- Project presentations and papers due:
  Thursday, MARCH 15, 2007,
  1030-1220

Lab today in Genome Sciences
Foege Building, S448.
Meet 1:30 in lobby

South Central Campus Map

Reading

- Today
  - 2 papers on lysozyme evolution
- Tuesday
  - 2 papers on influenza evolution

Convergent Evolution
Two examples

- Lysozyme
- Antifreeze proteins

What is convergent evolution?

Convergence versus parallel evolution

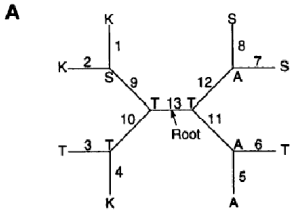


FIG. 1.—Examples of convergent and parallel changes. *A*, Convergent changes, parallel changes, and uniquely shared sites. There are convergent changes on branches 8 and 9 (A→S and T→S, respectively). There are parallel changes on branches 7 and 8 (A→S). When a

Convergent Evolution in foregut fermentation

- Evolved multiple independent times
- Is it the same process every time?

Ruminants

- Stomach with fermentative chamber
- Microbes grow in stomach
- Microbes lysed and digested as pass through the gut
  - Prevents loss of nutrients assimilated by microbes

Stomach environment

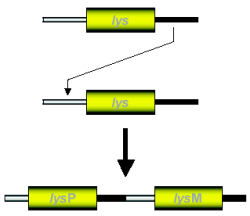
- Low pH
  - Activity of enzymes need to be optimal at low pH
- Presence of trypsin
  - Cleaves at basic residues
- Other reactive products
  - Reactive with Arg residues.

Lysozyme

- Found in virtually all organisms
- Expressed in macrophages, tears, saliva, milk, etc..
- Two major groups arose by gene duplication.
  - Conventional and Calcium binding

Gene duplications in lysozyme

- In ruminants, lysozyme gene has been duplicated ~10 times and is expressed less in extra-intestinal tissues
- Original gene duplication through unequal crossing-over in Alu-like B2 middle repetitive elements



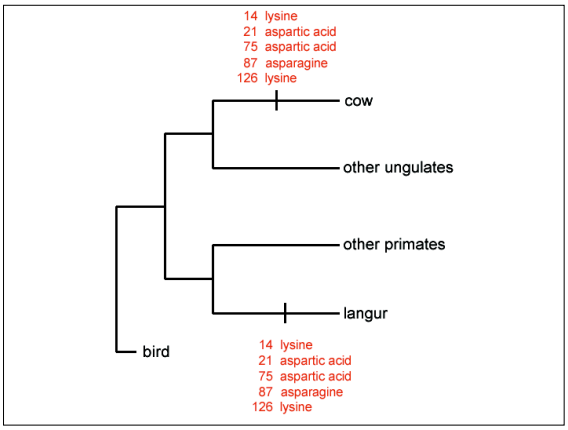
lysozyme
breaks down bacteria cell wall ---> lysis
saliva, blood, tears, milk

ruminants (cows etc), leaf-eating monkeys (langur)
new form of lysozyme
digest stomach bacteria
bacteria digest cellulose

environment of stomach lysozyme is more acidic

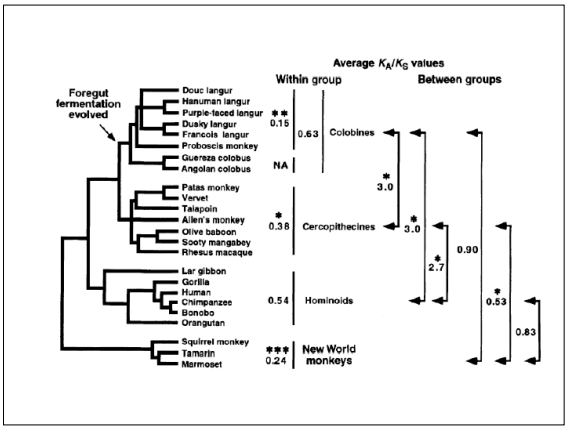
Check for convergence

- Sequence stomach lysozyme from langur and compare to cow.
- Look for residues that are convergent



Has there been adaptive evolution leading to colobine lysozyme

- Sequence from multiple species
- Estimate dn/ds pairwise within and between clades
- Estimate dn/ds among different lineages.



Conclusions from pairwise

- Evidence for positive selection between clades
- Evidence of purifying selection within clades.

How to determine when selection occurred?

- Lineage based test
- Use Maximum Likelihood and Maximum Parsimony to ***estimate*** ancestral sequences.
- Estimate dn/ds ratio between each ancestral sequence to get lineage dn/ds
- Look for adaptive evolution leading to foregut fermentation.

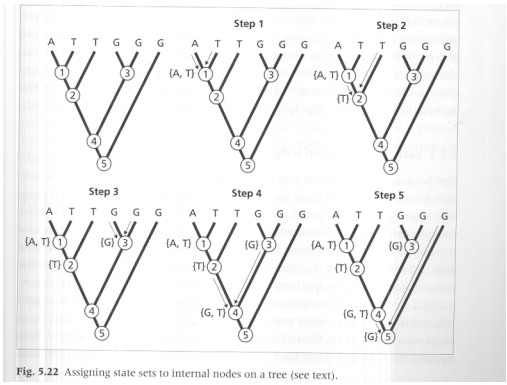


Fig. 5.22 Assigning state sets to internal nodes on a tree (see text).

Likelihood vs parsimony

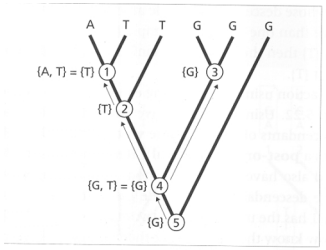
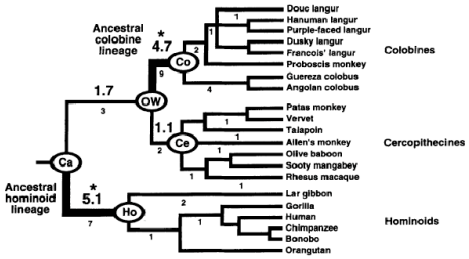
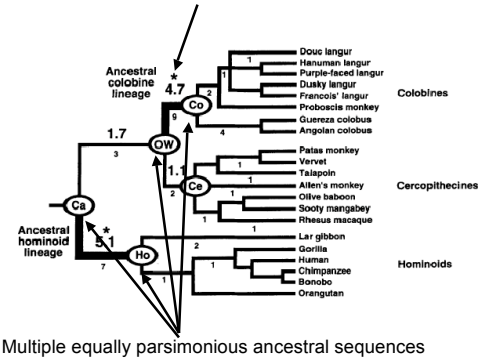


Fig. 5.23 Resolving the ambiguity in ancestral reconstructions by going back up the tree and assigning to ambiguous nodes the intersection of its state set with that of its ancestor.

Did you believe these results?



Range 1.5 - 5.3: is this really > 1?



Multiple equally parsimonious ancestral sequences

If you want to do this type of analysis:

- Ancestral sequences can be reconstructed using PAUP* or codeml
- Estimate dn/ds between all pairwise comparsions using codeml.
- Compare free estimate to estimate with dn/ds fixed at 1.

Lineage based tests

- Showed two lineages with dn/ds > 1
  - Lineage leading to hominoids
    - Selective pressure unknown.
  - Lineage leading to colobine
    - Foregut fermentation?
- Is this significant?

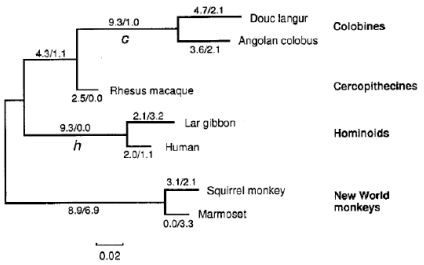
Likelihood based tests

- Incorporate uncertainty in estimating ancestral sequence
  - Average over all possible ancestral sequences weighted to likelihood of occurence
- Model transistion/transversion ratio and codon bias.

Lineage based analyses

- Compare likelihood of nested models
  - "free ratio" to one ratio across all lineages
  - Ratio along Human and coloning versus background
  - Test if dn/ds is significantly > 1.

Using a subset of the species, still see burst of evolution along Colonine (c) and Hominoid (h) lineage



Using likelihood ratio tests - how can we determine if this is Significant variation between lineages?

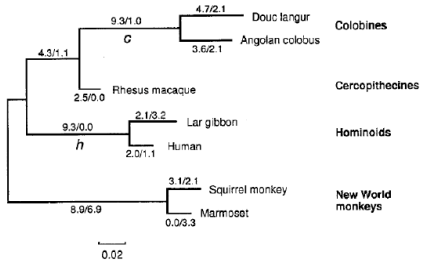


Table 1
Log Likelihood Values and Parameter Estimates Under Different Models

Model	p	ℓ	$\hat{\kappa}$	$\hat{\omega}_0$	$\hat{\omega}_H$	$\hat{\omega}_C$
Large data set (n = 19)						
A. One ratio: $\omega_0 = \omega_H = \omega_C$	35	−1043.84	4.157	0.574	$=\hat{\omega}_0$	$=\hat{\omega}_0$
B. Two ratios: $\omega_0 = \omega_H$, ω_C	36	−1041.70	4.163	0.489	$=\hat{\omega}_0$	3.383
C. Two ratios: $\omega_0 = \omega_C$, ω_H	36	−1039.92	4.186	0.484	∞	$=\hat{\omega}_0$
D. Two ratios: ω_0, $\omega_H = \omega_C$	36	−1037.59	4.199	0.392	7.166	$=\hat{\omega}_H$
E. Three ratios: ω_0, ω_H, ω_C	37	−1037.04	4.196	0.392	∞	3.516
F. Two ratios: $\omega_0 = \omega_H$, $\omega_C = 1$	35	−1042.50	4.074	0.488	$=\hat{\omega}_0$	1
G. Two ratios: $\omega_0 = \omega_C$, $\omega_H = 1$	35	−1042.29	4.058	0.484	1	$=\hat{\omega}_0$
H. Two ratios: ω_0, $\omega_H = \omega_C = 1$	35	−1040.32	3.974	0.392	1	1
I. Three ratios: ω_0, ω_H, $\omega_C = 1$	36	−1037.92	4.101	0.392	∞	1
J. Three ratios: ω_0, $\omega_H = 1$, ω_C	36	−1039.49	4.063	0.392	1	3.448
Small data set (n = 7)						
A. One ratio: $\omega_0 = \omega_H = \omega_C$	13	−906.02	4.540	0.807	$=\hat{\omega}_0$	$=\hat{\omega}_0$
B. Two ratios: $\omega_0 = \omega_H$, ω_C	14	−904.64	4.561	0.686	$=\hat{\omega}_0$	3.506
C. Two ratios: $\omega_0 = \omega_C$, ω_H	14	−903.08	4.568	0.675	∞	$=\hat{\omega}_0$
D. Two ratios: ω_0, $\omega_H = \omega_C$	14	−901.63	4.605	0.540	7.263	$=\hat{\omega}_H$
E. Three ratios: ω_0, ω_H, ω_C	15	−901.10	4.598	0.540	∞	3.646
F. Two ratios: $\omega_0 = \omega_H$, $\omega_C = 1$	13	−905.48	4.437	0.686	$=\hat{\omega}_0$	1
G. Two ratios: $\omega_0 = \omega_C$, $\omega_H = 1$	13	−905.38	4.413	0.675	1	$=\hat{\omega}_0$
H. Two ratios: ω_0, $\omega_H = \omega_C = 1$	13	−904.36	4.312	0.543	1	1
I. Three ratios: ω_0, ω_H, $\omega_C = 1$	14	−902.02	4.465	0.541	∞	1
J. Three ratios: ω_0, $\omega_H = 1$, ω_C	14	−903.48	4.435	0.541	1	3.559

NOTE.—p, number of parameters in the model not including the nine parameters for codon frequencies (π_j's in eq. 1). Parameters ω_H, ω_C, and ω_0 are the d_N/d_S ratios for branches h, c, and all other branches, respectively (see figs. 1 and 2). Estimates of branch lengths are not shown.

Table 2
Likelihood Ratio Statistics ($2\Delta\ell$) for Testing Hypotheses

Null Hypothesis Tested	Assumption Made	Models Compared	Large Data Set (n = 19)	Small Data Set (n = 7)
A. $(\omega_H = \omega_C) = \omega_0$	$\omega_H = \omega_C$	A and D	12.50**	8.78**
B. $\omega_C = \omega_0$	$\omega_H = \omega_0$	A and B	4.28*	2.76
C. $\omega_C = \omega_0$	ω_H free	C and E	5.76*	3.96*
D. $\omega_H = \omega_0$	$\omega_C = \omega_0$	A and C	7.84**	5.88*
E. $\omega_H = \omega_0$	ω_C free	B and E	9.32**	7.08**
A′. $(\omega_H = \omega_C) \leq 1$ . .	$\omega_H = \omega_C$	D and H	5.46*	5.46*
B′. $\omega_C \leq 1$	$\omega_H = \omega_0$	B and F	1.60	1.68
C′. $\omega_C \leq 1$	ω_H free	E and I	1.76	1.84
D′. $\omega_H \leq 1$	$\omega_C = \omega_0$	C and G	4.74*	4.60*
E′. $\omega_H \leq 1$	ω_C free	E and J	4.90*	4.76*

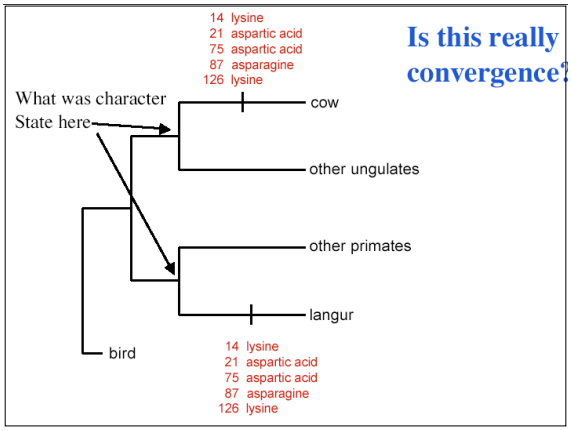
* Significant ($P < 5\%$; $\chi^2_1 = 3.84$).
** Extremely significant ($P < 1\%$; $\chi^2_1 = 6.63$).

Likelihood summary

- Significant variation in selective pressure between lineages.
- Lineage to hominoids dn/ds significantly >1
  - Statistical support for adaptive evolution
- Lineage to colobines dn/ds > 1, but not significantly > 1
  - Lack of constraint, but convergent changes suggest adaptive evolution

Why is there a difference between analyses?

- Ancestral state reconstruction
- Transistion/transversion bias
- Codon bias



Is this really convergence?

Convergence versus parallel evolution

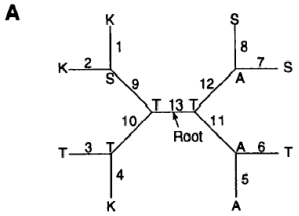
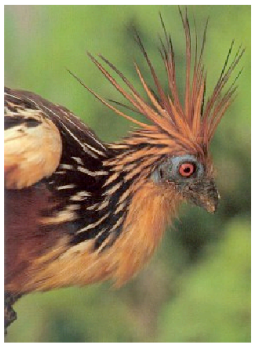


FIG. 1.—Examples of convergent and parallel changes. A, Convergent changes, parallel changes, and uniquely shared sites. There are convergent changes on branches 8 and 9 (A→S and T→S, respectively). There are parallel changes on branches 7 and 8 (A→S). When a

Look at another lineage with the evolution of foregut fermentation

- Do we see similar changes as in colobines.
- What type of amino acid changes occur?



Hoatzin (*Opisthocomus hoatzin*)

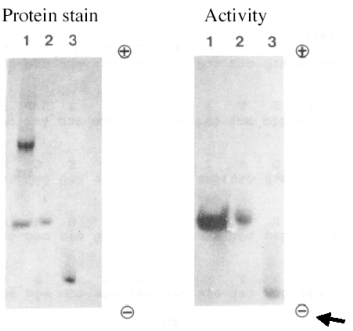
eats leaves
only identified
avian foregut fermenter

stomach lysozyme
similar to ruminants?

Hoatzin lysozyme prediction

- Should be expressed in foregut
- Should be more acidic
- Should show low pH optimium for activity

Native gels show Hoatzin has acidic lysozyme with lytic activity

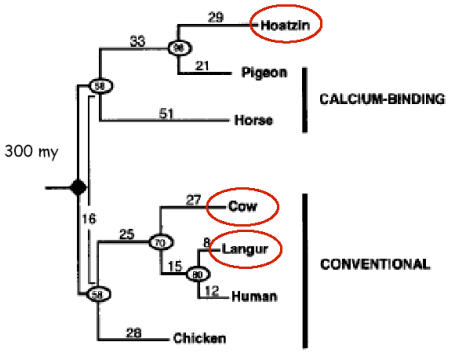


Gel run towards cathode at pH4.3

Identification of lysozyme gene

- Clone gene from cDNA extracted from stomach.
- Use PCR to get entire coding sequence
- Generate phylogenies with other lysozyme
- Look for convergent changes identical to mammalian foregut lysozymes

Hoatzin stomach lysozyme is from different clade than mammalian



Kornegay et al. 1994 Mol. Biol. Evol. 11:921

To check for convergence, compare residues identified in Langur to convergent residues in Hoatzin.

Note K126 not considered convergent

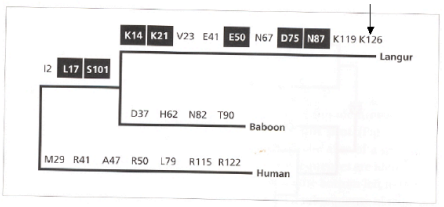


Fig. 5.2 Independent evolution of amino acid replacements in cows and langur monkeys. Although langur monkey lysozyme is phylogenetically closely related to other primate lysozymes it has independently acquired several amino acid substitutions in common with cow lysozyme (these are indicated by the black squares). Redrawn from Li and Graur (1991).

Now compare hoatzin to mammalian

```
                10        20        30        40        50        60
Hoatzin  EIIPRCELVKILREHGFEGFRGTTIADWICLVQHESDYNTEAYNNNG-P-SRDYGIPQINSKYWC
Pigeon   KD.........R.....V.K.V.N.V...K...G.R.T.P....-.N..............
Horse    KVFSK...AHK.KAQEMD..G.YSL.N.V.MAEY..NF..R.F.GKNANG.S...L..L.N.W..
Cow      KVFE....ART.KKL.LD.YK.VSL.N.L..TKW..S...K.T.Y.PSSE.T.........W..
Langur   K.FE....ART.KKL.LD.YK.VSL.N.V..AKW..G.....T.Y.PGDE.T.........R...
Human    KVFE....ART.KRL.MD.YR.ISL.N.M..AKW..G...R.T.Y.AGDR.T.........R...
Chicken  KVFG....AAAMKR..LDNYR.YSLGN.V.AAKF..NF..Q.T.R.T-DG.T....L....RW..

                70        80        90        100       110       120       130
Hoatzin  NDGKTSGAVDGCHISCSELMTWDLEDDIKCAKKIARDAHGLTPWYGWKNHCEGRDLSSYVKGC--
Pigeon   .....R.SKNA.N.N..K.RDDNIA...Q.......E.R.....VA..KY.Q.K.....R.--
Horse    K.N.-RSSSNA.N.M..K.LDENID...S...RVV..PK.MSA.KA.VK..KDK...E.LAS.NL
Cow      .....PN..D...V......EW.IAKAVA...H.VSE-Q.I.A.VA..S..RDH.V....E..TL
Langur   .N....P...DA......A.LQWNIA.AVA...RVVS.PQ.IRA.VA.R...QNK.V.Q..K..GV
Human    .....P...NA..L...A.LQDNIA.AVA...RVV..PQ.IRA.VA.R.R.QN..VRQ..Q..GV
Chicken  ...R.P.SRNL.N.P...A.LSS.ITASVN.....VS.GN.MNA.VA.R.R.K.T.VQAWIR..RL
```

Structural adaptations in stomach lysozyme

	lysozyme type		
characteristic	hoatzin	mammalian	egg-white
low pH optimum	+	+	-
isolelectric point	~6	6.2 – 7.7	11.2
total arginines	5	3 – 6	11
aginine to lysine ratio	0.63	0.27 – 0.67	1.83
adaptive residues:			
14 E/K	+	+	-
21 E/K	+	+	-
75 D	+	+	-
87 N	+	+	-
126 E/K	+	+	-

Could we test these with sites models?

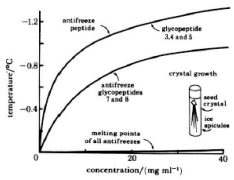
Hoatzin conclusions

- Lysozyme from different family recruited for foregut fermentation
- Similar amino acid changes lend support to convergent evolution.
- No tests for adaptive evolution, but functional convergence suggests positive selection.

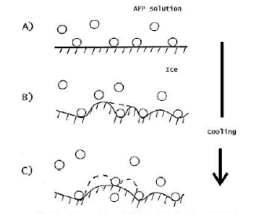
Convergent evolution of "antifreeze" proteins

Fish in freezing temperatures need to produce antifreeze to survive
Compare evolution of Artic and Antarctica fish AFPs.

RESISTANCE TO LOW TEMPERATURE: Antifreeze Proteins of Antarctic Fishes



Freezing and melting points of peptide and glycopeptides at several concentrations. The large glycopeptides and the polypeptides have similar antifreeze activity, low molecular weight glycopeptides (numbers 7 and 8) have less action in lowering freezing points. Melting points are the same for all sizes of antifreezes. [DeVries (34b).]



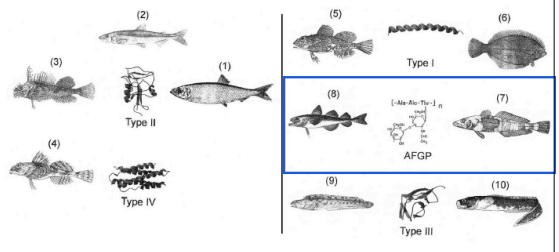
Inhibition of ice crystal growth following AFP adsorption. (A) AFP (open circles) in solution and in contact with the ice front (hatched line) at T °C. (B) The ice-water interface at an undercooling approaching the non-equilibrium freezing point, where the curvature between bound AFPs leads to ice growth inhibition by the Kelvin effect. The dotted line represents the ice growth of a bound AFP. (C) Binding of an AFP from solution to the point of nonzero growth in (B). Here the dotted line represents subsequent inhibition of the ice-water interface around the newly bound AFP. This figure is based on the two-dimensional representations of ice-growth inhibition displayed in Knight & DeVries (15) and Knight (17).

These unusual polypeptides actually interfere with the production of ice crystals in living tissues at very low temperatures.

Evolution of serum antifreeze glycoproteins

- Fish that live in polar waters have serum antifreeze glycoproteins (AFGPs) which allow them to tolerate temperatures of as low as –1.9°C
- It has been shown that fish from the north and south poles have evolved very similar AFGPs independently:
  - AFGP of Antarctic fish, made up of a simple tripeptide repeat: evolved by recruitment of the 5' and 3' ends of an ancestral trypsinogen gene (secretory signal and 3' UTR) and *de novo* amplification of a 9bp Thr-Ala-Ala motif
  - Arctic cod also have a Thr-Ala-Ala tripeptide repeat-based AFGP but this has no relationship with the trypsinogen gene
  - Threonines are O-linked to galactosyl-*N*-acetylgalactosamine and periodicity of repeats matches periodicity of water molecules
- Convergent evolution of the tripeptide-based AFGP

RESISTANCE TO LOW TEMPERATURE: Antifreeze Proteins of Antarctic Fishes



The diversity of antifreeze proteins and antifreeze glycoproteins in marine fishes. Shown are the structures of four types of AFPs (I–IV) and the AFGP found in Antarctic notothenioids and Arctic cod. AFPs I–IV are drawn as ribbon structures. The tripeptide repeat (–ala–ala–thr–) and carbohydrate moiety (galactosyl-*N*-acetylgalactosamine) of AFGPs illustrate the key element of AFGP structures in notothenioid and Arctic cod.

Type I AFPs are alanine-rich, amphipathic alpha helical proteins with characteristic repeat structures involving long stretches of alanines. Molecular masses of Type I AFPs are between 3.3 and 5 kDa. Like the AFGPs, Type I AFPs are extended molecules in which the polar side-chains are oriented along a common

from. Type II AFPs are globular proteins with varied secondary structures. Type III AFPs are composed of small beta-strand elements and a single helix turn that gives the protein a flat-faced folded surface. Type IV AFP is conjectured to be a helix-bundle protein. (See Fletcher et al., 2001, for review.) Species: (1) Atlantic herring (*Clupea harengus harengus*). (2) American smelt (*Osmerus mordax*). (3) sea raven (*Hemitripterus americanus*). (4) longhorn sculpin (*Myoxocephalus octodecimspinosis*). (5) shorthorn sculpin (*Myoxocephalus scorpius*). (6) winter flounder (*Pseudopleuronectes americanus*). (7) *Dissostichus mawsoni* (Antarctic notothenioid). (8) Arctic cod (*Boreogadus saida*). (9) ocean pout (eelpout) (*Macrozoarces americanus*). (10) Antarctic eelpout (*Lycodichthys dearborni*). (Figure modified after Fletcher et al., 2001.)

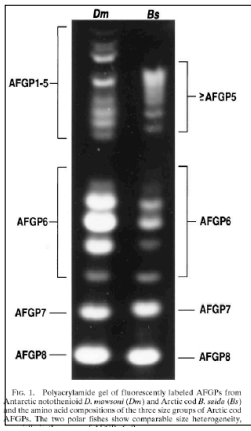


FIG. 1. Polyacrylamide gel of fluorescently labeled AFGPs from Antarctic notothenioid *D. mawsoni* (*Dm*) and Arctic cod *B. saida* (*Bs*) and the amino acid compositions of the three size groups of Arctic cod AFGPs. The two polar fishes show comparable size heterogeneity,

SDS gels show very similar AFP profiles from Artic and Antarctica fish

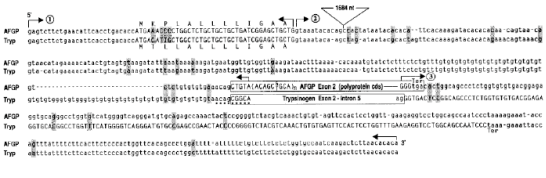
Amino acid composition show high Thr Ala content for both

Is this convergence?

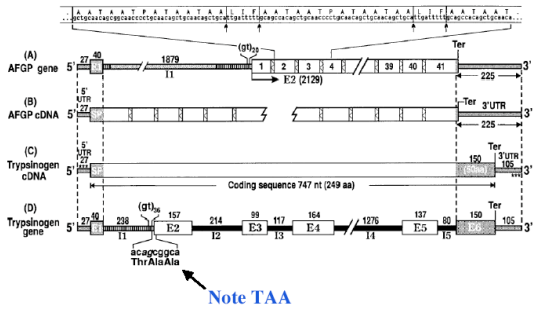
First, clone Antarctica AFP

- Make cDNA from liver
- Screen with probe from different species AFP
- Blast to genbank for comparison
- Align best matches and deduce origin of AFP.

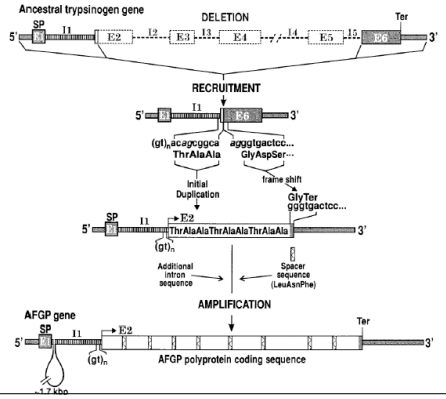
Sequence comparisons show similarity of AFP and Typrsinogen gene untranslated regions and signal sequence



Schematic of overlapping regions:



Proposed scheme for evolution of Antarctica AFP



Summary

- Antarctica version of AFP evolved from a trypsinogen gene with capture of intron sequence.
- How does this gene compare to artic version of AFPs?

No similarity between orgins of Artic and Antarctica AFPs
Artic did not evolve from tripsinogen precursor.

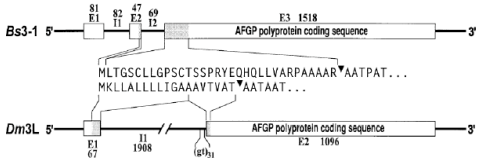


FIG. 3. Exon–intron organizations of *B. saida* and *D. mawsoni* AFGP genes, showing also the dissimilar signal peptide sequences of the two AFGP polyproteins. The coding region for the signal peptide in each gene is indicated by the shaded areas. ▼, Putative cleavage site for the signal peptide.

Summary

- Arctic and Antarctica AFPs look very similar and have similar function
- The two evolved from different ancestral genes.
- Example of convergent molecular evolution.

Lab today in Genome Sciences
Foege Building, S448.
Meet 1:30 in lobby

South Central Campus Map

Questions

- What is convergent evolution?
- What is the difference between parallel change and convergent change?
- Given a tree with bases at terminal nodes, reconstruct ancestral states.
- What is the hypothesized selective pressure driving the evolution of lysozyme in the Colobines?
- Name 2 reasons parsimony and likelihood could give different ancestral reconstructions.