phylogeography

Pavel Hulva
- Studying mechanisms influencing geographical distribution of genealogies
- Historical, phylogenetic component of spatial distribution of genealogies
Avise, 2000
**PHYLOGENETICS**

*phyle* = tribe, race and

*genetikos* = relative to birth, from *genesis* = birth

**PHYLOGENETICS**

revealing routes of phylogeny

**EVOLUTIONARY BIOLOGY**

revealing mechanisms of evolution
molecular biology

phenotype
often studied directly
proximate mechanisms
focused to extreme phenotypes
(loss of function)

genetic variability and genotype
simple approach to variation
(wild type / mutant etc.)

molecular evolution

phenotype
seldom studied directly

genetic variability and genotype
complex approach to variability
mechanisms of its origin and
maintenance
complicated mathematical apparatus
Divergences among species

Large genetic distances among samples

Easy to find well supported tree

trees – phylogenies

Tree topology important for hypothesis formulations

Nodes represent speciation events

divergences within species (phylogeography)

Small genetic distances among samples

Difficult to find well supported tree

trees – genealogies

Tree topology less important for hypothesis formulations

Nodes represent coalescent events
FIGURE 6.3 (a) Phylogeny for five extant biological species (A-E) and two geographically separated populations (C1 and C2) of C (after Avise and Wollenberg, 1997). The widths of branches (cross-hatched channels) are proportional to population sizes through time, and also indicate a geographic orientation. Thus, A is a peripheral isolate from C1, and B arose within the range of C2. The population sundering agents are extrinsic barriers to gene flow (white areas), intrinsic RIBs (black areas), or both in temporal order of appearance (white then black). (b) Simplified "stick" representation of the phylogeny in (a).
FIGURE 6.6  Same phylogeny as in Fig. 6.3 but here depicting the complete organismal pedigree through 21 discrete generations leading to the present (after Avise and Wollenberg, 1997). The two lines tracing from each male (filled square) or female (open circle) in any generation identify parents of that individual. They also describe the geographic dispersal of offspring (assumed to be limited by distance) and the mating events.
properties of mtDNA

small, compact, clonal, stable (two mitochondrial membranes, more copies), high mutation rate – high intraspecific variation - haplotypes, particular regions exhibit different mutation rate, lack of recombination, matrilinear

exceptions: recombination, heteroplazmy, paternal leakage
FIGURE 1.10 Structure and organization of genes in mammalian mtDNA. $O_H$ and $O_L$ are origins of replication of the two strands of the molecule. MtDNA molecules normally display identical gene orders within and among related animal species, but occasional rearrangements have occurred in evolution and can be exploited as markers to identify deep branches in macroevolutionary trees (Smith et al., 1993; Boore et al., 1995).
FIGURE 1.11 Mitochondria (black dots) exist as large populations of organelles within the cytoplasmic lineages of somatic cells and germ cells. In most animals, about 20–50 germ-cell generations intervene between organismal generations.
FIGURE 2.1 Schematic representation of the lineage sorting process within a population. Shown are matrilineal pathways connecting mothers and daughters across more than 40 generations. Lines that cross do not imply any genetic event such as recombination. Note that all seven founding lineages except one have gone extinct during this interval, and that all extant lineages (top) coalesce to a common female parent 10 generations earlier. Note also that these matrilineal pathways collectively describe a hierarchical and branched gene tree reflecting the hereditary transmission routes traversed by mtDNA molecules.

Avise, 2000

mtDNA
nonreticulate genealogy

x

nucDNA
anastomose genealogy
mitochondrial Eve

(mitochondrial mrca
 – most recent common ancestor)

Brown et al. (1980) – small divergence in mtDNA in recent populations of Homo sapiens (p=0.0036) – divergence 180 – 360 000 years, bottleneck (Garden of Eden, Noah’s Ark)

Cann et al. (1987) – hypothesis out of Africa
Y-chromosome Adam

(Y-chromosome mrca)
plant mtDNA – high rate of rearrangements, slow sequence evolution (ca 100x lower mutation rate compared to animal mtDNA)
molecular clock

Brown and Wright, 1975 – parthenogenetic lizards
Brown et al., 1979 – primates
  calibration: 2% between taxa per $10^6$ years
  1% per lineage per $10^6$ years
nDNA

methodological complications – diploid (problems with haplotype isolation in heterozygotes, gene families – crossamplifications

biological properties – relatively slow evolution, recombinations
FIGURE 2.28 Coalescent process (heavy lines) for an autosomal gene within an organismal pedigree spanning more than 40 generations. This pedigree is an amalgamation of its matrilineal and patrilineal components as presented separately in Figs. 2.1 and 2.26, with additional pathways connecting each daughter to her father and each son to his mother.
**Amia calva** (bowfin)
southeastern U.S.
- eastern drainages
- western drainages

**Thunnus albacares** (yellowfin tuna)
circumglobal
- rare haplotypes
- shared in similar frequencies by Indo-Pacific and Atlantic populations
landscape genetics
maximum difference Monmonier algorithm
The Womble approach
Geneland – Poisson - Voronoi tesselation
molecular demography
skyline plots
**ALLOPATRY AND DIVERGENCE**

examples of divergent evolution in isolation - islands

*Table 2.1* A simple classification of island types distinguishing: (1) classic types of ‘real’ island, being land surrounded by open water, from (2) habitat islands, for which the contrast between the ‘island’ and the surrounding matrix is less stark but still sufficient to represent a barrier or filter to population movements. Australia, given its huge size, is essentially continental in character and in practice is not treated as an island in the present work.

<table>
<thead>
<tr>
<th>Type of island</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Land surrounded by water</strong></td>
<td></td>
</tr>
<tr>
<td>Island continent</td>
<td>Australia</td>
</tr>
<tr>
<td>Oceanic islands</td>
<td>Hawaii, Canaries</td>
</tr>
<tr>
<td>Continental fragments</td>
<td>Madagascar, New Caledonia</td>
</tr>
<tr>
<td>Continental shelf islands</td>
<td>British Isles, Newfoundland</td>
</tr>
<tr>
<td>Islands in lakes or rivers</td>
<td>Isle Royale (Lake Superior), Barro Colorado island (Lake Gatún),</td>
</tr>
<tr>
<td></td>
<td>Gurupá island (River Amazon)</td>
</tr>
<tr>
<td><strong>Habitat islands</strong></td>
<td></td>
</tr>
<tr>
<td>Patches of a distinct terrestrial habitat isolated by a hostile matrix</td>
<td>Great-Basin (USA) mountain tops surrounded by desert</td>
</tr>
<tr>
<td></td>
<td>Woodland fragments surrounded by agricultural land</td>
</tr>
<tr>
<td></td>
<td>Thistle heads in a field</td>
</tr>
<tr>
<td></td>
<td>Continental lake (Baikal, Titicaca)</td>
</tr>
<tr>
<td><strong>Marine habitat islands</strong></td>
<td></td>
</tr>
<tr>
<td>The fringing reef around an isolated oceanic island</td>
<td></td>
</tr>
<tr>
<td>Coral reefs separated from other reefs by stretches of seawater</td>
<td></td>
</tr>
<tr>
<td>Seamounts (submerged or not yet emerged mountains below sea level)</td>
<td></td>
</tr>
<tr>
<td>Guyots (submerged flat-topped former islands, i.e. a type of seamount)</td>
<td></td>
</tr>
</tbody>
</table>
neutral processes – often connected with demographic changes (bottlenecks) during island colonization, sudden speciation scenarios

- founder effect (Mayr 1954) – partial transfer of ancestral polymorphism in course of colonization event

- founder-flush model (Carson 1968) – founder effect followed by population expansion with reduced selection

- transilience model (Templeton 1980) – rapid shift to different peak in adaptive landscape during short period of inbreeding following colonization event, without substantial reduction of genetic variability
- genetic drift
insular model of population structure

F statistics

impact of population subdivision on heterozygosity

\[ H_I = \text{average observed heterozygosity in subpopulation} \]

\[ H_S = \text{average expected heterozygosity in panmictic subpopulation} \]

\[ H_T = \text{expected heterozygosity per individuum in panmictic population} \]
**F statistics**

**Inbreeding coefficient**
Average reduction of individual $H$ due to assortative mating in subpopulation

\[
F_{IS} = \frac{(H_S - H_I)}{H_S}
\]

**fixation index**
Average reduction of subpopulation $H$ (in relation to total population) due to genetic drift in subpopulation

\[
F_{ST} = \frac{(H_T - H_S)}{H_T}
\]

**total fixation index**
Average reduction of individual $H$ in relation to total population

\[
F_{IT} = \frac{(H_T - H_I)}{H_T}
\]

\[
(1 - F_{IT}) = (1 - F_{IS}) (1 - F_{ST})
\]
isolation by distance
STUDY of statistical differences among local populations is an important line of attack on the evolutionary problem. While such differences can only rarely represent first steps toward speciation in the sense of the splitting of the species, they are important for the evolution of the species as a whole. They provide a possible basis for intergroup selection of genetic systems, a process that provides a more effective mechanism for adaptive advance of the species as a whole than does the mass selection which is all that can occur under panmixia.

RANDOM DIFFERENTIATION UNDER THE ISLAND MODEL

Mathematical consideration requires the use of simple models of population structure. The simplest model is that in which the total population is assumed to be divided into subgroups, each breeding at random within itself, except for a certain proportion of migrants drawn at random from the whole. Since this situation is likely to be approximated in a group of islands, we shall refer to it as the island model.

The gene frequency \( q \) of a subgroup tends to vary about a certain equilibrium point \( \bar{q} \) in a distribution curve \( \phi(q) \) determined by the net systematic pressure (measured by \( Dq \), the net rate of change of gene frequency per generation from recurrent mutation, immigration, and selection) in conjunction with the cumulative effects of accidents of sampling (random deviation \( \delta q \), variance per generation \( \sigma^2_{\delta q} \)) (Wright 1929, 1931, 1942).

\[
\phi(q) = \left( \frac{C}{\sigma_{\delta q}^2} \right) \exp \left[ 2 \int (\Delta q / \sigma_{\delta q}^2) dq \right].
\]

Let \( N \) be the effective size of the subgroup, \( m \) the effective proportion of its population replaced in each generation by migrants, and \( q \), the gene frequency in the total population. The rate of change of gene frequency per generation in a subgroup, taking account only of immigration pressure, is \( \Delta q = -m(q - \bar{q}) \). In a random breeding population \( \sigma_{\delta q}^2 = q(1 - q)/2N \). Substitution in (1) gives the following, choosing \( C \) so that \( \int_0^1 \phi(q) dq = 1 \) (Wright 1931, 1942).

\[
\phi(q) = \frac{\Gamma(4Nm)}{\Gamma(4Nmq) \Gamma[4Nm(1 - q)]} q^{4Nmq - 1} (1 - q)^{4Nm(1 - q) - 1}
\]

\[
\bar{q} = \int_0^1 q \phi(q) dq = q
\]
Drift

Selection

Flow
Eco-evolutionary problems connected with islands:

natural selection, adaptive radiation - Charles Darwin, A.R. Wallace
insular model of population structure – Sewall Wright
allopatric speciation, founder effect – Ernst Mayr
small population paradigm in conservation biology – Graeme Caughley
metapopulation – Ilka Hanski, Michael Gilpin
taxon cycle – Edward O. Wilson
community assembly rules – Jared Diamond
Island biogeography theory, r - K strategy – Robert MacArthur, E.O. Wilson
Mediterranean
**VITAL SIGNS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotspot Original Extent (km²)</td>
<td>2,486,292</td>
</tr>
<tr>
<td>Hotspot Vegetation Remaining (km²)</td>
<td>98,009</td>
</tr>
<tr>
<td>Endemic Plant Species</td>
<td>11,700</td>
</tr>
<tr>
<td>Endemic Threatened Birds</td>
<td>9</td>
</tr>
<tr>
<td>Endemic Threatened Mammals</td>
<td>11</td>
</tr>
<tr>
<td>Endemic Threatened Amphibians</td>
<td>14</td>
</tr>
<tr>
<td>Extinct Species †</td>
<td>5</td>
</tr>
<tr>
<td>Human Population Density (people/km²)</td>
<td>111</td>
</tr>
<tr>
<td>Area Protected (km²)</td>
<td>90,242</td>
</tr>
<tr>
<td>Area Protected (km²) in Categories HV*</td>
<td>20,751</td>
</tr>
</tbody>
</table>

† Recorded extinctions since 1500. *Categories I-IV afford higher levels of protection.*

- chaparral
- maquis
- matorral
- fynbos
- mallee
- kwongan

Areas with mediterranean climate
## DIVERSITY AND ENDEMISM

<table>
<thead>
<tr>
<th>Taxonomic Group</th>
<th>Species</th>
<th>Endemic Species</th>
<th>Percent Endemism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plants</td>
<td>22,500</td>
<td>11,700</td>
<td>52.0</td>
</tr>
<tr>
<td>Mammals</td>
<td>226</td>
<td>25</td>
<td>11.1</td>
</tr>
<tr>
<td>Birds</td>
<td>489</td>
<td>25</td>
<td>5.1</td>
</tr>
<tr>
<td>Reptiles</td>
<td>230</td>
<td>77</td>
<td>33.5</td>
</tr>
<tr>
<td>Amphibians</td>
<td>79</td>
<td>27</td>
<td>34.2</td>
</tr>
<tr>
<td>Freshwater Fishes</td>
<td>216</td>
<td>63</td>
<td>29.2</td>
</tr>
</tbody>
</table>
Ice ages
since Pliocene epoch, 2.56 Mya
periglacial steppe
1. Tibetan Plateau Blocks Monsoonal Flow
2. Deflection of Gulf Stream Southward
3. Scandinavian Ice Blocks Moisture Flow
4. Complete Ice Cover Seals North Atlantic
5. Lowered Sea Level Exposes Continental Shelf
Areál mamuta během pozdního pleistocénu a holocénu
Role of glaciations in distribution of genealogical lineages and speciation

Traditional paradigm of European phylogeography

• Glaciations pushed temperate components of interglacial communities into the south
• Owing to geomorphology of southern Europe, they survived in Mediterranean refuges (with divergent evolution and possible allopatric speciation)
• Recent (interglacial) northern ranges are outcome of postglacial recolonization from Mediterranean refugia
• In case of secondary contacts of allopatrically evolved lineages (suture zones) – situation reflecting degree of evolution of reproductive isolating mechanisms (reinforcement, hybrid zones, introgression)
• cf. distribution of European sister taxa
Figure 1
The three major Mediterranean refugial and differentiation centres of Southern Europe during the last ice-age (R1: Atlantic-Mediterranean, R2: Adriatic-Mediterranean, R3: Pontic-Mediterranean) and the geographic location of the five most important contact and hybridisation areas where different biota got into secondary contact during the postglacial range expansion processes (H1: Pyrenees, H2: Alps, H3: western Central Europe, H4: eastern Central Europe, H5: Central Scandinavia. (Based on Taberlet et al. [9] and Hewitt [5]).
Fig. 2 Three paradigm European postglacial colonization patterns, the grasshopper, the hedgehog and the bear, and three similar species patterns, the alder, the oak and the shrew (see Hewitt 1999 for detailed discussion and references).
**SYMPATRY AND HYBRIDIZATION**

**hybrid** – descendant of mating between heterospecifics

**integrade** – descendant of mating between conspecifics

- Definition depends upon species concept used. Recently usually general term for descendant of mating between different lineages
introgression

backcrossing

<table>
<thead>
<tr>
<th>Parental</th>
<th>AABB</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F₁ x P</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Backcross</th>
<th>AaBb</th>
<th>Aabb</th>
<th>aaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>b</td>
<td>B</td>
<td>b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>AB</th>
<th>Ab</th>
<th>aB</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Frequency</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
</tr>
</tbody>
</table>
**introgression** – exchange of genes between two species mediated via hybridization and backcrossing

two possible directions

In later generations of unidirectional backcrossing the proportion of latter genome decreases (mtDNA capture, ghost of hybrid past)

different intensity in different loci – islands of genomic differentiation
EXAMPLES

Erinaceus europaeus

E. concolor s.l. (recently E. roumanicus + E. concolor)
Fig. 1 Map of the geographical distribution of *Erinaceus europaeus* (vertical lines) and *E. concolor* (horizontal lines) showing the locations studied.
Fig. 6 Phylogeography of the different haplotypes emphasizing the east/west division of *Erinaceus europaeus* and *E. concolor* and the more recent dichotomies in each of them.
E1
E2
E3
C1
C2
Seddon et al., 2001
**Fagus sylvatica** (Magri et al., 2006)

---

**Fig. 1** Modern distribution of *Fagus sylvatica* (light grey; modified from Pott, 1997) and *Fagus orientalis* (dark grey; redrawn from Atalay, 1994) in Europe. Mountain ranges mentioned in the text: Cantabrian Mountains (CM), Basque Mountains (BM), Pyrenees (P), French Massif Central (FM), western Alps (WA), Vosges (V), northern Prealps (NP), Italian Prealps (IP), eastern Alps (EA), southern Bohemia (SB), southern Moravia (SM), Sudety (SU), western Carpathians (WC), eastern Carpathians (EC), Apuseni (A), southern Carpathians (SC), Dinaric Alps (DA), Pirin Mountains (PM), Rhodopes (R), Strandza Mountains (S), southern Apennines (SA), central Apennines (CA), and northern Apennines (NA). Asterisks indicate the location of the four longest European pollen records: (1) Bouchets/Praclaux, (2) Valle di Castiglione, (3) Ioannina and (4) Tenaghi Philippon.
Fig. A: Geographic distribution of life. The distribution is high in Asia, followed by Europe, and then North America. The lowest concentration is in South America.
Fig. 5 Geographical distribution of (a) chloroplast haplotypes detected using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and (b) microsatellites (data for the Italian Peninsula were taken from Velluto et al., 2004).
Fig. 6 Spatial analysis of variance (SAMOVA) on isozyme data. (a) geographical distribution of groups; (b) phylogenetic tree obtained using the neighbour-joining approach on the $F_{CT}$ distance matrix for SAMOVA isozyme groups.
Fig. 9 Tentative location of refuge areas for *Fagus sylvatica* during the last glacial maximum and main colonization routes during the postglacial period.
Figure 1 Three types of microrefugia, according to the relative position with respect to the macrorefugium: (a) distal or remote, (b) widespread or diffuse, and (c) proximal or ecotonal.
Diagrammatic representation of selected species responses to glacial–interglacial climatic changes, in terms of distribution and abundance. Distributions (except where otherwise indicated): blue, glacial; red, interglacial. (a) Classical refugia. (b) Cryptic refugia. Distributions: pale blue, glacial (low abundance, not necessarily continuous). (c, d) Altitudinal shift. Purple: continuous. (e) Tropical refugia. Early-Holocene refugia have a similar pattern, but within the Holocene. (f) Interglacial refugia. (g) Stepwise expansions through successive glacial–interglacial oscillations. Four stages indicated: blue, to light blue, to pink, to red. (h) Stasis, with separated populations. Purple: continuous. (i) Separated into two disjunct populations by ancient vicariance, each of which expands after the LGM. (j) Speciation from one LGM species (blue) into multiple Holocene species (other colours).
**Dryas octopetala**

Fig. 1 Geographic origin of the 52 *Dryas* populations analysed and their grouping according to the Bayesian clustering analyses (structure) based on 155 amplified fragment length polymorphism (AFLP) markers. The geographical distribution of *Dryas octopetala* is redrawn from Hultén & Fries (1986), and the maximum limits of the late Weichselian/Wisconsin ice sheets are redrawn from Abbott & Brochmann (2003) and Brochmann et al. (2004).
Fig. 2 Intrapopulation variation in *Dryas* measured as Nei’s unbiased diversity based on 155 amplified fragment length polymorphism (AFLP) markers. Populations with less than five individuals are excluded.
Schematic map showing some types of refugia for Europe and western Asia. Interglacial refugia for cold-adapted species are shown in blue, glacial refugia for temperate species in red. Long-term refugia, indicated by dark blue/red, are a subset of all refugia that are inhabited throughout at least one full glacial/interglacial cycle. The areas shown in paler colour are refugia in the sense that they are inhabited during the contraction phase, but are not inhabited during the expansion phase owing to the spread of ice sheets during glacials (cold-adapted species), or excessive temperatures and/or too high aridity during interglacials (temperate species). Also shown, in yellow, are interglacial refugia along the oceanic/continental gradient, with a continental refugium in the east and cryptic refugia further west. The ice sheet for the Last Glacial Maximum is taken from Ehlers & Gibbard (2004). The diagram is schematic; not all of the refugia would have been occupied simultaneously, but the ranges are based on real examples.