Orthologs conserve function and paralogs change function

- Introduction
  - what is function: "aspects of function" (what do you mean function, why does it happen)
  - Orthology conjencture

- Evolution (change) of function after gene duplication
  - Subfunctionalization
  - Neofunctionalization

- Orthologs tend to generally conserve function
  - Phylogenetic profiles
  - Comparative interactomics
    - Conserved interaction not always means conservation of "mode" of interaction (neutrality/variation within conservation)
 - Evolution of subcellular localization

Evolution of function

- Both this lecture and the lecture on "Phylogenetic profiles exceptions and evolution of regulation" deal with evolution of function.

- When / how do functions of proteins evolve (change) their function. My too simplistic summary: genes change some aspect of function after duplication and much less aspects of function after speciation.

- But also specify what you mean by function
  - "aspects of function" (what do you mean function, why does it happen)
    - Molecular function (homology), Module (pathway/complex), Expression regulation,
    - e.g. two proteins maintain their interaction but the interaction surface by which they interact evolves

"the orthology conjecture"

- Function is "the same" for orthologs and different for "paralogs"
- … it is not just "time" (i.e. outparalogs vs orthologs)

CKK
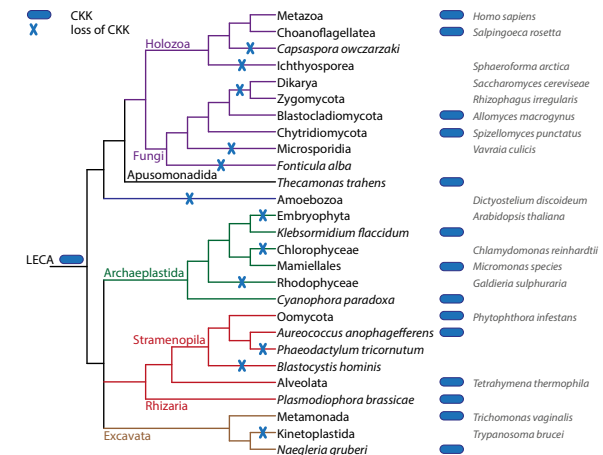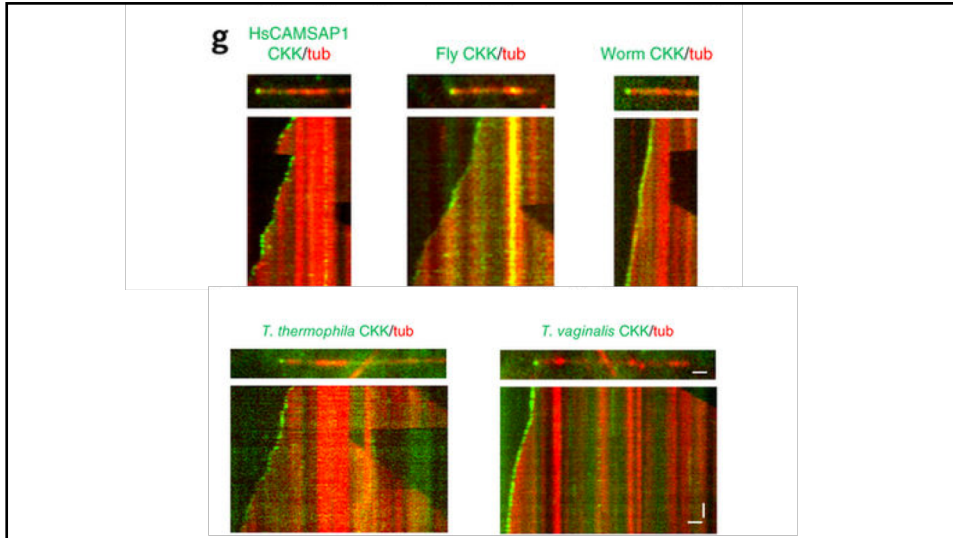
The Ortholog Conjecture Is Untestable by the Current Gene Ontology but Is Supported by RNA Sequencing Data

Xiaoshu Chen, Jianzhi Zhang*

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America

Abstract

The ortholog conjecture posits that orthologous genes are functionally more similar than paralogous genes. This conjecture is a cornerstone of phylogenomics and is used daily by both computational and experimental biologists in predicting, interpreting, and understanding gene functions. A recent study, however, challenged the ortholog conjecture on the basis of experimentally derived Gene Ontology (GO) annotations and microarray gene expression data in human and mouse. It instead proposed that the functional similarity of homologous genes is primarily determined by the cellular context in which the genes act, explaining why a greater functional similarity of (within-species) paralogs than (between-species) orthologs was observed. Here we show that GO-based functional similarity between human and mouse orthologs, relative to that between paralogs, has been increasing in the last five years. Further, compared with paralogs, orthologs are less likely to be included in the same study, causing an underestimation in their functional similarity. A close examination of functional studies of homologs with identical protein sequences reveals experimental biases, annotation errors, and homology-based functional inferences that are labeled in GO as experimental. These problems and the temporary nature of the GO-based finding make the current GO inappropriate for testing the ortholog conjecture. RNA sequencing (RNA-Seq) is known to be superior to microarray for comparing the expressions of different genes or in different species. Our analysis of a large RNA-Seq dataset of multiple tissues from eight mammals and the chicken shows that the expression similarity between orthologs is significantly higher than that between within-species paralogs, supporting the ortholog conjecture and refuting the cellular context hypothesis for gene expression. We conclude that the ortholog conjecture remains largely valid to the extent that it has been tested, but further scrutiny using more and better functional data is needed.
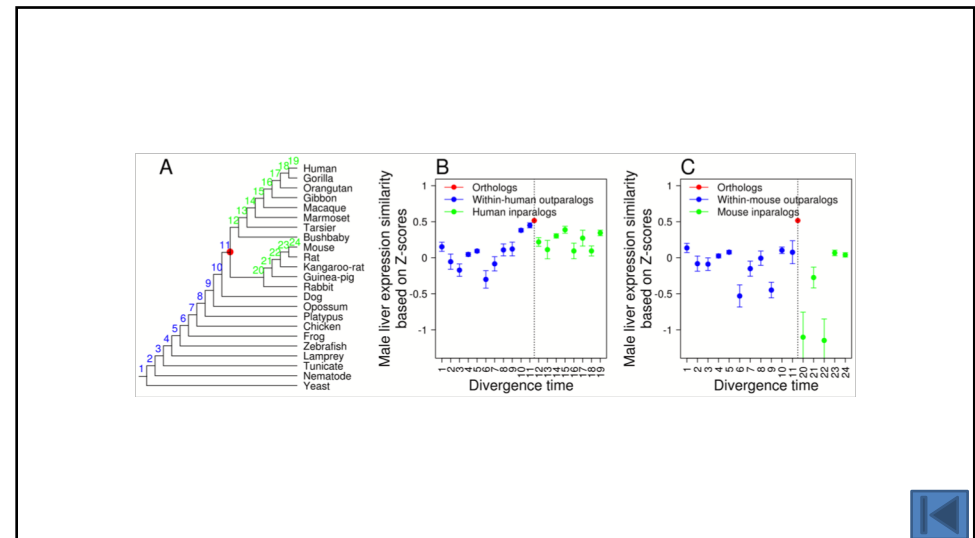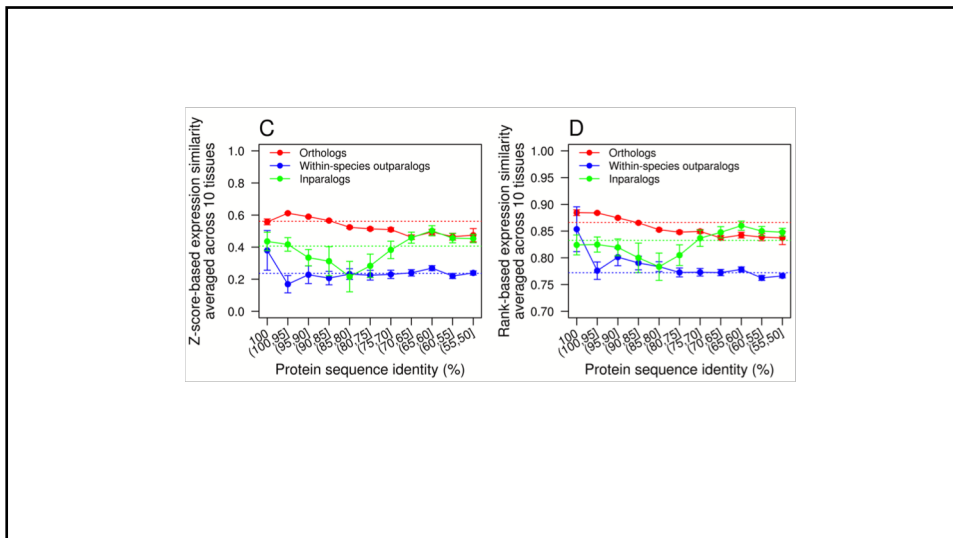
What happens after gene duplicaton?

- Gene Loss
- Subfunctionalization, e.g. Duplication Degeneration Complementation (DDC) modelions
  – One ancestral protein **specializes** into **sub**functions
  – Initially "neutral"
- Neofunctionalization,
  – One paralog is free to evolve a "new function"
  – adaptive

---

subfunctionalization: example in terms of protein complexes 1 (and ancestral multifunctional protein)
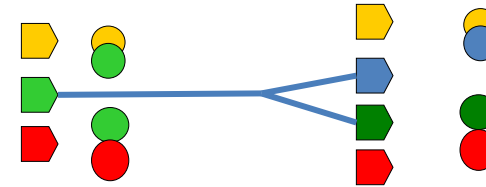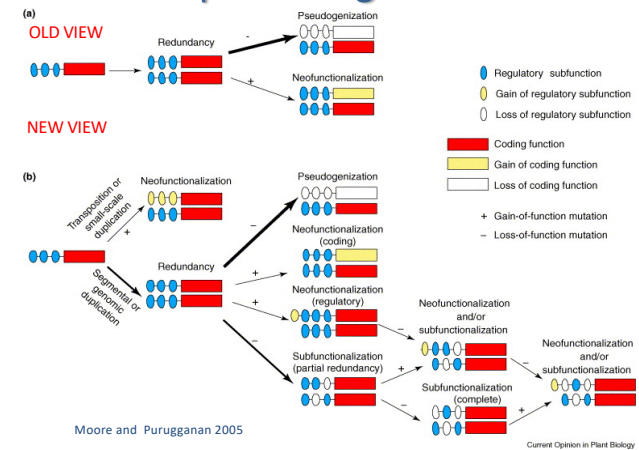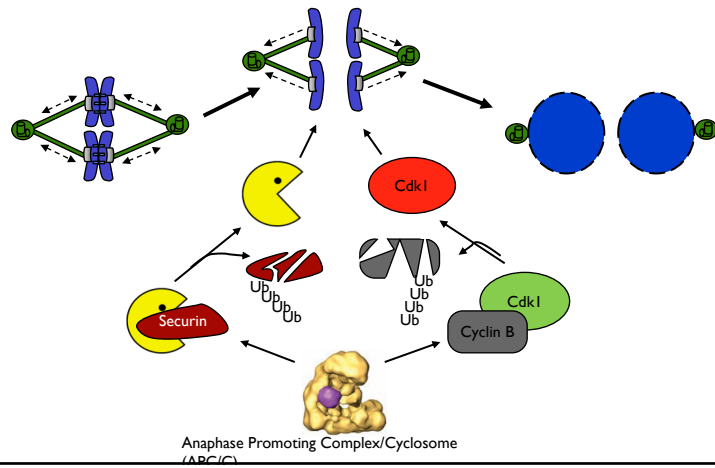

Duplication Degeneration Complementation: https://www.ncbi.nlm.nih.gov/pubmed/10101175

---

neofunctionalization: example in terms of protein complexes

Subfunctionalization vs neofunctionalization in a transcriptional regulation context
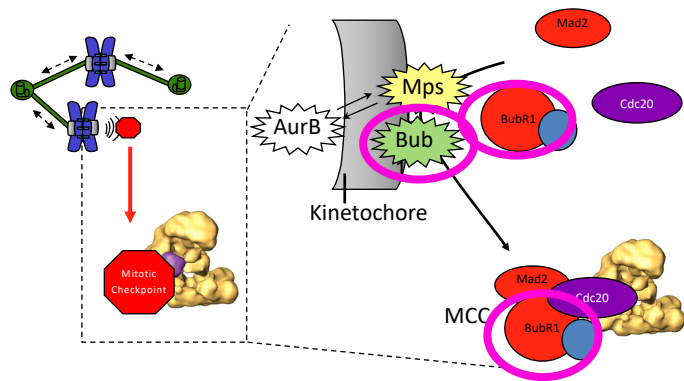

Moore and Purugganan 2005

Case story Spindle Assembly CheckpointInitiating Anaphase

Cdk1

Securin

Ub Ub Ub Ub

Ub Ub Ub Ub

Cdk1

Cyclin B

Anaphase Promoting Complex/Cyclosome
(APC/C)

The Spindle Assembly Checkpoint

Securin

Cdk1

Cyclin B

Mitotic Checkpoint
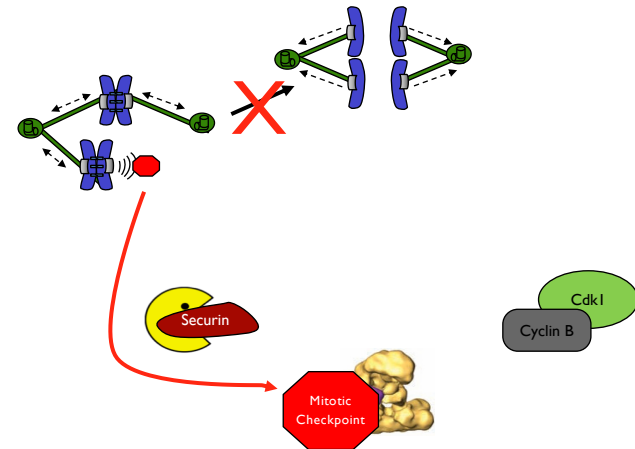
The Spindle Assembly *Complex*

Mad2

Mps

AurB

Bub

BubR1

Cdc20

Kinetochore

Mitotic Checkpoint

Mad2

Cdc20

MCC

BubR1

scMad3p
(fungi)

KEN   TPR   BUB3-binding

hsBubR1
(vertebrates)

KEN   TPR   BUB3-binding   Kinase
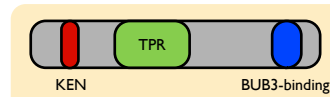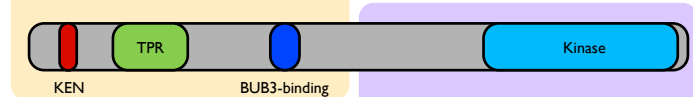
ScBub1
HsBub1
(fungi & vertebrates)

TPR   BUB3-binding   CDI   Kinase

Slide 1

HsBUB1 / ScBub1p — TPR GLEBS KINASE
HsBUBR1 — KEN TPR GLEBS KINASE
ScMad3p — KEN TPR GLEBS
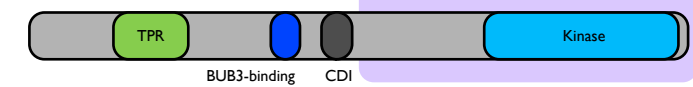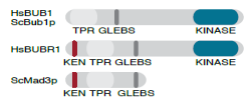
9 independent duplications.
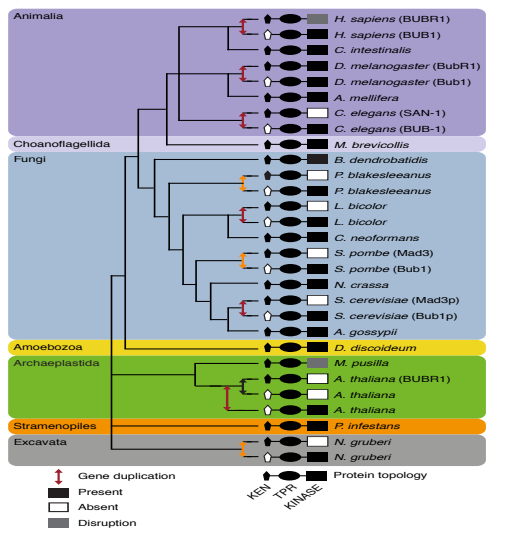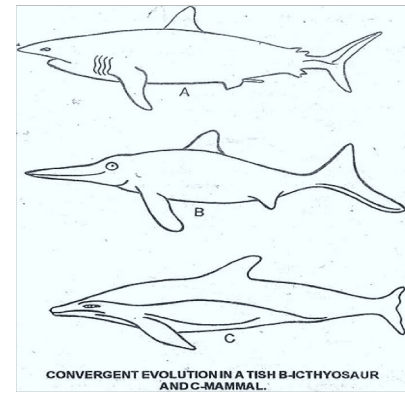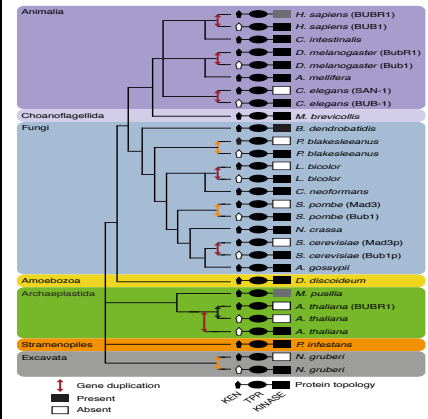7 cases where a mad3-like and a bub1-like protein arose out of a bubmad-like ancestor.

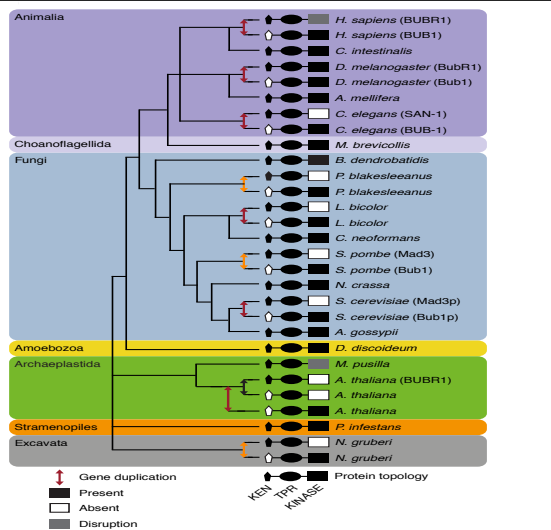Suijkerbuijk *et al.*, Snel and Kops *Dev Cell* 2012

Animalia
H. sapiens (BUBR1)
H. sapiens (BUB1)
C. intestinalis
D. melanogaster (BubR1)
D. melanogaster (Bub1)
A. mellifera
C. elegans (SAN-1)
C. elegans (BUB-1)
Choanoflagellida
M. brevicollis
Fungi
B. dendrobatidis
P. blakesleeanus
P. blakesleeanus
L. bicolor
L. bicolor
C. neoformans
S. pombe (Mad3)
S. pombe (Bub1)
N. crassa
S. cerevisiae (Mad3p)
S. cerevisiae (Bub1p)
A. gossypii
Amoebozoa
D. discoideum
Archaeplastida
M. pusilla
A. thaliana (BUBR1)
A. thaliana
A. thaliana
Stramenopiles
P. infestans
Excavata
N. gruberi
N. gruberi

Gene duplication
Present
Absent
Disruption
Protein topology
KEN TPR KINASE

Slide 2

Recurrent (convergent/parallel) evolution in molecular systems!

CONVERGENT EVOLUTION IN A TISH B-ICTHYOSAUR AND C-MAMMAL.

Slide 3

HsBUB1 / ScBub1p — TPR GLEBS KINASE
HsBUBR1 — KEN TPR GLEBS KINASE
ScMad3p — KEN TPR GLEBS

What about the kinase domain in human (and fly)

Slide 4

What about the kinase domain in human bubr1?

	Gly-rich loop I	II	III	Catalytic loop VI-B	VII
Consensus	G x G x x G x V	V A I K	E	H R D x K x x N	D F G
PKA	G T G S F G R V	Y A M K	E	Y R D L K P E N	D F G
HsBUB1	G E G A F A Q V	F V L K	E	H G D I K P D N	D L G
HsBUBR1	C E D Y K L F –	T V I K	D	H G D L S P R C	D F S

degeneration of motifs essential for catalysis

Further experiments showed vertebrates are not exception. The kinase domain of BubR1 lacks enzymatic activity.

"This explained the field's inability to identify substrates of BubR1, and dispelled a leading theory of SAC silencing based on inactivation of BubR1 after kinetochore-microtubule attachment."

Panel 1 (top left)

9 independent duplications.
8 cases where a bub1-like protein and a protein without a (functional) kinase arose: mad3, bubr1

Suijkerbuijk *et al.*, Snel and Kops *Dev Cell* 2012

Panel 2 (top right)

Extreme case of subfunctionalization as **sublocalization**: the rootlet

F. Schematic localization of SF components

https://elifesciences.org/articles/17557

Panel 3 (bottom left)

ROOTLETIN?

Chlamydomonas SFA
Tetrahymena SFAL
GROUP 1

CROP1
KDF3
DisA
GROUP 2p

BBC39
KDF4
KdD6
BBC29
KDF1
MDLG1
GROUP 2d

Panel 4 (bottom right)

F. Schematic localization of SF components

Subfunctionalization in space

Ancestor, green algae & other ciliates

Paramecium

"Paralog switching is a widespread mechanism that modulates protein complex composition" Subfunctionalization in time/condition

Genome Biology

RESEARCH   Open Access

Spatiotemporal variation of mammalian protein complex stoichiometries

Alessandro Ori[1,4†], Murat Iskar[1,5†], Katarzyna Buczak[1], Panagiotis Kastritis[1], Luca Parca[1], Amparo Andrés-Pons[1], Stephan Singer[1,2], Peer Bork[1,3*] and Martin Beck[1*]
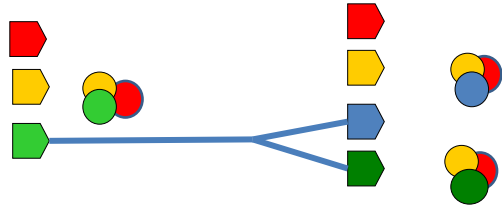
Abstract

Background: Recent large-scale studies revealed cell-type specific proteomes. However, protein complexes, the basic functional modules of a cell, have been so far mostly considered as static entities with well-defined structures. The co-expression of their members has not been systematically charted at the protein level.

Results: We used measurements of protein abundance across 11 cell types and five temporal states to analyze the co-expression and the compositional variations of 182 well-characterized protein complexes. We show that although the abundance of protein complex members is generally co-regulated, a considerable fraction of all investigated protein complexes is subject to stoichiometric changes. Compositional variation is most frequently seen in complexes involved in chromatin regulation and cellular transport, and often involves paralog switching as a mechanism for the regulation of complex stoichiometry. We demonstrate that compositional signatures of variable protein complexes have discriminative power beyond individual cell states and can distinguish cancer cells
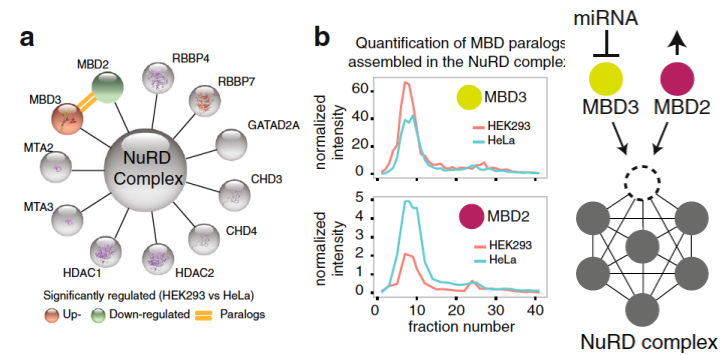

Paralog switching

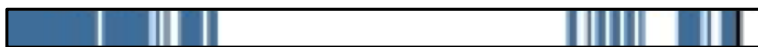## subfunctionalization: example in terms of protein complexes 2

Example NurRD

a NuRD Complex — MBD2, RBBP4, RBBP7, GATAD2A, CHD3, CHD4, HDAC2, HDAC1, MTA3, MTA2, MBD3

Significantly regulated (HEK293 vs HeLa)
● Up- ● Down-regulated ▬ Paralogs

b Quantification of MBD paralogs assembled in the NuRD complex

MBD3 — HEK293, HeLa
MBD2 — HEK293, HeLa

normalized intensity vs fraction number

miRNA
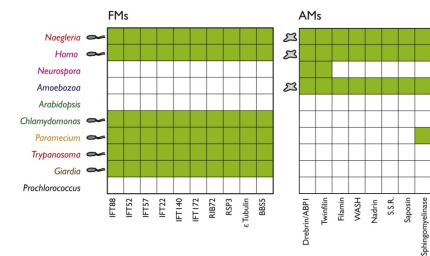MBD3 MBD2
NuRD complex

---

presence/absence patterns, (co-)occurrence, phyletic patterns

PHYLOGENETIC PROFILES OF ORTHOLOGS

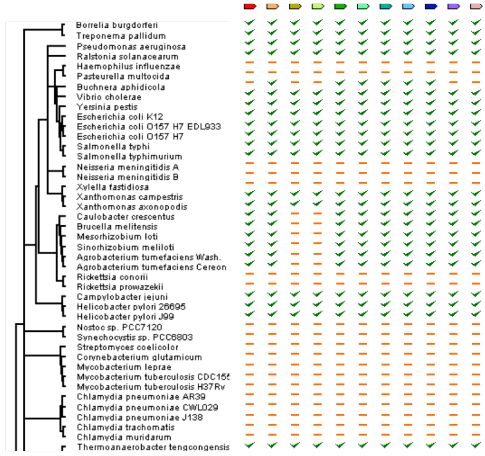## Phylogenetic profiles allow us to see co-evolving modules

• Phylogenetic profile= Presence and absence of genes (orthologs) across species

FMs: IFT88, IFT52, IFT57, IFT22, IFT140, IFT172, RIB72, RSP3, ε-Tubulin, BBS5

AMs: Drebrin/ABP1, Twinfilin, Filamin, WASH, Nadrin, S.S.R., Saposin, Sphingomyelinase

Naegleria, Homo, Neurospora, Amoebazoa, Arabidopsis, Chlamydomonas, Paramecium, Trypanosoma, Giardia, Prochlorococcus

Co-occurrence of genes across genomes as prediction for interaction / association

Borrelia burgdorferi
Treponema pallidum
Pseudomonas aeruginosa
Ralstonia solanacearum
Haemophilus influenzae
Pasteurella multocida
Buchnera aphidicola
Vibrio cholerae
Yersinia pestis
Escherichia coli K12
Escherichia coli O157 H7 EDL933
Escherichia coli O157 H7
Salmonella typhi
Salmonella typhimurium
Neisseria meningitidis A
Neisseria meningitidis B
Xylella fastidiosa
Xanthomonas campestris
Xanthomonas axonopodis
Caulobacter crescentus
Brucella melitensis
Mesorhizobium loti
Sinorhizobium meliloti
Agrobacterium tumefaciens Wash.
Agrobacterium tumefaciens Cereon
Rickettsia conorii
Rickettsia prowazekii
Campylobacter jejuni
Helicobacter pylori 26695
Helicobacter pylori J99
Nostoc sp. PCC7120
Synechocystis sp. PCC6803
Streptomyces coelicolor
Corynebacterium glutamicum
Mycobacterium leprae
Mycobacterium tuberculosis CDC155
Mycobacterium tuberculosis H37Rv
Chlamydia pneumoniae AR39
Chlamydia pneumoniae CWL029
Chlamydia pneumoniae J138
Chlamydia trachomatis
Chlamydia muridarum
Thermoanaerobacter tengcongensis

- i.e. two genes have the same presence/ absence pattern over multiple genomes:
- AKA phylogenetic profiles
- NB complete genomes absence -> needed for absence
- Correction for phylogenetic signal needed → events

b

Predicting function of a disease gene protein with unknown function, frataxin, using co-occurrence of genes across genomes / phylogenetic profiles
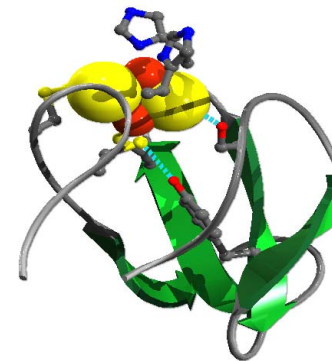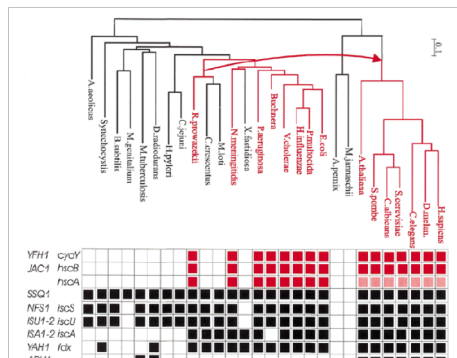
- Friedreich's ataxia
- No (homolog with) known function
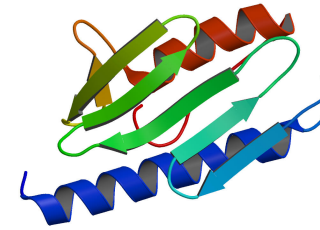
Predicting function of a disease gene protein with unknown function, frataxin, using co-occurrence of genes across genomes

- Friedreich's ataxia
- No (homolog with) known function

Frataxin has co-evolved with hscA and hscB indicating that it plays a role in iron-sulfur cluster assembly

---

Iron-Sulfur (2Fe-2S) cluster in the Rieske protein

The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly

Martijn A. Huynen*, Berend Snel[1], Peer Bork and Toby J. Gibson[1]

Biocomputing, EMBL/Max-Delbrueck-Center fur molecular medicin, Berlin-Buch and [1]Biocomputing, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received July 3, 2001; Revised and Accepted July 30, 2001

~Confirmation:

JACS

ARTICLES

Published on Web 04/26/2003

Iron−Sulfur Cluster Biosynthesis. Characterization of Frataxin as an Iron Donor for Assembly of [2Fe-2S] Clusters in ISU-Type Proteins

Taejin Yoon and J. A. Cowan*

Contribution from Evans Laboratory of Chemistry, The Ohio State University,
100 West 18th Avenue, Columbus, Ohio 43210

Received August 1, 2002; E-mail: cowan@chemistry.ohio-state.edu

---
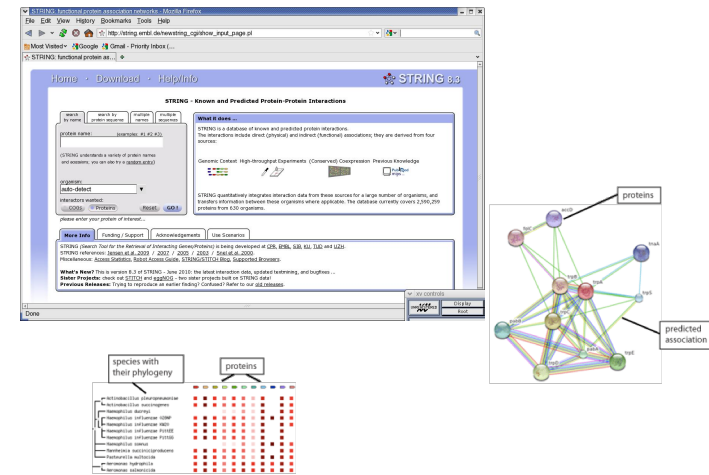
The fact that phylogenetic profiles also work in eukaryotes shows that independent gene loss is not random

---
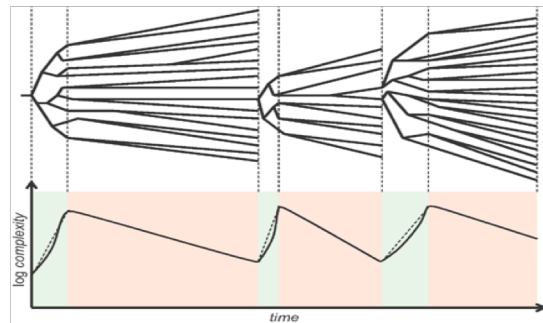
Presence-absence matrix of kinetochore proteins: recurrent/independent loss of ancestral kinetochore proteins
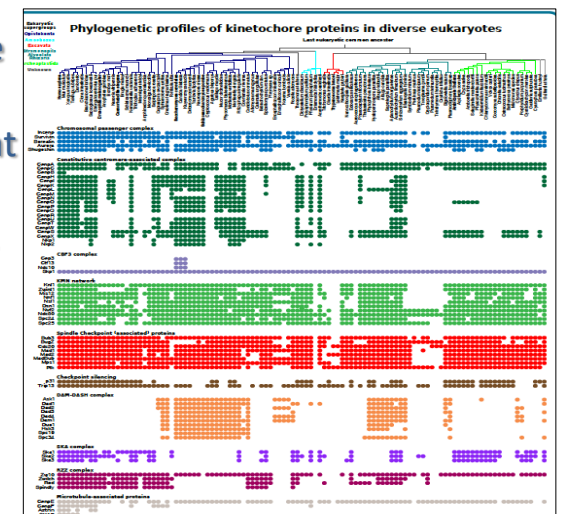
t-SNE Projection of matrix: Recurrent loss is not random: co-loss, dispensability of (sub-)complexes
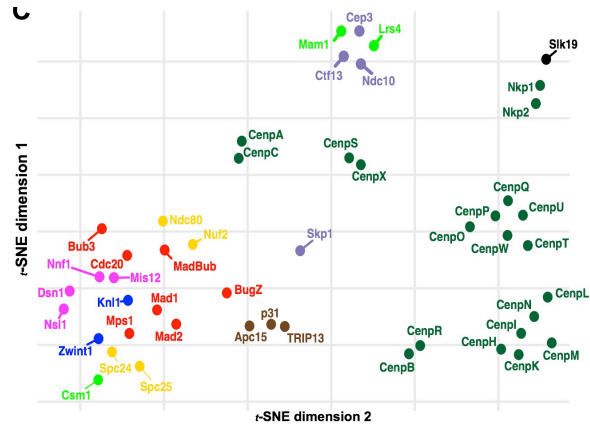
## Which cellular systems have similar phylogenetic profiles? (co-evolve?)

SUMMARY

The availability of diverse genomes makes it possible to predict gene function based on shared evolutionary history. This approach can be challenging, however, for pathways whose components do not exhibit a shared history but rather consist of distinct "evolutionary modules." We introduce a computational algorithm, clustering by inferred models of evolution, distribution of pathways and thereby highlight model organisms for experimental studies. Such evolutionary analyses may also teach us about the environmental niches within which they evolved. Importantly, correlated gains and losses can help to predict the function of unstudied genes and also reveal alternative functions for genes considered to be well characterized.

Pioneering work introduced the concept of "phylogenetic profiling" to chart the phylogenetic distribution of genes and
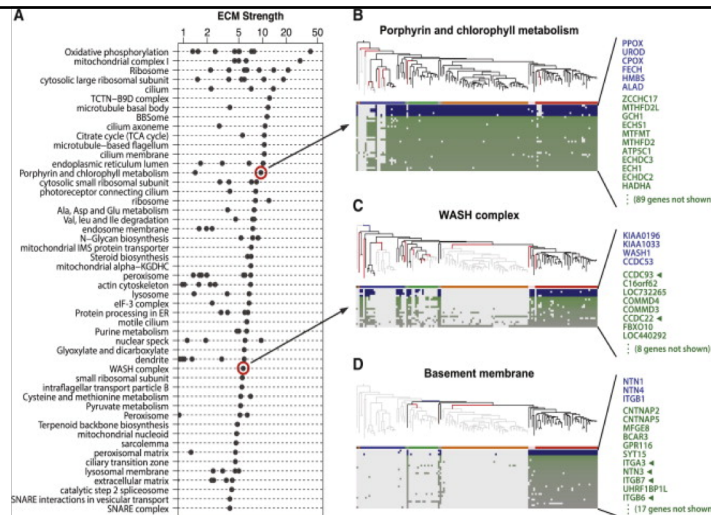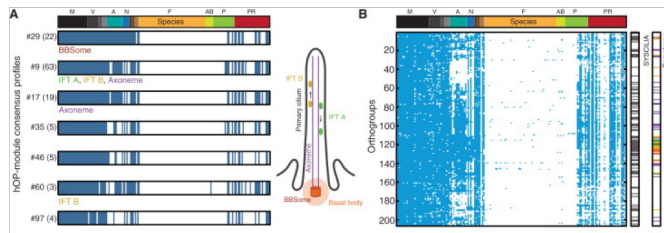
The happy families
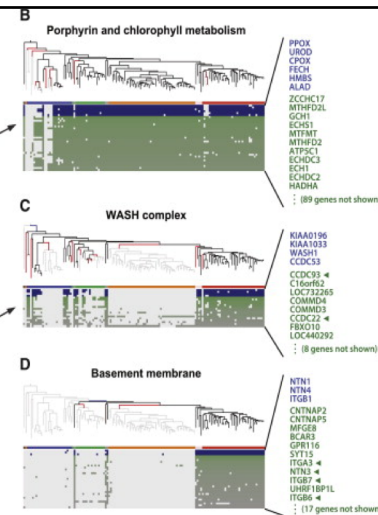
What can we learn about the evolution of function from understanding phylogenetic profiles of basal cellular processes?

Slide 1: Occurrence of similar phylogenetic profiles: evolution of function (vs conservation?)

- It seems as if for these systems, e.g. flagellar proteins, their function and interactions are largely conserved: "orthology conjecture holds"
- Some lineages lost need for it, or were in fact more fit without it. *Evolution by loss.* **No** evolutionary change in "function". **Conservation.**

Slide 2

NB (for later) Not all functional modules are perfectly co-evolving; limited cohesiveness; / disrupted co-evolution / discordant phylogenetic profiles

peptidoglycan biosynthesis pathway (highly cohesiveness, far from perfect)

ribose phosphate metabolism (not cohesive at all)

What is going on with the function of these genes? ("orthology conjencture"?) Why?

Slide 3: comparative genomics of high throughput data between species and evolution of function

- Evolution of function
- Orthology conjecture
  - for what aspects of function from model organism to e.g. human is orthology equals function "true"
- Some studies suggest interactions evolve quite rapidly between species, e,g, only 10% overlap fly-yeast (Suthram et al. Nature 2005)
- (What happens to the function of duplications)
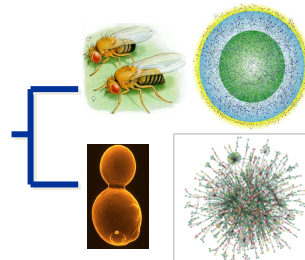
Slide 4: Integration between species / conservation

orthologs
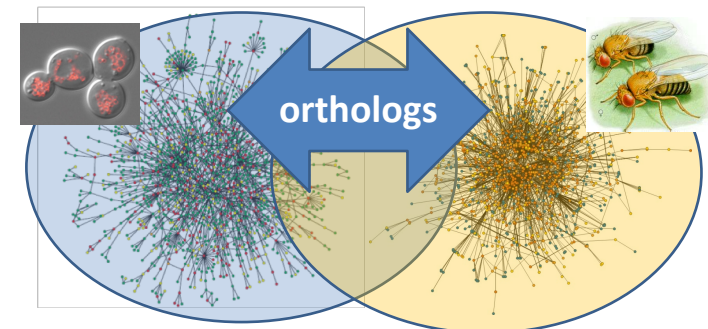
Accuracy of Y2H and how to improve it BUT coverage: real divergence?
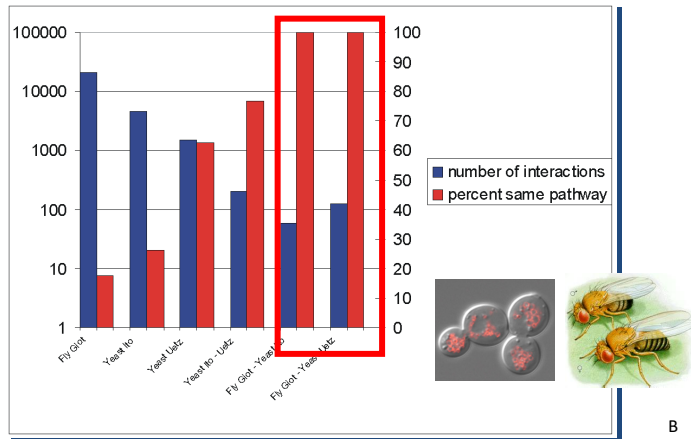
false positives (FP) and false negatives (FN), noise / incomplete knowledge, are stacked against detecting conservation

- Genes falsely selected as interacting in species A lower the conservation level, while genes falsely *not* selected (i.e. FN) do not
- Genes falsely selected to be *not* interacting in species B lower the conservation level (FN). Detecting absence?
  - E.g. strict co-expression threshold leads to many false negatives

How to perform comparative genomics of interactions (networks/interactome)

orthology
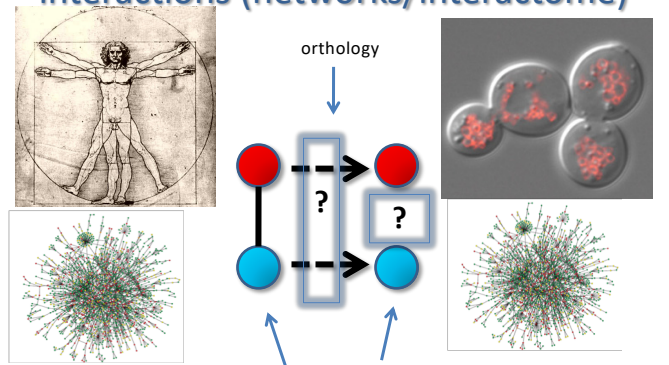
Reliability and coverage of data (false positives, false negatives)
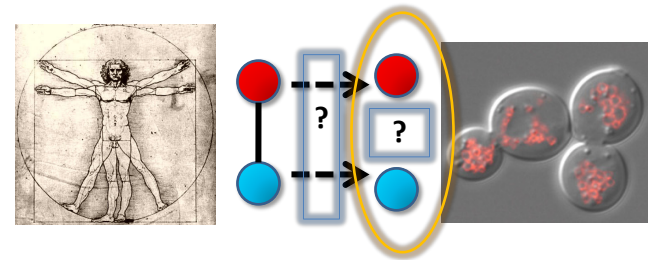
If two proteins are part of the same complex in human how often are they also part of the same complex in yeast

Interactions?

- - stable interactions such as in complexes like ribosomes and proteosomes or between subunits of an enzyme, etc.
- - labile interactions such as between kinases to their substrates, phophatase to their substrates

- Because of quality of expert curation (to use as source or reference) and most prolific HTP data (complex purification) complexes

---

Absence of interaction … ?

Conservation =

$$\frac{\text{number of interactions conserved}}{(\text{number of interactions conserved} + \text{number of interactions NOT conserved})}$$

TAP-MS data from krogan, and gavin unprecedented coverage so that failure to report co-purification might really mean absence of co-complex membership

---

Estimating absence of interactions from HTP data -> yeast; what do we call an absence of interaction data
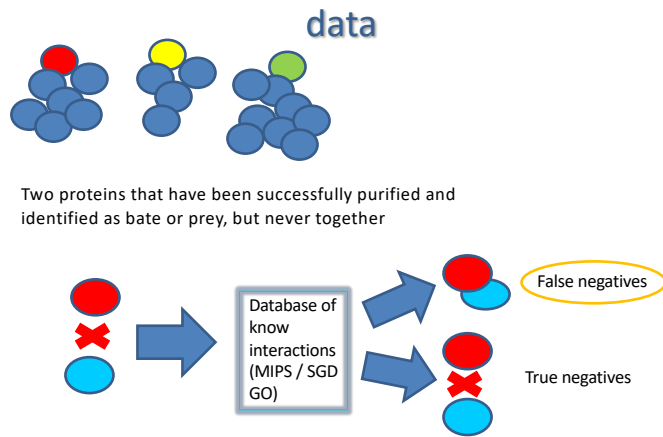
Two proteins that have been successfully purified and identified as bate or prey, but never together

Database of know interactions (MIPS / SGD GO)

False negatives

True negatives

The **false negative rate** (FNR) is the proportion of positive instances that were erroneously reported as negative.

Proten pairs known to be part of the same protein complex

Datasets	FNR	#FN	#TP
Gavin et al.	0.23	1226	4083
Krogan et al.	0.32	2209	4644

Positive instances: co-complex relation in MIPS *and* SGD-GO. Similarly negative instances,: two proteins known to be involved in complexes in MIPS and SGD-GO but in either ref never together
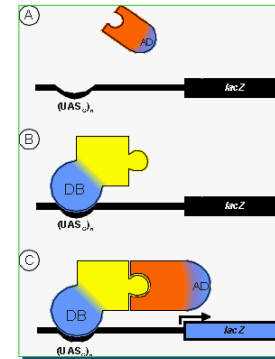
The **false negative rate** (FNR) is the proportion of positive instances that were erroneously reported as negative.

Proten pairs known to be part of the same protein complex

| Datasets	FNR	#FN	#TP
Gavin et al.	0.23	1226	4083
Krogan et al.	0.32	2209	4644
Intersection	0.11	517	4396
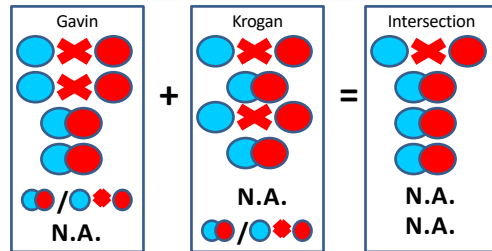
... but Y2H != TAP-MS

How do we know which bait prey pairs (hybridizations) have been "properly" tested?

Only count as not interactions pairs where both proteins have been successful as bait and prey

Datasets	FNR	#FN	#TP
Gavin et al.	0.23	1226	4083
Krogan et al.	0.32	2209	4644
Intersection	0.11	517	4396
Uetz et al.	0.66	91	46
Ito et al.	0.92	822	76
Uetz et al. strict	0.1	5	46
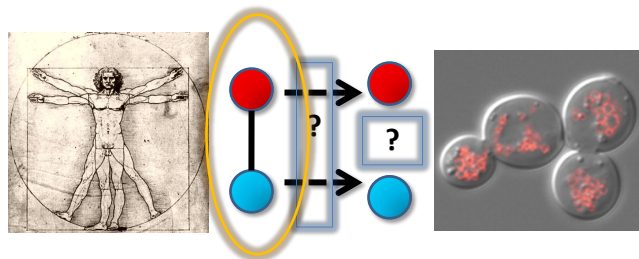
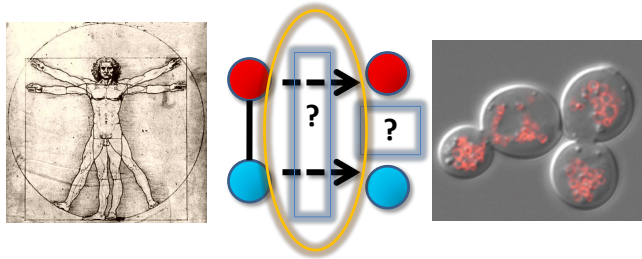## In human less htp and less curation = less coverage
(reason for assymetrry)

- High quality, non comprehensive literature curation: reactome direct complex: 5960 co-complex pairs
- Some 2h but even worse than yeast 2h
- new HTP data:

Large-scale mapping of human protein–protein interactions by mass spectrometry

Rob M Ewing[1,2], Peter Chu[1,5], Fred Elisma[3], Hongyan Li[1,6], Paul Taylor[1,7], Shane Climie[1,8], Linda McBroom-Cerajewski[1,9], Mark D Robinson[1,10], Liam O'Connor[1,11], Michael Li[1,12], Rod Taylor[1], Moyez Dharsee[1,2], Yuen Ho[1,13], Adrian Heilbut[1,14], Lynda Moore[1,15], Shudong Zhang[1], Olga Ornatsky[1,16], Yury V Bukhman[1,17], Martin Ethier[3], Yinglun Sheng[3], Julian Vasilescu[3], Mohamed Abu-Farha[3], Jean-Philippe Lambert[3], Henry S Duewel[1,18], Ian I Stewart[1], Bonnie Kuehl[1,19], Kelly Hogue[1,20], Karen Colwill[1,21], Katharine Gladwish[1], Brenda Muskat[1,22], Robert Kinach[1,16], Sally-Lin Adams[1,23], Michael F Moran[1,7], Gregg B Morin[1,15], Thodoros Topaloglou[1,4] and Daniel Figeys[1,3,*]

Orthology: complication

Y1 Y1.1 H1 Y2 H2

● Speciation
■ Duplication

H1 → Y1 Y1.1

?

H1 and H1.1 are "Inparalogs"
H1 and H1.1 are "Co-orthologous" to Y1

BBH only (inparanoid's main ortholog)

H1 H1.1 Y1 Y 1 H2

● Speciation
■ Duplication

What can happen to an interaction in evolution

? ?

Slide 1
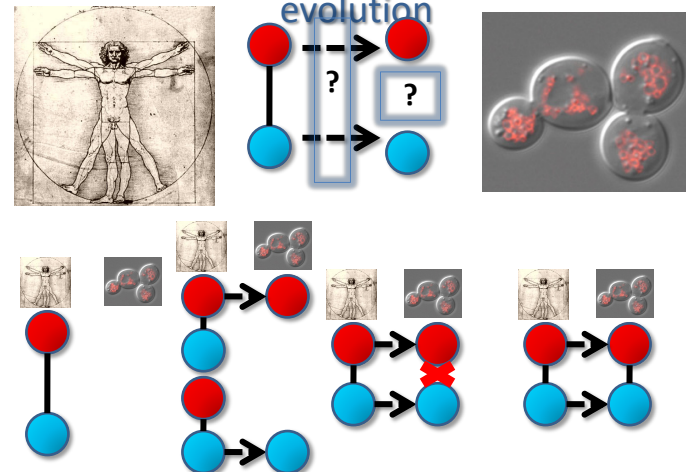
Many interactions are not "conserved" because the genes themselves are not conserved

BestHit	2276	1417	2267
InParanoid	1916	1448	2596
Ensembl	2216	1828	1916

Slide 2
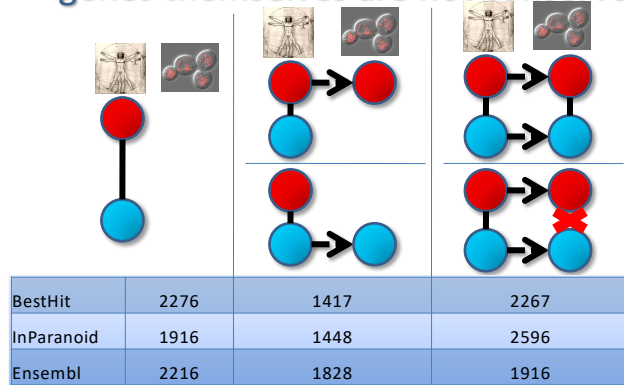
If both genes are conserved the interactions also tends to be conserved

Dataset	int	Non-int	Conservation	Coverage
Gavin	1305	226	85.2%	68.1%
Krogan	1547	328	82.5%	80.7%
Intersection	1392	75	94.9%	72.7%
Uetz	21	63	25.0%	1.1%
UetzInt	21	4	84.0%	1.1%

Slide 3
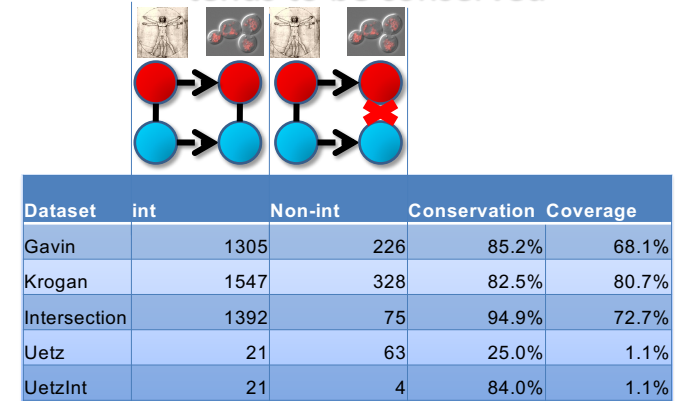
Human HTP data (Ewing, IP-HTMS)

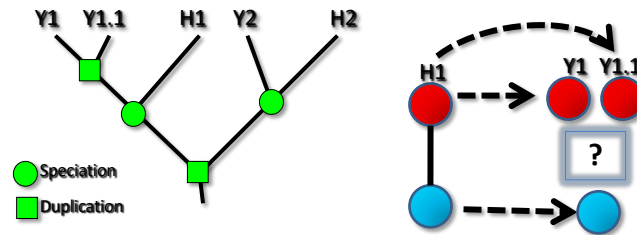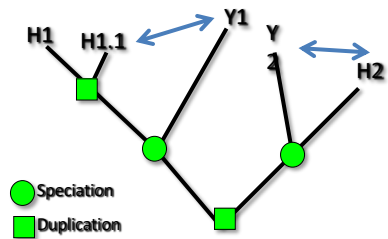| Ewing cut-off	int	non-int	conservation
0	117	245	32.32
0.3	78	59	56.93
0.5	20	3	86.96

Yeast data set = intersection

Large-scale mapping of human protein–protein interactions by mass spectrometry

Rob M Ewing[1,2], Peter Chu[1,5], Fred Elisma[3], Hongyan Li[1,6], Paul Taylor[1,7], Shane Climie[1,8], Linda McBroom-Cerajewski[1,9], Mark D Robinson[1,10], Liam O'Connor[1,11], Michael Li[1,12], Rod Taylor[1], Moyez Dharsee[1,2], Yuen Ho[1,13], Adrian Heilbut[1,14], Lynda Moore[1,15], Shudong Zhang[1], Olga Ornatsky[1,16], Yury V Bukhman[1,17], Martin Ethier[3], Yinglun Sheng[3], Julian Vasilescu[3], Mohamed Abu-Farha[3], Jean-Philippe Lambert[3], Henry S Duewel[1,18], Ian I Stewart[1,2], Bonnie Kuehl[1,19], Kelly Hogue[1,20], Karen Colwill[1,21], Katharine Gladwish[1], Brenda Muskat[1,22], Robert Kinach[1,16], Sally-Lin Adams[1,23], Michael F Moran[1,7], Gregg B Morin[1,15], Thodoros Topaloglou[1,4] and Daniel Figeys[1,3,*]

Slide 4

- Some conserved interactions are missed when taking the sequence-wise most similar ortholog cf. Notebaart2005 / Ideker 2006: i.e. limits of Bidirectional Best Hits
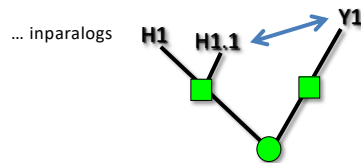
Different results of orthologies

Dataset	Orthology	int	Non-int	Conservation	Coverage
Intersection	BBH	1429	141	91.0%	63.0%
Intersection	Ensembl	1392	75	94.9%	72.7%
Intersection	InParanoid	1761	84	95.4%	67.8%

Inparanoid / ensembl similar conservation percentages despite different absolute values, But BBH lower

… inparalogs   H1   H1.1   Y1

Non-conserved interactions …

- Curation errors in reactome
- potential false negatives in HTP data as literature in yeast says the two do interact.
- Our high level of conservation is underestimation?
- few cases of genuine evo divergence … (e.g. new paralog in human involved in a new complex, human PCBP1 & yeast XAB2)
- flexibility resides in duplications cf. inparalogs

Cell   Resource

A Protein Complex Network of *Drosophila melanogaster*

K.G. Guruharsha,[1,4] Jean-François Rual,[1,4] Bo Zhai,[1,4] Julian Mintseris,[1,4] Pujita Vaidya,[1] Namita Vaidya,[1] Chapman Beekman,[1] Christina Wong,[1] David Y. Rhee,[1] Odise Cenaj,[1] Emily McKillip,[1] Saumini Shah,[1] Mark Stapleton,[2] Kenneth H. Wan,[2] Charles Yu,[2] Bayan Parsa,[2] Joseph W. Carlson,[2] Xiao Chen,[2] Bhaveen Kapadia,[2] K. VijayRaghavan,[3] Steven P. Gygi,[1] Susan E. Celniker,[2] Robert A. Obar,[1,*] and Spyros Artavanis-Tsakonas[1,*]

- "Our data support models of protein network evolution that are driven by the acquisition or loss of protein complex members rather than rewiring of existing components (van Dam and Snel, 2008 and Yamada and Bork, 2009)."

Summarizing conclusions

- Most interactions are not conserved because of acquisition / loss subunits but if two proteins *are* present they tend to interact (supports orthology conjencture)
- Despite issues, >> 10% previously implied
- Function prediction from model organism to man is justified w.r.t. co-complex membership
- Differences between species reside perhaps more in genome evolution than in new stable protein-protein interactions …
- Genome and network evolution are tightly connected and should not be studied independently (e.g. the simple distinction between loss/gain of interaction with existing protein vs loss/gain of interactor.)
- However also non-co-occurring proteins!

Slide 1

Conservation of interaction does not always mean conservation of interaction mode/surface (i.e. change & evolution) (also relevant for literature discussion paper)

Resource

Cell

A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human

Graphical Abstract

FissionNet — 2,278 Interactions
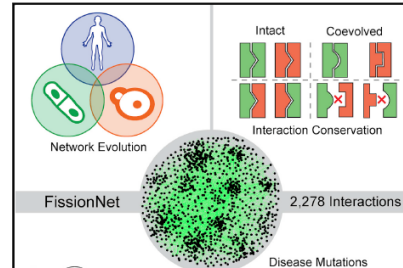
Authors
Tommy V. Vo, Jishnu Das, Michael J. Meyer, ..., Jeffrey A. Pleiss, Yu Xia, Haiyuan Yu

Correspondence
haiyuan.yu@cornell.edu

In Brief
FissionNet is a proteome-wide binary interactome network for *S. pombe*. Comparative analyses of FissionNet with protein networks in budding yeast and human reveal how protein networks evolve, principles of gene repurposing

Slide 2

Another aspect of function: subcellular localization

Research

Open Access

Expansion of the human mitochondrial proteome by intra- and inter-compartmental protein duplication
Radek Szklarczyk and Martijn A Huynen

1-to-1 human-yeast orthologs have conserved ancestral subcellular localization.
Gene duplication relaxes this constraint
•Quite some **intra**-mitochondrial duplications
•And **inter**-compartmental duplications create novel mitochondrial localization of the protein encoded by one of the daughter genes

Slide 3

1-to-1 human-yeast orthologs have conserved ancestral subcellular localization

• Use high quality data in localization: experimental identification, bioinformatics analysis, and literature curation

• "Of 143 one-to-one orthologous pairs localized to mitochondria in either of the two species, we find that 124 proteins (87%) are found in this organelle in both species and only 19 proteins localize to mitochondria in one species, but not the other"

Slide 4

intra-mitochondrial duplications are most frequent & gain of mitochondrial localization after gene duplication

Table I

Duplications in gene families with products localized to the mitochondria

Human localization of gene family	Yeast localization of gene family	Number of families	Number of human proteins
Mitochondrial	Mitochondrial	53	118
Mitochondrial and non-mitochondrial	Non-mitochondrial	26	101
Other	Other	25	55

'Mitochondrial' denotes mitochondrial localization for all genes from this family in a species; 'non-mitochondrial' indicates a localization to another subcellular compartment; 'mitochondrial and non-mitochondrial' indicates with both mitochondrial and non-mitochondrial paralogs. See also Table S4 in Additional data file I for other duplication classes.
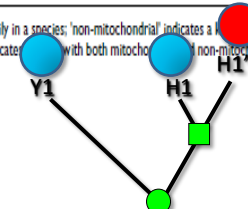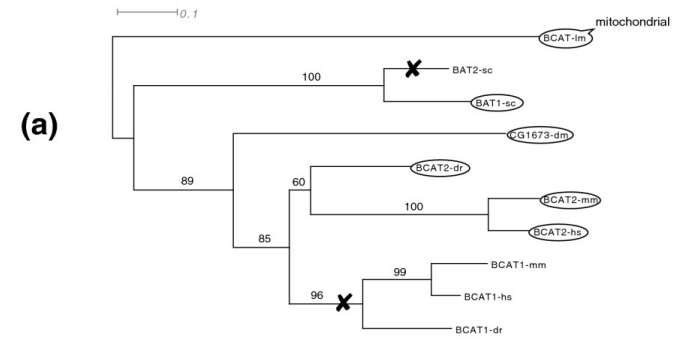
"Parallel evolution"

Table 3

Independent duplications and parallel relocalizations in the human and yeast lineages have happened multiple times during evolution

Family	Human		Yeast	
	Mitochondrial	Non-mitochondrial	Mitochondrial	Non-mitochondrial
Thioredoxins	TXN, TXN2	TXNDC2	TRX3	TRX1, TRX2
Glutaredoxins	GLRX2	GLRX, GLRXL	GRX2	GRX1 (nucleus)
Isocitrate dehydrogenases [NADP]	IDH2	IDH1	IDP1	IDP2, IDP3 (peroxisome)
Branched-chain-amino-acid aminotransferases	BCAT2	BCAT1	BAT1	BAT2
Serine hydroxymethyltransferases	SHMT2	SHMT1	SHM1	SHM2

(a)

"Parallel evolution through rapid parallel loss"

(b)