“Home depot” model of prokaryotic evolution by horizontal gene transfer

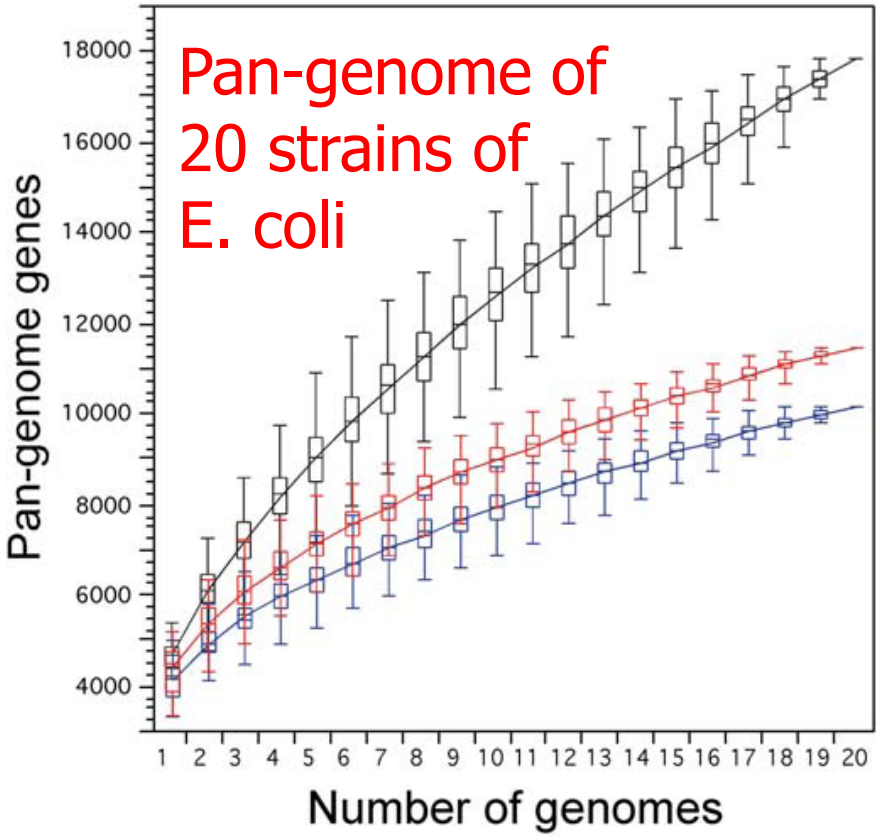
Sergei Maslov

Brookhaven National Laboratory

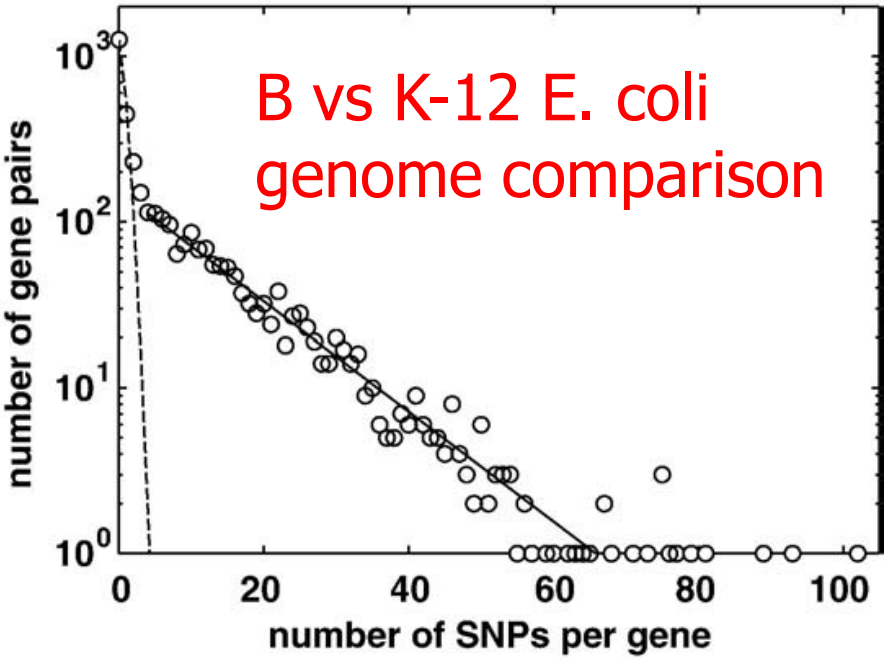
Different modes of evolution

- Genes – tools, Genomes - toolboxes
- Working genes acquire adaptive mutations that fine-tune them to particular environments (common)
  sharpen scissors
- Completely new functions of genes evolve de novo (rare)
  invent scissors from two knives
- Horizontal Gene Transfer in prokaryotes lets acquire entire sets of genes evolved in other organisms (common)
  buy scissors in Home Depot if you need them

Bacterial genomes are full of genetic material from other strains and species



M Touchon et al. PLoS Genetics (2009)



FW Studier, P Daegelen, RE Lenski, S Maslov, JF Kim, JMB (2009)

PNAS

Toolbox model of evolution of prokaryotic metabolic networks and their regulation

Sergei Maslov[a,1], Sandeep Krishna[b], Tin Yau Pang[a,c], and Kim Sneppen[b]

[a]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973; [b]Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark; and [c]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800

Edited by David J. Lipman, National Institutes of Health, Bethesda, MD, and approved April 16, 2009 (received for review March 23, 2009)

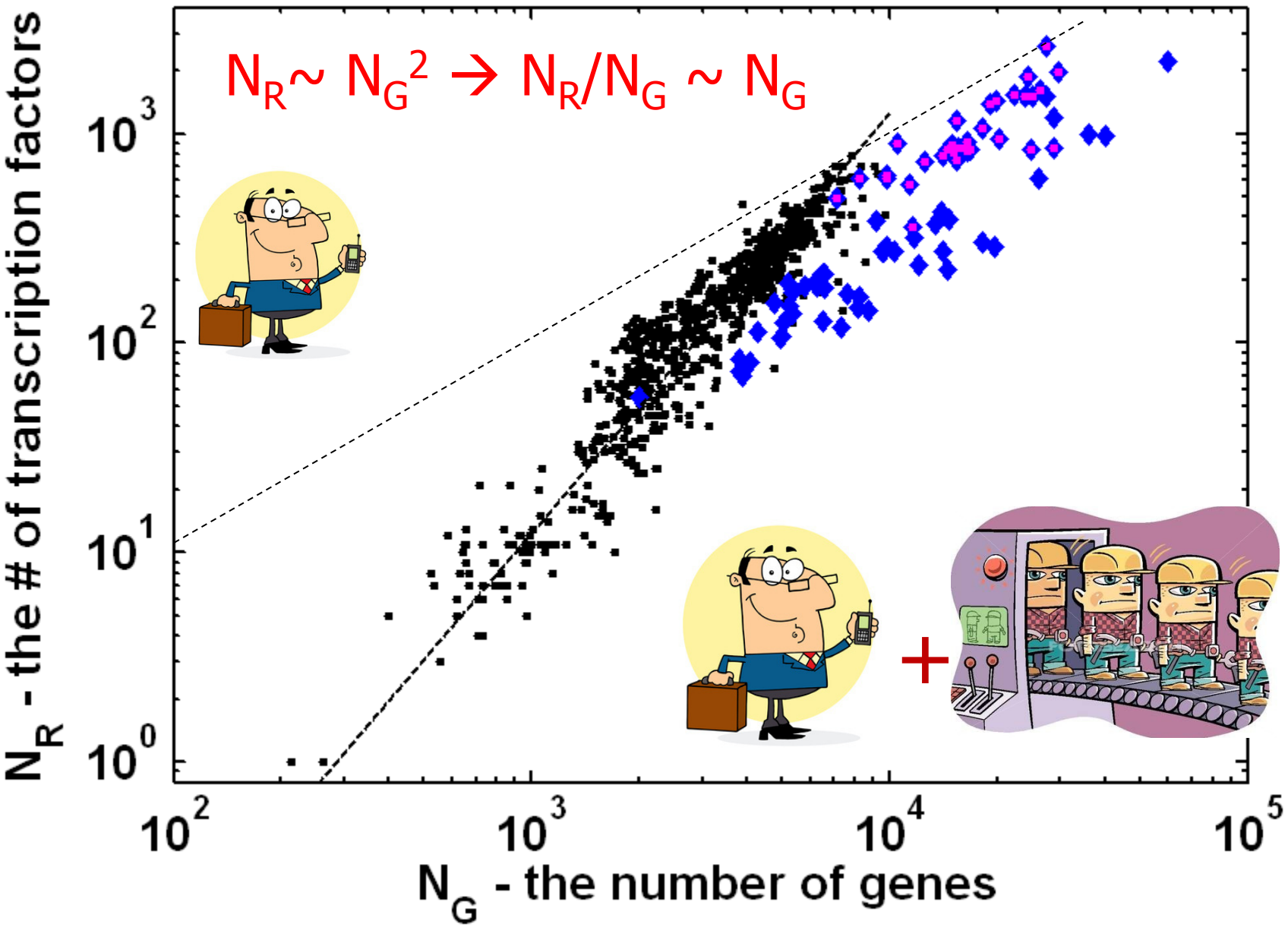
It has been reported that the number of transcription factors encoded in prokaryotic genomes scales approximately quadratically with their total number of genes. We propose a conceptual explanation of this finding and illustrate it using a simple model in which metabolic and regulatory networks of prokaryotes are

A simple evolutionary model explains both these empirical observations in a unified framework based on modular functional design of prokaryotic metabolic networks and their regulation.

Toolbox View of Metabolic Networks

Disclaimer: authors of this study (unfortunately) received no financial support from Home Depot, Inc.; Leroy Merlin; Homebase, LTD; or Obi, GMBH

Stover *et al.*, Nature (2000) van Nimwegen, TIG (2003)



The Economist

Parkinson's Law

The total of those employed inside a bureaucracy grew by 5-7% per year "irrespective of any variation in the amount of work (if any) to be done."

Parkinson explains the growth of bureaucracy by two forces:

- "An official wants to multiply subordinates, not rivals"
- "Officials make work for each other."

Nov 19th 1955 | From *The Economist* print edition

Is this what happens in bacterial genomes?

Probably not!

Economies of scale in genome evolution?

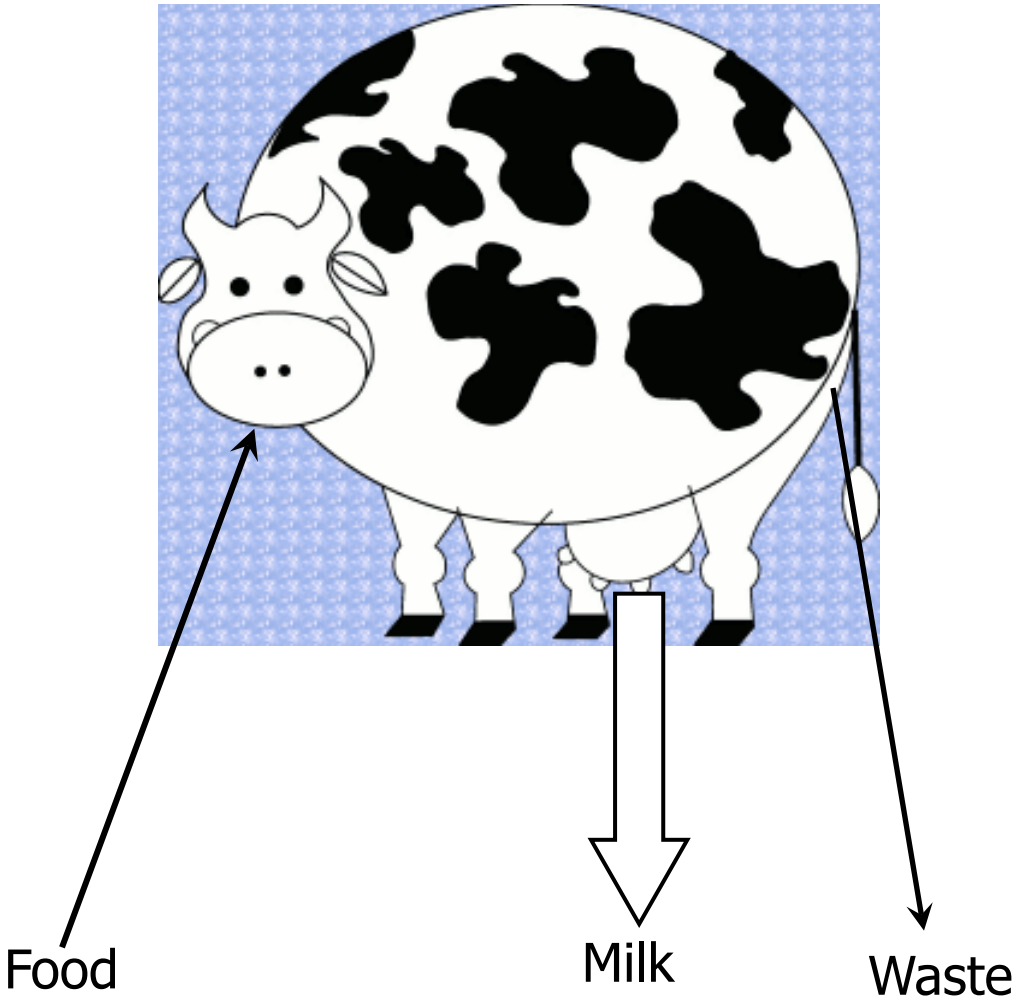
- $N_R = N_G^2/80{,}000$ --> $\Delta N_R = \Delta N_G \, 2N_G/80{,}000$
- When a new regulated function is added
  $\Delta N_R = +1$
  $\Delta N_G = 40{,}000/N_G$
- Economies of scale: as genome gets larger newly added regulated pathways get shorter

Toolbox argument (inspired by my personal experience as a homeowner)

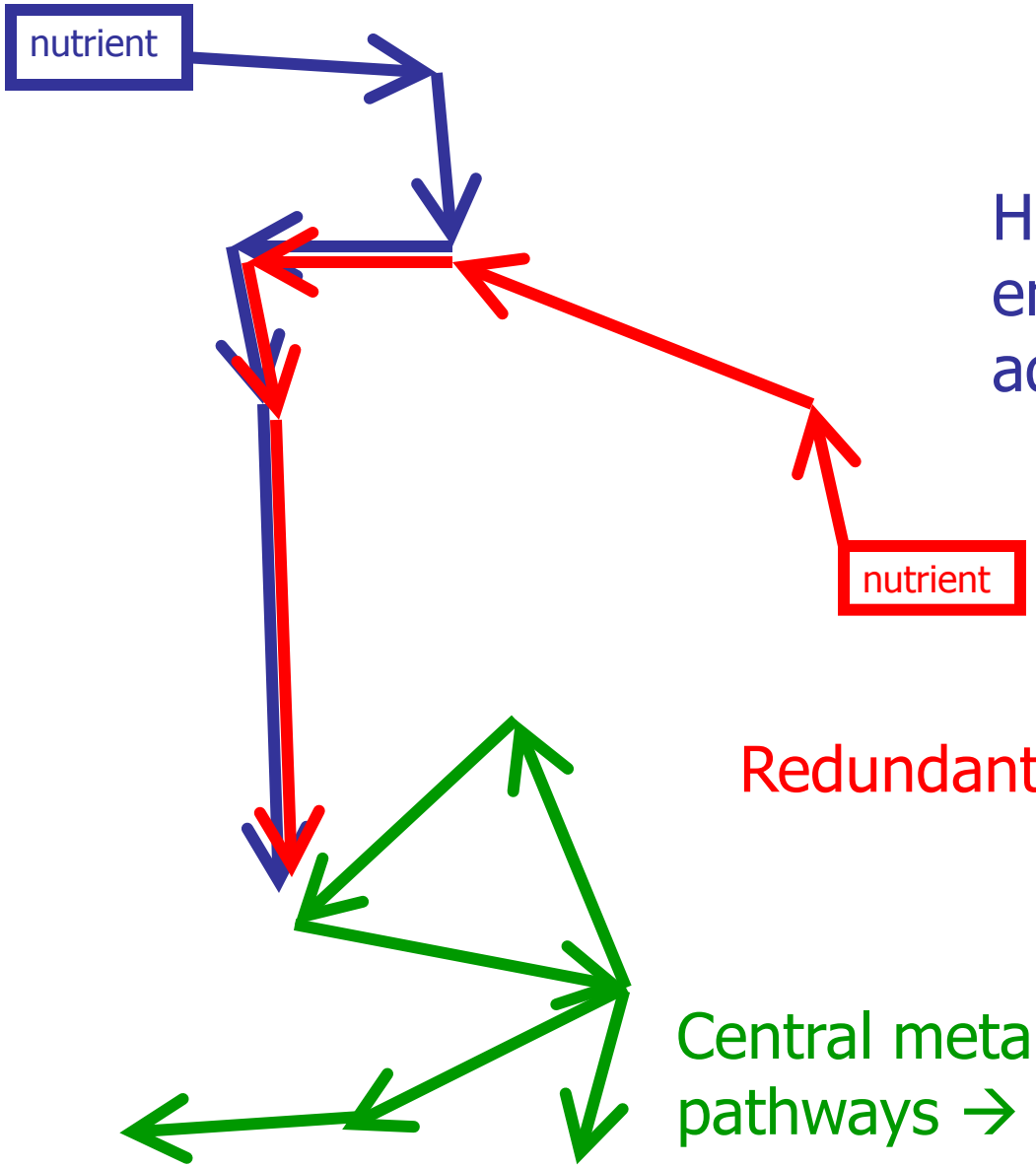
- Sets of tools are bought to accomplish functional tasks e.g. to fix a leaking faucet
- As your toolbox grows you need to get fewer and fewer new tools to accomplish each new task
- Duplicate tools are returned to "Home Depot"

- Bacteria have tools encoded by non-regulatory "workhorse" genes (e.g. metabolic enzymes)
- Entire pathways (sets of tools) are routinelyacquired from other bacteria by Horizontal Gene Transfer
- Regulators control these pathways (one TF per pathway)
- Redundant genes are promptly deleted (in prokaryotes)
- (# of new tools) per task gets smaller → FASTER THAN LINEAR SCALING

Spherical cow model of metabolic networks



Pathways could be also removed



Horizontal gene transfer: entire pathways could be added in one step

Redundant enzymes are removed

Central metabolic core → anabolic pathways → biomass production

Simple argument for quadratic scaling

- New pathways come from the "universal metabolic network" of size N_U : the union of all reactions in all organisms

- The current size of the toolbox (~# of genes ~ # of enzymes ~ # of metabolites): N_G

- Probability to merge with existing pathway:

  $p_{merge} = N_G / N_U$

- Length before merger: $L_{added\ pathway} = 1/p_{merge} = N_U/N_G$

- Assume one regulator per function/pathway:

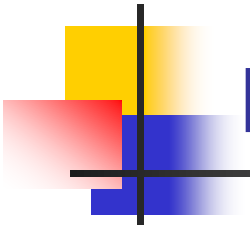
$\Delta N_G / \Delta N_R = L_{added\ pathway} + 1 \sim N_U/N_G \rightarrow$

$\rightarrow$ Quadratic law: $N_R = N_G^2 / 2N_U$

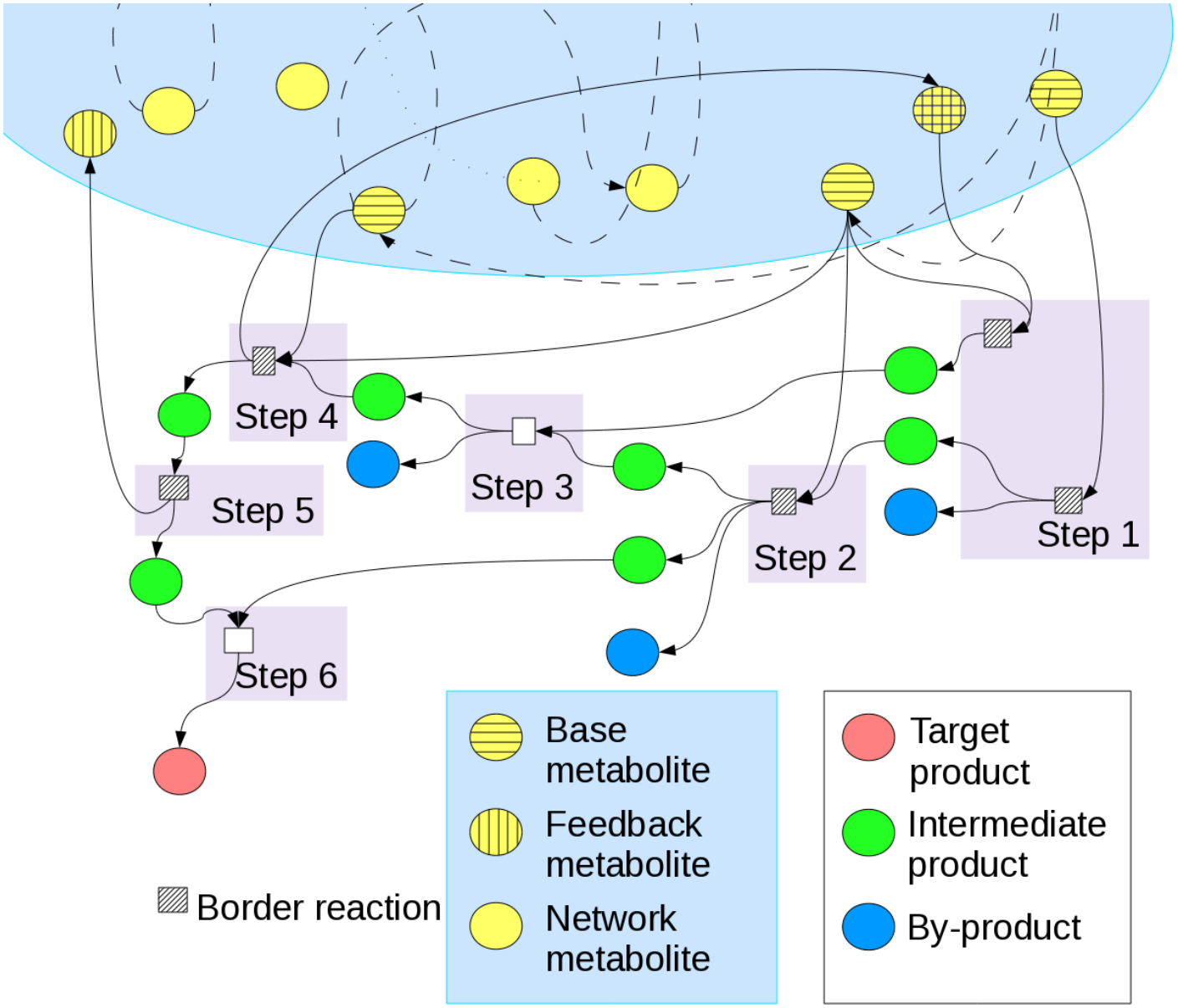
Different universal networks give the same result

- Random branched network: analytically solved to give $N_R \sim N_{met}^2$
- Universal network - union of all metabolic reactions in the KEGG database
- Real reactions and pathways are more complicated than linear pathways in my simple model

Logical structure of metabolic networks

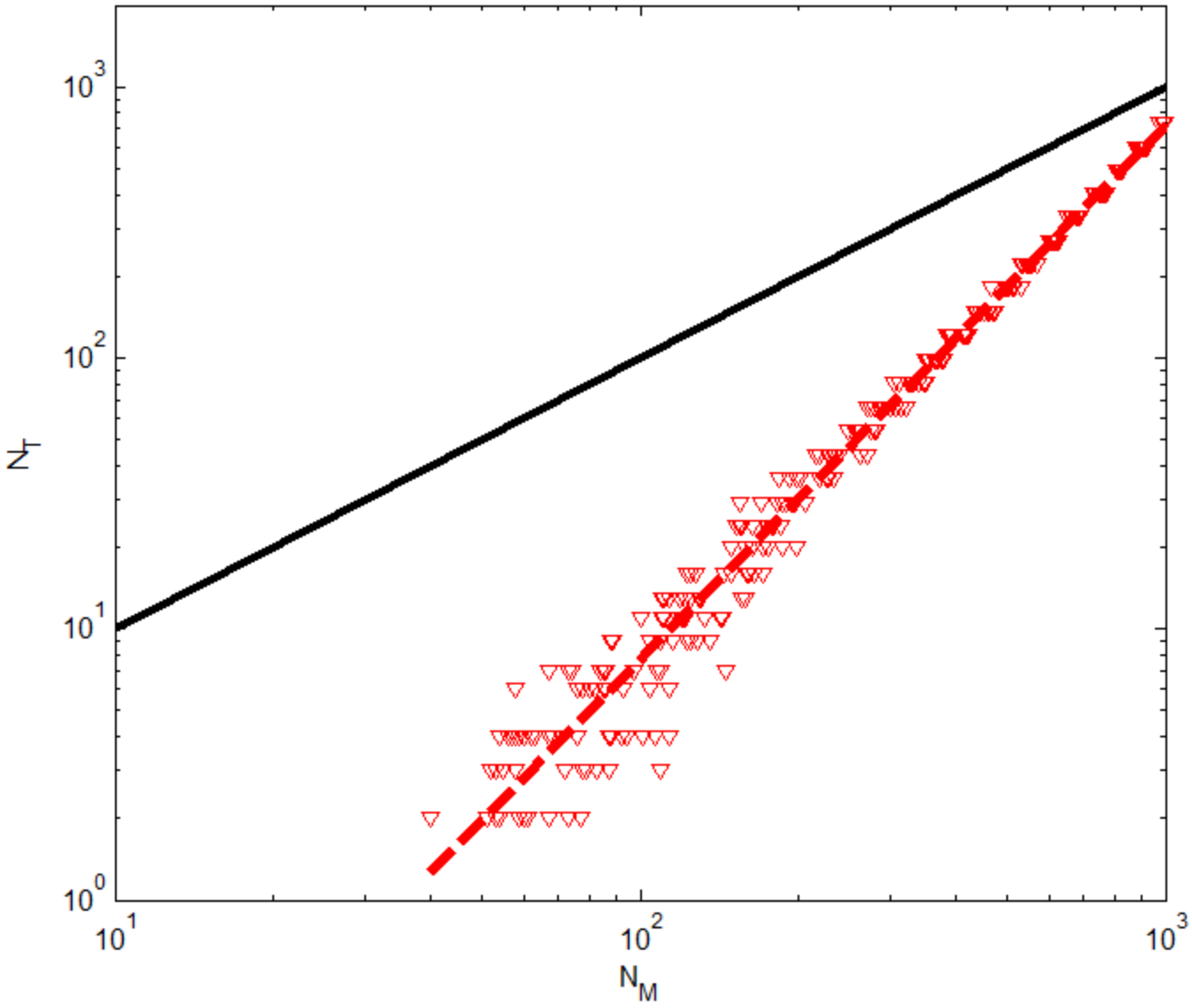


- Two types of nodes: reactions & metabolites
- Multiple substrates & products
- AND function acting on multiple substrates (inputs) of reactions
- OR function acting on metabolites
- No small-world properties!



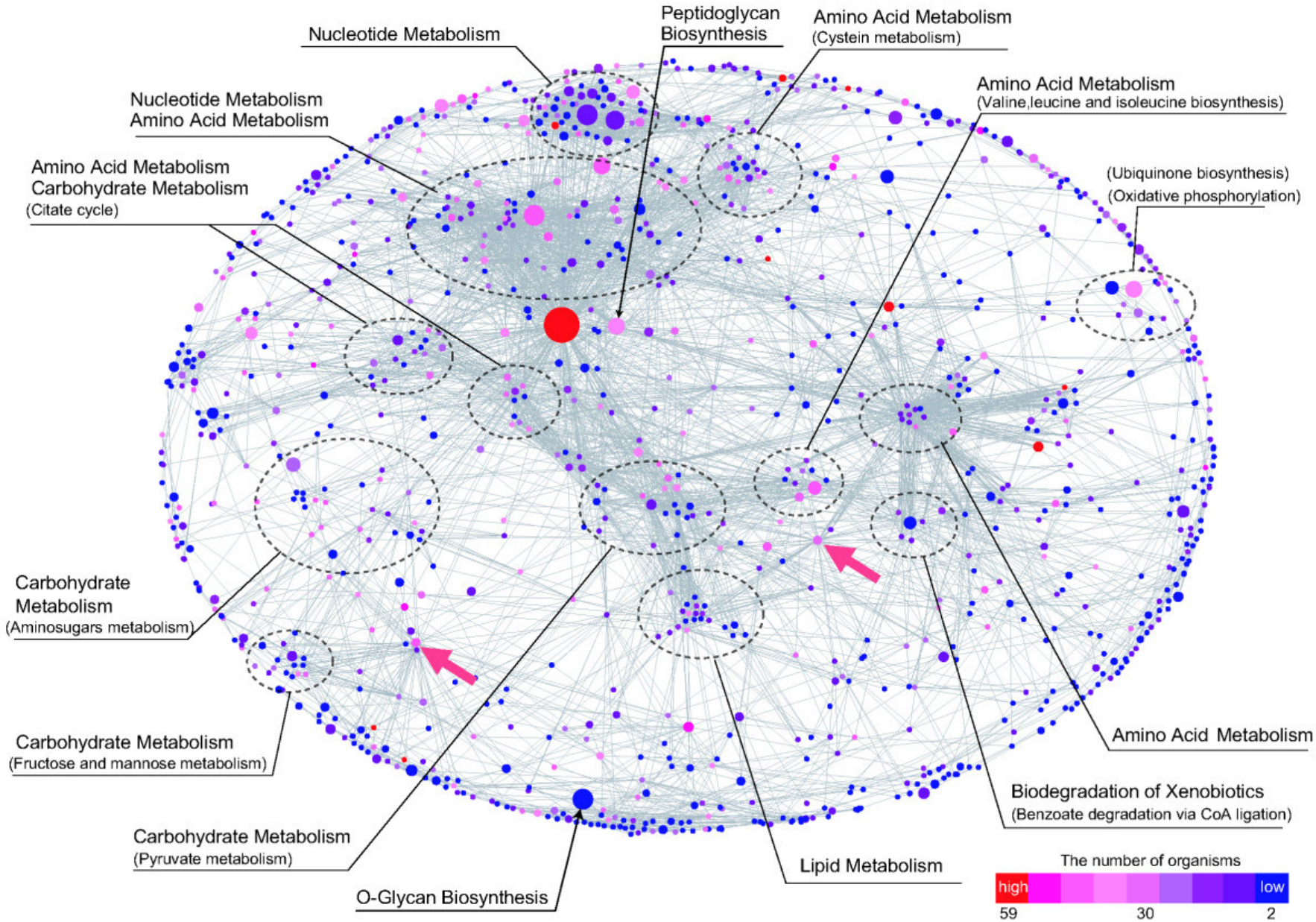
Inspired by "scope-expansion" algorithm
by Reinhart Heinrich and collaborators

KEGG model with realistic multi-substrate & multi-products reactions



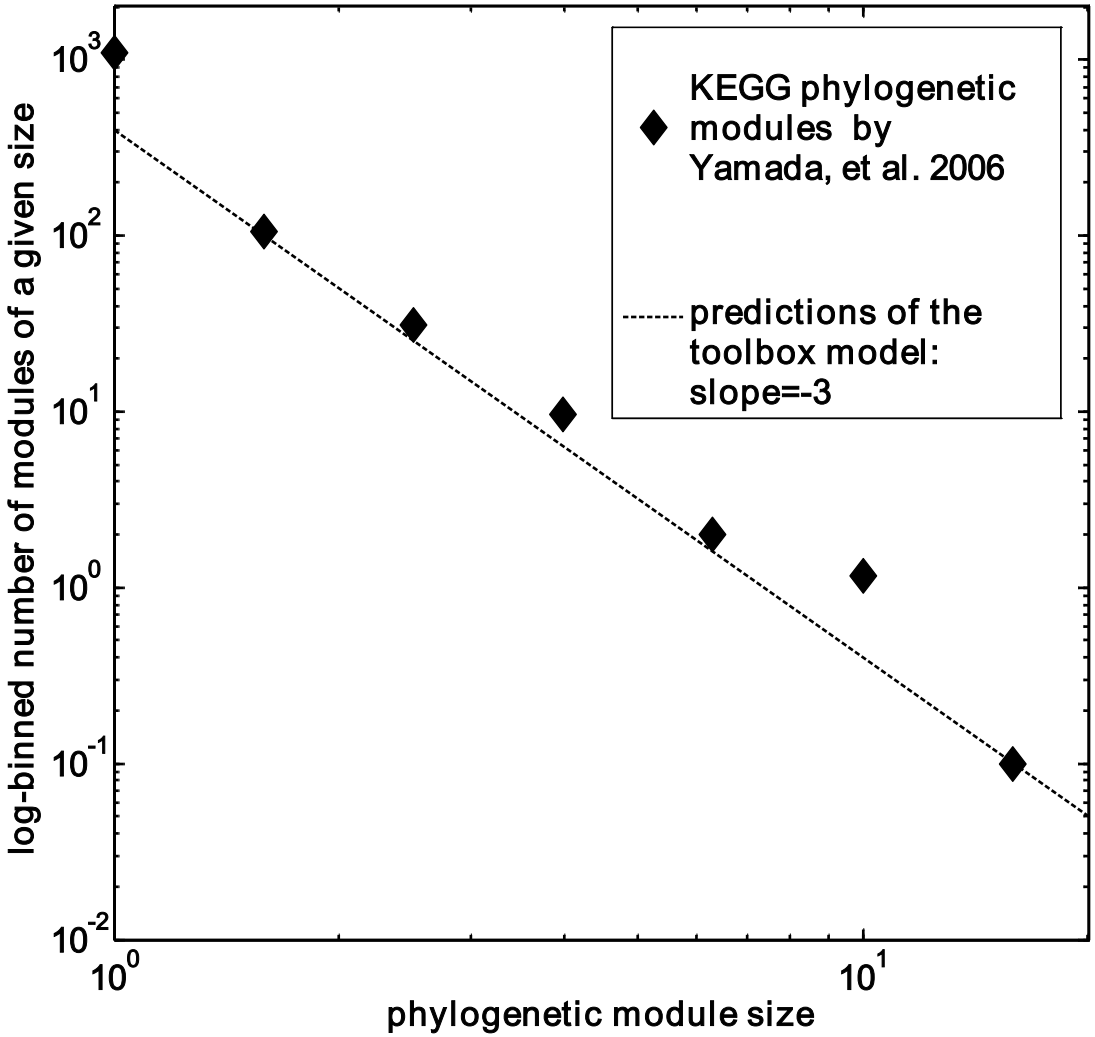
Test of the model

- Model predicts specific distribution of pathway sizes S: $P(S) \sim S^{-3}$
- Pathways are defined phylogenetically and NOT functionally
- Phylogenetic analysis of metabolic enzymes was done by Yamada, Kanehisa, Goto (KEGG) 2004 & 2006



1130 phylogenetic modules containing ~1600 enzymes

from Fig. 4 from Yamada, Kanehisa, Goto, BMC Bioinformatics 2006

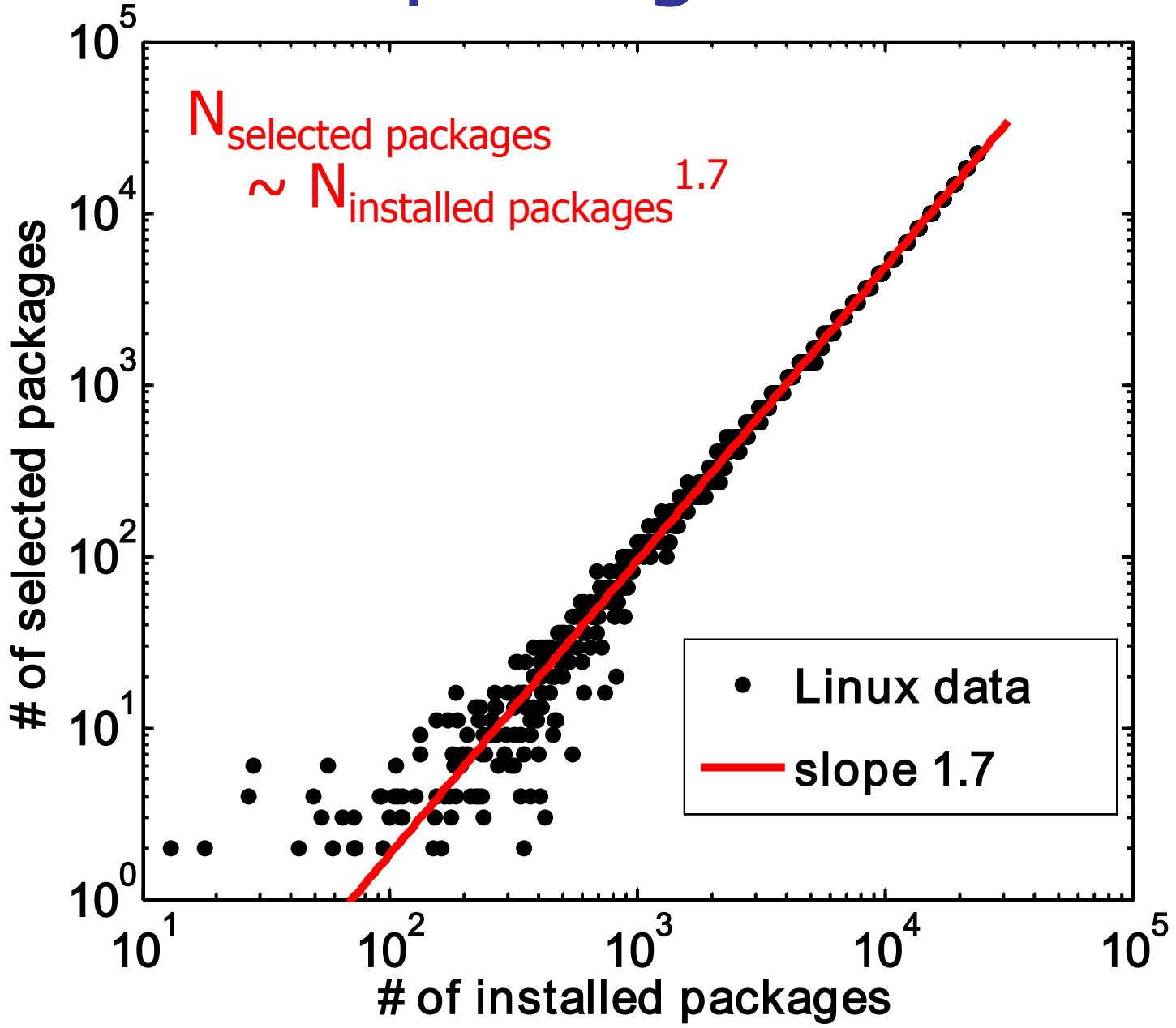


Phylogenetic modules in metabolic networks from Yamada, Kanehisa, Goto, BMC Bioinformatics 2006

Software subroutines are similar to metabolic reactions

- Two types of nodes:
  - subroutines (analog of reactions) with multiple inputs & outputs
  - variables (analog of metabolites)
- AND function acting on inputs of subroutines
- OR function acting on variables

Software packages for Linux



Collaborators and support

- Kim Sneppen (Center for Models of Life, Niles Bohr Institute, Copenhagen, Denmark)
- Sandeep Krishna (National Centre for Biological Sciences, Bangalore, India)
- Tin Yau Pang (Physics Department @ Stony Brook U)

- DOE Office of Basic Energy Sciences
- DOE program on Systems Biology Knowledgebase (joint funding with Mark Gerstein @ Yale)

I have 1-2 postdoc positions to work on metabolic and regulatory networks in bacteria.

If you interested contact me (maslov@bnl.gov).