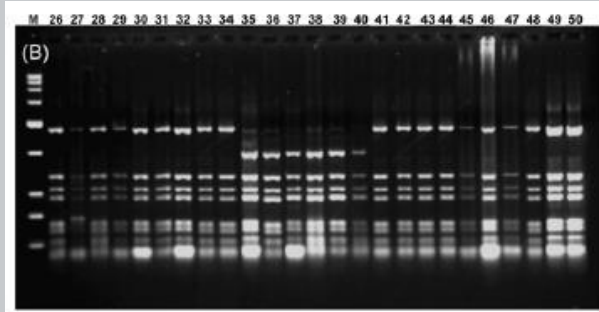
- RFLP + hybridization
- PCR-RFLP

- cpDNA
- rDNA



- interspecific and intergeneric relationships
- hybridization – possibility to detect more ITS copies without cloning

Interspecific relationships

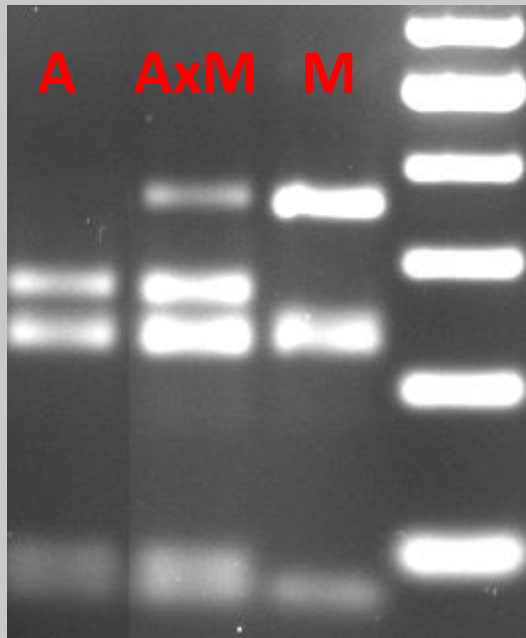


Hybridization – restriction of ITS region (PCR-RFLP)

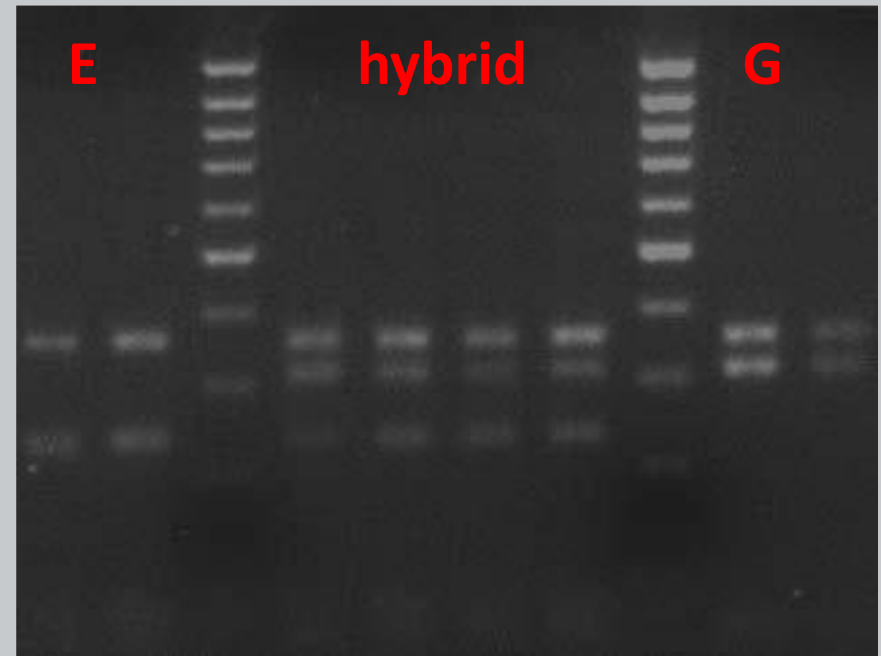


C. enneaphylos

C. glanduligera



Curcuma, unpubl.



Cardamine, Lihová et al. 2007

Population study

Tarayre M. (1997): The spatial genetic structure of cytoplasmic (cpDNA) and nuclear (allozyme) markers within and among populations of the gynodioecious *Thymus vulgaris* (Labiatae) in southern France. *American Journal of Botany* 84(12): 1675-1684



Systematic study

Segraves K.A. et al. (1999): Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology* 8:253-262



Literature

Wicke S. & Schneeweiss G.M. (2015): *Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics*. In: Hörandl E. & Appelhans M.S. (eds.): Next-generation sequencing in plant systematics. IAPT.

Wicke S. et al. (2011): *The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function*. Plant Molecular Biology 76: 273-297.

Palmer J.D. (1986): *Isolation and structural analysis of chloroplast DNA*. Methods in Enzymology 118:167-186

McCauley D.E. (1995): *The use of chloroplast DNA polymorphism in studies of gene flow in plants*. Trends in Ecology & Evolution 10(5): 198-202

Newton A.C. et al. (1999): *Molecular phylogeography, intraspecific variation and the conservation of tree species*. Trends in Ecology & Evolution 14(4):140-145

Petit R.J. & Vendramin G.G. (2007): *Plant phylogeography based on organelle genes: an introduction*. In: Weiss S. & Ferrand N. (eds.), Phylogeography of Southern European Refugia, pp. 23-97. Springer

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

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