Molecular markers in plant systematics and population biology

5. RFLP, cpDNA

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

total DNA

endonuclease

restriction fragments

transfer to the membrane

Southern blotting

hybridization

probe

restriction profile

electrophoresis
Probe preparation

• bacterial cloning of particular part of DNA (gene, operon)
• (non)radioactive product labelling – $^{32}\text{P},^{35}\text{S}$, 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

[Diagram showing hybridizing part of DNA and 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 materialas (detection etc.)
PCR-RFLP

- or – CAPS (Cleaved Amplified Polymorphic Sequence)

- 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
Chloroplast genome (cpDNA)

- circular molecule
- 70-200 kb
- 30-100 in plastide
- homoplasy
- 2 subunits – LSC, SSC
- IR – inverted repeats
- genes for
  - photosystem – *ps*
  - tRNA – *trn*
  - RUBISCO – *rbc*
  - ...

[Diagram of chloroplast genome with labels and genes indicated]
Chloroplast genome (cpDNA)

Lolium perenne L.
135 282 bp
Specificity of cpDNA

- relatively *conserved* – low mutation rate
- 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
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
PCR-RFLP cpDNA

1. PCR

2. Restriction

endonuclease

electrophoresis
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 geographical distribution of gene lineages

- maximum glaciation – 20,000-18,000 BP
- maximum (concentrated) variability in the Mediterranean
- 3 major refugia – Iberian Peninsula, the Appenines, 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.

http://www.pierroton.inra.fr/Fairoak
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 microsatellites and 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
Analysis of phylogeographical data – statistical parsimony

network of haplotypes

- **software** – TCS 1.21


Analysis of phylogeographical data – *nested clade analysis*

- software – GeoDis


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

- **testing** – $F_{ST}$ vs. $N_{ST}$ vs. 0
- **software** – PERMUT, SPAGeDi

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

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

Silene alba (McCauley 1994)

- cpDNA
  - \( F_{st} = 0.67 \)
- allozymes
  - \( F_{st} = 0.14 \)
Proportion of pollen and seed transfer to the gene flow

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

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

\[
\text{seed migration} \approx \frac{1}{F_{STm}} - 1
\]

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

• RFLP + hybridization
• PCR-RFLP
• cpDNA
• rDNA

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

Citrus, Jena et al. 2009
Hybridization – restriction of ITS region (PCR-RFLP)

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

Literature


