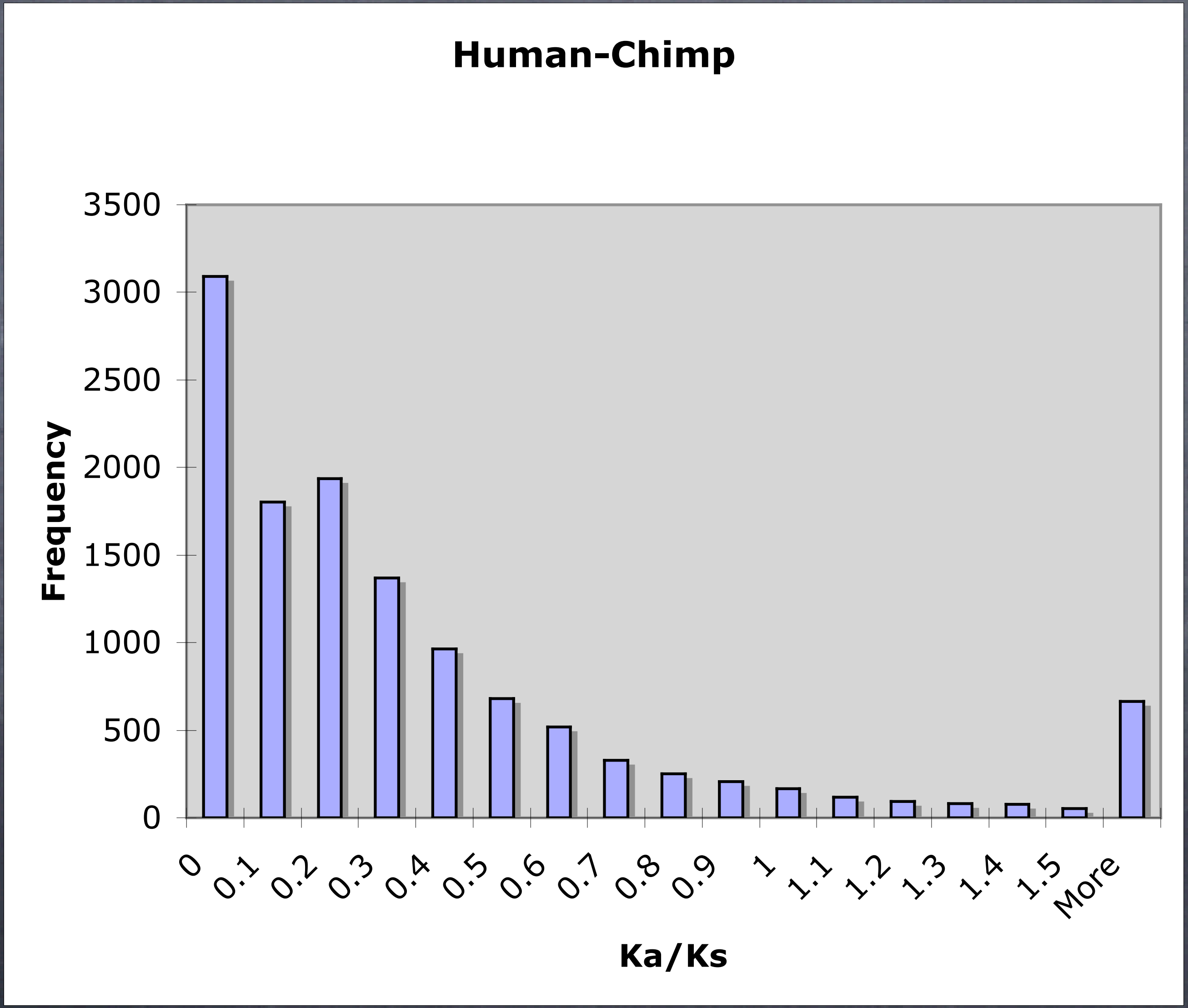
Evolution in genetic networks

Matthew Hahn
Department of Biology
& School of Informatics
Indiana University

The Big Question

- Why do some genes evolve faster than others?
- (Why do some traits evolve faster than others?)

Variation among genes

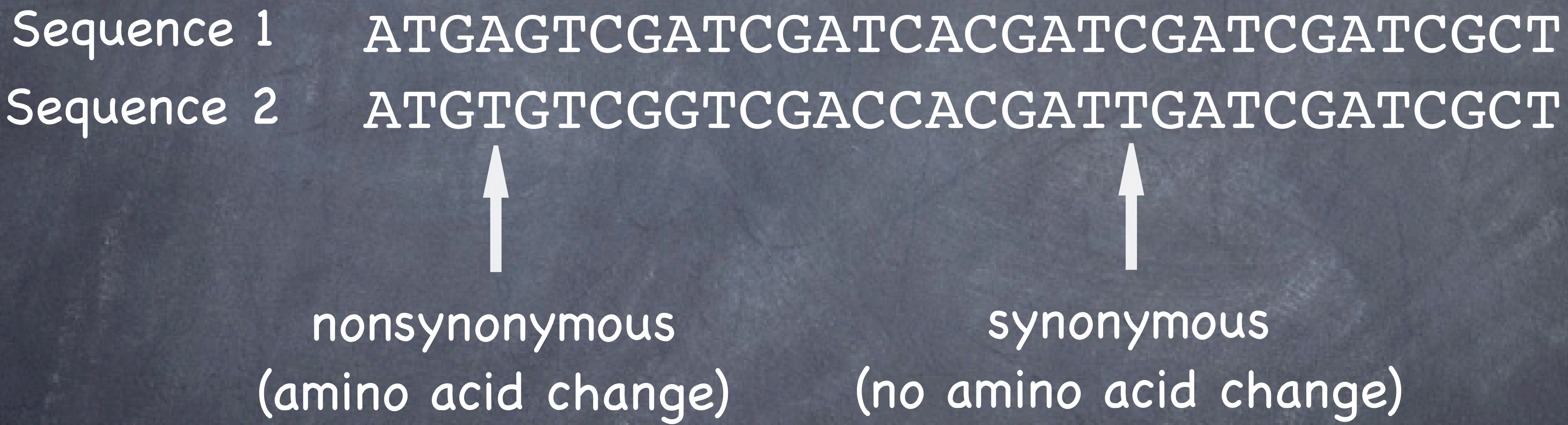


Chimp Genome Sequencing Consortium 2005

Outline

I. Background (Why networks?)
II. Networks and gene divergence
III. Networks and essential genes

Measuring molecular evolution



Sequence 1 ATGAGTCGATCGATCACGATCGATCGATCGCT
Sequence 2 ATGTGTCGGTCGACCACGATTGATCGATCGCT

nonsynonymous
(amino acid change)

synonymous
(no amino acid change)

Measuring molecular evolution

Sequence 1	ATGAGTCGATCGATCACGATCGATCGATCGCT
Sequence 2	ATGTGTCGGTCGACCACGATTGATCGATCGCT

K_a = nonsynonymous changes/site

K_s = synonymous changes/site

Measuring molecular evolution

mutation rate

$$K_a = \mu * f_0$$

substitution rate

fraction neutral mutations

$f_0 \sim 0.20$ for nonsynonymous

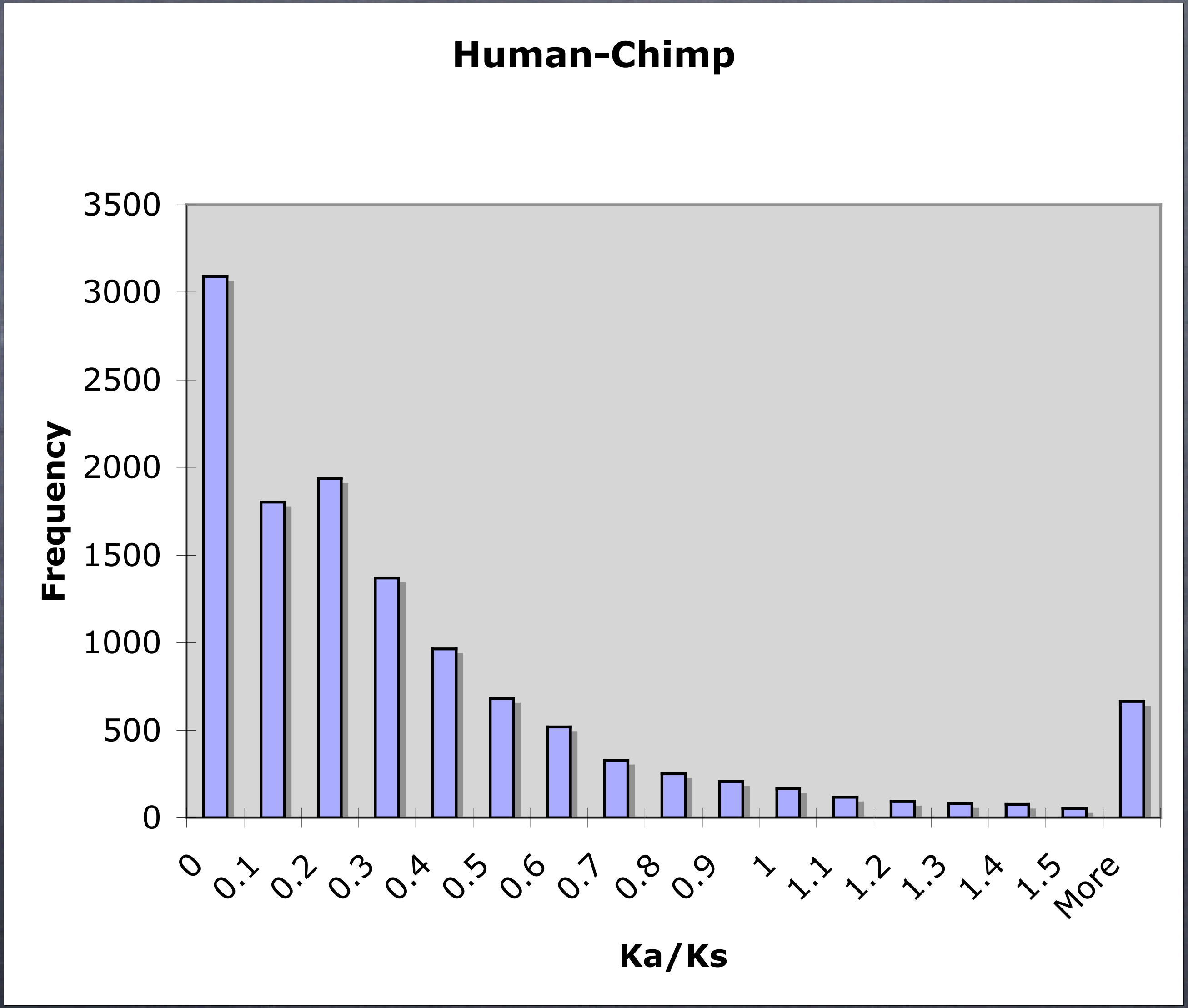
$K_a = \mu * f_0$ nonsynonymous

$K_s = \mu * f_0$ synonymous

($f_0 = 1$ for synonymous)

$$K_a/K_s = f_0$$

Variation among genes



Chimp Genome Sequencing Consortium 2005

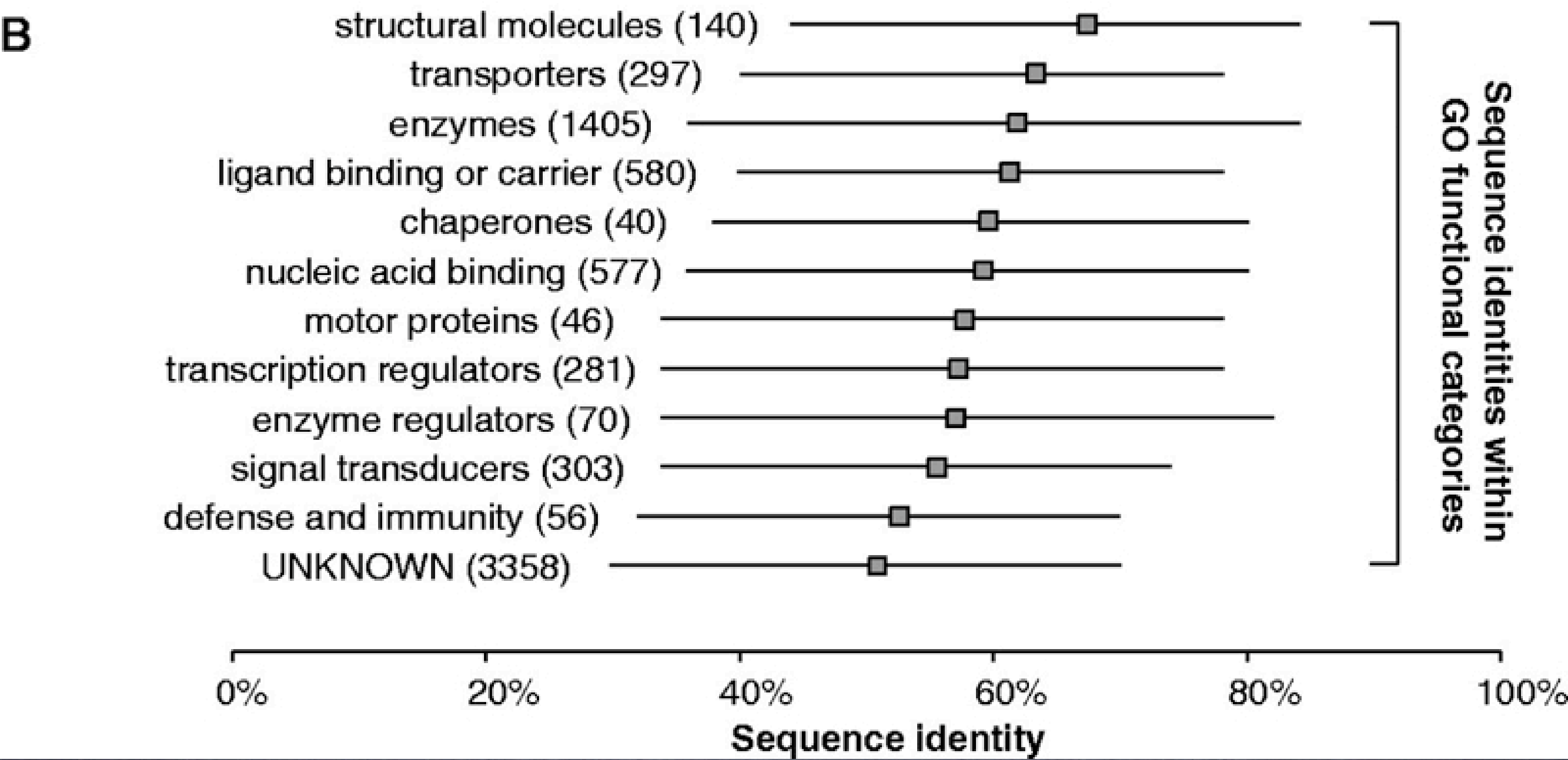
Variation among genes

Gene	Ka/Ks
actin	0.003
myosin	0.05
insulin	0.07
apolipoprotein	0.40
interferon	0.56

Li 1997

Why is there variation in rates of evolution?

- Mutation
- Structure (type of protein: enzyme, tx factor, etc.)
- Function (immunity, sex, housekeeping, etc.)
- Expression level



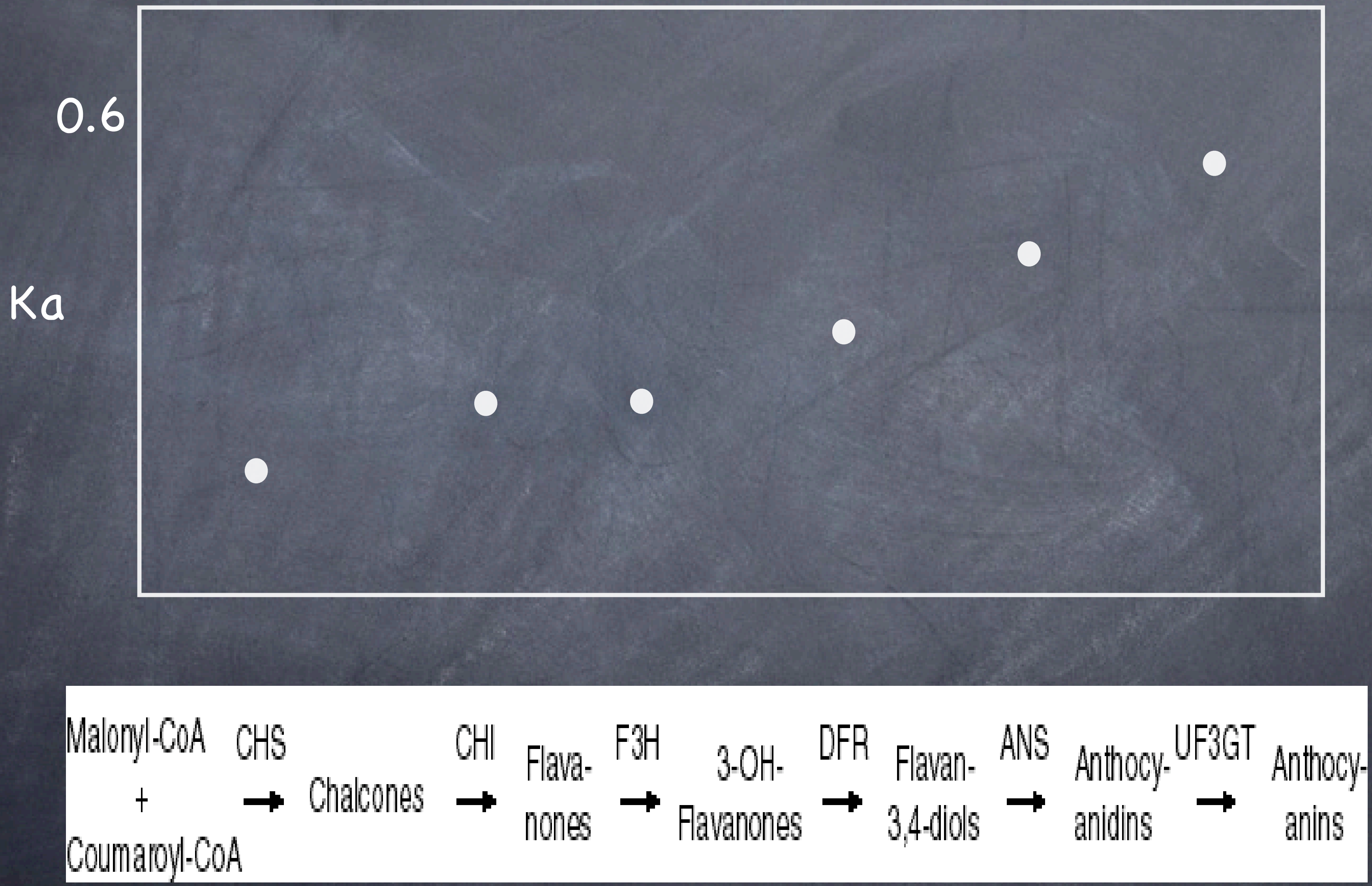
Both structure and function have effects on evolutionary rate.

Why is there variation in rates of evolution?

Position in genetic network?

Hypothesis:
The position of genes in pathways or networks affects their rate of evolution

Rausher, Miller, and Tiffin 1999



Malonyl-CoA + Coumaroyl-CoA →(CHS) Chalcones →(CHI) Flavanones →(F3H) 3-OH-Flavanones →(DFR) Flavan-3,4-diols →(ANS) Anthocyanidins →(UF3GT) Anthocyanins

Why Networks?

Genes more central to a network have a greater number of pleiotropic effects on other genes and biological processes (Promislow 2004).

Why Networks?

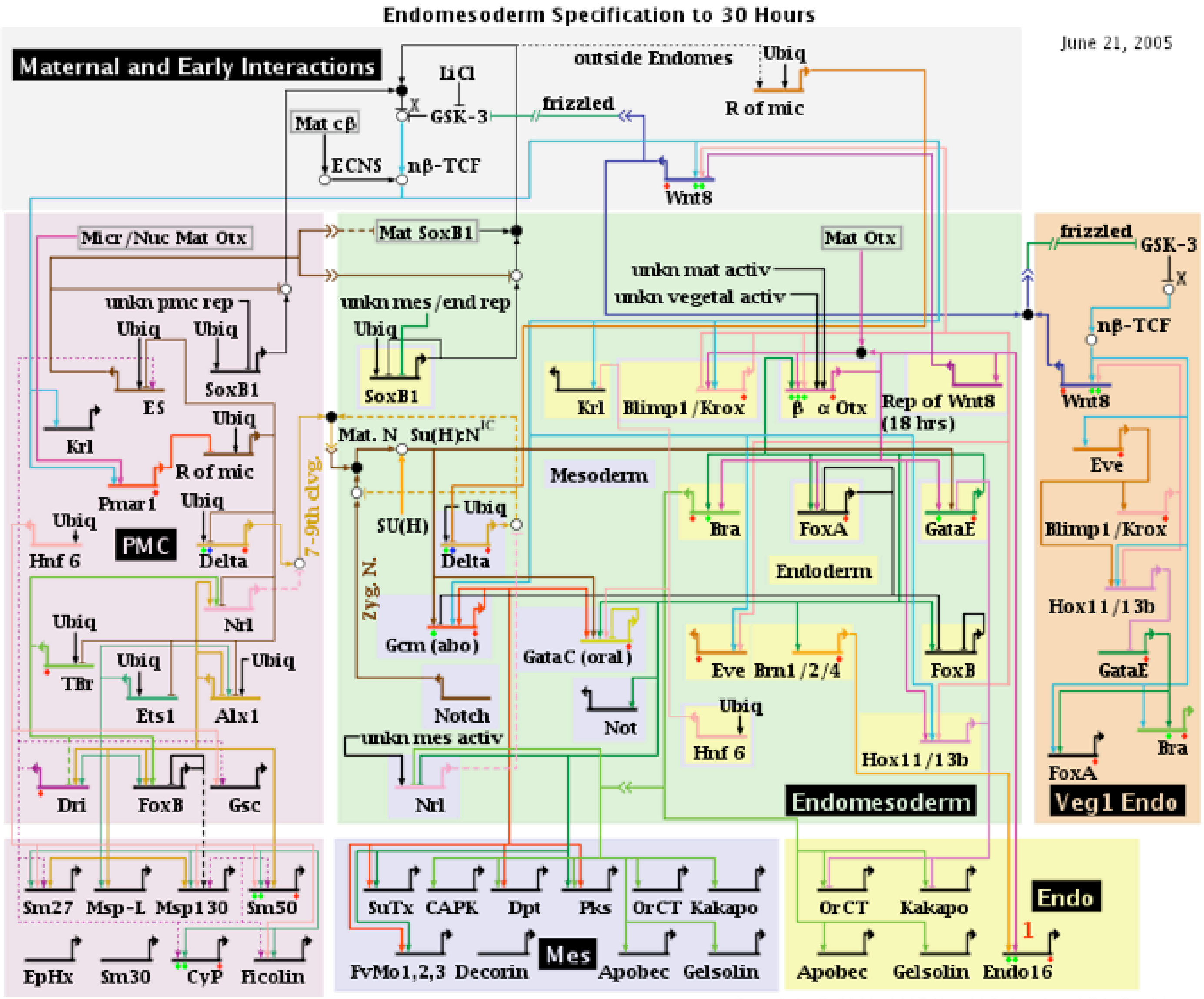
Darwin (1859): Nothing

Fisher (1930): Pleiotropy
constrains evolution

Genetic Networks

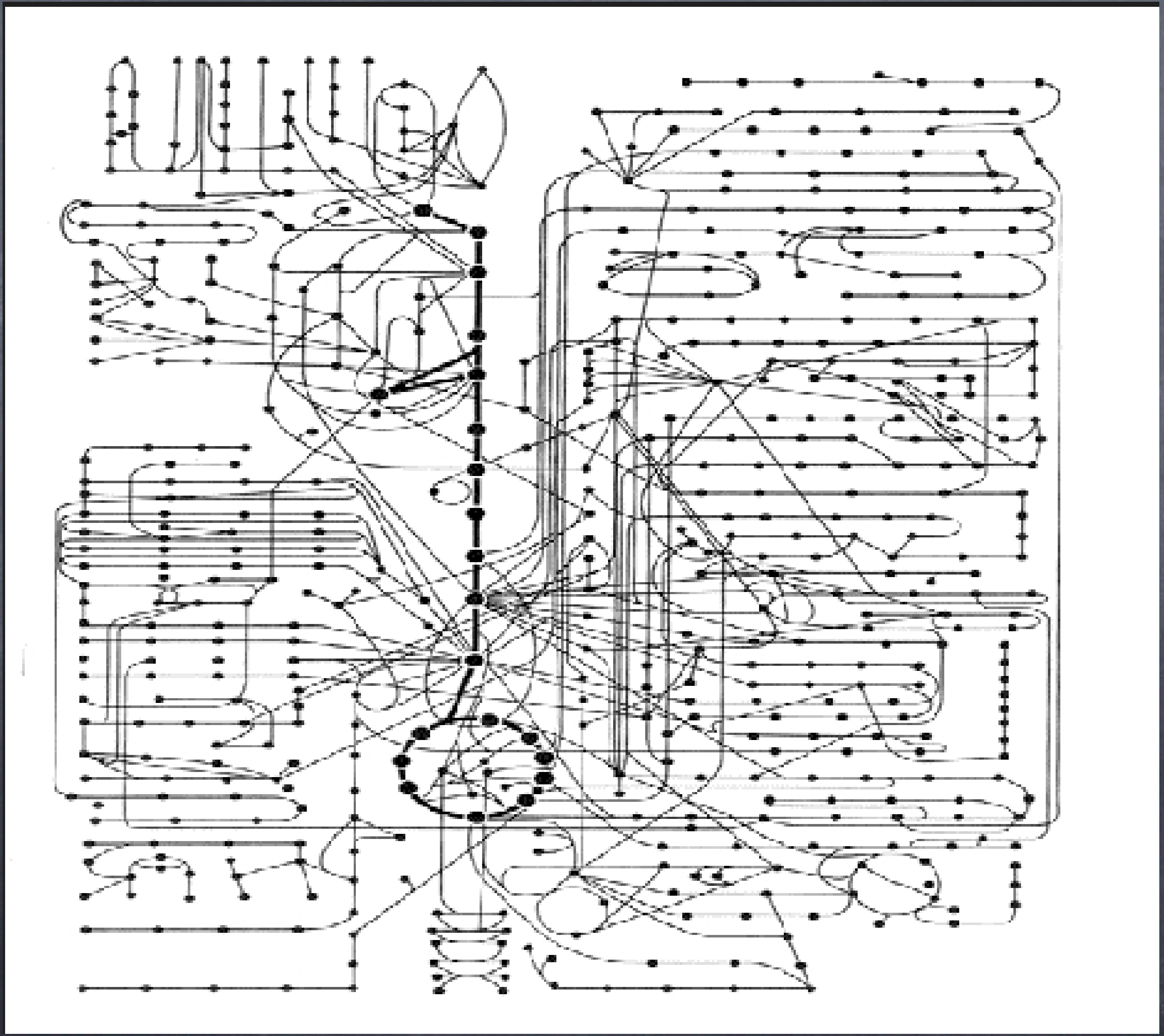
- Regulatory
- Metabolic
- Protein-interaction

Regulatory Network

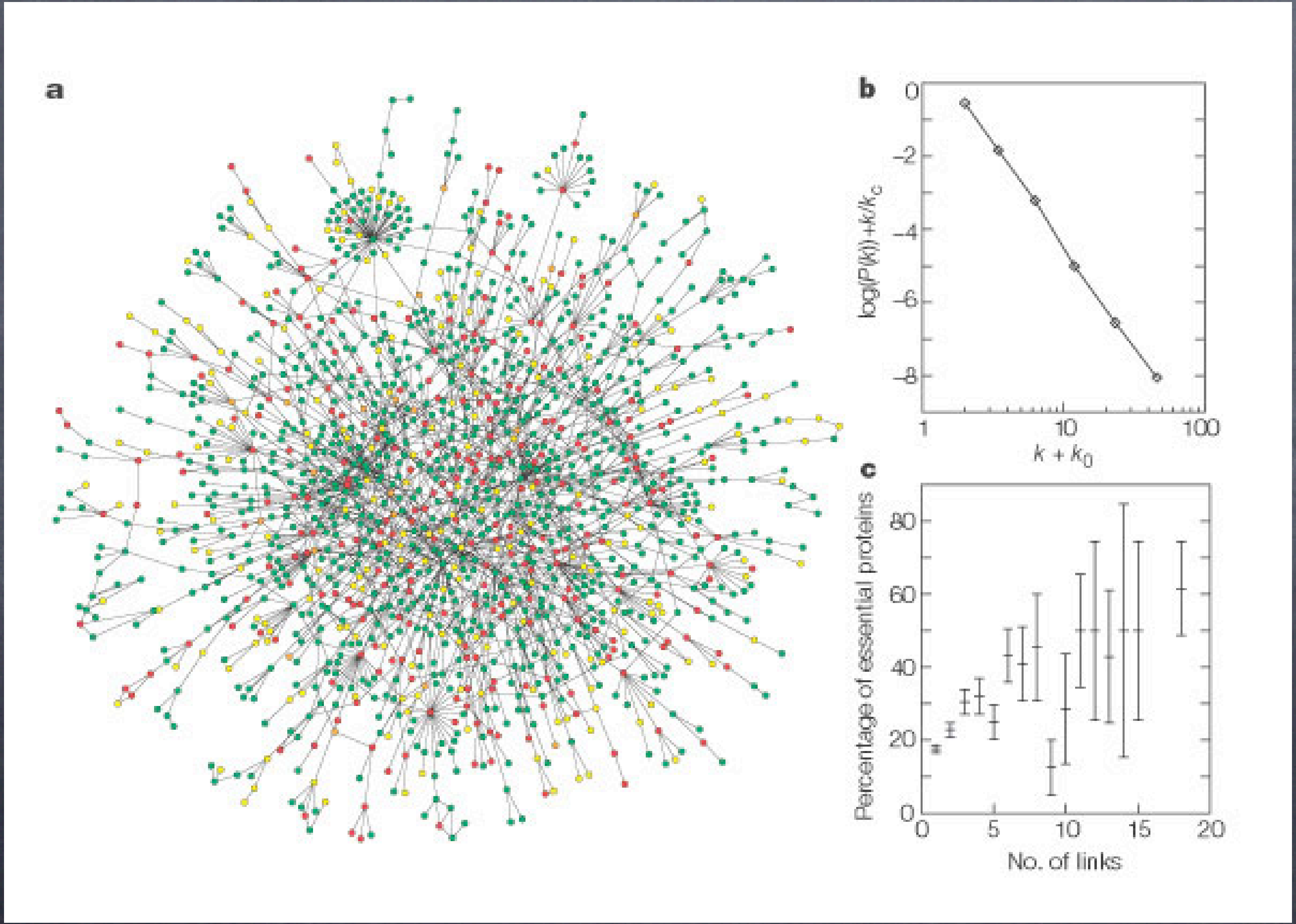

Endomesoderm Specification to 30 Hours

June 21, 2005

Metabolic Network



Protein-interaction Network



Yeast protein-interaction network

Yeast Protein Network

Genes that are more “central” are more likely to…

- … be lethal when knocked out (Jeong et al. 2001)
- … evolve more slowly (Fraser et al. 2002; Krylov et al. 2003; Hahn et al. 2004)

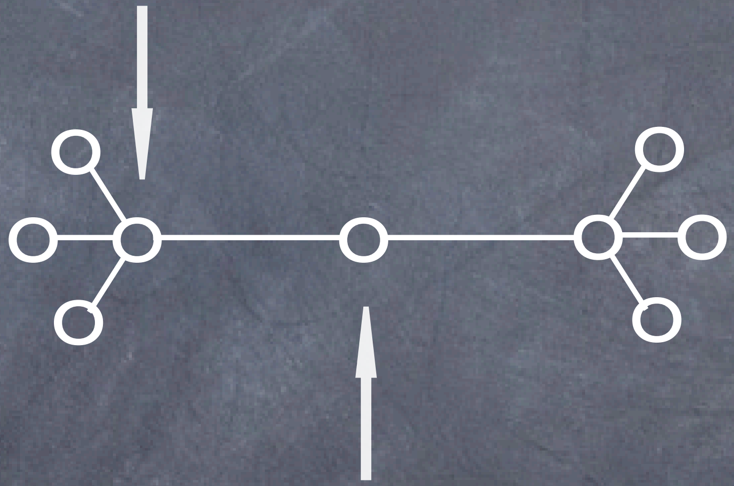
Problems with previous analyses

Measured centrality as the "connectivity" of a protein: simply the total number of interactors.

- Experimental methods are biased towards more highly expressed genes (Bloom and Adami 2003)
- Greater connectivity means more direct contacts (Fraser et al. 2002)

Centrality?

- 1-D Connectivity (or "degree"): the number of direct interactors a protein has
- 2-D Betweenness: the frequency with which a protein lies on the shortest path between other proteins
- 2-D Closeness: the average distance to all other proteins



High connectivity

High betweenness

Our Work

Protein-interaction networks, divergence, and KO effects from:

- S. cerevisiae (S. paradoxus)
- D. melanogaster (D. pseudoobscura)
- C. elegans (C. briggsae)

Hahn and Kern 2005

Networks used

Protein-interaction data from the GRID database (Breitkrutz et al. 2003):

- S. cerevisiae (20252 interactions, 2434 proteins)
- D. melanogaster (16002 interactions, 5082 proteins)
- C. elegans (5977 interactions, 1997 proteins)

All networks and statistics were calculated with "Pajek" (Batagelj and Mrvar 1998).

KO Data

- KO data from S. cerevisiae (Giaever et al. 2002)
- Lethals in D. melanogaster (Flybase)
- RNAi phenotypes in C. elegans (Maeda et al. 2001; Kamath et al. 2003)

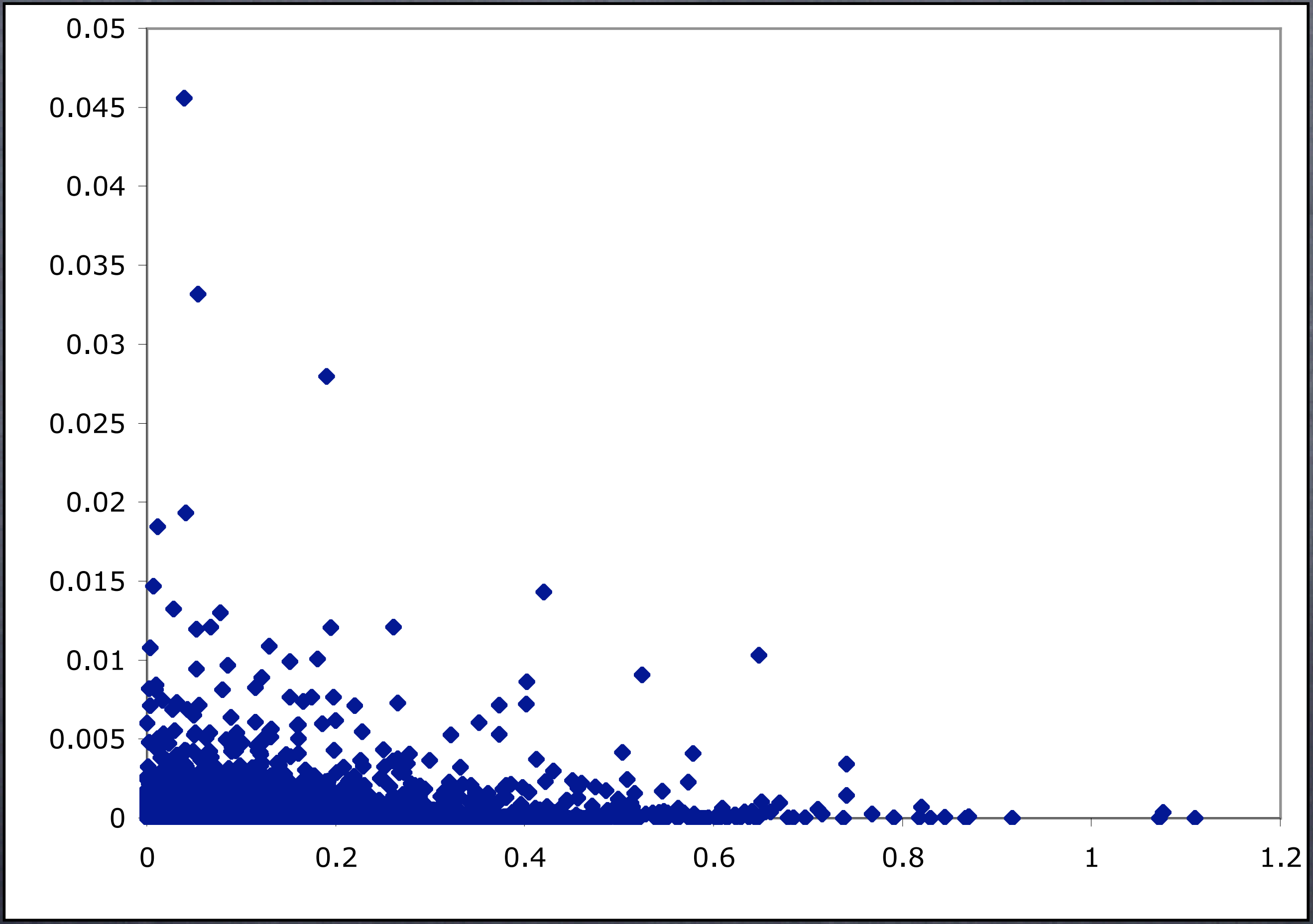
Measures of Centrality

	Yeast	Worm	Fly
Con-Bet	0.21	0.96	0.94
Bet-Close	0.69	0.54	0.78
Con-Close	0.24	0.55	0.84

Spearman's rho: all significant at 0.0001

Results...

D. melanogaster-D.pseudoobscura



Betweenness

Ka

Centrality and Divergence

	Yeast	Worm	Fly
Ka-Bet	-0.17	-0.12	-0.07
Ka-Conn	-0.09	-0.11	-0.06
Ka-Close	-0.16	-0.03*	-0.05

*not significant

Centrality and Divergence

For yeast and fly we can estimate independent effects of measures of centrality:

Betweenness and connectivity both have significant effects in a multiple regression (not closeness)

Centrality and Divergence

Preliminary conclusion:

-Proteins at the center of networks--regardless of the number of direct interactors--evolve more slowly

-This means that observed correlations are not due simply to the number of physical contacts a protein makes (Fraser et al. 2002), and

-Correlations cannot be due to bias in detecting more interactions for more highly expressed genes (Bloom and Adami 2003)

Centrality and Essentiality

Are essential genes more central in all three networks?

and, if so,

Which measure of centrality are they correlated with?

Centrality and Essentiality

		Yeast	Worm	Fly
Betweenness	Essential	0.0009	0.0017	0.0007
	Non-essential	0.0007	0.0009	0.0004
Connectivity	Essential	19.3	8.2	9.8
	Non-essential	15.8	5.6	5.7
Closeness	Essential	0.244	0.183	0.238
	Non-essential	0.239	0.175	0.221

Wilcoxon two-sample test: all significant at 0.001

Centrality and Essentiality

Essential genes are more likely to be central in all three networks!

Betweenness has an independent effect on the probability of being essential in both yeast and fly (LRT $P<0.0001$)

Essentiality and Divergence

It has previously been shown in E. coli, yeast, and worm that essential genes evolve more slowly than non-essential genes (Jordan et al. 2002; Hirsh and Fraser 2001; Stein et al. 2003)

If essential genes are found in the center of the network, then this may explain the correlation between centrality and divergence...

Essentiality and Divergence

Ka

	Yeast	Worm	Fly
Essential	0.031	0.102	0.096
Non-essential	0.044	0.143	0.137

Wilcoxon two-sample test: all significant at 0.001

Essentiality and Divergence

Essential genes evolve more slowly in all three networks,
and are more likely to be central,
but...

Looking only within non-essential genes there is still a
correlation between divergence and centrality!
($P<0.01$ for all three networks)

(Same for only essential genes.)

The 70% rule

Strangely, essential genes evolve at 70% the rate of non-essential genes in all three networks:

Ka	Yeast	Worm	Fly
Essential	0.031	0.102	0.096
Non-essential	0.044	0.143	0.137
	70.5%	71.3%	70.1%

In E. coli, they're 30% the rate of non-essentials (Jordan et al. 2002)

What does it all mean?

-Genes evolve at different rates because of their location in the protein-interaction network

-Genes evolve at different rates because they are essential to survival

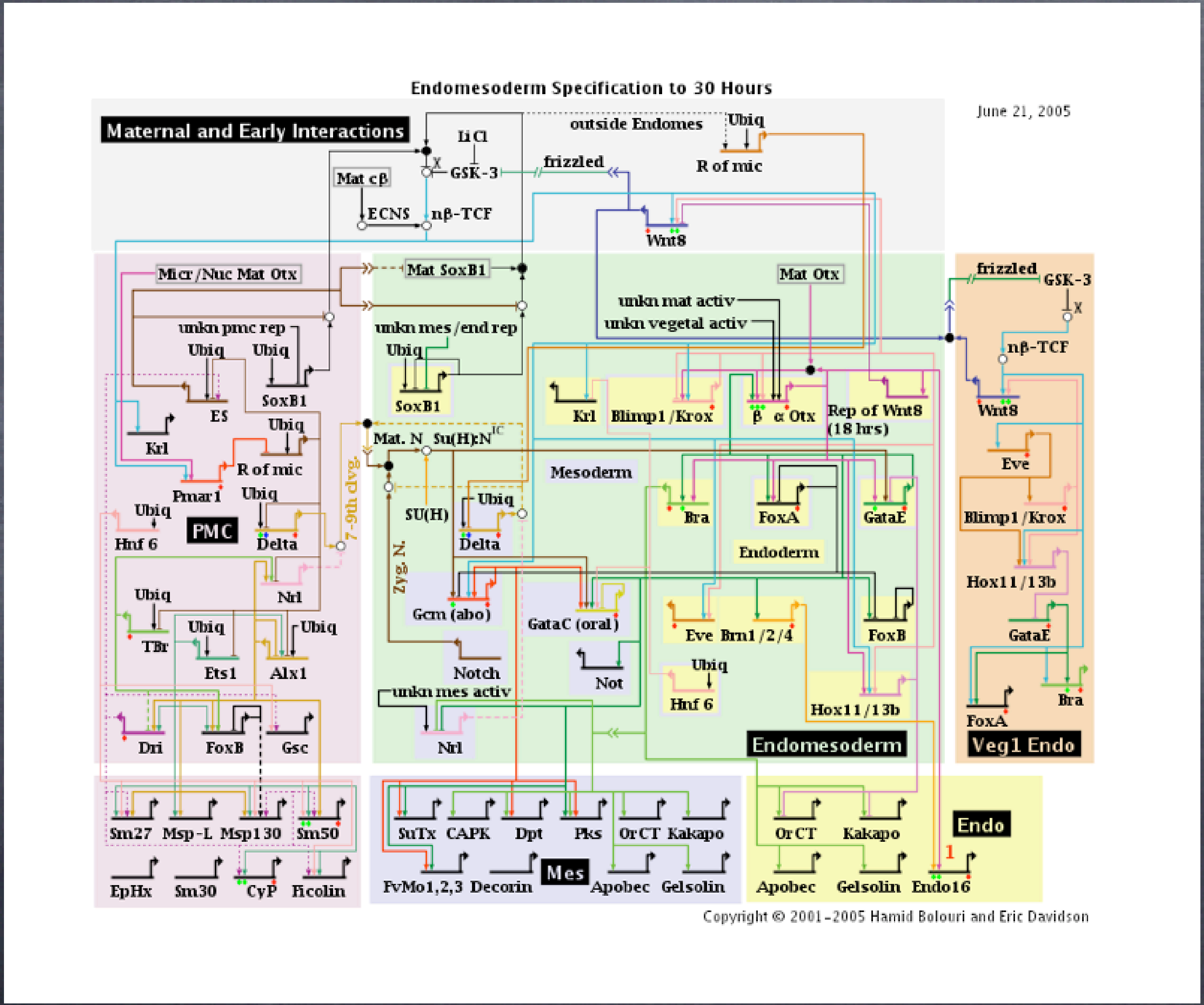
Maybe Fisher was right after all...

The Future

-Networks have finer structure than what we’ve considered here (”motifs”)

-How does the network itself change over time within and between species?

“The” Network



Balhoff and Wray 2005

“The” Network

- S. cerevisiae & S. paradoxus: 6000 genes
- C. elegans & C. briggsae: 19000 genes
- D. melanogaster & D. pseudoobscura: 13000 genes

Hahn et al. 2005, Demuth et al. in prep.

Thanks

Andy Kern (UCSC)
Andreas Wagner (UNM)

Dave Begun
Corbin Jones
Jason Mezey
Leonie Moyle
Jason Stajich
Kevin Thornton