Molecular markers in plant systematics and population biology

3. Isoenzyme analysis

Tomáš Fér

tomas.fer@natur.cuni.cz

What are enzymes ?

- proteins – more than 100 aminoacids connected with peptide bond
- function – catalysts of chemical processes (enable substrate transformation)
- more than 5,000 enzymes known
- *isoenzymes* (isozymes) – enzymes with the same metabolic function, catalyzing the same reaction, but with different (primary) structure
- *allozymes* – products coded by different alleles of the same gene (locus) – very similar to each other

Enzyme structure

- aminoacids – positive, negative or neutral charge (depends on pH)
- *primary structure* – sequence of aminoacids, determined genetically
- *secondary* and *tertiary structure* (molecule shape) – influenced by molecule size, charge and polarity (hydrophility) – stabilized by covalent disulfide bonds, non-covalent hydrogen bonds, ionic bonds and hydrophobic interactions
- *quarternary structure* – formation of functional enzyme from more subunits (monomeric, dimeric, tetrameric enzymes)

What we get with isoenzyme analysis?

- geneticaly based (inherited) differences
  - i.e., differences in the primary structure
- differences are reflected by
  - total charge of the molecule
  - shape and size of the molecule
- i.e., different mobility of particular isoenzymes in the electric field

How to study isoenzymes

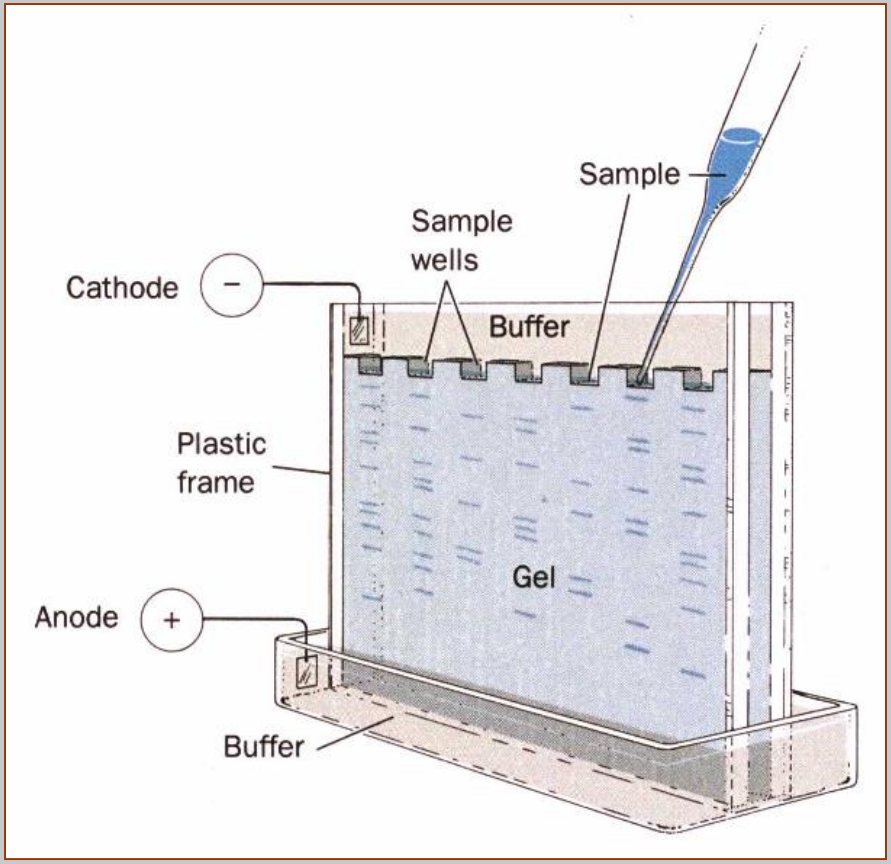
1. extraction
   - from the fresh material
   - homogenization with extraction buffer
   - centrifugation
   - supernatant can be stored frozen at -70 °C
2. separation – electrophoresis
3. detection

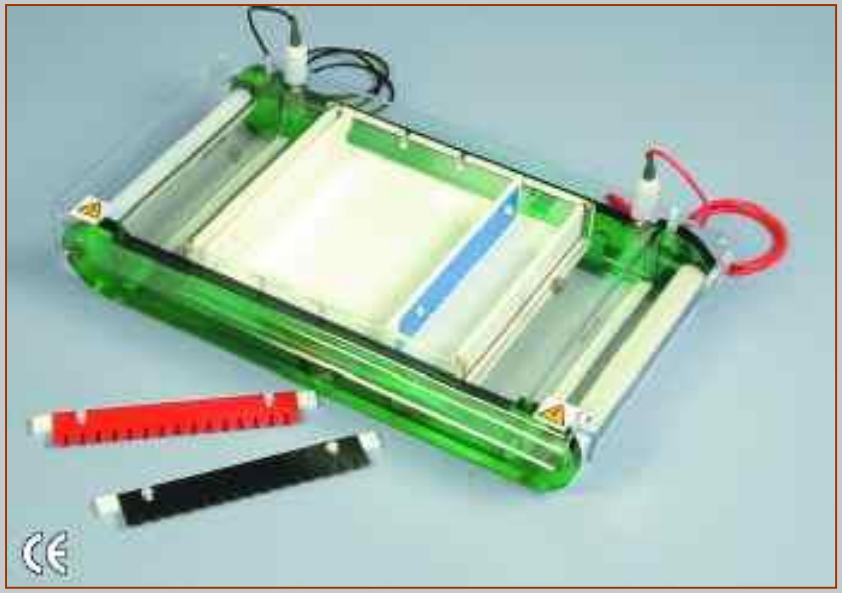
Electrophoresis

- separation of molecules according to their mobility in the electric field
- majority of aminoacids – negative charge in alkalic pH
- molecules move to anode (positively charged electrode)
- mobility is influenced by
  - shape and size of the molecule
  - molecule charge
- sensitive method – separation of molecules differing by one charge unit

Electrophoresis – techniques

- vertical – polyacrylamide gels



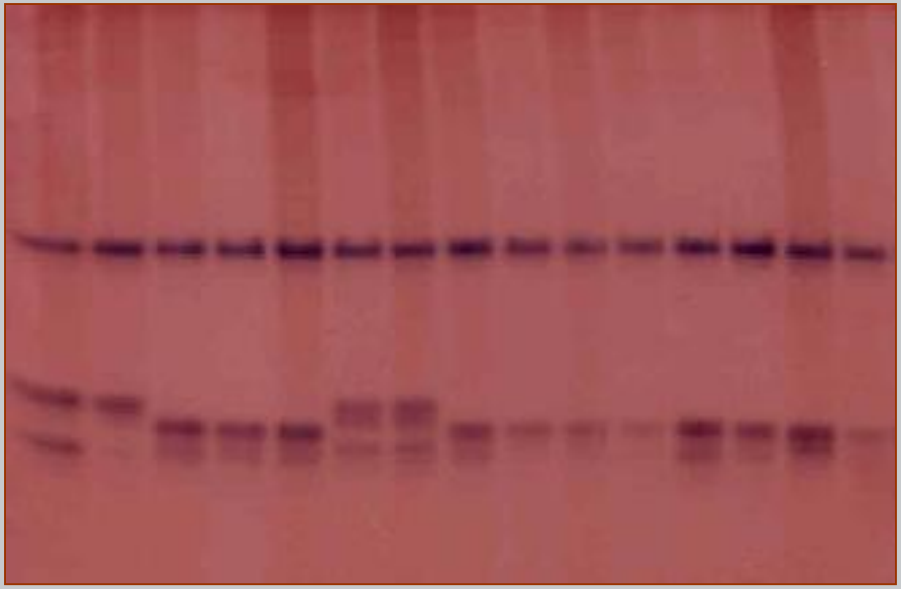


- horizontal – starch gel

Protein detection on the gel

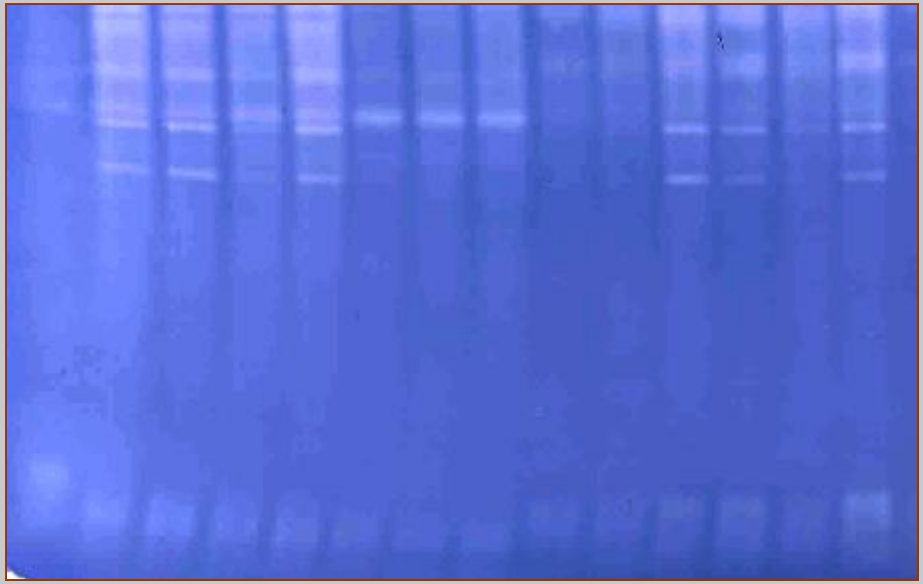
- nothing visible on the gel
- unspecific staining of all proteins (*Coomassie Brilliant Blue*)
- detection of enzymatic activity – specific staining – based on the reaction that is catalyzed by the particular enzyme
- different types of detection
  - *coloured product* – coloured band at the position of enzyme
  - *coloured substrate* – gel destained at the position of enzyme
  - *mixed reaction* – product not coloured but made visible with other reaction(s)

Examples of enzyme detection



LAP

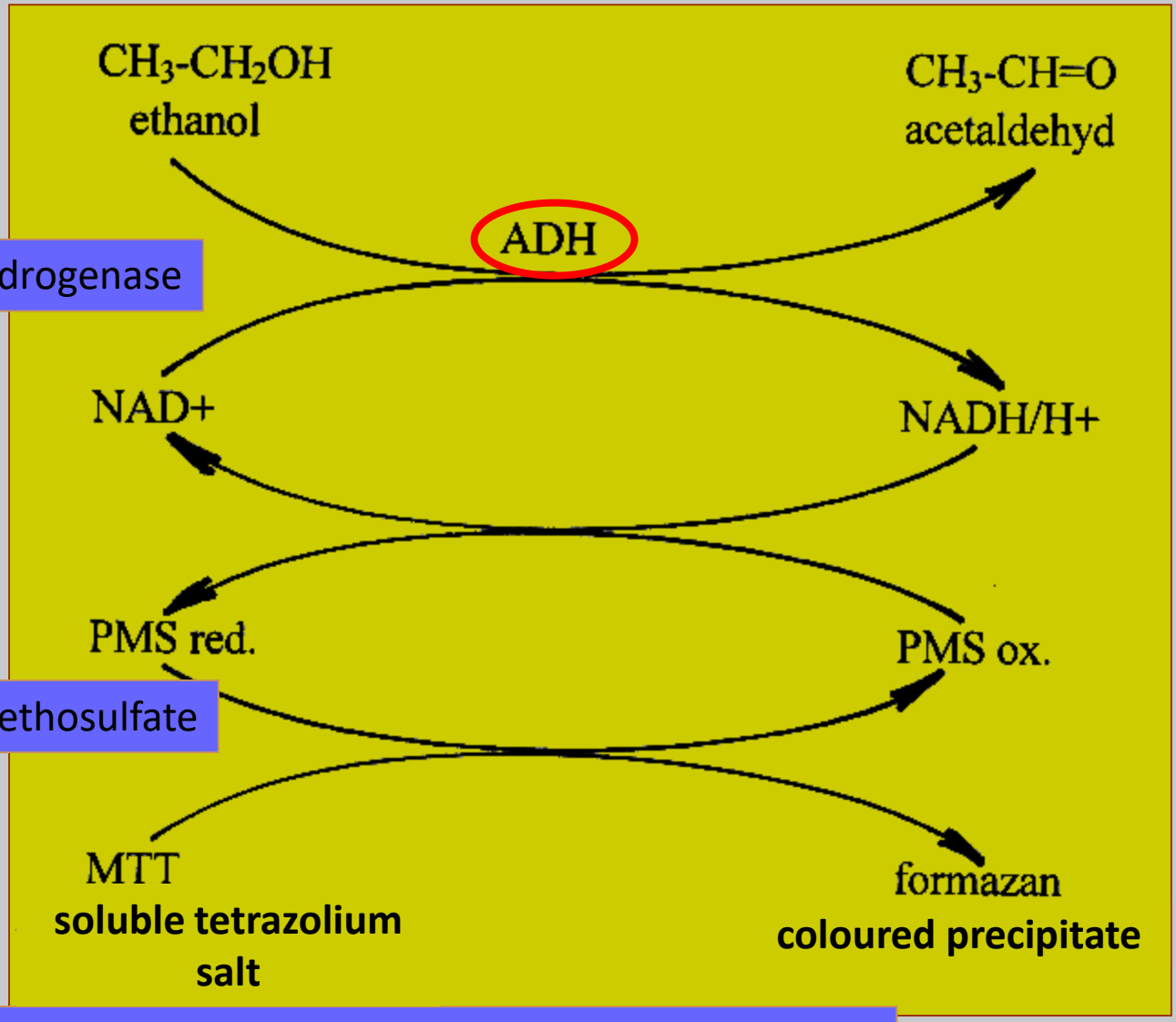
leucin aminopeptidase



SOD

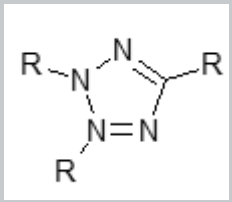
superoxid dismutase

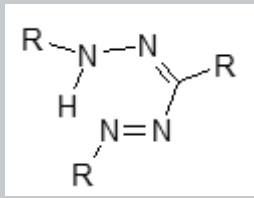
Detection of enzymatic activity



alkoholdehydrogenase

phenasine methosulfate





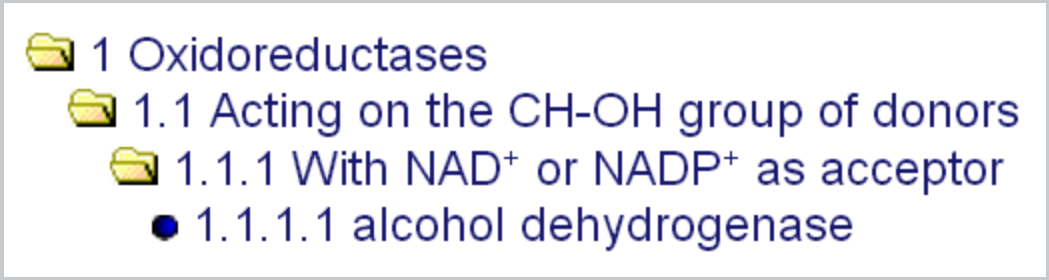
(3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

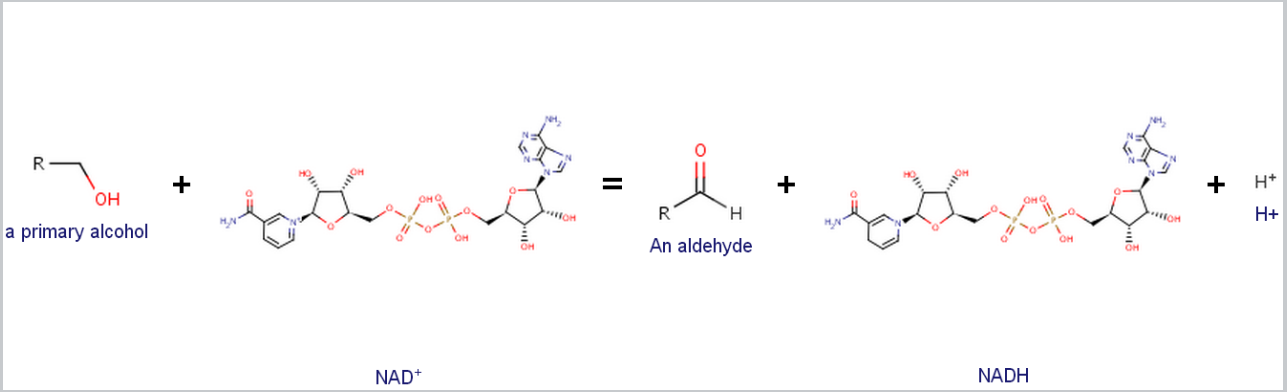
Enzyme classification

1. *oxidoreductases* – electrone transfer (oxidase, dehydrogenase)
2. *transferases* – transfer of functional group (monosacharide, phosphate, methyl, amine, acetyl…)
3. *hydrolases* – hydrolytic cleavage of C-O, C-N or C-C bond
4. *lyases* – cleavage of C-O, C-N or C-C bond
5. *isomerases* – change of geometric structure
6. *ligases* – linkage of two molecules

Enzyme example

- E.C. 1.1.1.1 : alcohol dehydrogenase

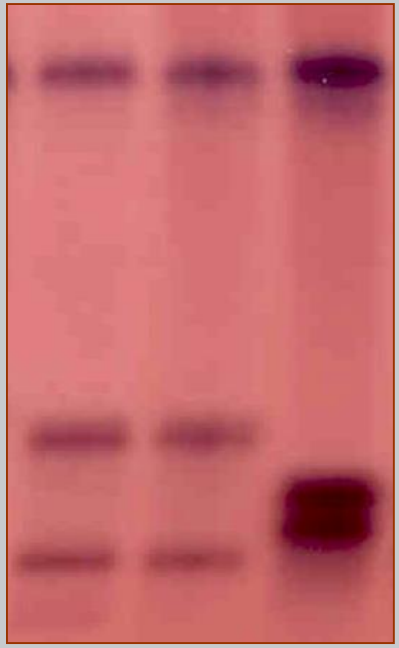




http://www.brenda-enzymes.info/

What we see on the gel

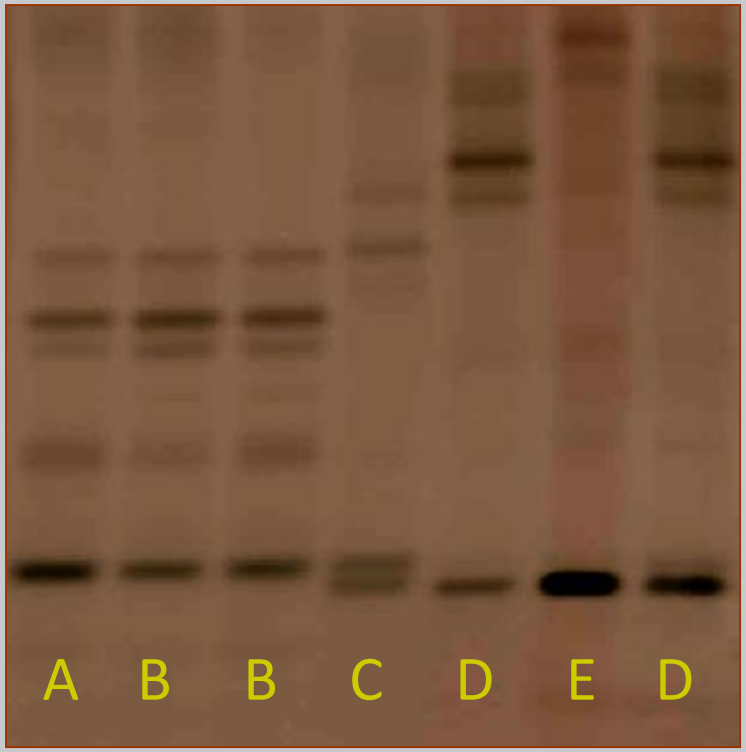
- *zymogram* – banding pattern
- isozyme bands – zones of enzymatic activity
- assumptions for interpretation
  - different mobility reflects difference in DNA (difference is inherited)
  - homology of comigrating bands
  - codominant expression
    - all alleles are exprimated
    - homozygotes and heterozygotes can be distinguished
  - quarternary structure known



Isozyme data evaluation

simple comparison of banding pattern

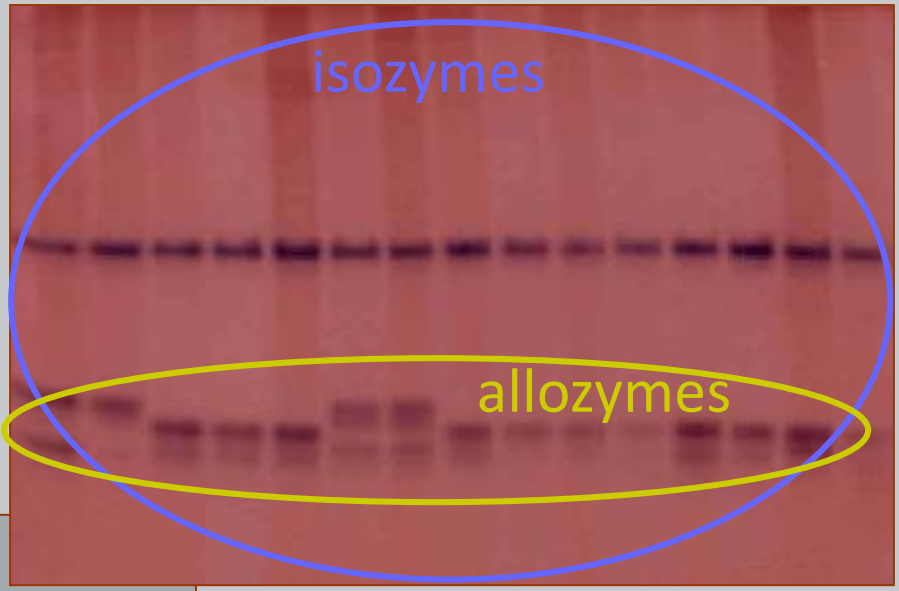
- entire congruence – clone identification



- limited variation…

Allelic evaluation of isozymes

1. determination of number of loci
   - different loci – isozymes might originate from different compartments (e.g., cytosol, chloroplast etc.)
2. determination of number of alleles per locus
   - codominance
   - quarternary structure
   - ploidy level

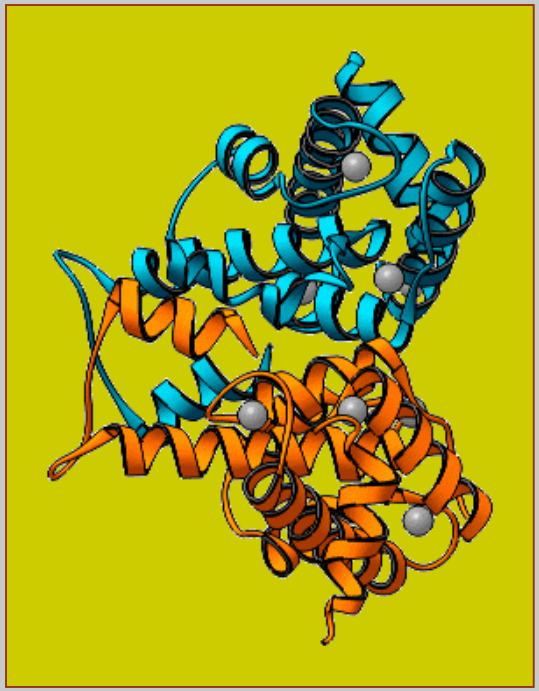


isozymes – catalyze the same reaction

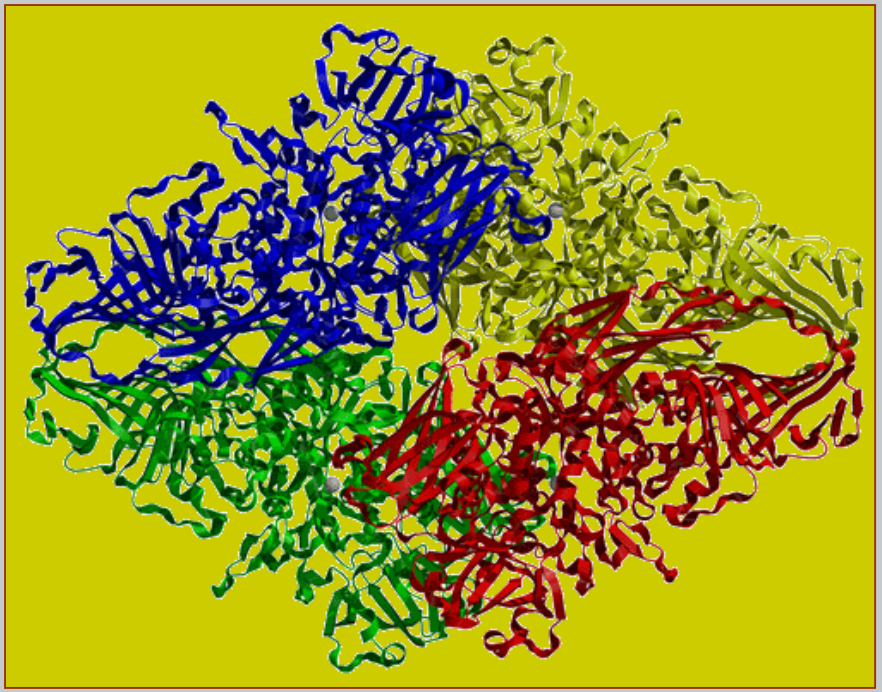
allozymes – products (alleles) of the same gene

Quarternary structure

number and arrangement of subunits into the functional enzyme



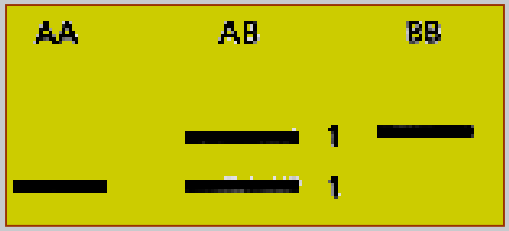
dimer



tetramer

Evaluation of heterozygotes at the locus

monomer

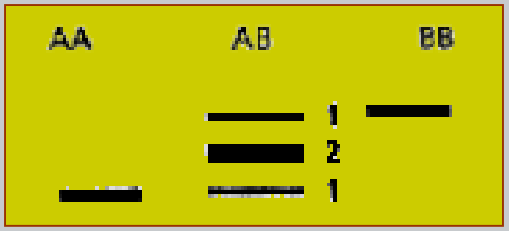


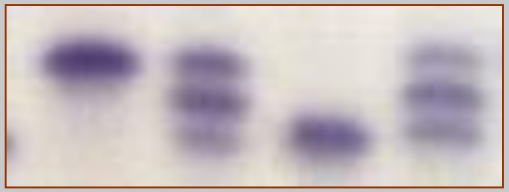
Leucine Aminopeptidase (**LAP**)

Phosphoglucomutase (**PGM**)

Shikimat Dehydrogenase (**SKDH**)

dimer





Amino Aspartate Transferase (**AAT**)

Alcohol Dehydrogenase (**ADH**)

Carboxylesterase (**EST**)

Glucose-6-Phosphate Isomerase (**GPI**)

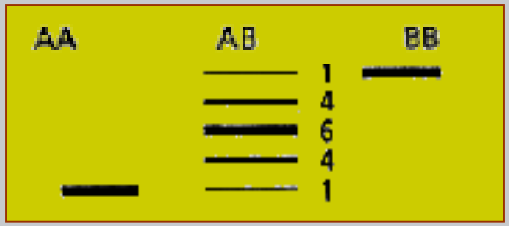
Isocitrate Dehydrogenase (**IDH**)

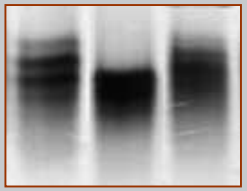
Malate Dehydrogenase (**MDH**)

6-Phosphogluconate Dehydrogenase (**6PGDH**)

Superoxide Dismutase (**SOD**)

tetramer





Glucose-6-Phosphate Dehydrogenase (**G6PDH**)

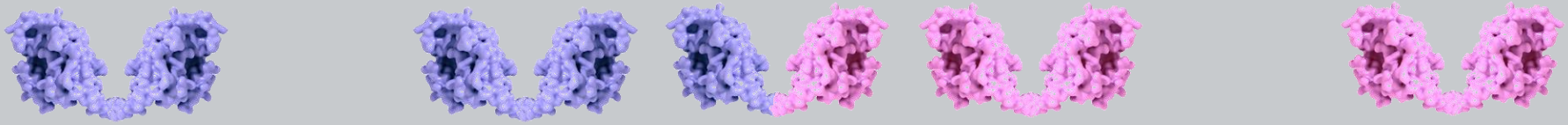
Malate Dehydrogenase NADP+ (**ME**)

Dimeric enzymes

homozygote

heterozygote

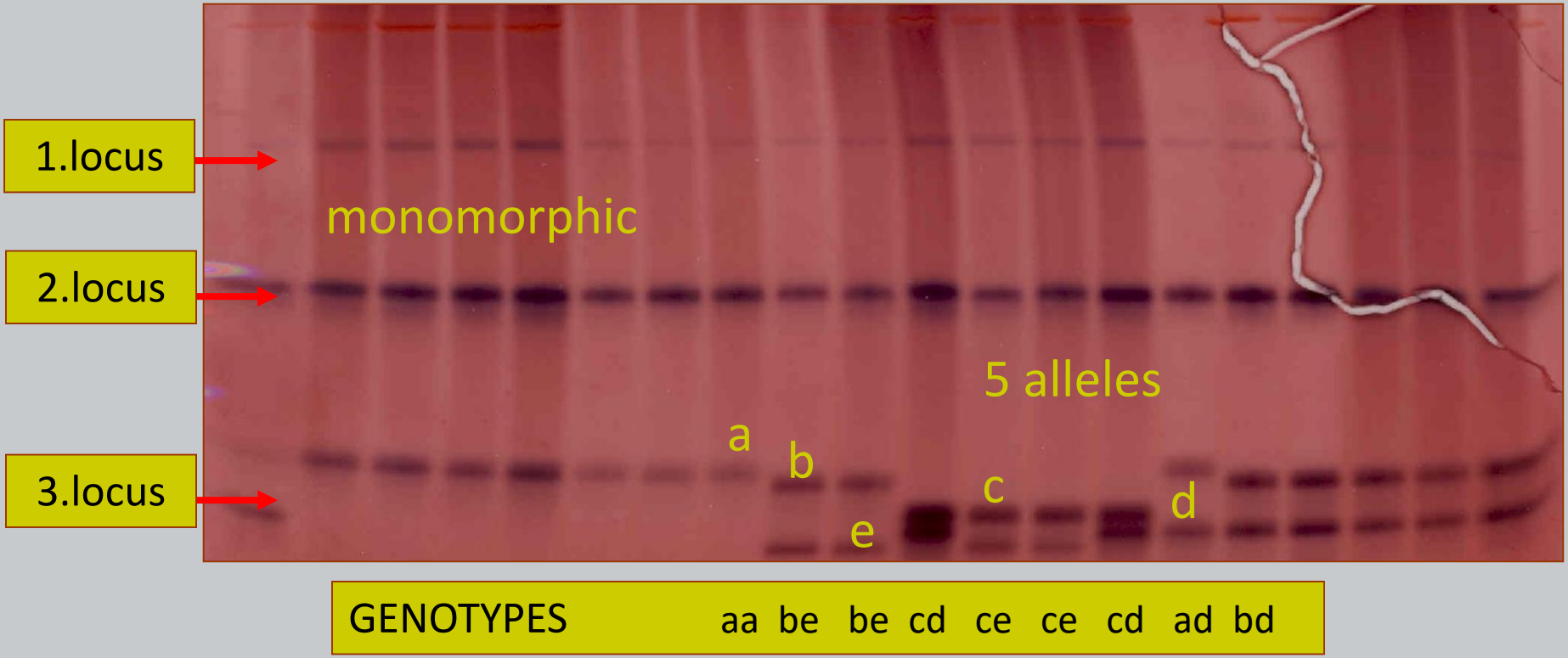
homozygote



1 : 2 : 1

LAP – monomeric enzyme

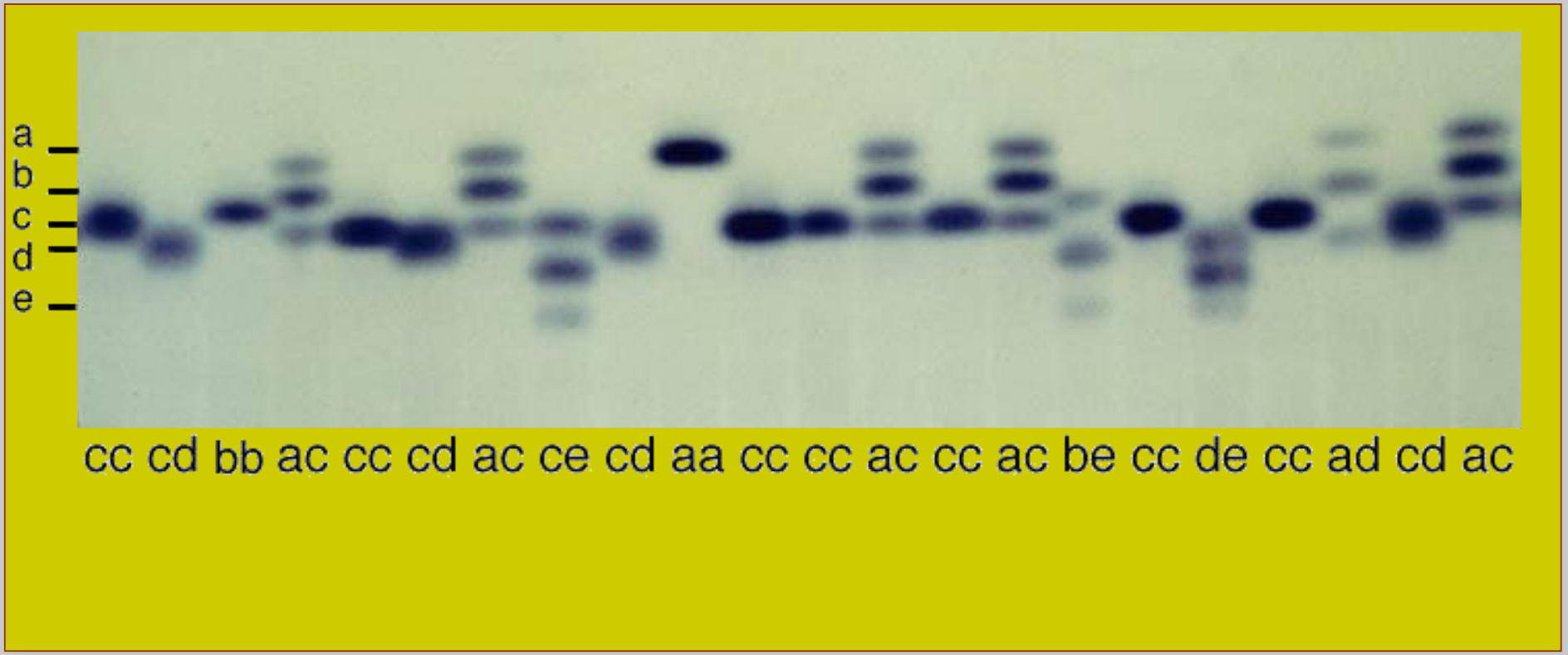
Sparganium erectum – diploid



GENOTYPES aa be be cd ce ce cd ad bd

6-PGDH – dimeric enzyme

Arceuthobium (Viscaceae) – diploid

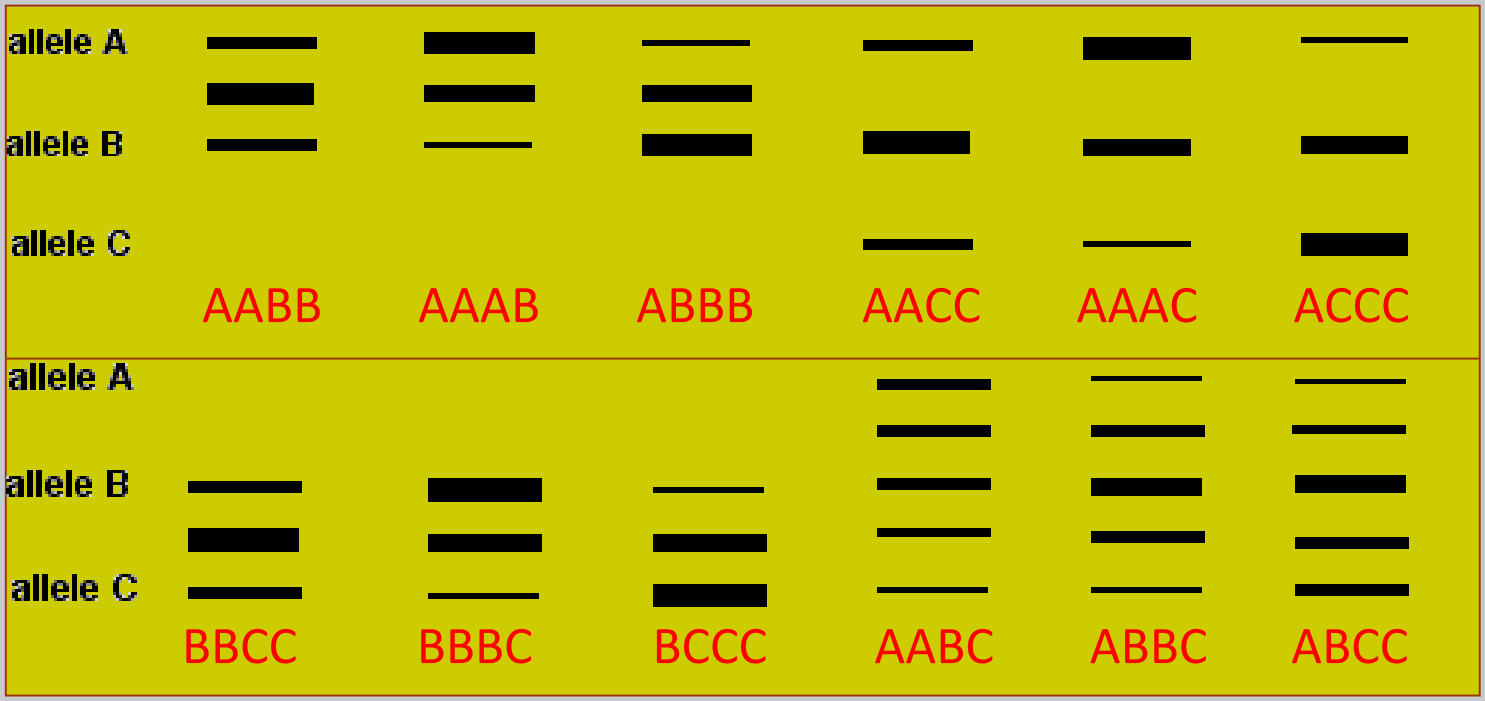


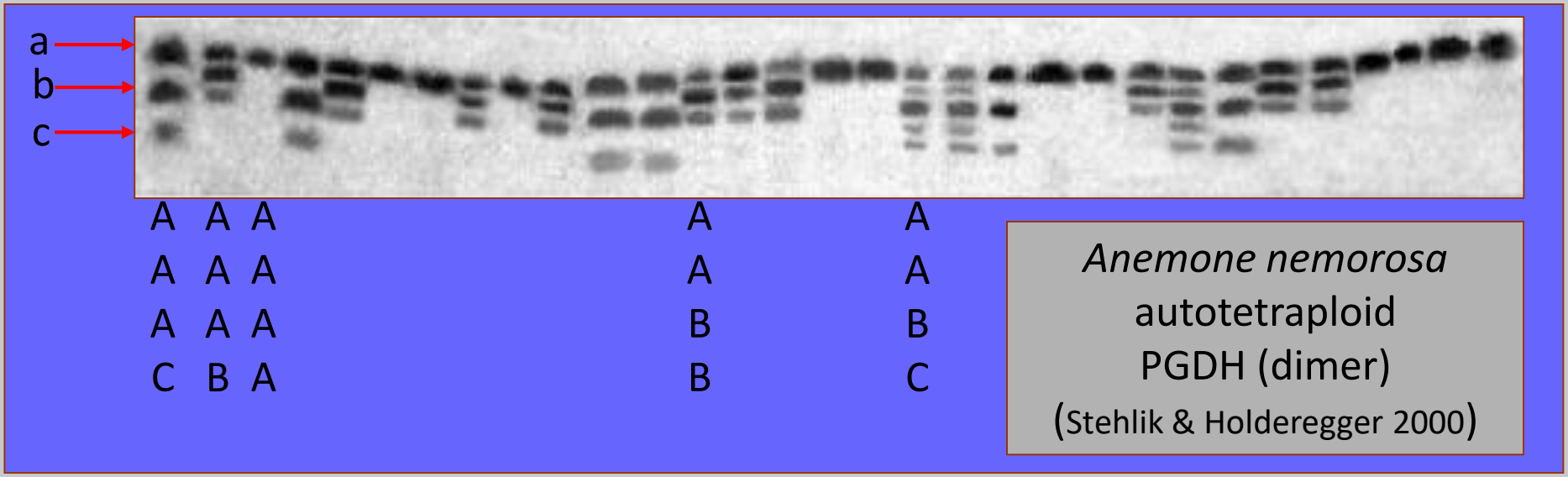
http://www.plant.siu.edu/PLB479/IsozymeTechniques/GelExercise.html

Tetraploid organisms

- autotetraploids
  - 2/2 heterozygotes – AAaa
  - 2 types of 3/1 heterozygotes – AAAa, Aaaa
  - *tetrasomic inheritance* – all combinations are equally possible
- allotetraploids
  - chromozomal and genetic differentiation of two parental genomes
  - *disomic inheritance* – fixed heterozygosity – AABB

Zymogram of tetraploid organisms

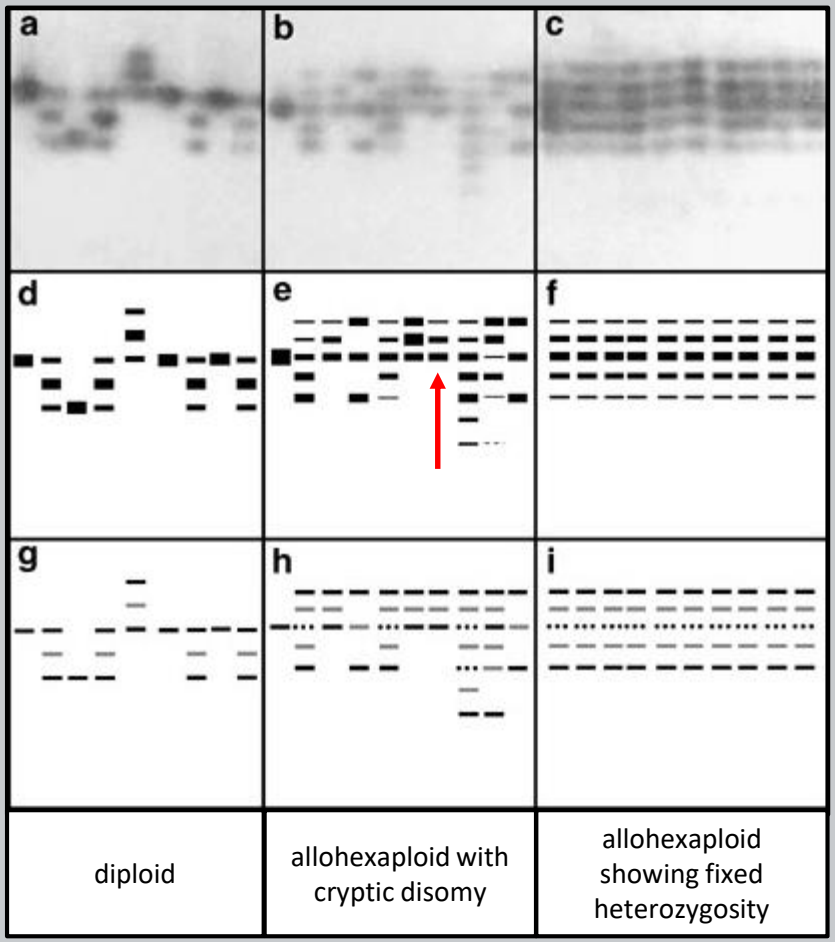




Anemone nemorosa
autotetraploid
PGDH (dimer)
(Stehlik & Holderegger 2000)

Allopolyploids

Mercurialis annua



glucose-6-phosphate isomerase (PGI, E.C. 5.3.1.9)

gel photo

interpretation of band presence and approximate intensity

allelic interpretation

Problems

- allelic dosage
  (→ *ffmmmm* or *fmmmmm*?)
- isoloci assignment (due to disomic pattern)
  (→ genotype *ff*, *mm*, *mm*, or *fm*, *fm*, *mm*?)

Obbard D. J. et al. 2006: Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. Heredity 97:296-303.

Isoenyzme analysis

pros

- fast method – possible to analyse many individuals simultaneously
- cheap technique (in comparison with DNA techniques) - ?
- data comparable among different studies
- codominant marker
- estimate allelic dosage in polyploids
- slow mutation rate (advantage against microsatellites)
  - 10^{-7} /locus*year

cons

- living material needed
- limited variability – low number of alleles per locus – often 2-4 only
- variability in coding part of the genome only
- detected variability
  - 10% of variability of DNA (Nei 1987)
  - only 1/3 of nucleotide substitutions is reflected by aminoacid changes
  - and only ca. 25% is detectable with electrophoresis

Evaluation of codominant data

- number of alleles per locus – *A*
- *allelic richness*
  - expected number of different allele
  - standardized for number of samples
- percentage of polymorphic loci – *P*
- heterozygosity
  - *observed* – H_o (proportion of heterozygotes)
  - *expected* – H_e
  - if Hardy-Weinberg equilibrium expected

    = *gene diversity* – *D*

$$D = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{k} p_i^2$$

- *m* – number of loci
- *k* – number of alleles per locus
- p_i – frequency of *i*-th alelle from *k*

  - probability that particular individual is heterozygote

Interpopulation variation

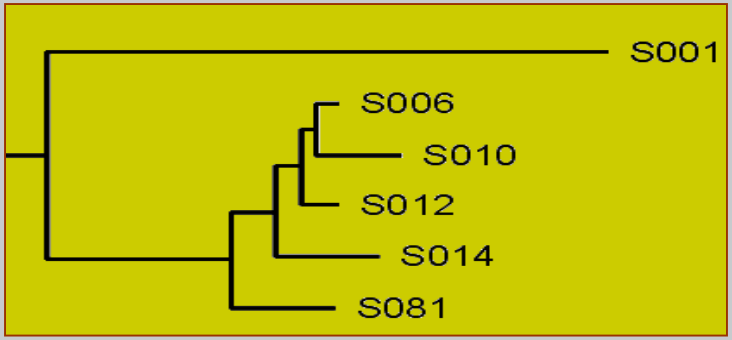
- coefficient of *genetic distance* or *genetic identity*
  - *Nei´s* coefficient

$$I = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 y_i^2}}$$

x_i, y_i	allele frequency in *i*-th locus in populations X and Y
$\sum x_i y_i$	probability of allele identity in populations X and Y
$\sum x_i^2$	probability of allele identity in population X

  - *Rogers´ genetic distance* – D_R – typical for isozymes
- dendrogram – based on the pairwise similarity matrix

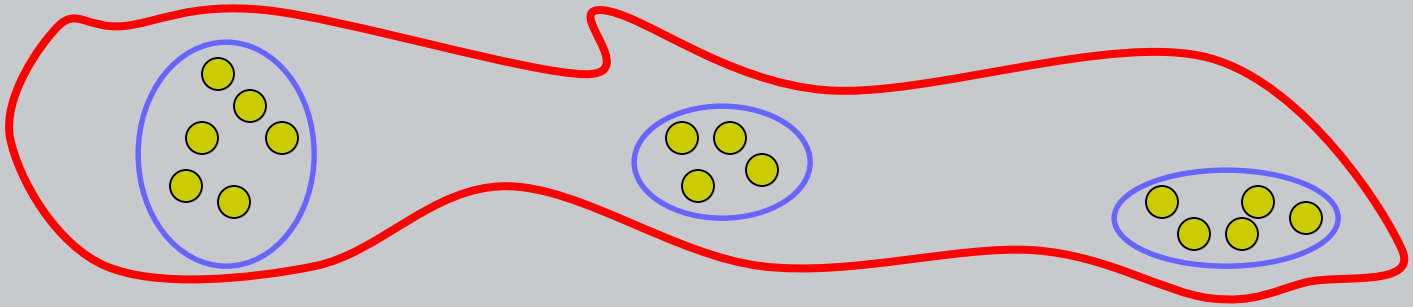
	S001	S006	S012	S008
S001	1.000	0.811	0.811	0.778
S006	0.811	1.000	0.977	0.876
S012	0.811	0.977	1.000	0.898
S008	0.778	0.876	0.898	1.000



  - UPGMA
  - *neighbour-joining* (NJ) – minimalizes tree length

F-statistics (Wright 1951)

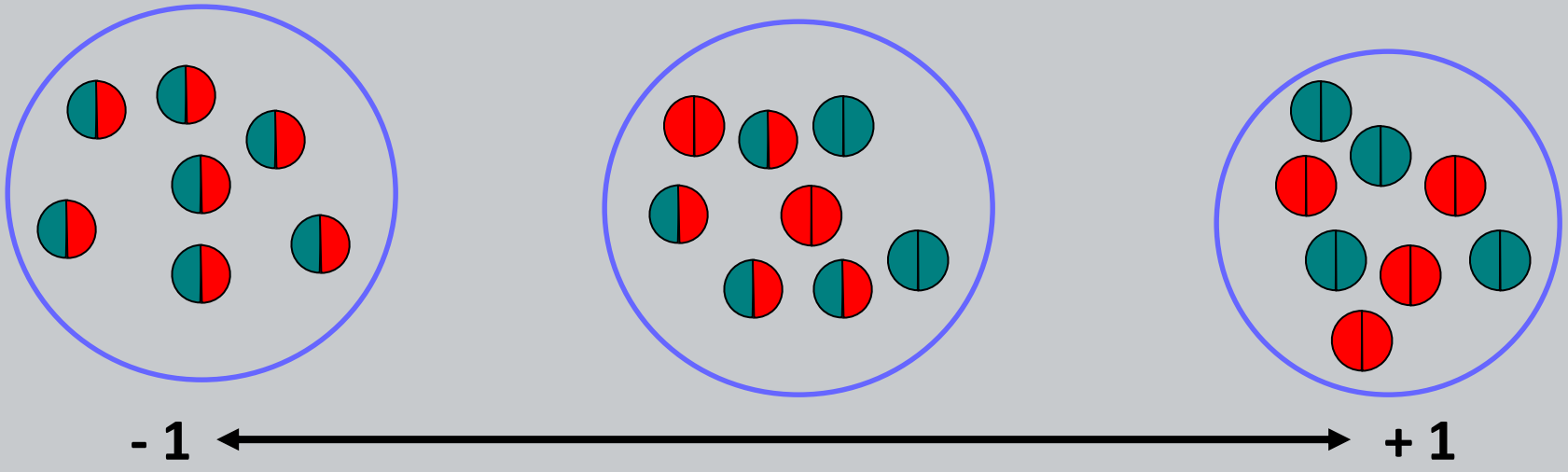
- partitioning of genetic variation
  - I-individual, S-subpopulation, T-total



- F_{IS} – level of inbreeding (*inbreeding coefficient*)
- F_{ST} – subpopulation differentiation
- F_{IT} – global H-W disequilibrium (deviation from random mating)
- $1- F_{IT} = (1- F_{IS})(1- F_{ST})$
- parameters estimation (Weir & Cockerham 1984)
  - correction for number of individual and populations
  - F (~ F_{IT}), θ (~ F_{ST}), f (~ F_{IS})

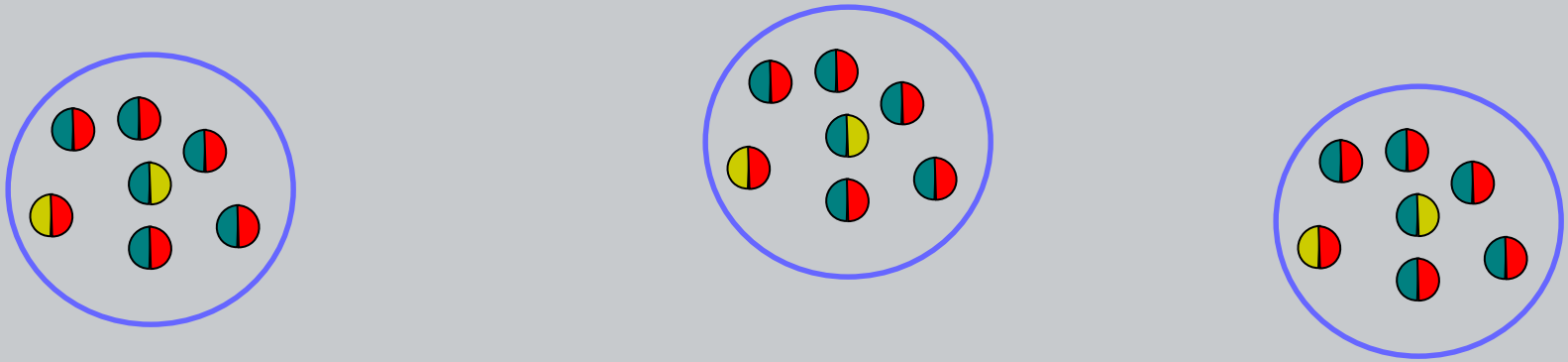
F_{IS} – level of inbreeding (*inbreeding coefficient*)

- - 1 – completely outbred population, i.e., no homozygotes
- 0 – no inbreeding
- +1 – completely inbred population, i.e. no heterozygotes

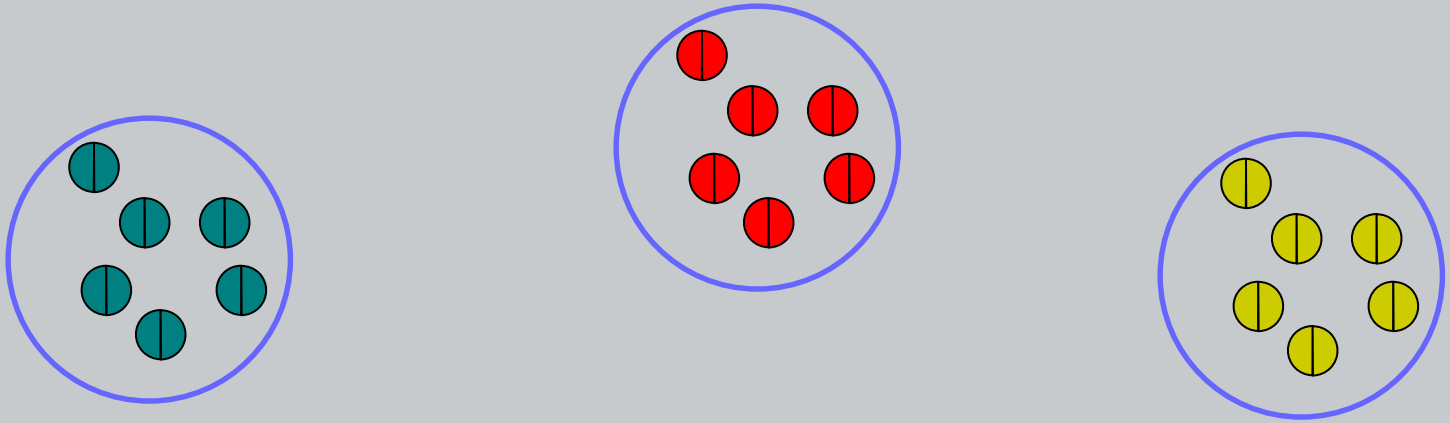


F_{ST} – subpopulation differentiation

- **0** – no genetic population structure
  (same allele frequencies in all populations)

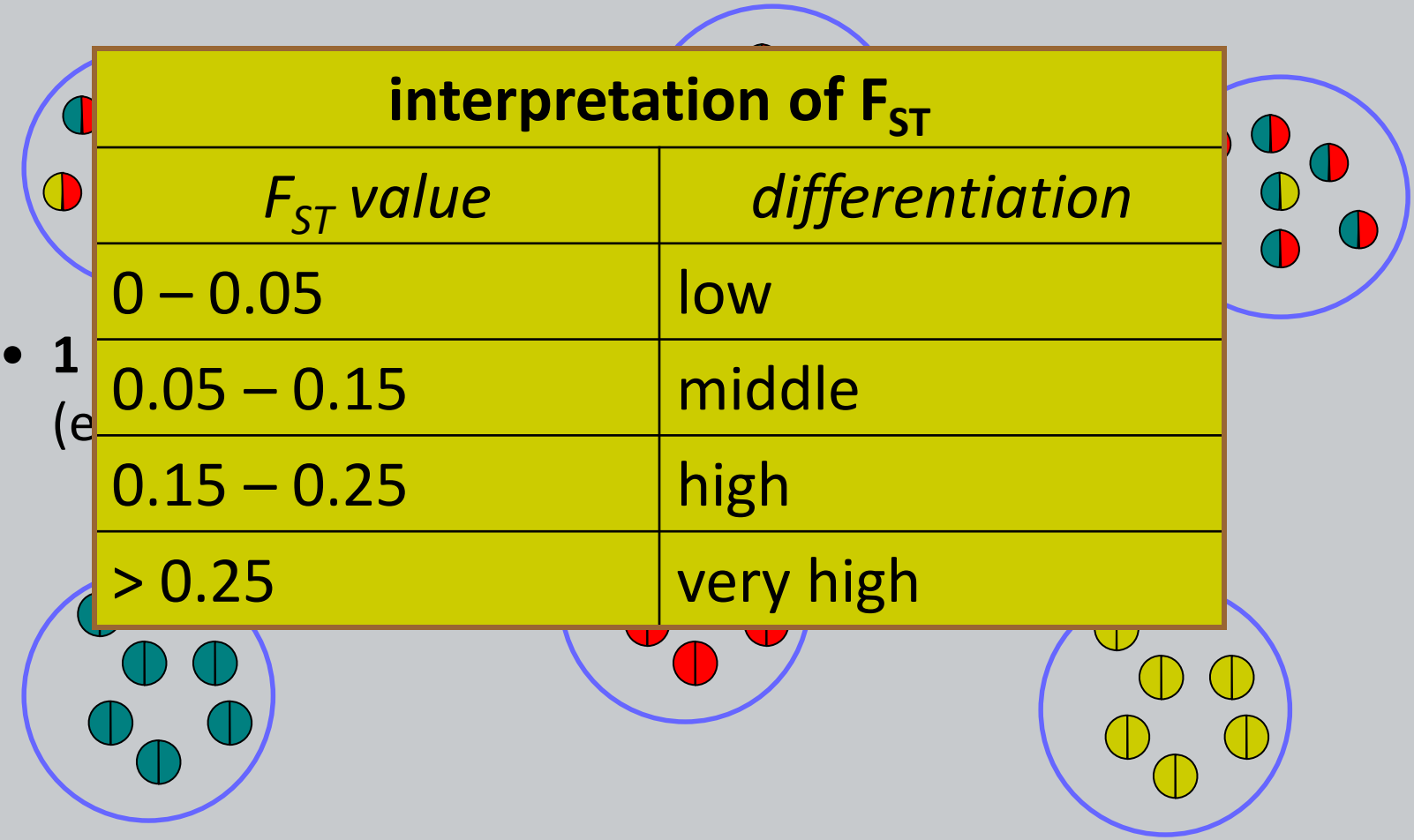


- **1** – maximum genetic population structure
  (each population fixed for different allele)



F_{ST} – subpopulation differentiation

- **0** – no genetic population structure
  (same allele frequencies in all populations)



- **1**
  (e

interpretation of F_{ST}	
F_{ST} value	*differentiation*
0 – 0.05	low
0.05 – 0.15	middle
0.15 – 0.25	high
> 0.25	very high

Example of F-statistics calculation

	Genotype							
	AA	**Aa**	**aa**	**N**	**p**	**Ho**	**He**	**F**
population 1	125	250	125	500	0.5	0.5	0.5	0
population 2	50	30	20	100	0.65	0.3	0.46	0.341
population 3	100	500	400	1000	0.35	0.5	0.46	-0.099

- allele frequency
  - p(A) = (2*AA + Aa) / (2*N)
  - p_1(A) = (2*125 + 250) / 1000 = *0.5*
  - q(a) = 1-p
- **H_o** – *observed heterozygosity*
  - proportion of heterozygotes, i.e., H_o = Aa / N
  - H_{o1} = 250 / 500 = *0.5*
- **H_e** – *expected heterozygosity*
  - 2pq
  - H_{e1} = 2*0.5*0.5 = *0.5*
- **F** – inbreeding coefficient in population
  - F = (H_e - H_o) / H_e

Example of F-statistics calculation II

	Genotype							
	AA	**Aa**	**aa**	**N**	**p**	**Ho**	**He**	**F**
population 1	125	250	125	500	0.5	0.5	0.5	0
population 2	50	30	20	100	0.65	0.3	0.46	0.341
population 3	100	500	400	1000	0.35	0.5	0.46	-0.099

- allele frequency across all populations
  - $\bar{p} = (p_1*N_1*2 + p_2*N_2*2 + p_3*N_3*2) / (N_1*2 + N_2*2 + N_3*2) = 0.4156$
- heterozygosity indices
  - **H_I** – observed heterozygosities in populations
    - $H_I = (H_{o1}*N_1 + H_{o2}*N_2 + H_{o3}*N_3) / N_{total} = 0.4875$
  - **H_T** – expected heterozygosities in populations
    - $H_T = (H_{e1}*N_1 + H_{e2}*N_2 + H_{e3}*N_3) / N_{total} = 0.4691$
  - **H_S** – expected heterozygosities across all populations
    - $H_S = 2*\bar{p}*\bar{q} = 0.4858$

Example of F-statistics calculation III

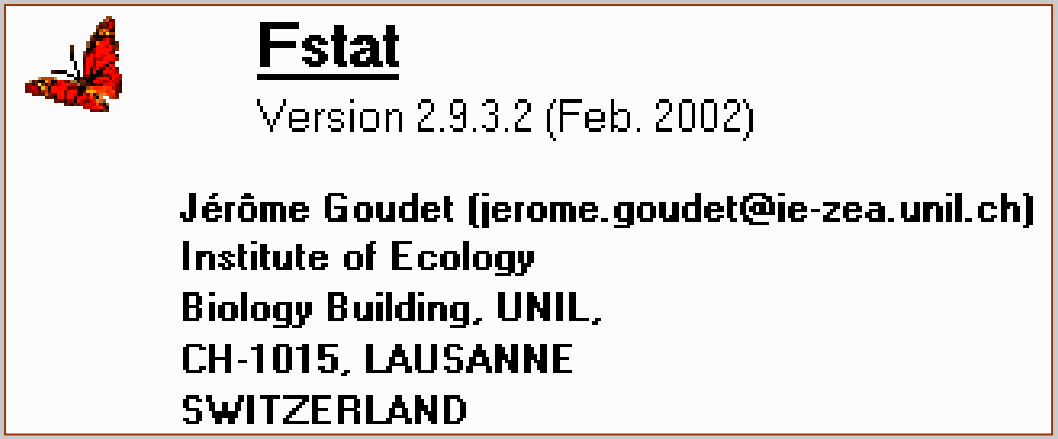
	Genotype							
	AA	**Aa**	**aa**	**N**	**p**	**Ho**	**He**	**F**
population 1	125	250	125	500	0.5	0.5	0.5	0
population 2	50	30	20	100	0.65	0.3	0.46	0.341
population 3	100	500	400	1000	0.35	0.5	0.46	-0.099

fixation indices **level of**

- $\mathbf{F_{IS}} = (H_S - H_I) / H_S = -0.0393$ inbreeding in populations
- $\mathbf{F_{ST}} = (H_T - H_S) / H_T = 0.0344$ subpopulation differentiation
- $\mathbf{F_{IT}} = (H_T - H_I) / H_T = -0.0036$ total inbreeding

Software for isozyme analysis

FSTAT



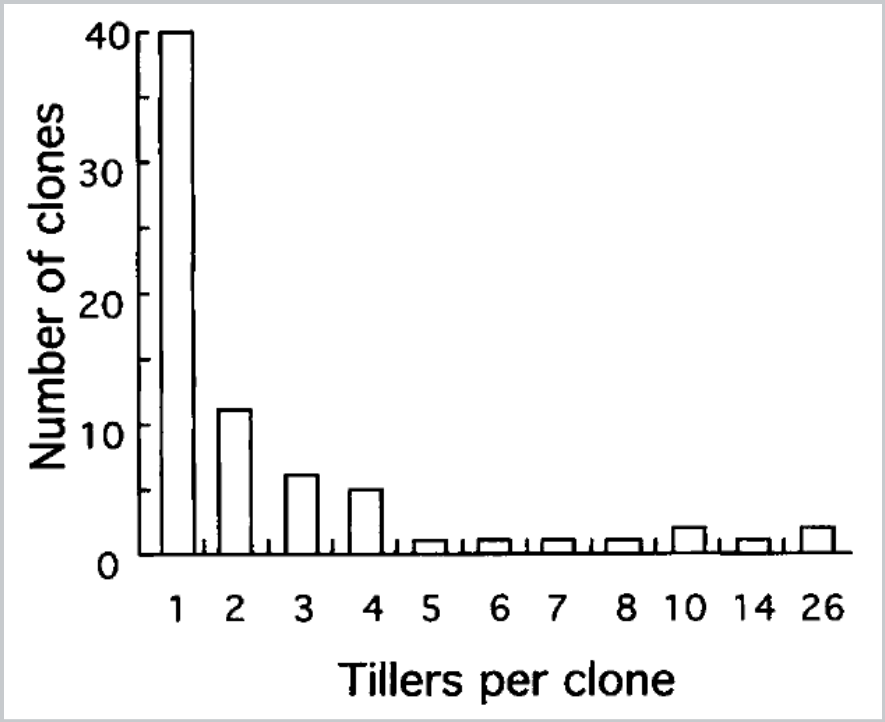
http://www2.unil.ch/izea/softwares/fstat.html

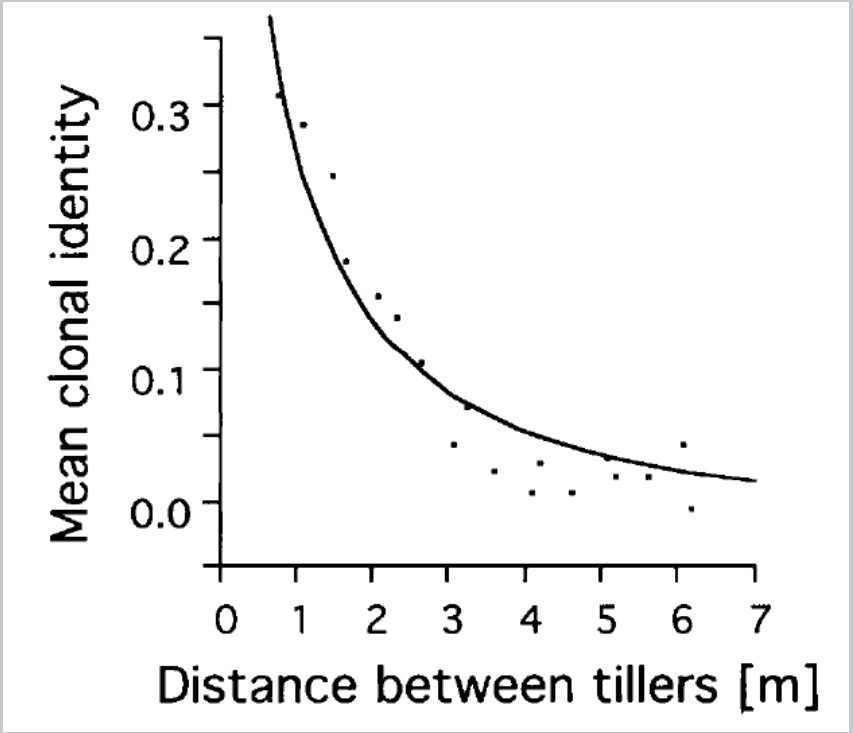
- allele frequency, heterozygosity
- F-statistics (Nei, Weir & Cockerham)
- H-W equilibrium testing

Application of isozyme analysis

- clone identification
  - comparision of zymograme pattern
  - limited variation – use variable DNA markers instead
- population level – population genetics …
- geographical variation
- hybrid identification, introgression
- phylogenetic relationships
  - at the level of closely related species
- evolutionary rate – molecular clock

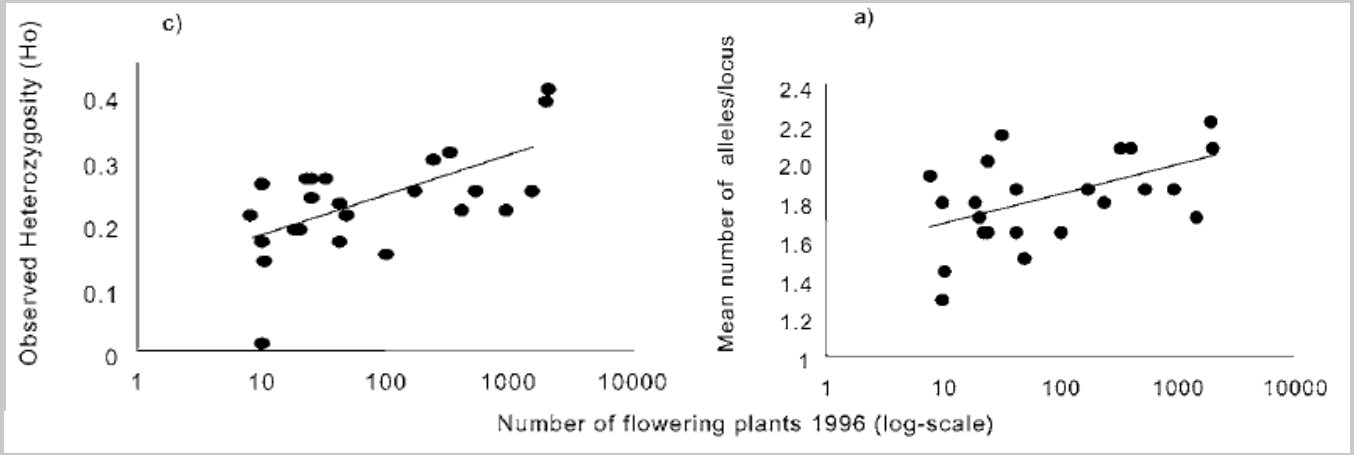
Clonal diversity

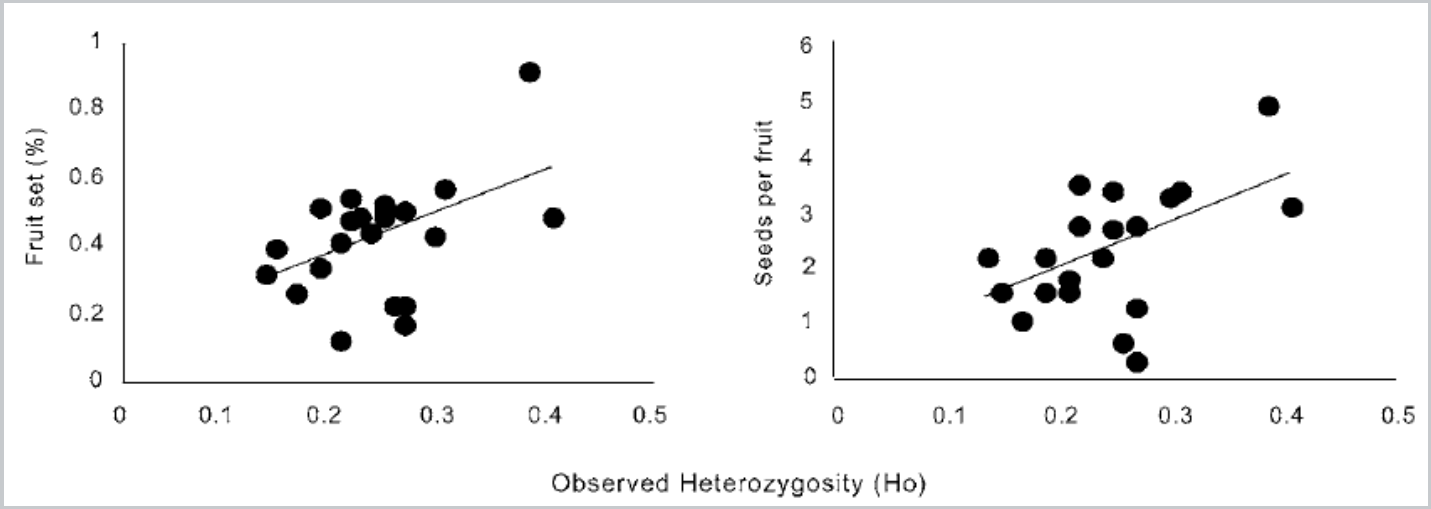




Brachypodium pinnatum, Schläpfer & Fischer 1998

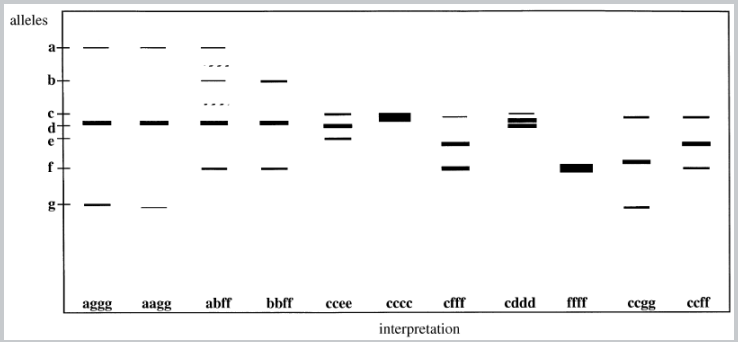
Population size, *H*, fitness

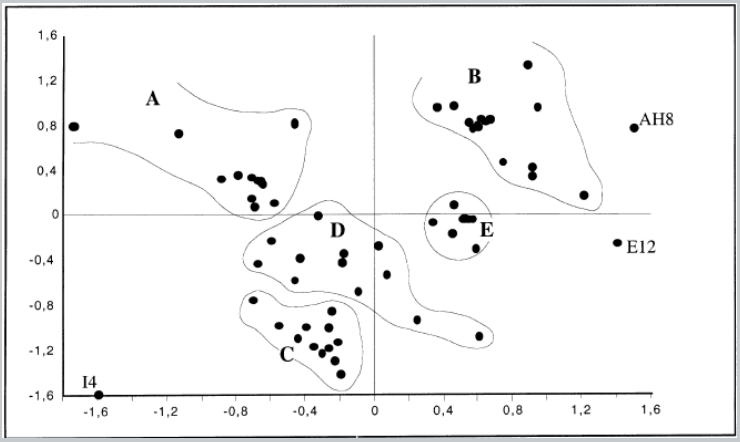


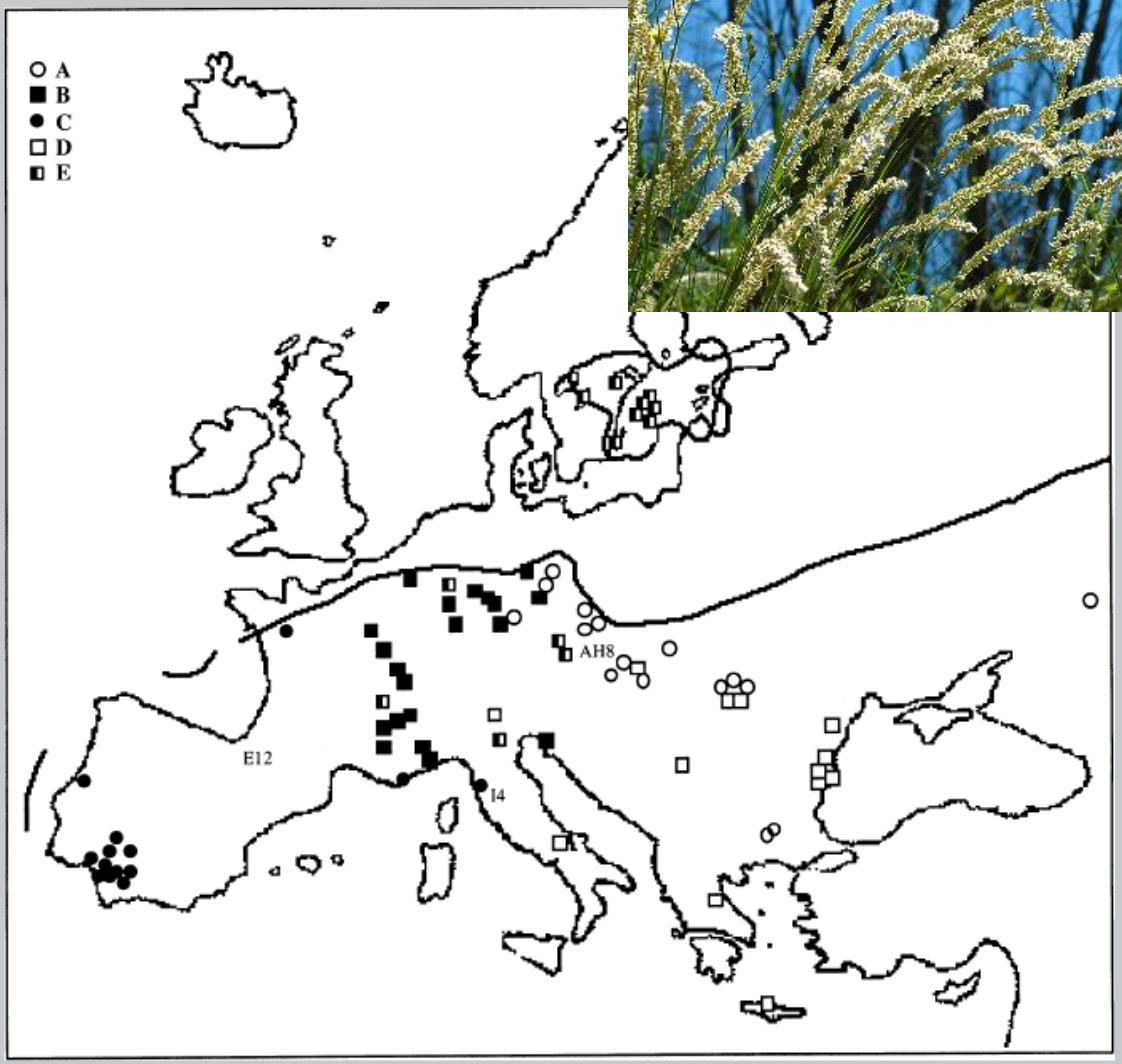


Cochlearia bavarica, Paschke et al. 2002

Geographical variation

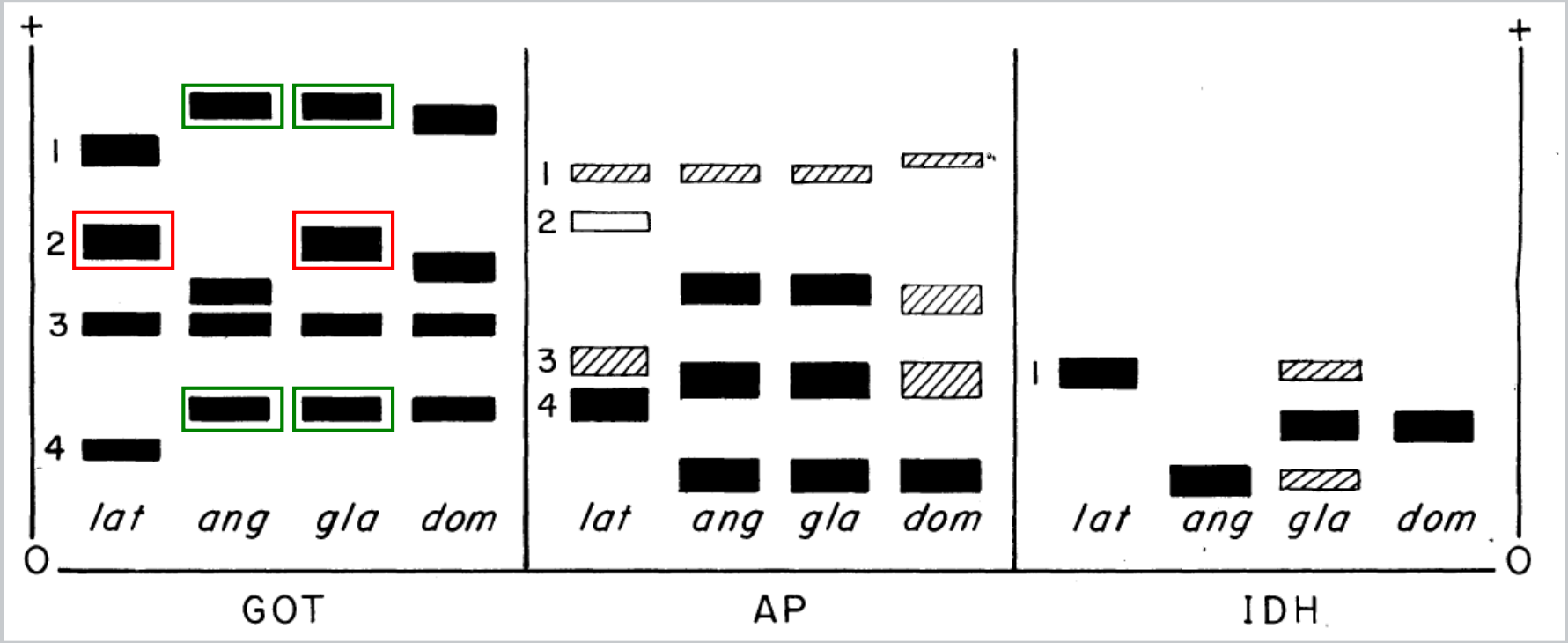






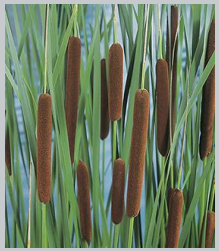
Melica ciliata, Tyler 2004

Hybridization



latifolia

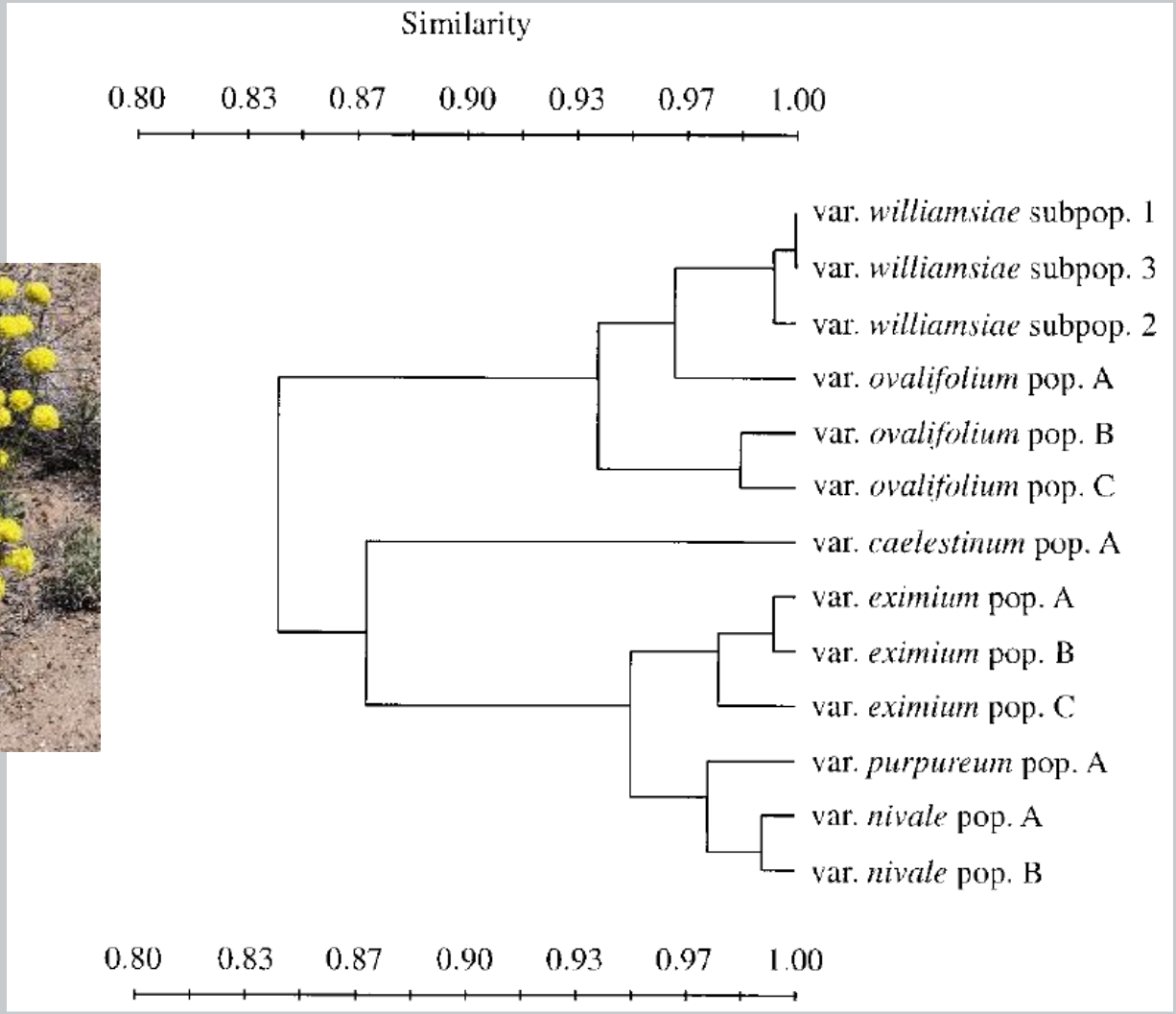
glauca



angustifolia

Typha, Sharitz et al. 1980

Relationships within a species



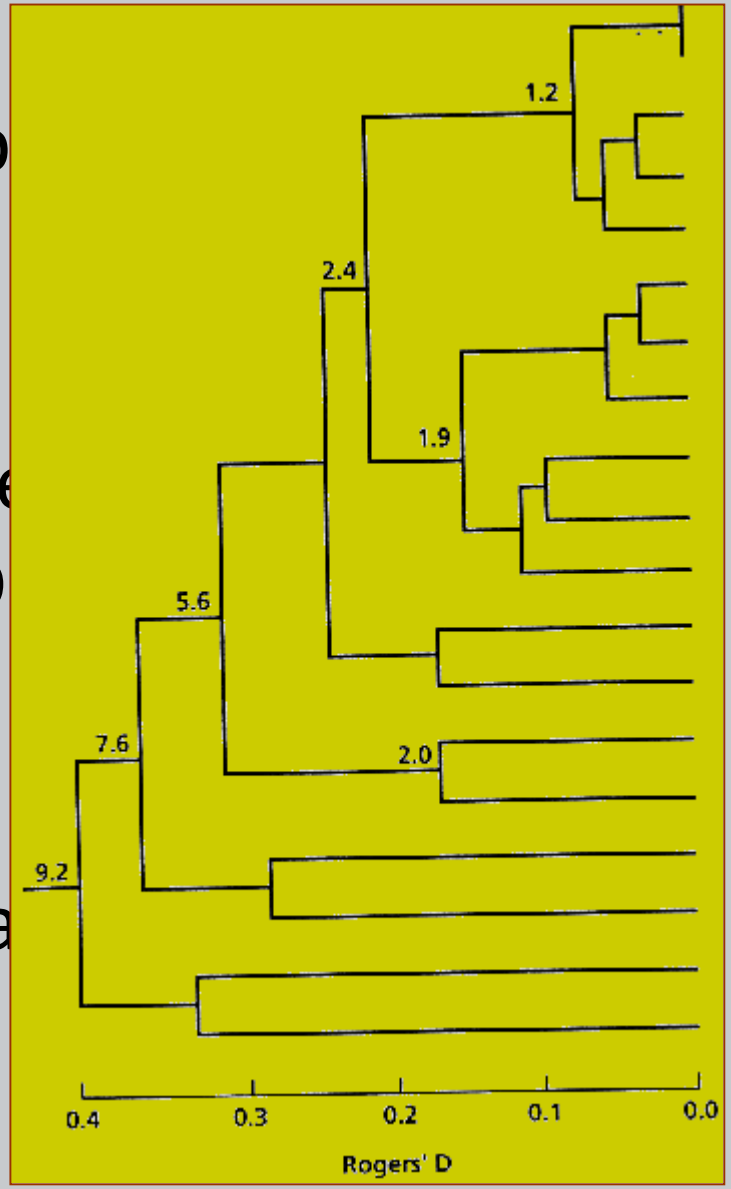
Eriogonum ovalifolium, Archibald et al. 2001

Evolutionary rate – molecular clock

- constant mutation rate expected
  - 10^{-7} /locus*year
  - might be very variable
- relationship between genetic distance (*D*) and divergence time (*t*) – $D=2\alpha t$
  - α – substitution rate
  - **$t = 5*10^6 D$**
- rough estimate, closely related species only

Evolutionary rate – molecular clock

- constant mutation rate exp
  - 10^{-7} /locus*year
  - might be very variable
- relationship between gene and divergence time (t) – D
  - α – substitution rate
  - **$t = 5*10^6 D$**
- rough estimate, closely rela



Species traits and allozyme diversity

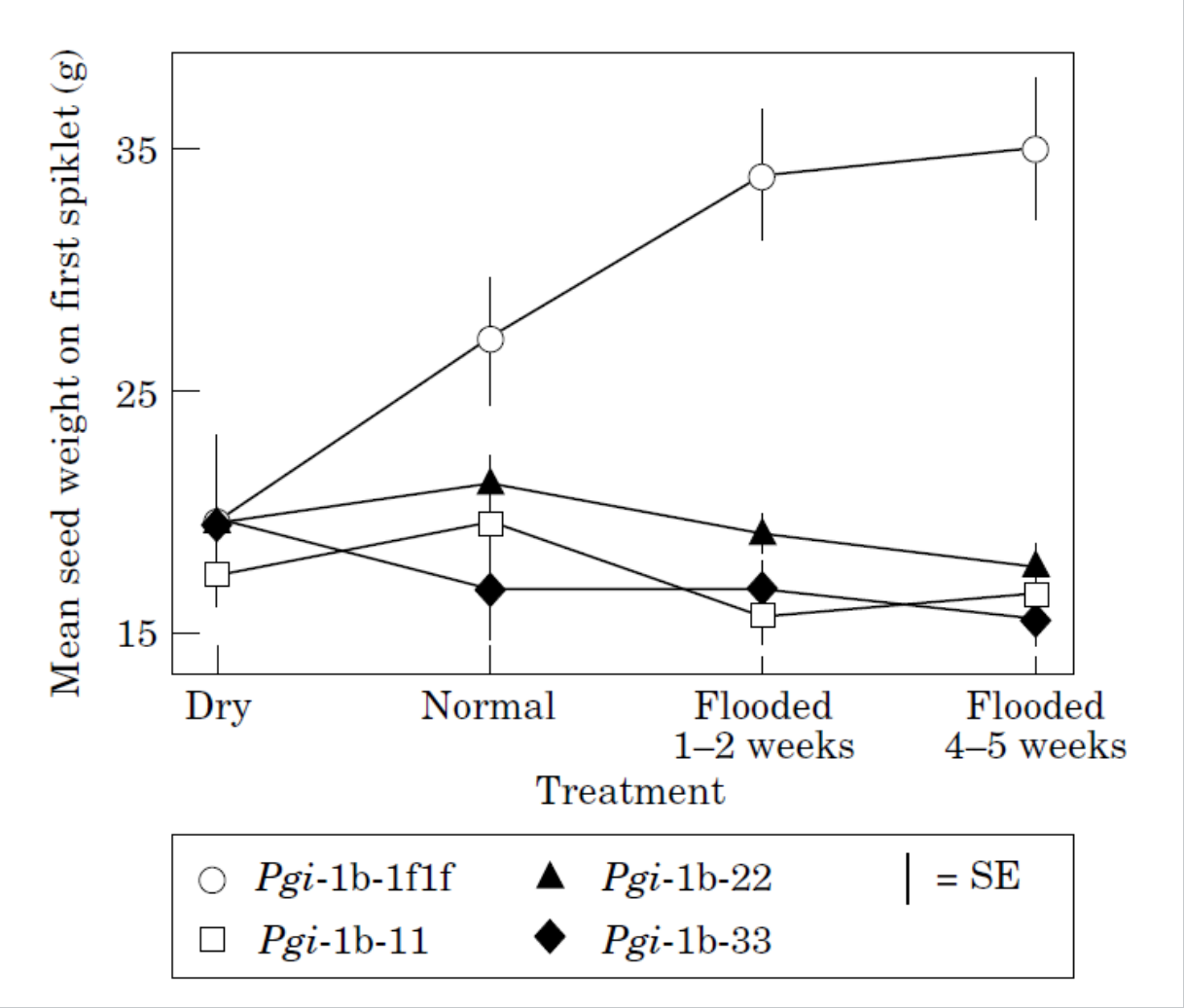
		proportion of genetic variation			
		within populations		among populations	
characteristics		*low*	*high*	*low*	*high*
life form	prennials, trees		•	•	
	annuals	•			•
area	large		•	n.s.	
	endemit	•			
reproduction system	allogamy, anemogamy		•	•	
	autogamy	•			•
seed dispersal	zoochory, anemochory		•	•	
	explosive	•			•

(Hamrick & Godt 1989: review from 449 species and 165 genera)

Selective neutrality of isozymes

- selective neutrality
  - *neutral alleles* – maintained by equilibrium between mutation (origin of new) and genetic drift (extinction)
  - i.e., isoenzymes are functionally equal – no allele has a selective advantage
- true for broad spectrum of species
- BUT: some loci could be correlated with fitness
  - allele frequency changed along an ecological gradient, e.g., elevation

Selective neutrality of isozymes ?



Bromus hordeaceus, Lönn et al. 1998

Population study

Tomimatsu H. & Ohara M. (2003): Genetic diversity and local population structure of fragmented populations of *Trillium camschatcense* (Trilliaceae). *Biological Conservation* 109: 249–258.

Systematic study

Ritland K., Meagher L.D., Edwards D.G.W. & El-Kassaby Y.A. (2005): Isozyme variation and the conservation genetics of Garry oak. *Canadian Journal of Botany* 83: 1478-1487

Literature

Soltis & Soltis [eds.] (1989): *Isozymes in plant biology*.

Baker A.J. (2000): *Molecular methods in ecology*. pp. 65-88

Murphy R.W. et al. (1996): *Proteins: Isozyme electrophoresis*. In: Hillis D.M. et al. [eds.]: Molecular systematics.

Karp A. et al. (1998): *Molecular tools for screening biodiversity*. pp. 73-81

Avise J.C. (2004): Molecular markers, natural history and evolution. pp. 57-63.

Hamrick, Godt, Murawski & Loveless (1991): *Correlations between species traits and allozyme diversity: Implications for conservation biology*. pp. 75-86. In Falk & Holsinger [eds.] Genetics and Conservation of Rare Plants

Hartl & Clark (2007): *Principles of Population Genetics*. 4th edition.

Hamilton M.B. (2009): Population genetics.

Allendorf & Luikart (2006): Conservation and the genetics of populations. pp. 47-62.

Internet resources

enzyme database: http://www.brenda-enzymes.info/

methodology, gel evaluation:

http://www.plantbiology.siu.edu/PLB479/IsozymeTechniques/GelExercise.html