Genetická diverzita

volně žijících druhů

Dr. Martina Žurovcová, Ph.D.

"Nothing in biology makes sense except in the light of evolution."

Theodosius Dobzhansky (1970)

"Very little in evolution make sense except in the light of genetics."

Lots of people

Poděkování/ Acknowlegments

Kromě vlastních výsledků jsou v přednášce použity práce, materiály a informace těchto kolegů:

Christian Stauffer (výzkum Ips typographus; Institute of Forest Entomology, Forest Pathology and Forest Protection, BOKU-University of Natural Resources and Applied Life Sciences,Vienna, Austria)

Gill McVean (přednášky z populační a evoluční genetiky, University of Oxford, Department of Statistics, UK) http://www.stats.ox.ac.uk/~mcvean/

Douglas Futuyma (učebnice evoluční biologie "Evolutionary Biology", 3rd edition, Sinauer, 1998)

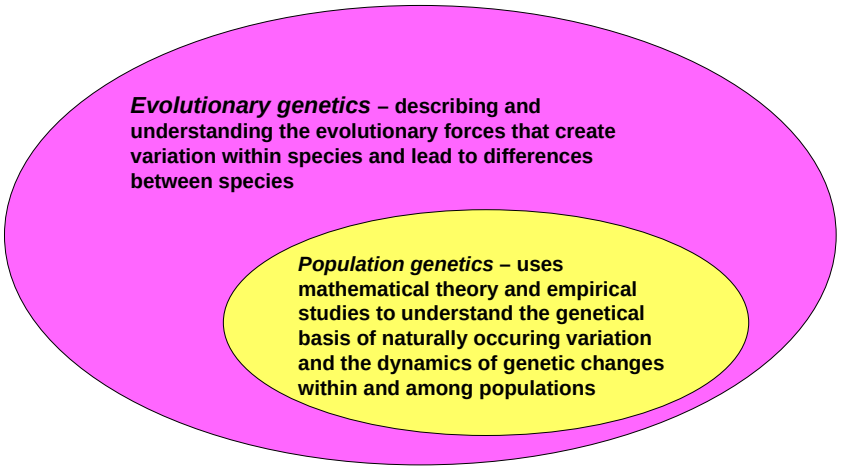
Jana Ždychová (bakalářská práce, Biologická fakulta JČU v Českých Budějovicích)
Lucie Kučerová (bakalářská práce, Biologická fakulta JČU v Českých Budějovicích)

Všem těmto kolegům děkuji za ochotu sdílet i nepublikovaná data.

Martina Žurovcová martina@entu.cas.cz

Evolution is a two-step process:

1) origin of the variability among individuals in populations
2) change in this variability from generation to generation (for this transfer, heritability is essential)



Studium vnitrodruhové variability lýkožrouta smrkového (*Ips typographus, L.*)

Lýkožrout smrkový jako škůdce lesních porostů

Šumava, Modrava

Bavorský les

Lýkožrout smrkový a jeho požerek

Studium vnitrodruhové variability lýkožrouta smrkového (*Ips typographus, L.*)

Hlavní cíle :

- druh *I. typographus* tvoří jedna velká populace anebo malé lokální populace ?

- můžeme od sebe lokální populace odlišit a odhadnout velikost migrace mezi nimi ?

Hledání vhodných markerů

alozymy - nedostatek heterozygotů a narušení HWE na lokální úrovni, rozlišení Skandinávie a střední Evropy, velká panmiktická populace

mtDNA - malá variabilita, podařilo se objasnit kolonizační cesty *I. typographus* po poslední době ledové, fylogenetika rodu *Ips*

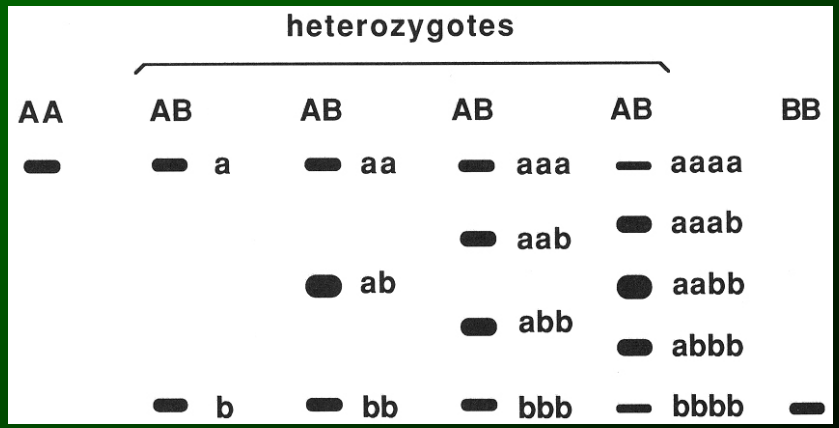
RAPD - variabilita příliš vysoká, menší stupeň migrace, fylogenetika rodu Ips

mikrosatelity - nepodařilo se izolovat dostatečný počet

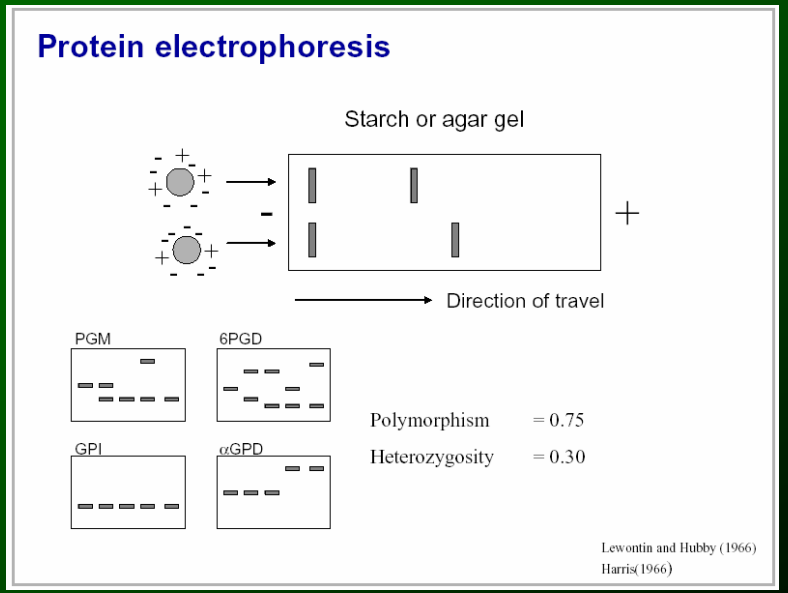
nDNA - isolace genů pro G6PD a Amy

Alozymová analýza

Alozym - enzym (protein) kódovaný jedním lokusem, který se vyskytuje ve více dědičných formách, které nazýváme alely. Tyto alely se od sebe liší svou primární strukturou a vykazují odlišnou pohyblivost na elektroforéze.



Protein electrophoresis



Bakalářská práce

Stanovení úrovně vnitrodruhové variability lýkožrouta smrkového (*Ips typographus, L.*) na základě alozymů

Autorka:

Lucie Kučerová

Cíl práce

- Vytypovat a experimentálně odzkoušet maximální počet polymorfních alozymových markerů.

- Prozkoumat genetickou stukturu populací lýkožrouta smrkového na nižší zeměpisné úrovni (na území ČR) metodou alozymové elektroforézy.

!!! Vyvarovat se předchozích chyb !!!!
- odběr vzorků přímo z napadených stromů
- správné rozlišení alel

Metodika

- Vertikální polyakrylamidová diskontinuální elektroforéza (PAG) nativních proteinů
- Specifické substrátové barvení na detekci enzymů
- Nespecifické barvení (CoomassieBlue) na detekci obecných proteinů

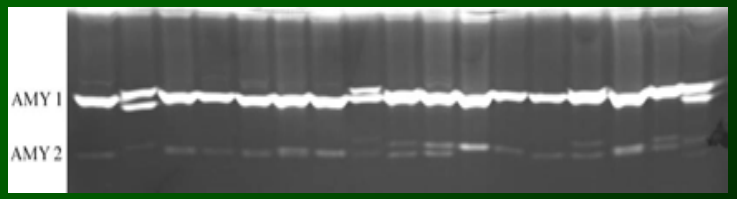
Vyzkoušené alozymové markery

Alozymy s žádnou nebo velice slabou aktivitou:
- Kataláza EC 1.11.1.6
- Glukokináza EC 2.7.1.2
- Hexokináza EC 2.7.1.1
- Izocitrát-dehydrogenáza EC 1.1.1.42
- Leucin-aminopeptidáza EC 3.4.11.1
- Laktát-dehydrogenáza EC 1.1.1.27
- Mannosa-fosfátizomeráza EC 5.3.1.8
- Fosfoglycerátkináza EC 5.4.2.2
- Fosforyláza EC 2.4.1.1

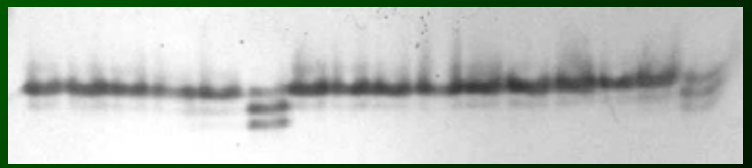
Úspěšně detekované alozymy :
- α-Amyláza EC 3.2.1.1 (2 lokusy)
- Glukóza-6-fosfátdehydrogenáza EC 1.1.1.49 (1 lokus)
- Malátdehydrogenáza EC 1.1.1.37 (1 lokus)
- Enzym kyseliny jablečné EC 1.1.1.40 (1 lokus)
- Fosfoglukoizomeráza EC 5.3.1.9 (1 lokus)
- Fosfoglukomutáza EC 5.4.2.2 (1 lokus)
- Proteiny bez enzymatické aktivity (PT) (3 lokusy)

Profily vybraných markerů

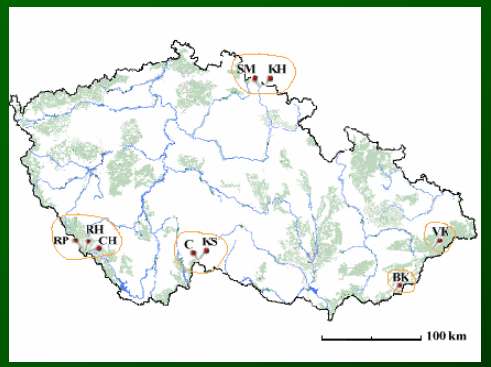
α-Amyláza 1 a 2 (monomery, 2 lokusy, Amy1 - 7 alel, Amy2 - 8 alel)



Malátdehydrogenáza (dimer, 1 lokus, 5 alel)



Studované lokality



Vzájemná vzdálenost : 2,65 - 356 km

Nadmořská výška : 521 - 1390 m n.m.

Čechy :

Krkonoše
SM - Špindlerův Mlýn
KH - Kozí Hřbety

Šumava -Modrava
RP - Roklanský průsek
RH - U cesty k Roklanské hájovně
CH - Černá hora

Jindřichohradecko
C - Číměř
KS - Kunějovské samoty

Morava :

Beskydy
VK - Velké Karlovice

Bílé Karpaty
BK - Vápenky

Hardy-Weinbergova rovnováha a testy homogenity

Na úrovni lokalit :

Kozí hřbety (Krkonoše) - nerovnováha a nedostatek heterozygotů
(F_i MDH = 0,231; F_i PT1 = 0,659)

Ostatní lokality - v rovnováze

Na úrovni oblastí :

- test homogenity zamítá příslušnost k jedné panmiktické populaci téměř ve všech oblastech, regionech i v celkové populaci

Vysvětlení

- Šumava - vysoké početní stavy lýkožrouta
- existence přirozených bariér (borové lesy na Jindřichohradecku, rozdíl v klimatu v Krkonoších)
- struktura krajiny
- možný vliv lokální selekce

Celkové charakteritiky variability

Průměrný počet alel na lokus A = 4,20
Podíl polymorfních lokusů (s 95% kritériem polymorfismu) P = 0,3
(s 99% kritériem polymorfismu) P = 0,8
Heterozygosita (observed heterozygosity) H_o = 0,108
Genetická diverzita (expected heterozygosity) H_e = 0,111

Hodnota v práci Pavlíčka et al. (1997) H_e = 0,184

Míra genetické diferenciace : F_{ST} = 0,0185

Genetický tok : Nm = 13,3 (9,57)

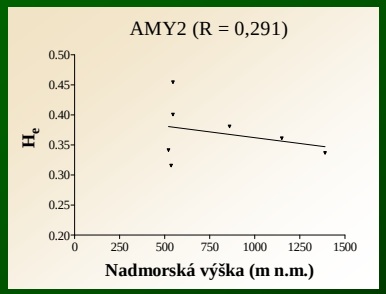
Počet detekovaných privátních alel = 4

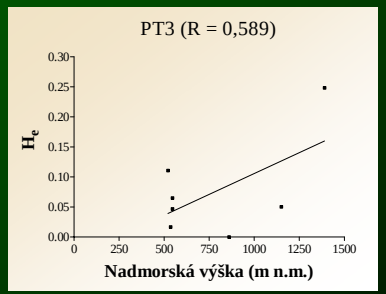
Závěr

Pokles heterozygosity během časového období 9 let způsobený fluktuacemi početnosti.

Klesá genetická diverzita s nadmořskou výškou?

- testování závislosti genetické diverzity na nadmořské výšce pomocí lineární regrese

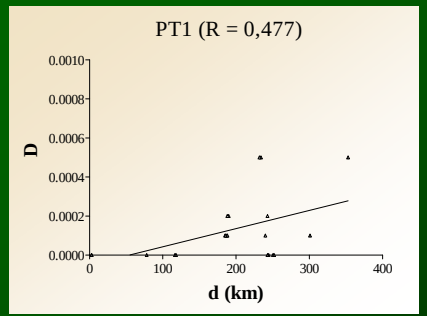




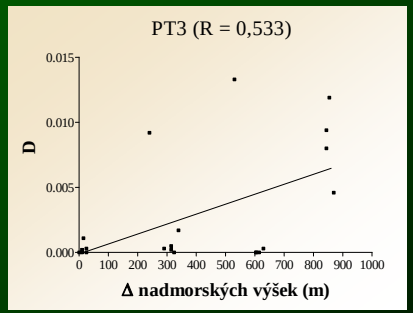
Závěr: Různé alozymové lokusy se chovají různým způsobem a celková diverzita nevykazuje žádný trend.

Genetická vzdálenost

Vliv geografické vzdálenosti



Vliv nadmořské výšky



Závěr

závislost genetické vzdálenosti na vzdálenosti geografické a na nadmořské výšce pomocí lineární regrese je průkazná jen v ojedinělých případech = možný "founder effect"

Závěr

- Byl zaznamenán pokles genetické diverzity, který můžeme vysvětlit fluktuacemi v početnostech lýkožrouta smrkového.
- Populace *Ips typographus* žijící na území České republiky je podrozdělena do menších subpopulací vlivem lokálních podmínek.
- Velký vliv na rozdělení do subpopulací má také fáze populačního cyklu (přemnožení, ústup)
- Shlukovací analýza od sebe odlišila různá místa původního rozšíření a oblasti, kam byl *Ips typographus* druhotně introdukován.

Budoucnost ????

The null model in evolutionary genetics

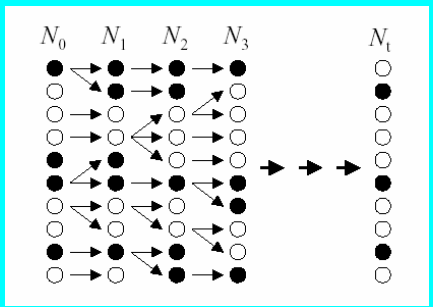
Let us suppose

Nothing interesting ever happens in biology.

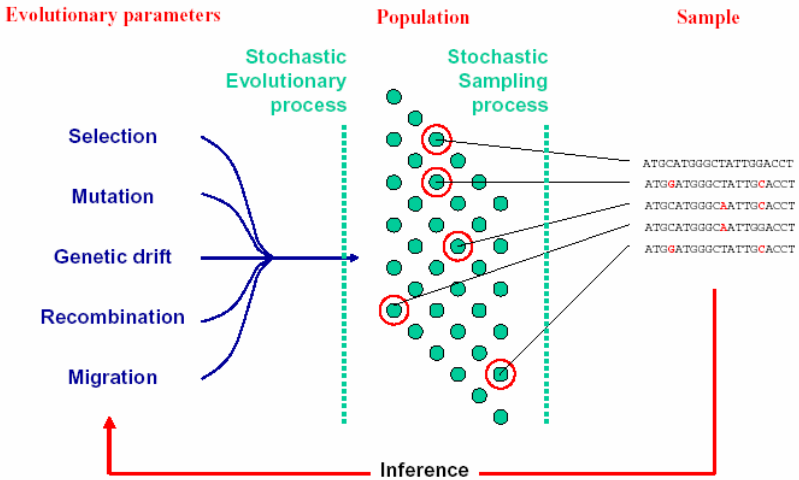
The Wright-Fisher population model

Assumptions

- **sexual reproduction without selfing**
- **random mating with respect to genotype**
- **non-overlapping generations**
- **constant diploid size of N (2N alleles)**
- **no migration or selection**
- **mutations occur at at constant rate**



Population genetic inference



Polymorphism

-at the gene level, a locus is said to be polymorphic if the most common allele has a frequency of less than 0.99

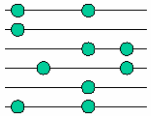
-in DNA sekvence, every nucleotide site can be considered to be an allele:

polymorphic site = existence of more than one nucleotide at this position within population (species)

divergence = means site which is monomorphic (fixed) within species but different between species.

Species - line	
D. melanogaster - USA	T T A T T T G G C
Africa	T T A G T T G G C
D. simulans - USA	G T A T C T T G C
Africa	G T A G T T T G C
D. yakuba - 1	G T G G T T G G C
2	G T G G T T G G C
Poly/Div	d d p p d

Statistics of polymorphism



Number of Segregating sites $S = 4$

Average pairwise differences $\pi = 1.9$

Seq	2	3	4	5	6
1	1	2	4	1	0
2		3	3	2	1
3			2	1	2
4				3	4
5					1

Number of distinct haplotypes $H = 5$

Patterns of variation at the DNA level

- Synonymous & nonsynonymous mutations

```
Arg Gln Val          Arg Gln Val
AGA CAA GTA          AGA CAA GTA
    A                    A
    ↓                    ↓
CAG CGA GTA          AGA CAG GTA
Arg Arg Val          Arg Gln Val
```

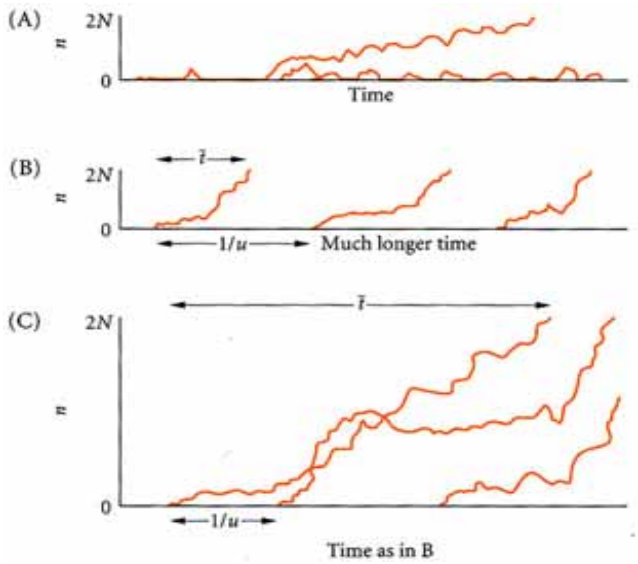
e.g. *D. simulans*

π_{total} = 0.010 per site
π_{silent} = 0.038
$\pi_{noncoding}$ = 0.023

- Nucleotide variation v. protein variation?

	Humans	*D. melanogaster*
Allozyme	6%	14%
Nucleotide	0.1%	1%

Fate of the new mutation in a population (Motoo Kimura – Neutral theory)



Features of the neutral theory

- The majority of changes in proteins and at the level DNA which are fixed between species, or segregate within species, are of no selective importance
- The rate of substitution is equal to the rate of neutral mutation

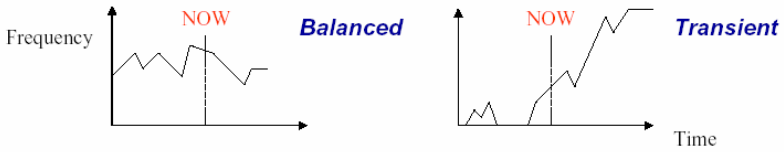
$$k = f_{neutral}\ \mu$$

- The level of polymorphism in a population is a function of the effective population size and the neutral mutation rate

$$\pi = 4N_e\mu$$

|s|< 1/(4N_e)

- Polymorphisms are transient rather than balanced

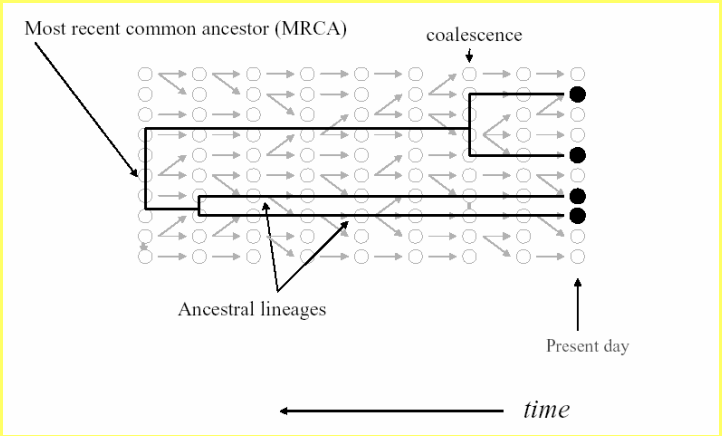


The coalescent theory

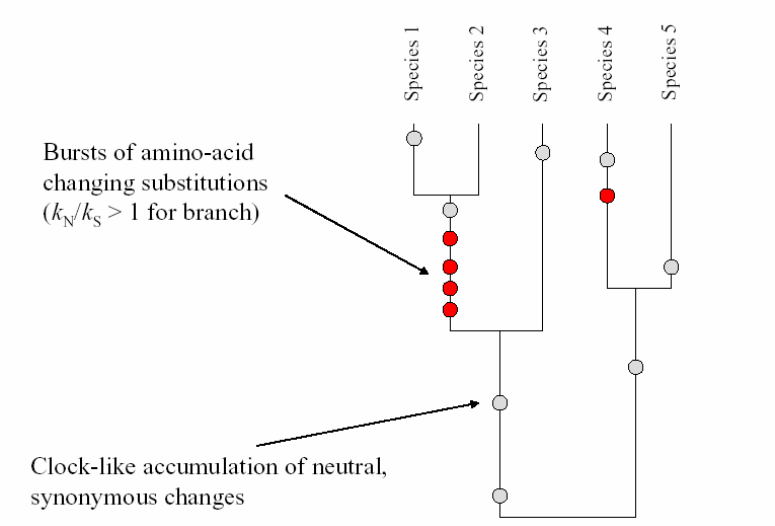
(J.F.C. Kingman)

- describes the genealogical relationships among individuals in a Wright-Fisher population

- all information about underlying evolutionary processes is in the underlying gene genealogy



Phylogenetic method

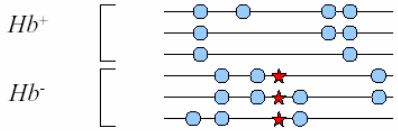


Selection: what are we looking for?

- What types of selection might we consider?
  - Directional selection (selective sweeps)
  - Balancing selection
  - Local adaptation (local selective sweeps)
- There are many ways of summarising data, how do we know which aspects will be most sensitive to the action of selection?
  - Models of selection
  - Formulation of test statistic
- How can we be sure that a deviation from the assumed model is due to selection?
  - Comparison to reference loci
  - Comparison across multiple populations
  - *a priori* knowledge

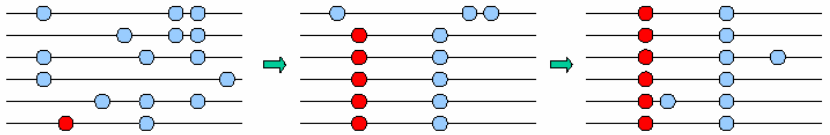
Balancing selection

- Alleles that are maintained at a constant frequency by natural selection are called balanced mutations
  - Mutations causing malaria-resistant haemoglobinopathies
  - Alcohol dehydrogenase in *Drosophila melanogaster*



- Neutral mutations that occur at linked sites on one background can only cross onto the other by recombination
- Balancing selection leads to an increase in local diversity and strong haplotype structure
  - Like structured populations

The hitch-hiking effect of beneficial mutations



A new advantageous mutation appears in the population

The mutation sweeps to high frequency, dragging linked mutations, and reducing variability in the region

Diversity recovers slowly; the first mutations to appear are at low frequency

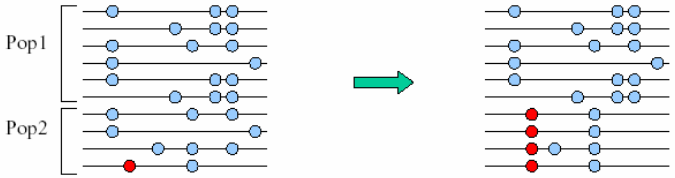
Standard neutral model applies

Strong linkage disequilibrium, high frequency derived mutations

Skewed allele frequency spectrum, low diversity

Local adaptation

- Localised selective sweep due to geographically restricted selection pressure
  - e.g. malaria, drug or pesticide treatment

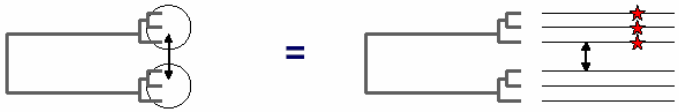


- Leads to local reduction of variability, strong linkage disequilibrium and strong geographical structuring of genetic variability around the locus of importance

Distinguishing selection from demographic effects

- Demographic processes can mimic selection
  - Population growth can look like selective sweeps

  - Population subdivision can look like balanced selection



- Differences between loci can distinguish between genome-wide effects and the local effect of natural selection
  - BUT need to know about variance of demographic processes……

Nucleotide variation in two developmental genes, *Idgf1* and *Idgf3*, of *Drosophila melanogaster*

Martina Zurovcova [1,2], Francisco J. Ayala[2]

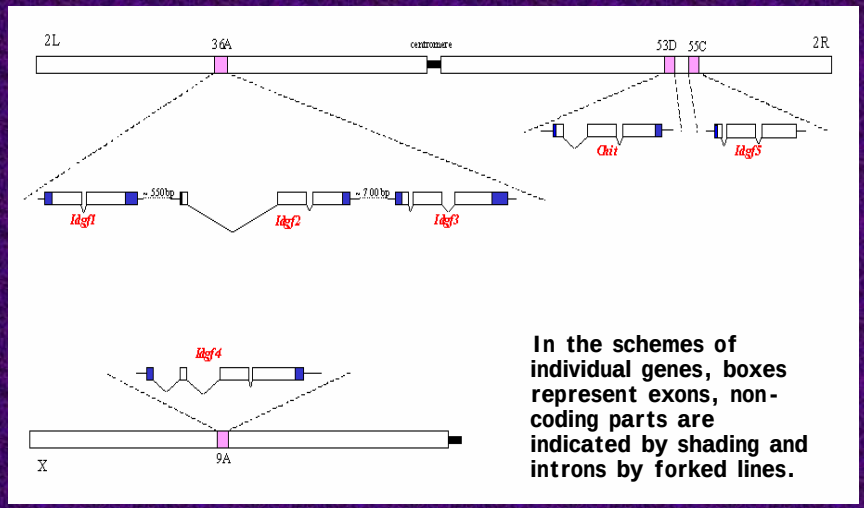
1) Institute of Entomology, Czech Academy of Sciences, Ceske Budejovice

2) Department of Ecology and Evolution, University of California, Irvine

Drosophila melanogaster IDGF genes and their products

- small family of developmental genes identified by Kirkpatrick et al. (1995) and Kawamura et al. (1999)
- *Idgf1-3* are localized in a tight cluster in region 36A, *Chit* is in 53D, *Idgf5* in 55C and *Idgf4* is in 9A
- products of those genes are small secreted chitinase-related proteins without chitinase enzymatic activity
- *Idgf1 and Idgf3*: - only ~4 kbp apart
  - intriguing organisation of introns (*Idgf3* has the first one shorter, *Idgf1* has only one intron)

Chromosomal localization and exon/intron structure of the *Idgf* genes



In the schemes of individual genes, boxes represent exons, non-coding parts are indicated by shading and introns by forked lines.

Methods and materials

➢ Flies :
D.melanogaster - 20 isogenic lines from one local population (Montblanc, Spain)

outgroups (from Drosophila Stock Center, Texas) - *D. simulans*
- *D. yakuba*

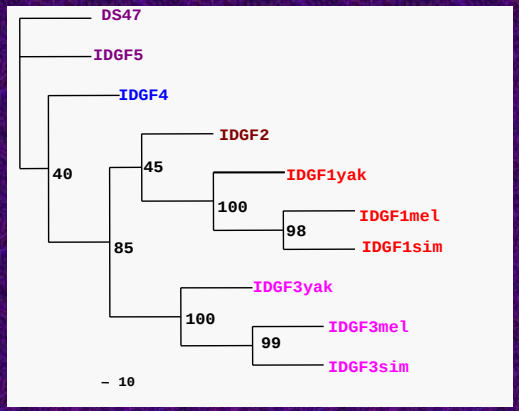
- ➢ PCR amplification of the targets (*Idgf1* and *Idgf3*)
- ➢ direct sequencing of the PCR product
- ➢ alignment of the sequences
- ➢ statistical analysis of the sequences (MEGA, DNAsp)

Alignment of sequences and formatting in Nexus

```
#NEXUS
[File generated by DnaSP Ver. 4.00.5, from file: idgf1.nex     IX 16, 2004]

begin data;
dimensions ntax=22 nchar=2032;
format datatype=DNA interleave gap=- missing=' ;
matrix
[               10        20        30        40        50        60        70        80        90]
[                *         *         *         *         *         *         *         *         *]
line01  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line08  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line13  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line15  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line25  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line29  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line33  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line34  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line36  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line37  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line39  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line40  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line45  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line46  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line47  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line48  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line52  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line58  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line63  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
line80  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
simid1  ATTCGTGAAACTGTATTAGCTGACAGGGCATCACAAACAAACGCAATAAAGTCGGAAAACATTGACAGGAACTACGTAATGCCTGAAAAC
yakid1  ATTCGTGAAACTGAATTAGCTGACAGGGCATCACAAACAAACGAAATAAAGTCGGAAAACATTGACAGAAACTACGTAATGCCTGGAAAC
```

What is the phylogenetic relationship among *Idgfs* ?



Phylogenetic tree of the *Idgf* family based on amino-acid sequences (maximum parsimony method, 10.000 boostrap replications; MEGA software)

Conclusions

- Phylogenetic analysis, codon bias and G+C content of *Idgf*s show that the genes in the tight cluster in 2L are much more similar to each other than to the rest, but this pattern does not hold for the genes in 2R.

- From the exon/intron structure it can be assumed that the ancestral gene had at least two introns and the loss of an intron in *Idgf1* is a recent event.

- Recovery of adequate homologs for at least two genes from this family from *D. simulans* and *D. yakuba* manifests that the origin of this gene family predates the split of these species.

Is there any difference in the amount of polymorphism and divergence between *Idgf1* and *Idgf3* ?

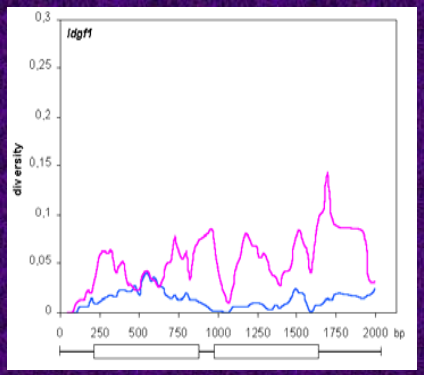
Levels of nucleotide diversity of *Idgf1* and *Idgf3*

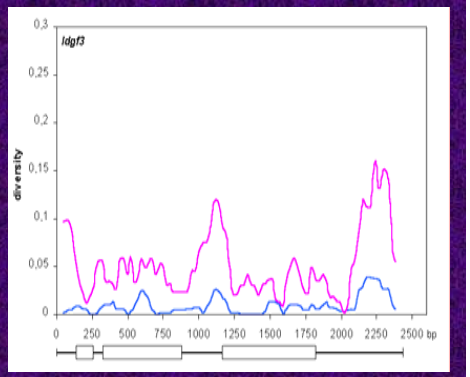
# of lines	*Idgf1* 20	*Idgf3* 20
bp (gaps excluded)	1905	2364
S (segregating sites)	69	64
π_{total} (per site)	0.01286	0.00852
θ_{total} (per site)	0.01021	0.00763
H (number of haplotypes)	17	14
Hd (haplotype diversity)	0.989	0.979
R (recombination rate per bp)	0.02520	0.01324
K_{total} (divergence)	0.04745	0.04762

Distribution of polymorphism and divergence

π — K —

(sliding window profile, window size =100, step =25; gaps excluded; boxes represent coding regions, horizontal lines noncoding regions)

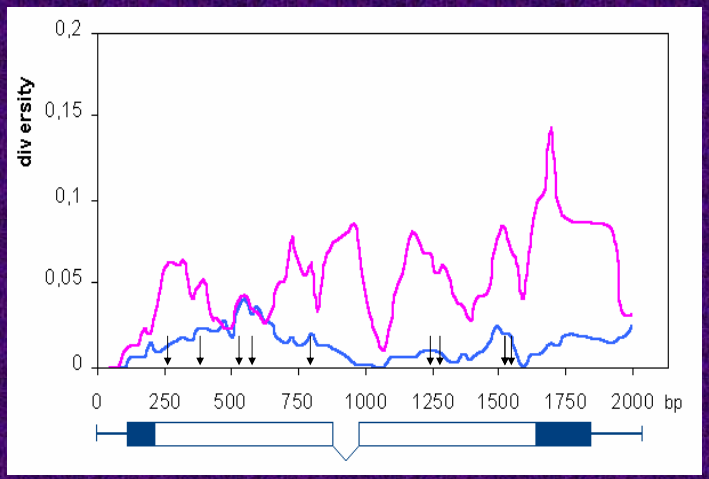




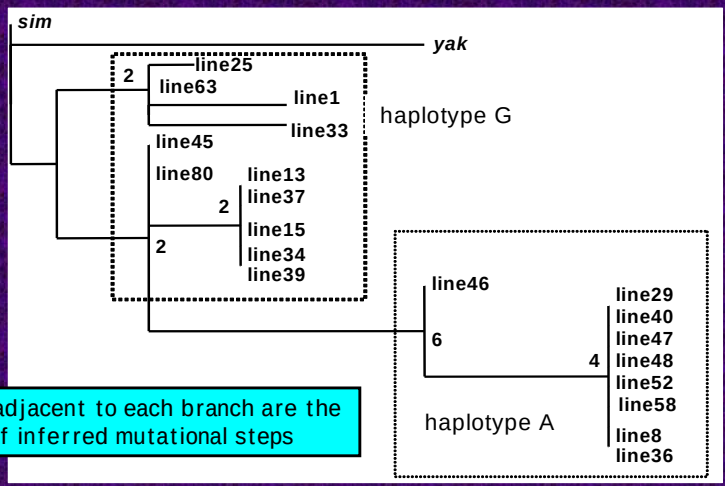
Tests of neutrality

Test	*Idgf1*	*Idgf3*
Ka/Ks	0.0902	0.0841
MK	ns	ns
HKA (silent sites)	*	ns
Tajima's D	1.05697 ns	0.47322 ns
Fu/Li D (with outgroup)	1.66939 *	0.35377 ns
F (with outgroup)	1.78931 *	0.48419 ns
pattern of nucleotide variability	excess of syn. polymorphisms	neutral
mode of evolution	balancing selection	neutral

Positions of the replacement polymorphisms in *Idgf1*



Phylogram of the most parsimonious tree for the "haplotype" region



Numbers adjacent to each branch are the numbers of inferred mutational steps

Conclusions

- *Idgf1* and *Idgf3* have higher levels of polymorphism than the average amount in *D. melanogaster* genes, and the distribution of polymorphisms across the loci is not uniform.

- The neutrality tests indicated that *Idgf3* evolves neutrally, while *Idgf1* is under balancing selection

- As *Idgf1* and *Idgf3* are only ~ 4 kbp apart, it is quite surprising that they are probably under different selective constraints.

Continuing research – what is the origin of the *Idgf* gene family and how are the individual genes shaped by selection ?

- Phylogenetic approach – screening more distant species – finding the original ancestral gene and dating the duplications

- Population genetics approach – screening the patterns of polymorphism and divergency of all *Idgfs* in two distant populations

Martina Žurovcová martina@entu.cas.cz