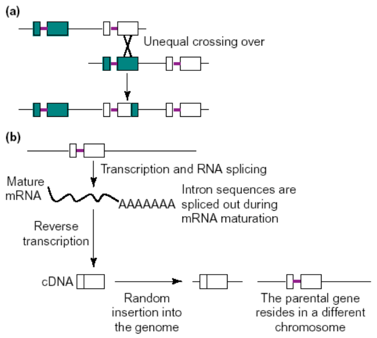
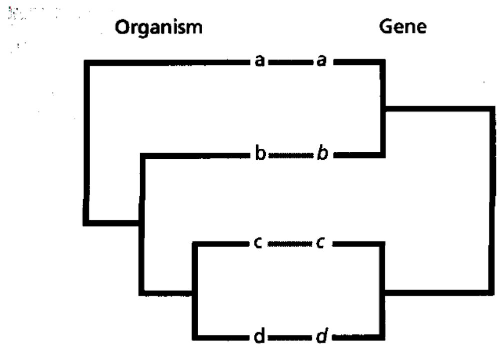
Gene duplication is common

	Total number of genes	Number of duplicate genes (% of duplicate genes)
Bacteria		
Mycoplasma pneumoniae	677	298 (44)
Helicobacter pylori	1590	266 (17)
Haemophilus influenzae	1709	284 (17)
Archaea		
Archaeoglobus fulgidus	2436	719 (30)
Eukarya		
Saccharomyces cerevisiae	6241	1858 (30)
Caenorhabditis elegans	18 424	8971 (49)
Drosophila melanogaster	13 601	5536 (41)
Arabidopsis thaliana	25 498	16 574 (65)
Homo sapiens	40 580[b]	15 343 (38)

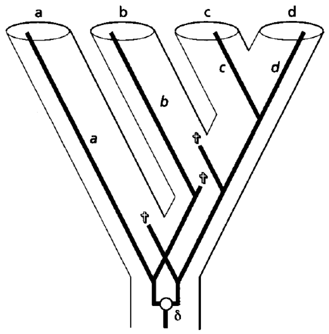
Two types of gene duplication:

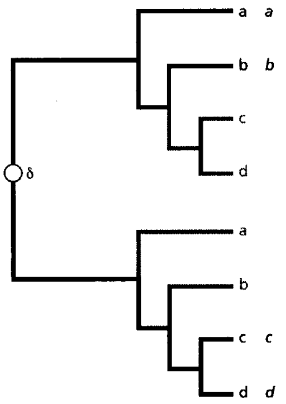


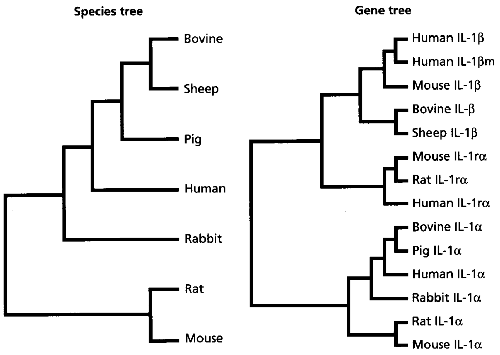
Gene tree and species tress may not be congruent

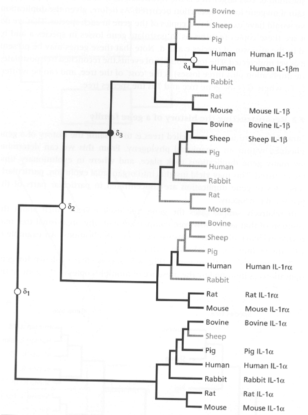


Gene duplication and loss can explain incorrect phylogeny

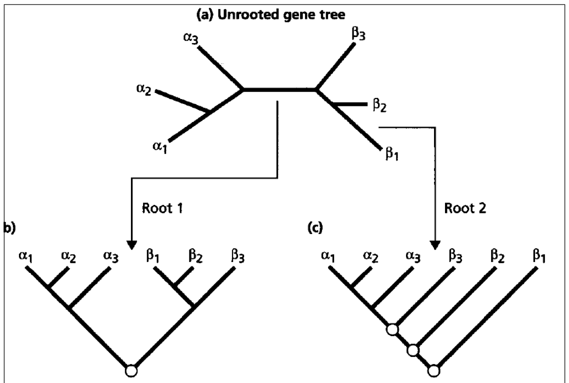








Duplicate genes can be used to root trees



What happens following gene duplication

- Specialization of function
  - Optimize two pre-existing function
- Obtain new function
- Obtain new expression pattern

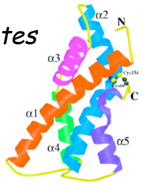
Specialization of function

- Ancestral gene has two function.
- Following duplication, adaptive evolution optimizes function for each gene.
- Genes then become non-redundant and fixed in population.

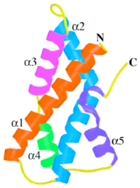
Sp18 and Lysin are Duplicates

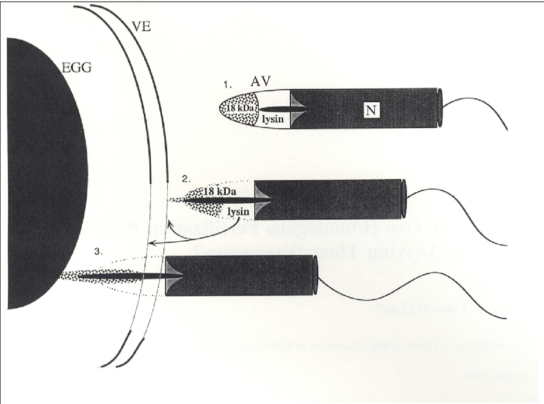
Fusion: Sp18

(≈146AA)

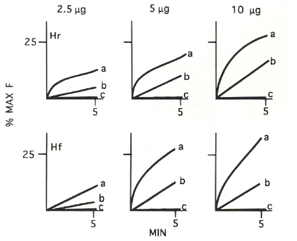


VE Dissolution : Lysin (≈155AA)



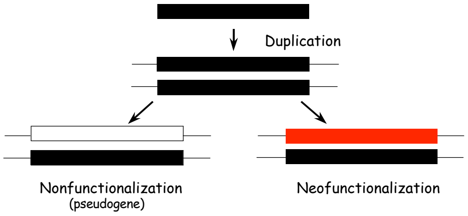


Sp18 better fusagen than lysin



What about origins of new functions?

The Classical Model for the Fates of Duplicate Genes



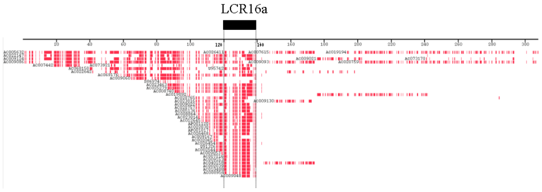
Duplication

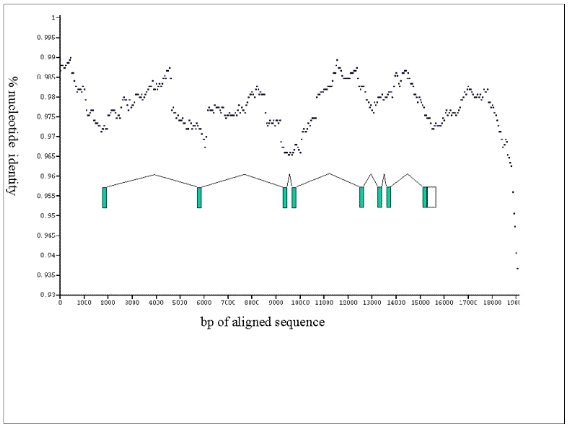
Nonfunctionalization
(pseudogene)

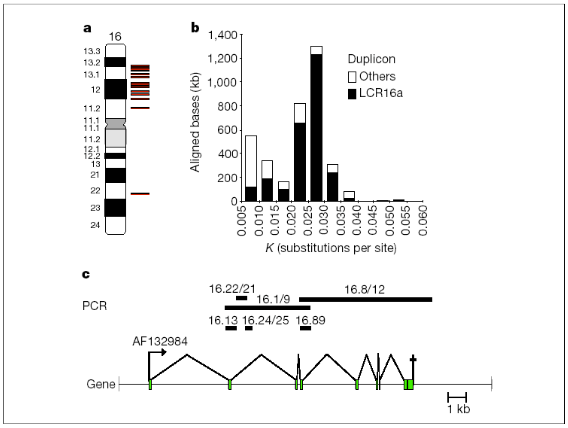
Neofunctionalization

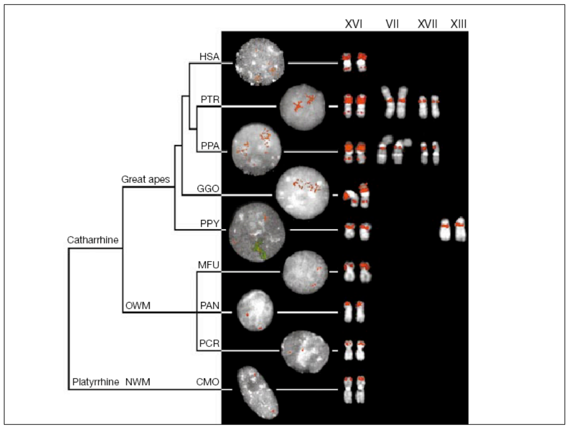
Example from morpheus gene family

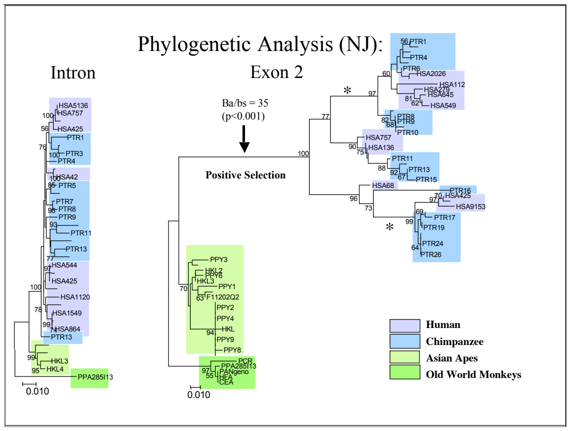
- From heirachical sequencing, complex duplications were found
- Blast these to the complete genome and find locations
- Look at duplications by FISH in other primates
- Perform phylogenetic and adaptive evolution studies.







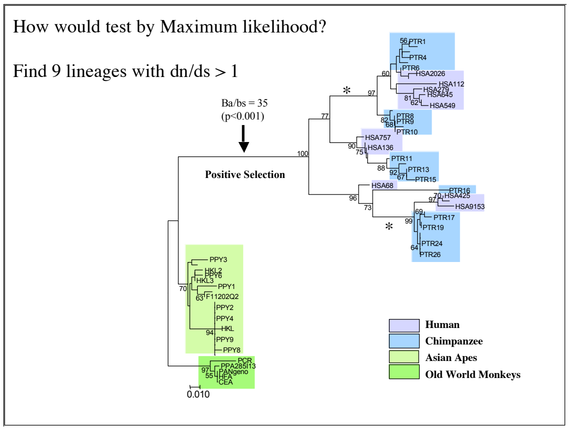




Phylogenetic Analysis (NJ):

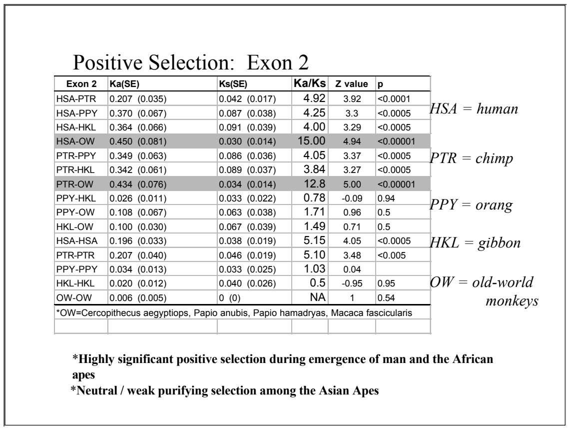
Intron

Exon 2



How would test by Maximum likelihood?

Find 9 lineages with dn/ds > 1



Positive Selection: Exon 2

Exon 2	Ka(SE)	Ks(SE)	Ka/Ks	Z value	p
HSA-PTR	0.207 (0.035)	0.042 (0.017)	4.92	3.92	<0.0001
HSA-PPY	0.370 (0.067)	0.087 (0.038)	4.25	3.3	<0.0005
HSA-HKL	0.364 (0.066)	0.091 (0.039)	4.00	3.29	<0.0005
HSA-OW	0.450 (0.081)	0.030 (0.014)	15.00	4.94	<0.00001
PTR-PPY	0.349 (0.063)	0.086 (0.036)	4.05	3.37	<0.0005
PTR-HKL	0.342 (0.061)	0.089 (0.037)	3.84	3.27	<0.0005
PTR-OW	0.434 (0.076)	0.034 (0.014)	12.8	5.00	<0.00001
PPY-HKL	0.026 (0.011)	0.033 (0.022)	0.78	-0.09	0.94
PPY-OW	0.108 (0.067)	0.063 (0.038)	1.71	0.96	0.5
HKL-OW	0.100 (0.030)	0.067 (0.039)	1.49	0.71	0.5
HSA-HSA	0.196 (0.033)	0.038 (0.019)	5.15	4.05	<0.0005
PTR-PTR	0.207 (0.040)	0.046 (0.019)	5.10	3.48	<0.005
PPY-PPY	0.034 (0.013)	0.033 (0.025)	1.03	0.04	
HKL-HKL	0.020 (0.012)	0.040 (0.026)	0.5	-0.95	0.95
OW-OW	0.006 (0.005)	0 (0)	NA	1	0.54
*OW=Cercopithecus aegyptiops, Papio anubis, Papio hamadryas, Macaca fascicularis					

HSA = human

PTR = chimp

PPY = orang

HKL = gibbon

OW = old-world monkeys

***Highly significant positive selection during emergence of man and the African apes**
***Neutral / weak purifying selection among the Asian Apes**

How would we analyze these data with ML?
Would any additional information be gained?

Exon 2	Ka(SE)	Ks(SE)	Ka/Ks	Z value	p
HSA-PTR	0.207 (0.035)	0.042 (0.017)	4.92	3.92	<0.0001
HSA-PPY	0.370 (0.067)	0.087 (0.038)	4.25	3.3	<0.0005
HSA-HKL	0.364 (0.066)	0.091 (0.039)	4.00	3.29	<0.0005
HSA-OW	0.450 (0.081)	0.030 (0.014)	15.00	4.94	<0.00001
PTR-PPY	0.349 (0.063)	0.086 (0.036)	4.05	3.37	<0.0005
PTR-HKL	0.342 (0.061)	0.089 (0.037)	3.84	3.27	<0.0005
PTR-OW	0.434 (0.076)	0.034 (0.014)	12.8	5.00	<0.00001
PPY-HKL	0.026 (0.011)	0.033 (0.022)	0.78	-0.09	0.94
PPY-OW	0.108 (0.067)	0.063 (0.038)	1.71	0.96	0.5
HKL-OW	0.100 (0.030)	0.067 (0.039)	1.49	0.71	0.5
HSA-HSA	0.196 (0.033)	0.038 (0.019)	5.15	4.05	<0.0005
PTR-PTR	0.207 (0.040)	0.046 (0.019)	5.10	3.48	<0.005
PPY-PPY	0.034 (0.013)	0.033 (0.025)	1.03	0.04	
HKL-HKL	0.020 (0.012)	0.040 (0.026)	0.5	-0.95	0.95
OW-OW	0.006 (0.005)	0 (0)	NA	1	0.54

*OW=Cercopithecus aegyptiops, Papio anubis, Papio hamadryas, Macaca fascicularis

HSA = human

PTR = chimp

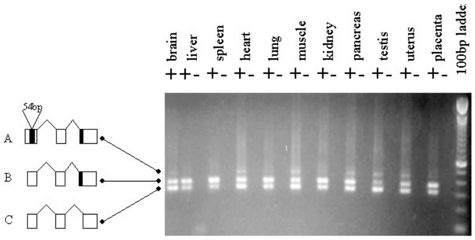
PPY = orang

HKL = gibbon

OW = old-world monkeys

***Highly significant positive selection during emergence of man and the African apes**
***Neutral / weak purifying selection among the Asian Apes**

Some copies are obtaining different tissue expression

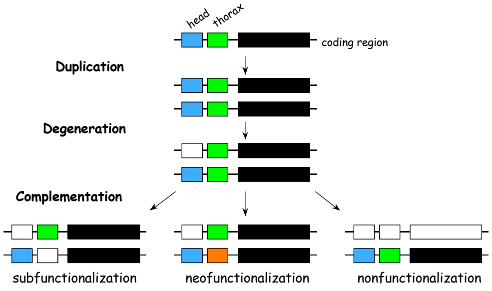


Conclusions

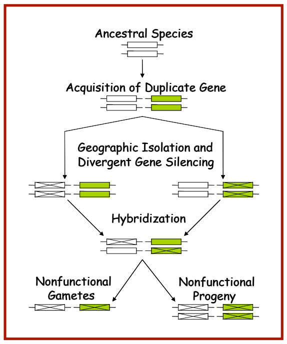
- Extraordinary plasticity of genome and gene
- Evidence for evolution of recent hominoid gene by duplication and adaptation.
  - Gene not detected in other organisms
- Additional examples expected.
  - 5-7% of all human sequences duplicated in last 20Myr.

Other models of duplication loss

The DDC Model



The stochastic birth and loss of duplicate genes leads to the passive origin of reproductive isolating barriers.



Ways to identify gene duplicates as orthologs

- Reciprocal best hit Blasts
- Synteny

Reciprocal best hits

- Blast gene of interest from genome A against genome B.
- Find best hit in genome B.
- Take best hit against from genome B and blast against Genome A.
- If oringinal gene of interest is found, then assume orthologs.