SOME PRINCIPLES OF NETWORK EVOLUTION:

The roles of:
single gene duplication, large scale duplications,
homodimerisation and domain rearrangements
in the evolution of transcription factor interaction networks.

Erich Bornberg-Bauer

University of Manchester (UK) ---> University of Münster (Ger)

Evolution of Gene(tic) Networks

1) Evolution of Complexity - Animals, Genetic Networks

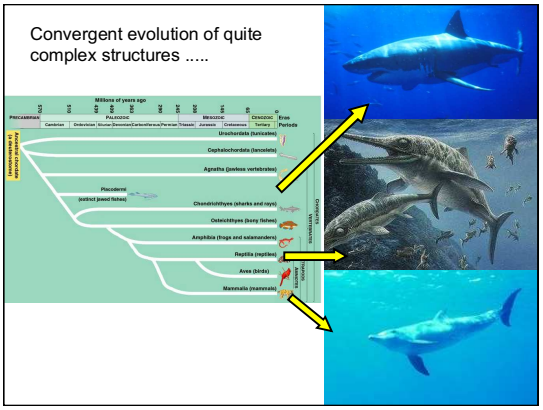
2) Evolution of interaction domains (Leucine zipper)

3) Domain - networks

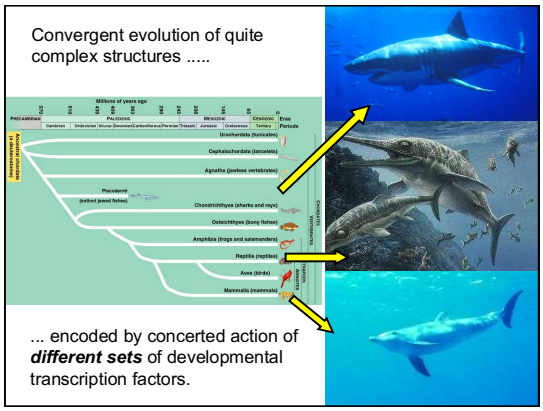
4) Evolution of interaction networks in transcription factors

1) Evolution of Complexity
1a) Phenotypic convergence

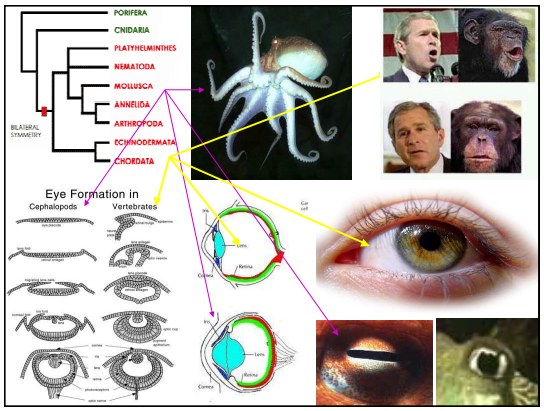
Convergent evolution of quite complex structures .....

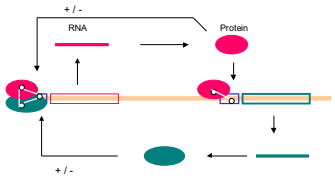


Convergent evolution of quite complex structures .....



... encoded by concerted action of ***different sets*** of developmental transcription factors.





PPI and P-DNA interactions - not an all or nothing process

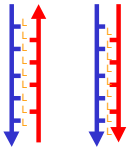
Complexity: integrated feed-back loops ...
is co-evolution or teleological „knowledge“ needed?

2) A molecular view on the Evolution of interaction domains

First molecular model of eukoryotic genetic regulation:

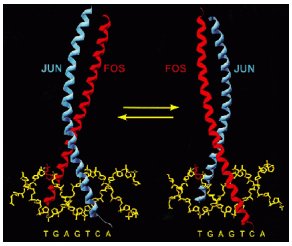
Leucine Zipper hypothesis (Landshulz, Science 1989)

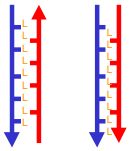
- 2 proteins dimerise through a ``Leucine Zipper'' domain
- coiled coil structure
- Leucine every 7th position (``heptad'')
- antiparallel or parallel dimerisation
- ``knobs into holes'' -- zips up
- LZ binds to DNA ...



First molecular model of eukoryotic genetic regulation:

Leucine Zipper structure (1991):

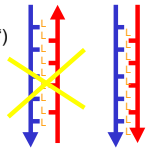




First molecular model of eukoryotic genetic regulation:

Leucine Zipper hypothesis (Landshulz, Science 1989)

- 2 proteins dimerise through a ``leucine zipper'' domain
- coiled coild structure
- Leucine every 7th position (``heptad'')
- antiparallel or parallel dimerisation
- ``knobs into holes'' -- zips up
- LZ binds to DNA ...



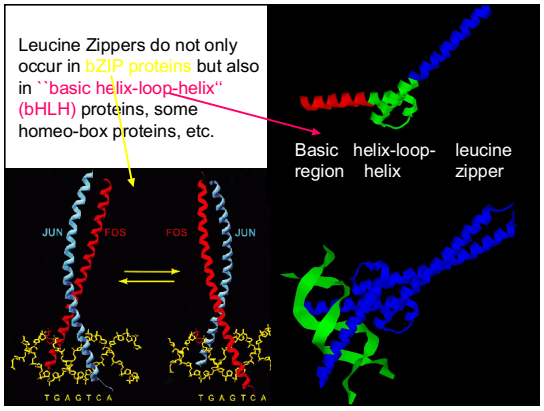
Impact: ~ 2500

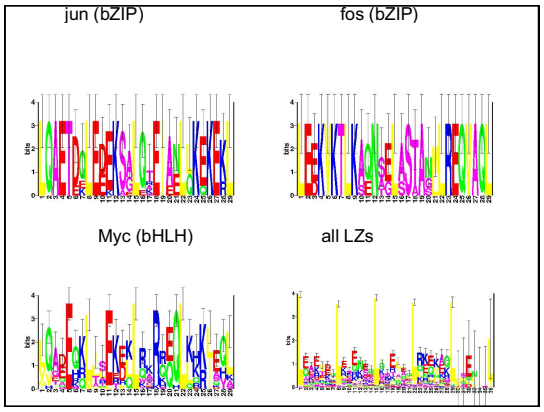
Detection of leucine zippers:

- PROSITE [Lx6]4 .... 600 false positives
- TRESSPASSER: some machine learning software using all LZs as training set
- HMMs with poor predictive power ...

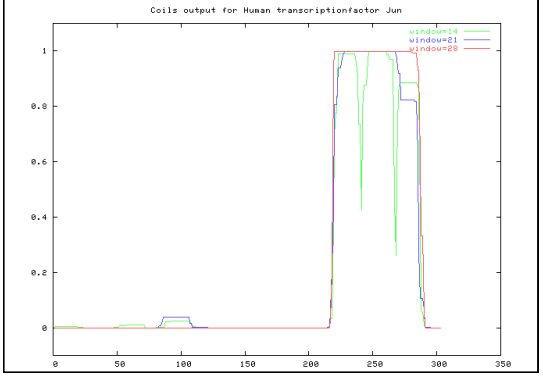
WHY?

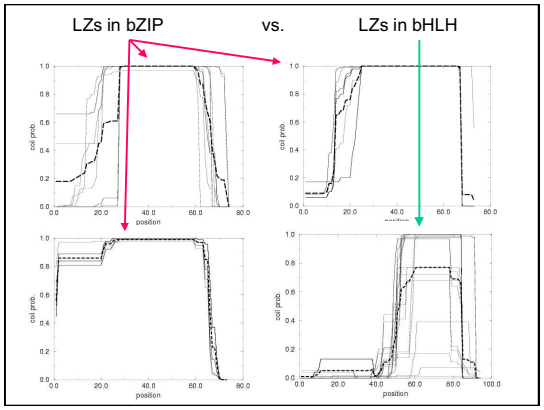
Leucine Zippers do not only occur in bZIP proteins but also in ``basic helix-loop-helix'' (bHLH) proteins, some homeo-box proteins, etc.





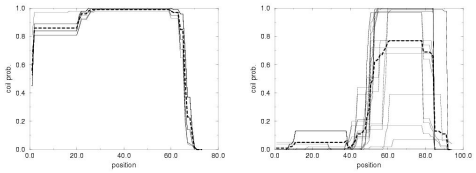
Coils: profile based prediction w/ shifting windows, yields probability





Leucine zippers in bZIP: dimerisation + specificity
Leucine zippers in bHLH: specificity only !! (less stable)
surprisingly there are not many similarities ....

..... Results of convergent evolution?
--> Use functional knowledge to redefine and detect LZs



Detection of Leucine Zippers:

- PROSITE [Lx6]4 .... 600 false positives
- TRESSPASSER: some machine learning software
  using all LZs as training set
- HMMs with poor predictive power ...

WHY?

- Problem: LZ used to sex up papers -->
  many wrong DB annotations (``LZ by similarity'')
  different function, different families?

Improved detection of leucine zippers:

2ZIP:

- **short** coiled coil
- relaxed leucine repeat

Worked OK, virtually no false positives when:

- literature used to overrule SPROT annotations and
- co-occurrence with DNA binding motif (basic region)
  used to evaluate results

BUT:

- some x00 new potentially spurious annotations since ... ;-(

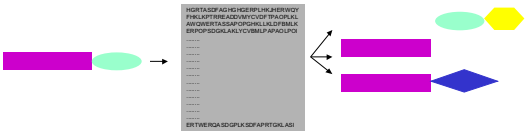
The lessons learnt so far:

- Functional + evolutionary insights can help to circumvent problems with classification and machine learning.
- Co-occurrence of domains more reliable than DB annotations
- DBs show error propagation
- Highly cited papers may be very ``misleading“
- Delineating PPIs automatically is problematic

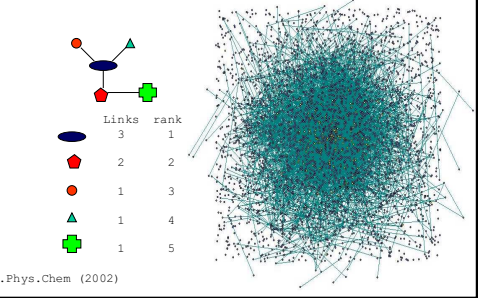
3) Domain networks

Leucine zipper domains do yet occur together with various other domains.

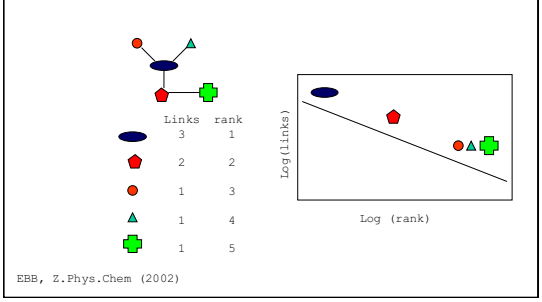
Reincarnation of the classical domain problem?



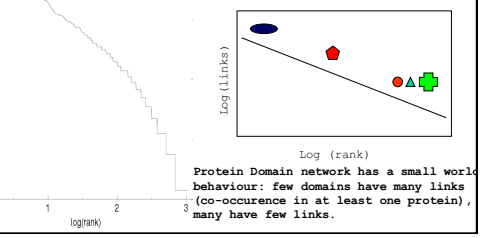
How mobile just are proteins?



EBB, Z.Phys.Chem (2002)

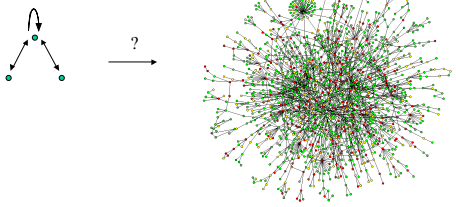


EBB, Z.Phys.Chem (2002)

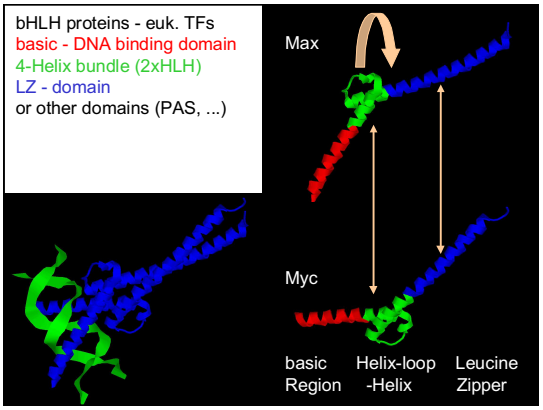


Protein Domain network has a small world behaviour: few domains have many links (co-occurence in at least one protein), many have few links.

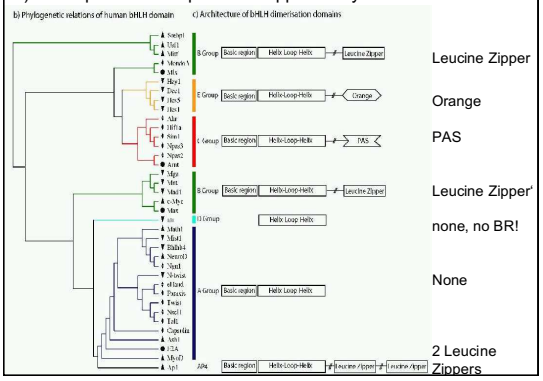
4) Evolutionary Analysis of Network Evolution
- bHLH Network Evolution
- Model for bHLH evolution
- Extension to NR and bZIP

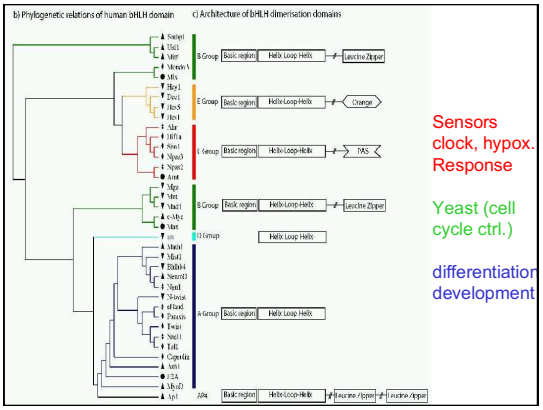


bHLH proteins - euk. TFs
basic - DNA binding domain
4-Helix bundle (2xHLH)
LZ - domain
or other domains (PAS, ...)



4a) bHLH proteins: deep nodes supported by domain-architecture





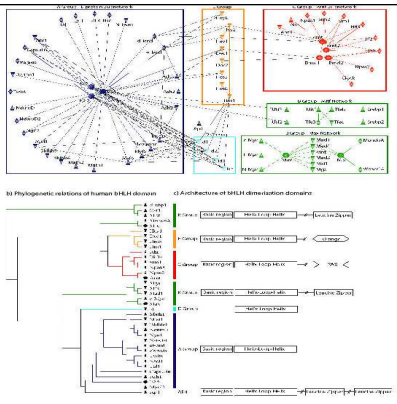
Sensors clock, hypox. Response

Yeast (cell cycle ctrl.)

differentiation development

Combine Phylogenies w/ confirmed PPI:

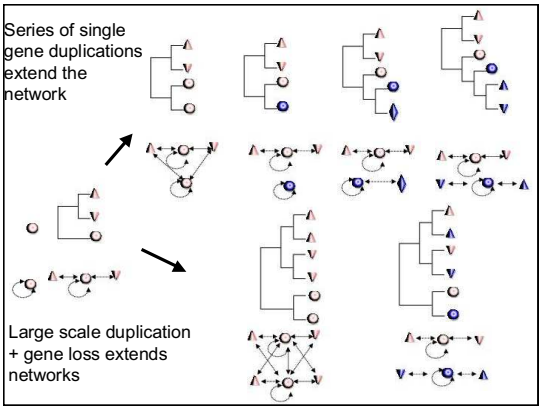
networks are basically separated and hub-like



What are the possible courses of evolution for PPI networks?

- Network duplication by large scale (genome, chromosome, ...) duplication (LSD), followed by gene loss
- Series of single gene duplications (SGD), followed by gene loss
- ``Rosetta Stone'' theory: repeated domain-fusions+fissions
- ``Random'' evolution of interfaces (gain/loss due to genetic drift)
- Homo-dimerisation --> gene duplication --> drift --> heterodimerisation
- Combinations of any or some of these

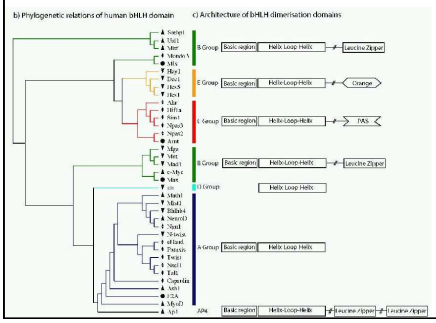
Series of single gene duplications extend the network



Large scale duplication + gene loss extends networks

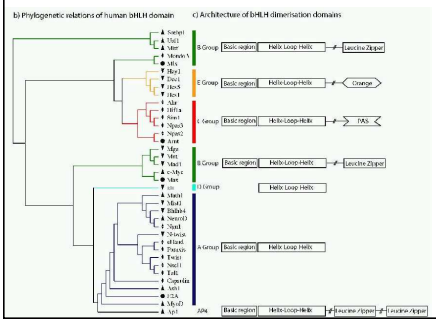
4b) Model for evolution of bHLH proteins:

- tree-topology suggests homo --> heterodimerisation as a pattern of evolution of specificity



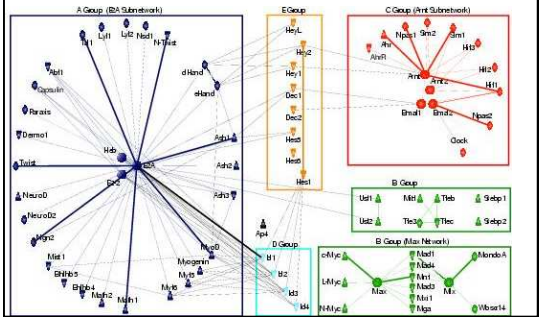
4b) Model for the evolution of bHLH proteins:

- SGD main driving force in NW emergence



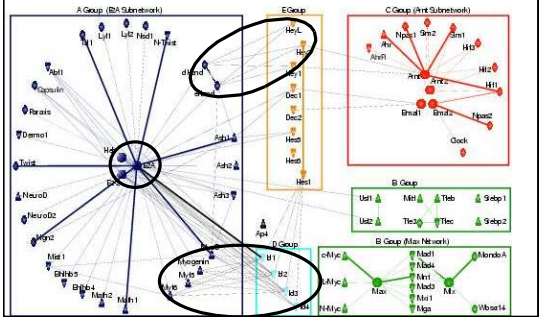
4b) Model for evolution of bHLH proteins:

- SGD main driving force in NW ***emergence***



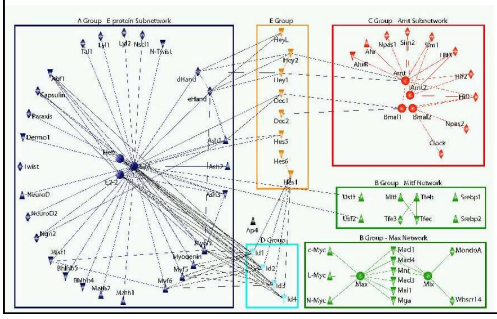
4b) Model for the evolution of bHLH proteins:

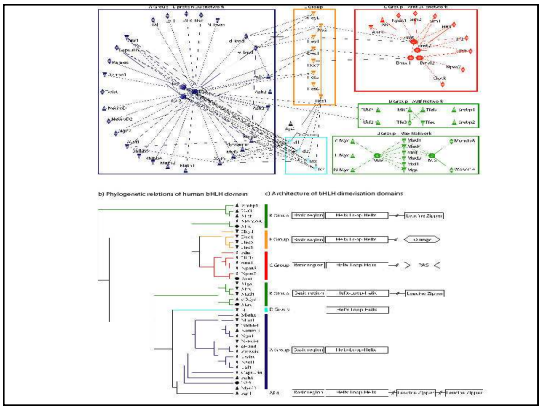
- However: LSDs may have increased the ***size*** of networks with ***existing topology***

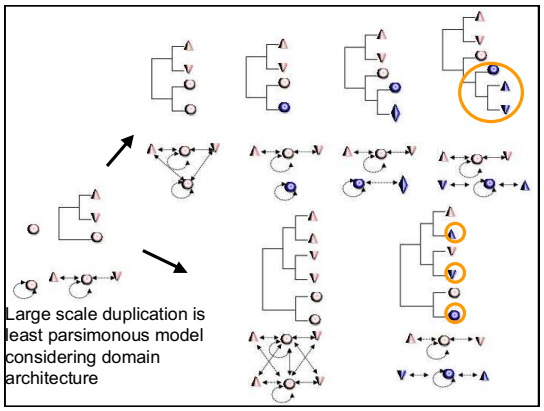


4b) Model for Evolution of bHLH proteins:

- 2 disconnected networks, similar topology !!!
  --> Convergent Evolution at a higher level

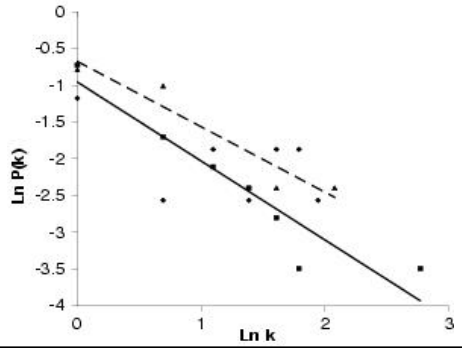


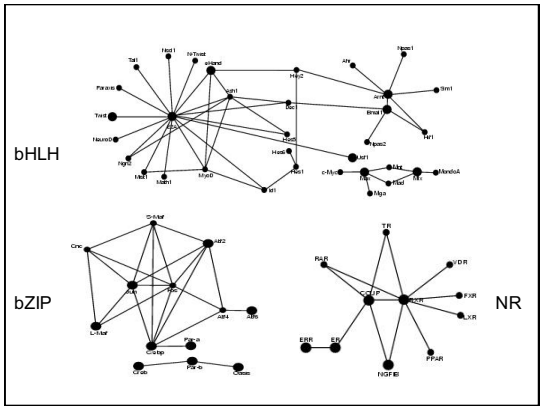




Large scale duplication is least parsimonous model considering domain architecture

NR (▲·····) and bZIP (●) proteins (bHLH: ■ —)





NR: contain additional dimerisation domains
scale free
homo-dimerising factors hubs and ancestral or not (yet) differentiated into a new network

bZIP: NOT scale free
no additional dimerisation domains

--> Different topologies for different families
--> Different stories for different families

Summary network evolution -- perspective from TFs:

- Homo --> heterodimerisation (bHLH, NR)
- Hubs: ancient, central, often homo-dimerising
- SGDs sufficient to generate complex networks (bHLH, NR)
- Domain rearrangements support emergence of sub-networks (bZIPs: no additional domains --> no separate networks ?)
- First (?) evidence of convergent evolution of complex organisational units
- No general rule for ``small world'' topology
  PPI-NW is a network of networks, one model won't fit all
- No evidence for selection towards network structure per se

Acknowledgements:

- LZ: Eric Rivals, Martin Vingron
- Networks: Greg Amoutzias,
  David Robertson,
  Steve Oliver (U o Manchester)

  Astra Zeneca, BBSRC

Literature:

- bZIPs: EBB, ER, MV, NAR 1998
- Domain networks: EBB Z. Phys. Chem. 2002,
  Cui, Wong, EBB, Chan PNAS 2002
- TF NW evolution: Amoutzias, DR, SO, EBB EMBO Rep 2004
  Amoutzias, DR, EBB, Comp. Funct. Gen 2004

THE END