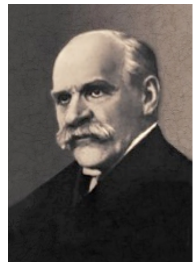
Gene Duplication and Evolution

Jan Aagaard
jaagaard@gs.washington.edu

- Readings
  - Required: Aguileta et al., 2004. JME 59:177.
  - Optional: Gonzalez et al., 2005. Science 307:1434.
- Genome Sciences seminar on gene duplication
  - Wednesday 3/9, 3:30 in Hitchcock 132
- Final exam question on gene duplication material!



J. B. S. Haldane
1892-1962
The Causes of Evolution. 1932

H. J. Muller
1890-1967
1935. Genetics 17:237

Overview

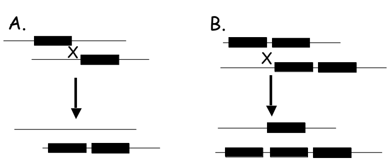
- Mechanisms and frequency of gene duplication
- Fate of duplicate genes
- Evolution duplicate genes
  - example of β-globin genes

Three Classes of Gene Duplication

1. Individual Gene Duplication
   Duplication of an individual gene
2. Segmental Gene Duplications
   Duplication of portions of chromosomes
3. Chromosomal or Whole Genome Duplications
   Duplication of entire individial chromosomes (aneuploidy) or the entire set of chromosomes (polyploidy)

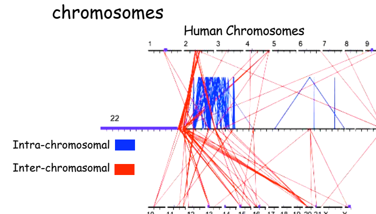
1. Individual Gene Duplications

Most commonly result of unequal crossing over between two chromosomes (sister chromatids or homologous chromosomes)



2. Segmental Gene Duplication

Result of transposition of segments of a chromosome (1 to >200 kb) within the chromosome or to other chromosomes

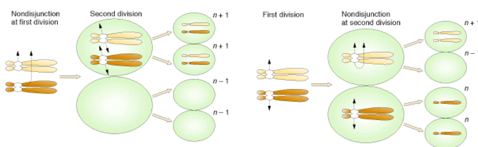


Bailey et al., 2002.
AJHG 70:83.

3. Changes in Ploidy

Result of incomplete partitioning of chromosomes during cell division -- may be partial (aneuploidy) or involve the entire chromosome complement (polyploidy)

Meiosis I Meiosis II

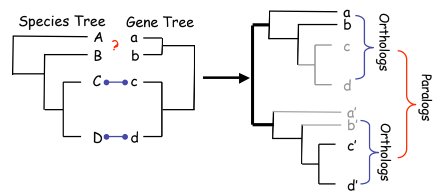


Modified from Griffiths et al., 2000. *An Introduction to Genetic Analysis.* Fig. 18-16

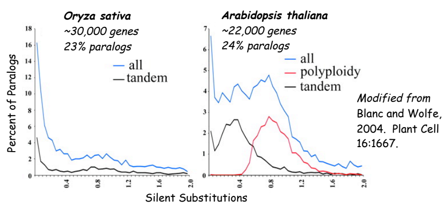
Identifying Duplicate Genes

Sequence, gene structure, physical proximity

Phylogenetic Analyses



Gene Duplication is Frequent...... loss of Duplicates Frequent, too!



Modified from Blanc and Wolfe, 2004. Plant Cell 16:1667.

1. ~ 1% of organism's genes duplicated per 1 My
2. 1/2-life of paralogs ~ *2-7 My*

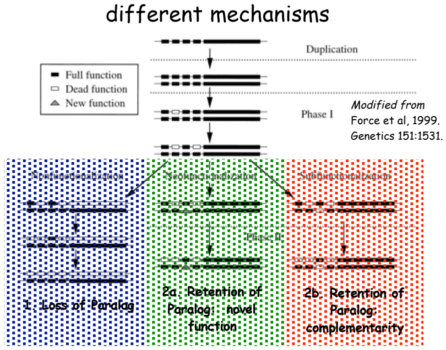
-may be much longer for polyploid paralogs (~ *23 My)?*

Summary to Here.......

- Gene duplication is common via several mechanisms
- Majority of duplicates rapidly lost

- *What about retained duplicates?*

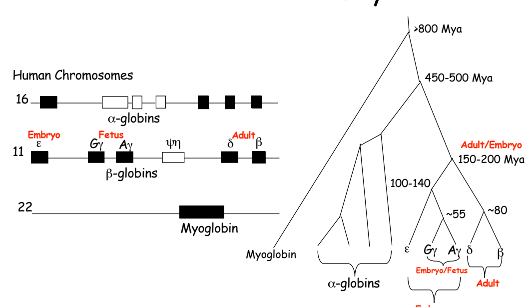
Two Possible Fates for Paralogs....... several different mechanisms



Modified from Force et al, 1999. Genetics 151:1531.

2c. Retention of Paralog: dosage

Globin Gene Family



Aguileta et al., 2004. JME 59:177.

1. **Neufunctionalization**

   Adaptive evolution following gene duplication (dN/dS or ω > 1)?

2. **Subfunctionalization**

   Purifying selection following gene duplication (dN/dS or ω < 1)?

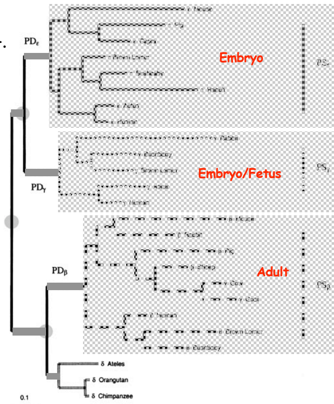


Table 1. Parameter estimates and likelihood scores in separate analyses of the β-, γ-, and ε-globin genes under site-specific models

Model	Parameter estimate(s)	ℓ
M0 (one-ratio)		
β	$\omega = 0.27$	−1676.08
γ	$\omega = 0.26$	−1609.76
ε	$\omega = 0.17$	−2137.83
M1 (neutral)		
β	$(\omega_0 = 0), f_0 = 0.60, (\omega_1 = 1), (f_1 = 0.40)$	−1621.00
γ	$(\omega_0 = 0), f_0 = 0.57, (\omega_1 = 1), (f_1 = 0.43)$	−1598.16
ε	$(\omega_0 = 0), f_0 = 0.54, (\omega_1 = 1), (f_1 = 0.46)$	−2145.53
M2 selection		
β	$(\omega_0 = 0), f_0 = 0.60, (\omega_1 = 1), f_1 = 0.36, \omega_2 = 3.58, (f_2 = 0.04)$	−1617.42
γ	$(\omega_0 = 0), f_0 = 0.52, (\omega_1 = 1), f_1 = 0.006, \omega_2 = 0.57, (f_2 = 0.47)$	−1592.80
ε	$(\omega_0 = 0), f_0 = 0.33, (\omega_1 = 1), f_1 = 0.11, \omega_2 = 0.16, (f_2 = 0.56)$	−2100.22
M3 discrete		
β	$\omega_0 = 0.02, f_0 = 0.65, \omega_1 = 0.57, f_1 = 0.26, \omega_2 = 2.02, (f_2 = 0.09)$	−1608.57
γ	$\omega_0 = 0.001, f_0 = 0.52, \omega_1 = 0.42, f_1 = 0.18, \omega_2 = 0.66, (f_2 = 0.31)$	−1592.80
ε	$\omega_0 = 0.04, f_0 = 0.66, \omega_1 = 0.27, f_1 = 0.24, \omega_2 = 0.89, (f_2 = 0.11)$	−2099.60
M7 beta		
β	$p = 0.11, q = 0.29$	−1612.60
γ	$p = 0.23, q = 0.55$	−1593.31
ε	$p = 0.34, q = 1.50$	−2101.40
M8 beta and ω		
β	$p = 0.16, q = 0.061, f_0 = 0.93, \omega_1 = 2.19, (f_1 = 0.07)$	−1608.76
γ	$p = 0.03, q = 0.64, f_0 = 0.57, \omega_1 = 0.60, (f_1 = 0.43)$	−1592.79
ε	$p = 0.95, q = 7.81, f_0 = 0.88, \omega_1 = 0.85, (f_1 = 0.12)$	−2099.67

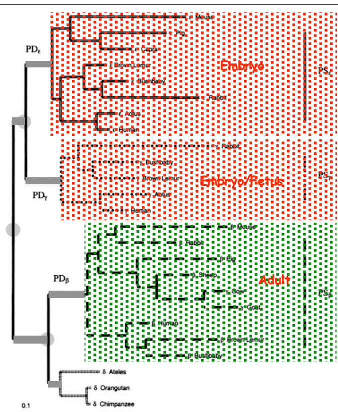
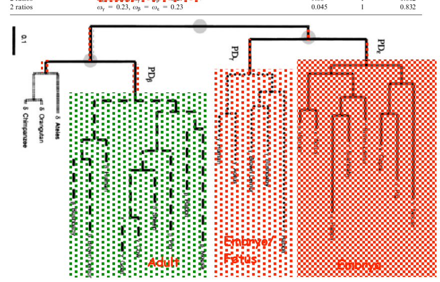


Table 3. Maximum likelihood estimates of ω ratios under branch-specific models and likelihood ratio test statistics when the model is compared with the null model M0 (one ratio)

Alternative model	Parameter estimate(s)	2δ	df	p value
Postduplication and postspeciation (PD–PS) models				
2 ratios	$\omega_{(PD)} = 0.34, \omega_{(PS)} = 0.23$	1.10	1	0.29
4 ratios	$\omega_{\beta(PD)} = 0.41, \omega_{\varepsilon(PD)} = 0.22, \omega_{\gamma(PD)} = 0.08, \omega_{(PS)} = 0.24$	1.10	3	0.78
Paralog models				
3 ratios	$\omega_\beta = 0.29, \omega_\gamma = 0.23, \omega_\varepsilon = 0.16$	11.66	2	0.003
2 ratios	$\omega_\beta = 0.28, \omega_\varepsilon = \omega_\gamma = 0.19$	7.89	1	0.005
2 ratios	$\omega_\varepsilon = 0.16, \omega_\gamma = \omega_\beta = 0.27$	9.88	1	0.002
2 ratios	$\omega_\gamma = 0.23, \omega_\beta = \omega_\varepsilon = 0.23$	0.045	1	0.832



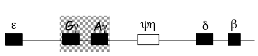


Table 4. Parameter estimates and log-likelihood scores for the γ-globin gene under different sites and branch-site models

Model	p	Parameter estimates	Positive selection	ℓ
Site-specific models				
M1 (neutral)	1	$(\omega_0 = 0.00), f_0 = 0.47 (\omega_1 = 1.00), (f_1 = 0.53)$	No	−3170.38
M3 (discrete)(K = 2)	3	$\omega_0 = 0.06, f_0 = 0.74 \omega_1 = 0.63, (f_1 = 0.26)$	No	−3094.02
Branch-site models				
Model A	3	$(\omega_0 = 0), f_0 = 0.45 (\omega_1 = 1), f_1 = 0.44 \omega_2 = 10, (f_{2+3} = 0.1)$	Yes	−3151.80
Model B	5	$\omega_0 = 0.05, f_0 = 0.60 \omega_1 = 0.63, f_1 = 0.22 \omega_2 = 4.58, (f_{2+3} = 0.18)$	Yes	−3084.69

Note. p is the number of parameters in the ω ratio distribution. The foreground branch in the branch-site models is the branch leading to simian primates.