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

- total DNA
- restriction fragments
- electrophoresis
- hybridization
- probe
- transfer to the membrane
  *Southern blotting*
- restriction profile
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
## RFLP

<table>
<thead>
<tr>
<th>pros</th>
<th>cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• highly reproducible pattern</td>
<td>• large amount of DNA necessary</td>
</tr>
<tr>
<td>• variability in particular part of DNA (e.g., cpDNA, rDNA ...)</td>
<td>• need for an amount of labelled probe</td>
</tr>
<tr>
<td></td>
<td>• expensive and complicated</td>
</tr>
<tr>
<td></td>
<td>• blotting equipment</td>
</tr>
<tr>
<td></td>
<td>• work with labelled materialas (detection etc.)</td>
</tr>
</tbody>
</table>
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
PCR-RFLP cpDNA

1. PCR
2. Restriction
**dCAPS** — derived Cleaved Amplified Polymorphic Sequence

- 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/](http://helix.wustl.edu/dcaps/)

<table>
<thead>
<tr>
<th>Type</th>
<th>Primer Sequence</th>
<th>Restriction Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>type1</td>
<td>GTGGAAGAAGACTCGATGAGGCTTTTGAGG</td>
<td>CCNNNNNNNNNGG</td>
</tr>
<tr>
<td>type2</td>
<td>GTGGAAGAAGACTCGATGAGGCTTTTGAAGG</td>
<td>CCAGGCTTTTG</td>
</tr>
</tbody>
</table>

BslI digestion:

<table>
<thead>
<tr>
<th>Type</th>
<th>Digestion Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>type1</td>
<td>GTGGAAGAAGACTCGA<strong>CCAGGCT</strong> / TTGGGG</td>
</tr>
<tr>
<td>type2</td>
<td>GTGGAAGAAGACTCGA<strong>CCAGGCTTTTGA</strong>AGG</td>
</tr>
</tbody>
</table>
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 (2×4)
  - ...

![Diagram of the chloroplast genome](image)
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* tranfer in angiosperms – via seeds
  - *paternal* transfer in gymnosperms – via pollen
Plastome structure

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

Wicke & Schneeweiß 2014
Plastome structure

- gene gain is extremely rare
  - three gained before the transition to land (\textit{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

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
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 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.

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

Magri et al. 2006

cpDNA microsatellites

MLP cpDNA
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)
    - 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)
    - 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
- Genetic structure without phylogeographical pattern
- No genetic structure

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

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

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

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

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

Silene alba (McCauley 1994)

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

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

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

<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><em>Quercus sp.</em></td>
<td>Wind</td>
<td>Bird</td>
<td>196</td>
<td>Kremer et al. (1991)</td>
</tr>
<tr>
<td><em>Pinus contorta</em></td>
<td>Wind</td>
<td>Wind</td>
<td>28</td>
<td>Dong and Wagner (1993)</td>
</tr>
<tr>
<td><em>Argania spinosa</em></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


