Evolution of function, beyond similar phylogenetic profiles and only functional change after gene duplication

- Exceptions phylogenetic profiles
  - Retention of functionally differentiated paralogs
  - Multi-functional proteins
  - Motif-protein co-evolution
  - Anti-correlating proteins
- Evolution of regulation
  - Evolution of Genetic interactions
  - Evolution of (co-)regulation
  - Evolution of phosphorylation & summary evolution of function
- Where do novelty/innovations come from some final thoughts

Explaining discordant phylogenetic profiles of proteins that interact:

- (we could also just say that evolution is flexible and proteins change function; which I am not going to argue with but (A) conservation of interaction and (B) this is a “just so”, non testable explanation)
- “Happy families are all alike; every unhappy family is unhappy in its own way.” (from Leo Tolstoy’s book Anna Karenina, which begins with this statement)
- Case stories and large scale studies
- And what does it tell us about evolution of function?

discordant phylogenetic profiles because of lineage/group specific duplications (inparalogs) that changed their function
Reconstructing Complex I evolution by mapping the variation onto a phylogenetic tree. After an initial "surge" in complexity (from 14 to 35 subunits in early eukaryotic evolution) new subunits have been gradually added and incidentally lost, most other loss is large scale.

Huynen et al., BBA 2009
Cardol, BBA 2012
Yip et al., JBC 2011

"Exceptions" in the perfect co-evolution of Complex I
Complex I loss is not always "complete", S.cerevisiae and S.pombe have retained 1 and 3 proteins

The Complex I assembly protein CI30 has been duplicated in the Fungi. This can explain the presence of a CIA30-homolog in Complex I-less S.pombe

This principle is also recognized for phylogenetic profile function prediction.
what do we learn about evolution of function from discordant phylogenetic profiles bc of lineage specific duplications that changed their function

• Change of function after duplication. (=evolution).
• For the original protein. Evolution by loss. No change in “function”

Discordant phylogenetic profiles because of multifunctional proteins

TOR1 complex

• Kinase
• Regulates growth
• Mutations of TOR1 components involved in Cancer
Evolution of TOR

- TOR2 is involved in rearrangement of cytoskeleton

Shared Protein Complex Subunits Contribute to Explaining Disrupted Co-occurrence

Abstract

The gene composition of present-day genomes has been shaped by a complicated evolutionary history, resulting in diverse distributions of genes across genomes. The pattern of presence and absence of a gene in different genomes is called the phylogenetic profile. It has been shown that proteins whose encoding genes have highly similar profiles tend to be functionally related. As these genes were gained and lost together, their encoded proteins can probably only perform their full function if both are present. However, a large proportion of genes encoding interacting proteins do not have matching profiles. In this study, we analyzed one possible reason for this, namely that phylogenetic profiles can be affected by multi-functional proteins such as shared subunits of two or more protein complexes. We found that by considering triplets of proteins, of which one protein is multi-functional, a large fraction of disturbed co-occurrence patterns can be explained.

Adrian Schneider1, Michael F. Seidler2, Berend Snel1,2,3

1Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands, 2Centre for Biocomputational Genomics, Wageningen, The Netherlands

References


Copyright © 2012 Schneider et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
What do discordant functional profiles caused by multifunctional proteins tell us about the evolution of function

- One of the functions was not necessary anymore. That function is part of one protein and another protein, those are lost.
- **Evolution by loss. No change in “function”**

Asymmetric functional/metabolic relations explain discordant phylogenetic profiles


What do discordant functional profiles caused by asymmetric functional/metabolic proteins tell us about the evolution of function

- Either functions is not always necessary so loss or (re-)gain through HGT
Lack of co-evolution (phylogenetic profile similarity) between TBP and MOT1/NC2

Organisms without MOT1/NC2 tend to lose one of the critical phenylalanines, this explains how they cope AND reveals co-evolution between presence/absence of a gene and residues in another gene.

Csm1 is a LECA kinetochore subunit of the Monopolin complex lost in higher animals that interacts with Dsn1.
The phylogenetic profiles of the motif Dsn1-N and Csm1 are highly similar

Disruption of phylogenetic profile similarity; what have we learned about function?

- The interaction/function is ancestral
- Orthologs differentiate in function by loss of interaction and the function associated with this interaction (cf. multifunctional proteins)
- Potentially useful tool to predict interaction motifs

Non-orthologous gene displacement/analogous proteins explain discordant phylogenetic profiles

- First systematic analysis on *M. genitalium* (Koonin et al., Trends Genet. 1997)
The opposite of co-occurrence: anti-correlation / complementary patterns: predicting analogous enzymes

Genes with complementary phylogenetic profiles could have a similar biochemical function.

Complementary patterns in thiamin biosynthesis predict analogous enzymes

Prediction of analogous enzymes is confirmed

Ska & Dam1: functional counterparts

- KT-MT attachments
  - Dependent on Ndc80
    - Interaction with loop?
    - Tracking of depolymerizing microtubules

→ Orthologs of Ska (3 subunits) and Dam1 (10 subunits) across 94 genomes
Ska & Dam1 across eukaryotes: intracomplex correlation and intercomplex anticorrelation

- Ska complex subunits in i.e. Metazoa, Chytridiomycota, Apusozoa, Archaeplastids and some SAR.
- Dam1 complex subunits in most fungal lineages, Filasteria, Amoebozoa, various Stramenopila, Rhizaria, red algae, Cryptophyta.

Alternative evolutionary scenarios

What do we learn about evolution of function from analogous enzymes

- The function is “conserved”, there is no evolution of function (for the network / organisms) (???)
- But there is evolution of protein/gene with similar functionality (and where does the analogous protein come from?) (but also perhaps a lot of evolution by loss)
- And why?
EVOLUTION OF GENETIC INTERACTIONS

Negative / syntetic lethal / aggravating
Positive / buffering / alleviating

Genetic interactions

"generate 774,309 double mutants"

But ...

“Our Sp map identified > 700 high-confidence gene-to-gene correlations indicative of genes with related functions”
We present a genetic interaction map of pairwise measures including ~40% of nonessential *S. pombe* genes. By comparing interaction maps for fission and budding yeast, we confirmed widespread conservation of genetic relationships within and between complexes and pathways.

*i.e. the data is of high enough quality to reliably (consistently) presence or absence of “function”*

**Example ESCRT**

- the endosomal sorting complex required for transport (ESCRT) genes in endosomal maturation
- Also a role in cytokinesis in pombe (and animals) but not in cerevisiae
- Extensive experimental validation
- *?* Loss of function in yeast
- **Different behavior for intra complex vs inter-complex interactions in evolution:** within module/complex interactions are conserved but regulation and role of module for the cell evolves
Lower amino acid similarity did not correlate with repurposing (Figure 2D, left), but lower percentage coverage (i.e., additional motifs or domains present in only one of the orthologs) did correlate with apparent repurposing.

“Co-regulation” is quite well conserved (if the genes are conserved) -> co-regulation indicates “same complex” “close together in a pathway.”

Change in function between orthologs does not seem to depend on sequence identity but does seem to depend on sequence domain/motif composition.

EVOLUTION OF (CO-)REGULATION

“Co-regulation” is quite well conserved (if the genes are conserved) -> co-regulation indicates “same complex” “close together in a pathway.”

Measuring the Evolutionary Rewiring of Biological Networks

http://academic.oup.com/nar/article/32/16/4725/1023281

Shou C, Bhardwaj N, Lam H, Yan K, Lam PM, Snyder M, Gerstein MB.


### Table S3

<table>
<thead>
<tr>
<th>Network Type</th>
<th>TF</th>
<th>S. cerevisiae</th>
<th>S. bayanus</th>
<th>D. melanogaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge change from Edge Gain</td>
<td>26</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge change from Edge Loss</td>
<td>53</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge change from Node Gain</td>
<td>60</td>
<td>12733</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge change from Node Loss</td>
<td>306</td>
<td>76543</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Regulatory evolution. Dynamic conservation?

- “We found that although the gene expression patterns characterizing the response to drugs were remarkably conserved between the two species, part of the underlying regulatory networks differed.”

EVOLUTION OF PHOSPHORYLATION
Shou C, Bhardwaj N, Lam HY, Yan X, Lam PM, Snyder M, Gerstein MB.

- position of most phosphorylation sites is not conserved in evolution; instead, clusters of sites shift position in rapidly evolving disordered regions.
- the regulation of protein function by phosphorylation often depends on simple nonspecific mechanisms that disrupt or enhance protein-protein interactions.
- Is similar to?

Functions of non-globular / disordered / unstructured regions

So how do they evolve? How should we think about that?
Diverged orthologous IDRs recapitulate S. cerevisiae IDR functions compared with the 5A mutant.

Other class of phosphosites

Other class of phosphosites: eEF2
Other class of phosphosites: Raf


“dynamic conservation” “neutral-rewiring & conserved output/function”

- Function / output is conserved but exact wiring / positions is not
- Also implied to play a large role in evolution of transcription factor binding sites.
- i.e. in normal (globular) protein sequence evolution conservation of function implies conservation of sequence/structure, neutrality means similar amino acids (or synonymous substitutions) but for other units of function it could be higher level (conservation of charge and length, conservation of co-expression*) and dynamics at lower level
- * When and why (role) a protein/module/complex does its thing will evolve a lot more than module-membership and module molecular activity

Evolution of function: grand summary

- Strong interplay between network and genome evolution
  - Within pathways/complexes (modules) evolve by loss and gain of genes (from the genome!) but little rewiring (as in loss or gain of co-expression/interaction)
    - Most differences in networks are due to gain and loss of genes from the genome!
  - Also gain (and “loss”) of module membership after duplication followed by rapid functional substitutions
- Regulatory relations “dynamic conservation”
  - At “shorter” evolutionary distances, change in wiring, but same output (“function”)
  - At longer distances repurposing of when / how modules are needed
    - Between module relations are less conserved than within module
    - (also “applies” to intrinsically disordered proteins, and a subset of phosphosites)
So where does “new stuff” come from (besides duplication)

• Duplication / invention of new genes, & domain-recombination
• Inflation-contraction / biphasic model of genome evolution: e.g. eukaryogenesis, origin of animals, origin of vertebrates (mix of duplication, innovation, vertical inheritance)
• Constructive neutral evolution
• Function evolution is often episodic: rapid emergence of new functions, long periods of conservative evolution
• Exception: Arms-race processes (genetic conflict, host-pathogen) adaptive evolution is much more frequent

“new proteins” from duplication & domain recombination

Accumulation of complexity: a neutral explanation

- Neospora mitochondrial genome encodes several introns which require a tyrosyl tRNA synthetase (TyrRS) to splice.
- “to compensate for structural defects acquired by the intron sequences”
- BUT Introns with defects arising -> negative selection
- ? Reverse: first binding (fortuitously or for reason unrelated to splicing) -> accumulation of mutations in the intron that inactivate splicing, if TyrRS not bound.
- Because the compensatory / suppressive activity exists before mutation “presuppression,”
- the protein dependence by the intron could be selectively neutral (or slightly disadvantageous)
“Constructive neutral evolution”

• Suggested that many taxon specific subunits (taxon specific proteins that are a subunit in a complex) are regulatory subunits

• Hypothesis: neutrally added but necessary subunits could have been appropriated as regulatory subunits or “assembly” factors?

• “Finally, and to me most interestingly, how can we combine multi-level selection theory with reasoning about introns as adaptations (Doolittle, 1987, Cold Spr Hbr Symp Quant Biol 52: 907–913)? It may well be that multicellular eukaryotes of a certain type (us, for instance) have gained considerable evolvability (and consequent diversity) from having alternatively spliceable introns. But clearly, introns were not added to the genome of LECA so that more than a billion years later this advantage could be realized. Authors are (although too circumspectly in my opinion) down on such teleological rationalizing, but might we imagine such evolvability to be an adaptation at some much higher level (clades above species, Doolittle 2017; Phil Sci 84: 275–295)?"