Evolutionary rate & saturation

Torsten Struck – t.h.struck@nhm.uio.no
NHM UiO

© nhm.uio.no

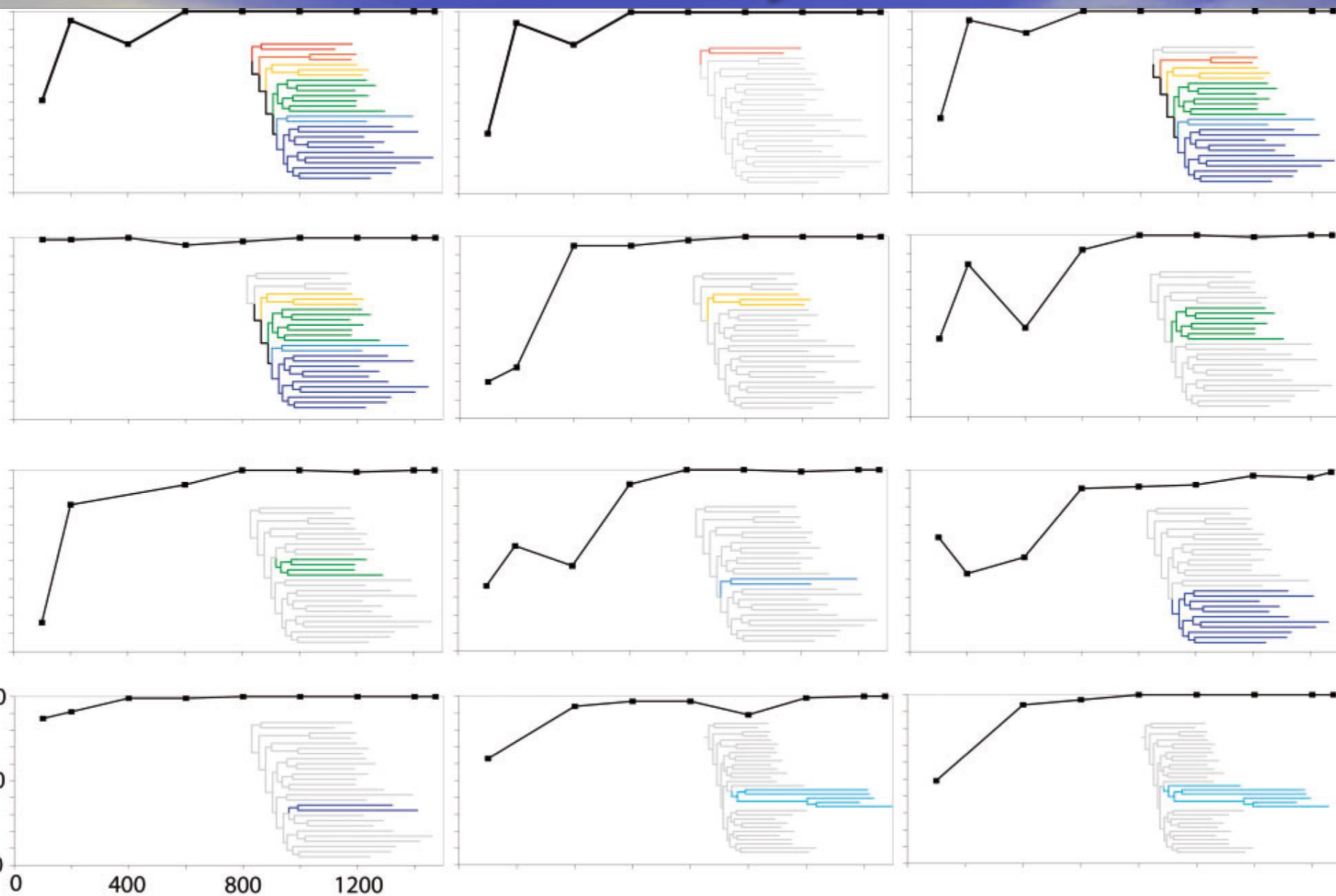# *Increased chance of substitution*
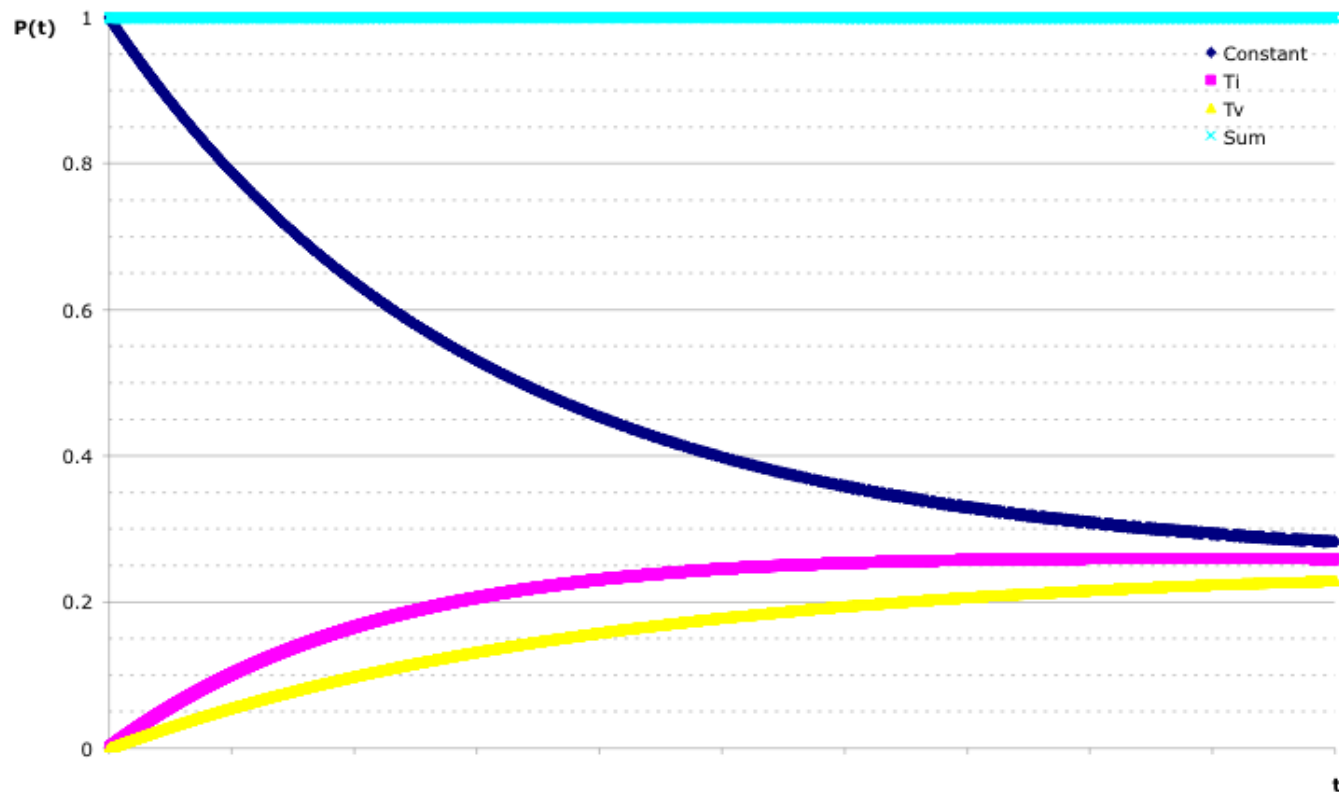
μ = 0.25
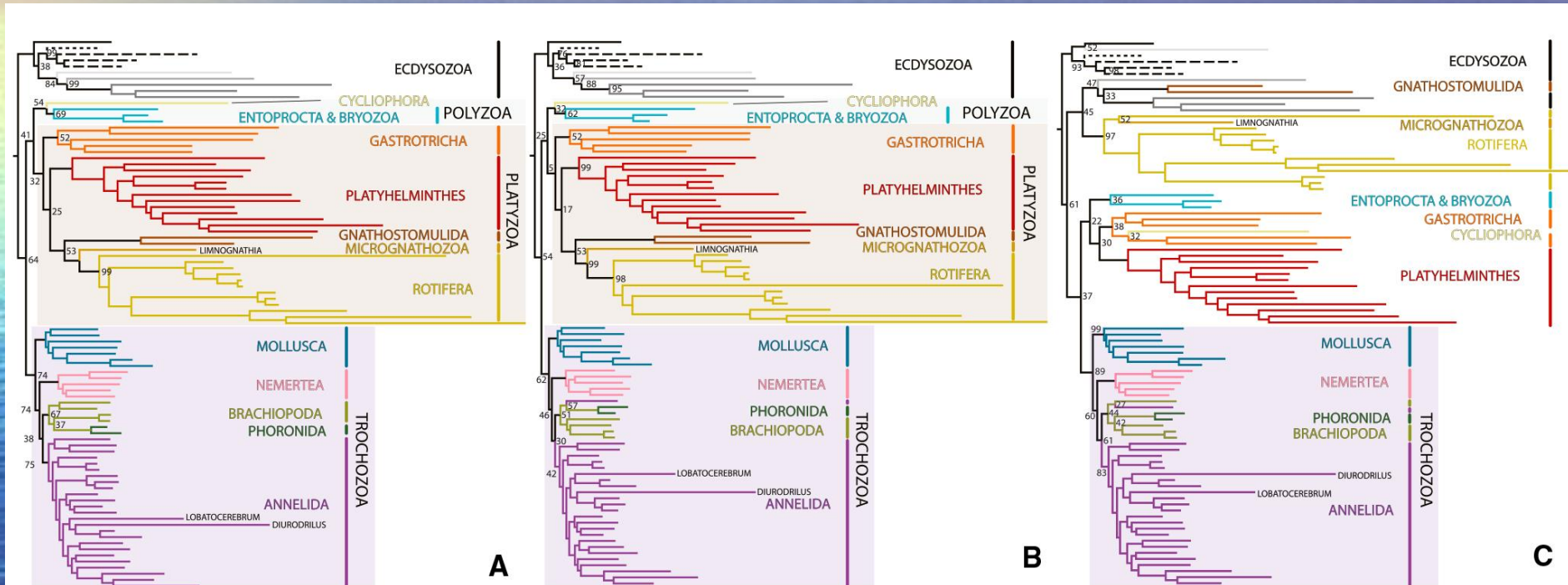κ = 2

Evolutionary rate

Evolutionary rate

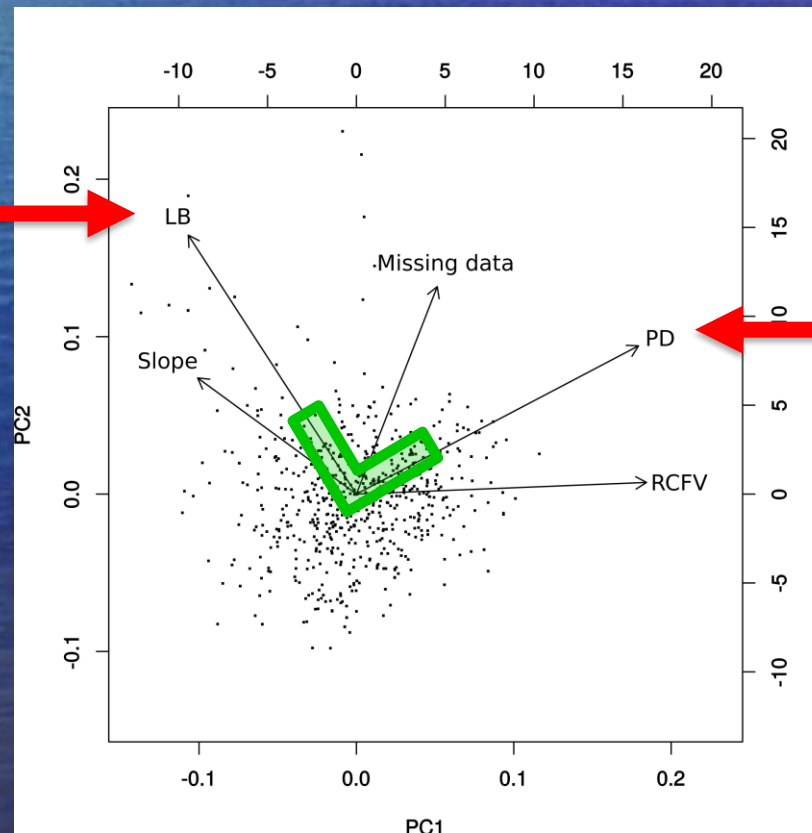**All**   **75% slowest**   **75% fastest**

Laumer et al. (2015) Curr. Biol.

Evolutionary rate &
Long branch attraction

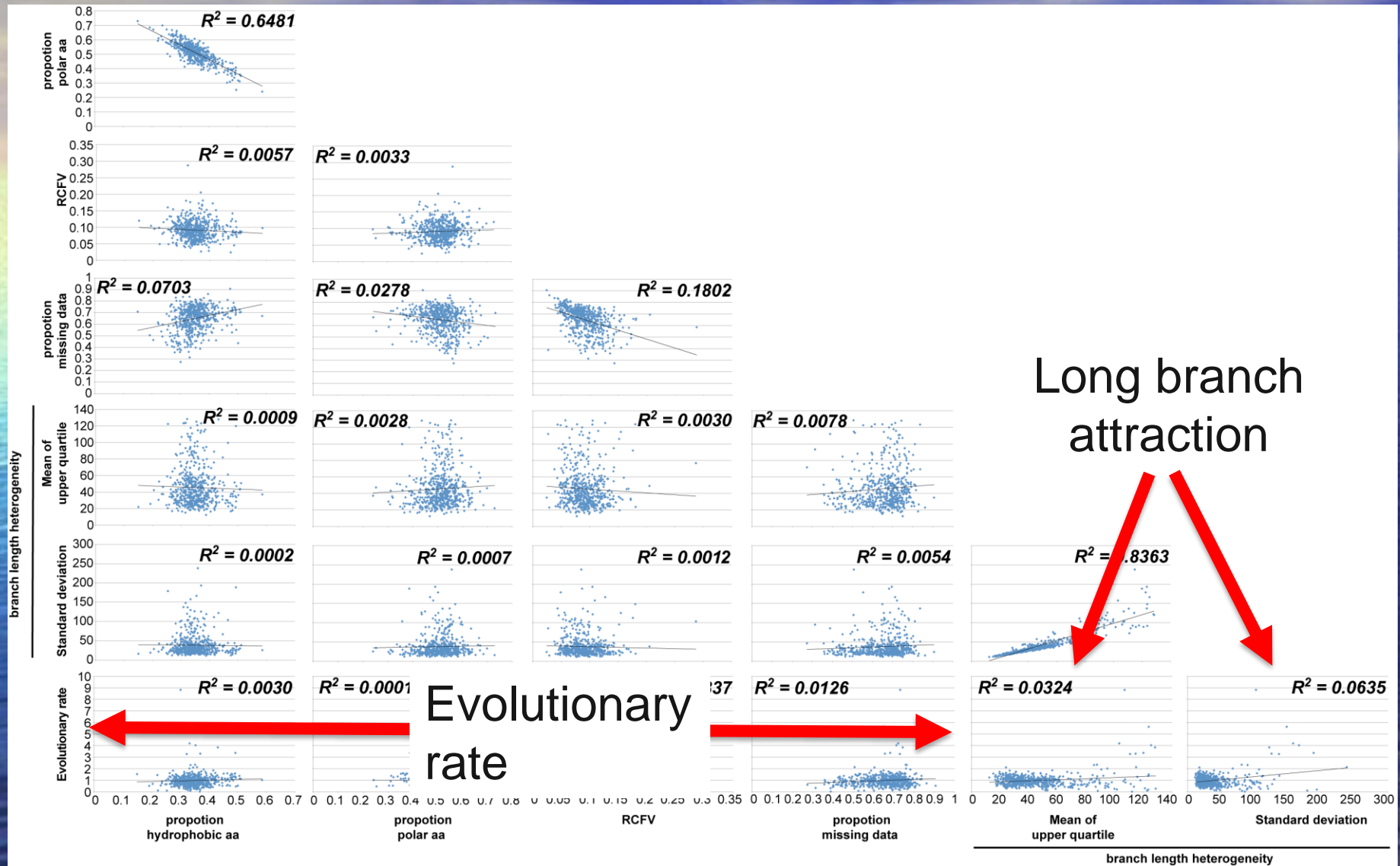Over all evolutionary rate of a gene is often taken as a proxy for long branch attraction.

Long branch attraction

Evolutionary rate

Kocot et al. (2016)

Evolutionary rate & Long branch

Struck et al. (2014) MBE

Evolutionary rate & Long branch attraction

Long branch attraction is caused by differences in branch lengths and NOT by an overall increased substitution rate.

Saturation

How can substitutions saturate?

Transitions are more likely to occur than transversions.

At a certain observed genetic divergence between sequences multiple substitutions become increasingly likely.

Because transitions occur more often than transversions this effect will be seen much earlier in transitions leading to saturation:
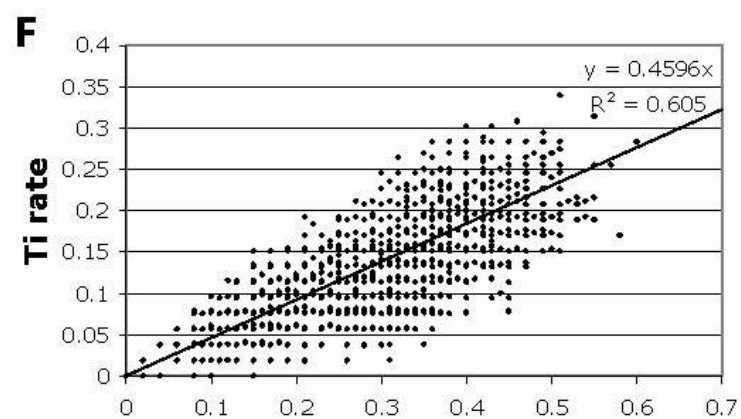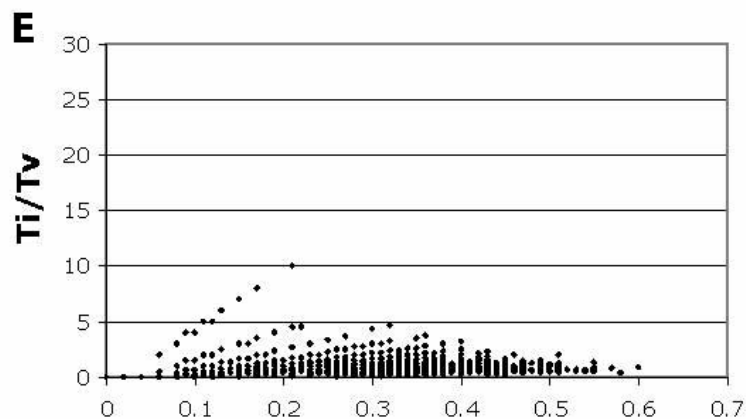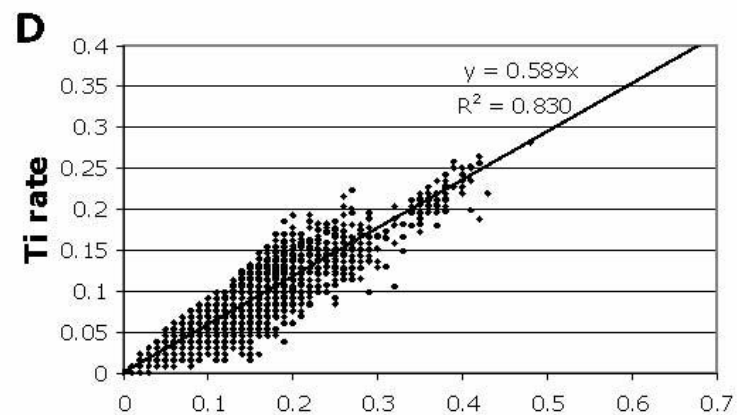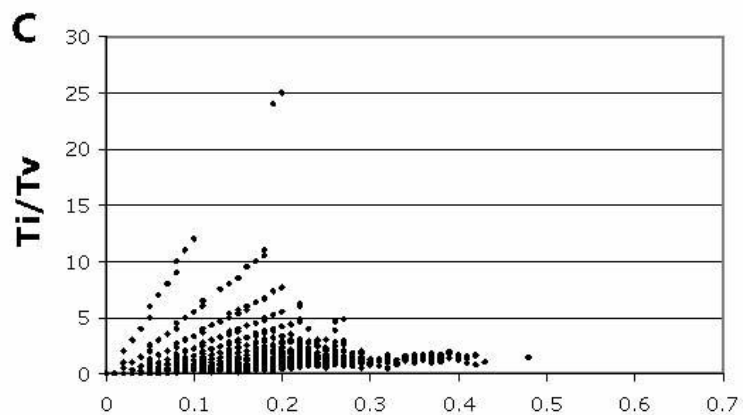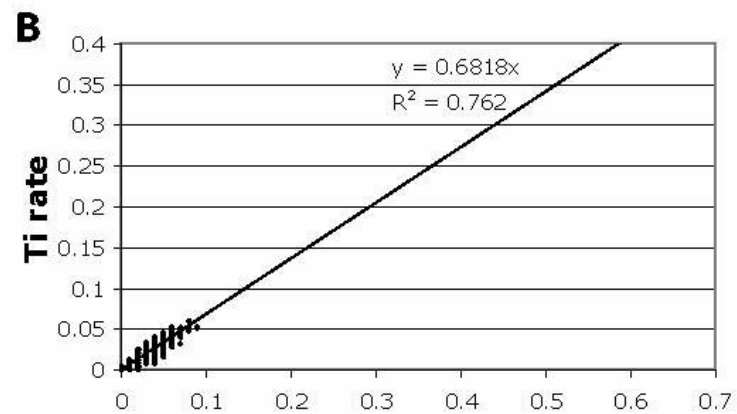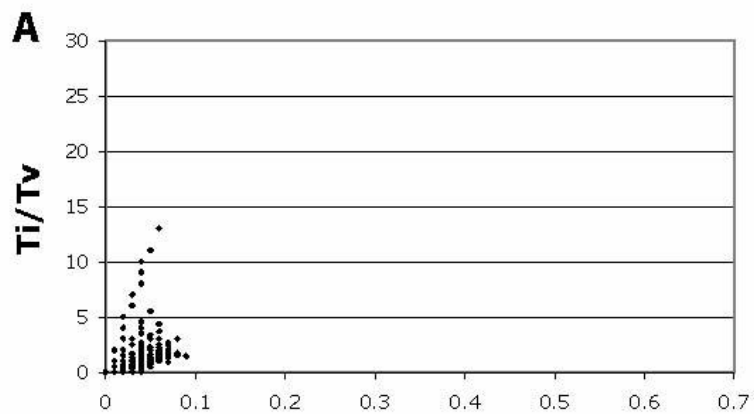
While the genetic divergence is still increasing the transition rate is not.

Saturation
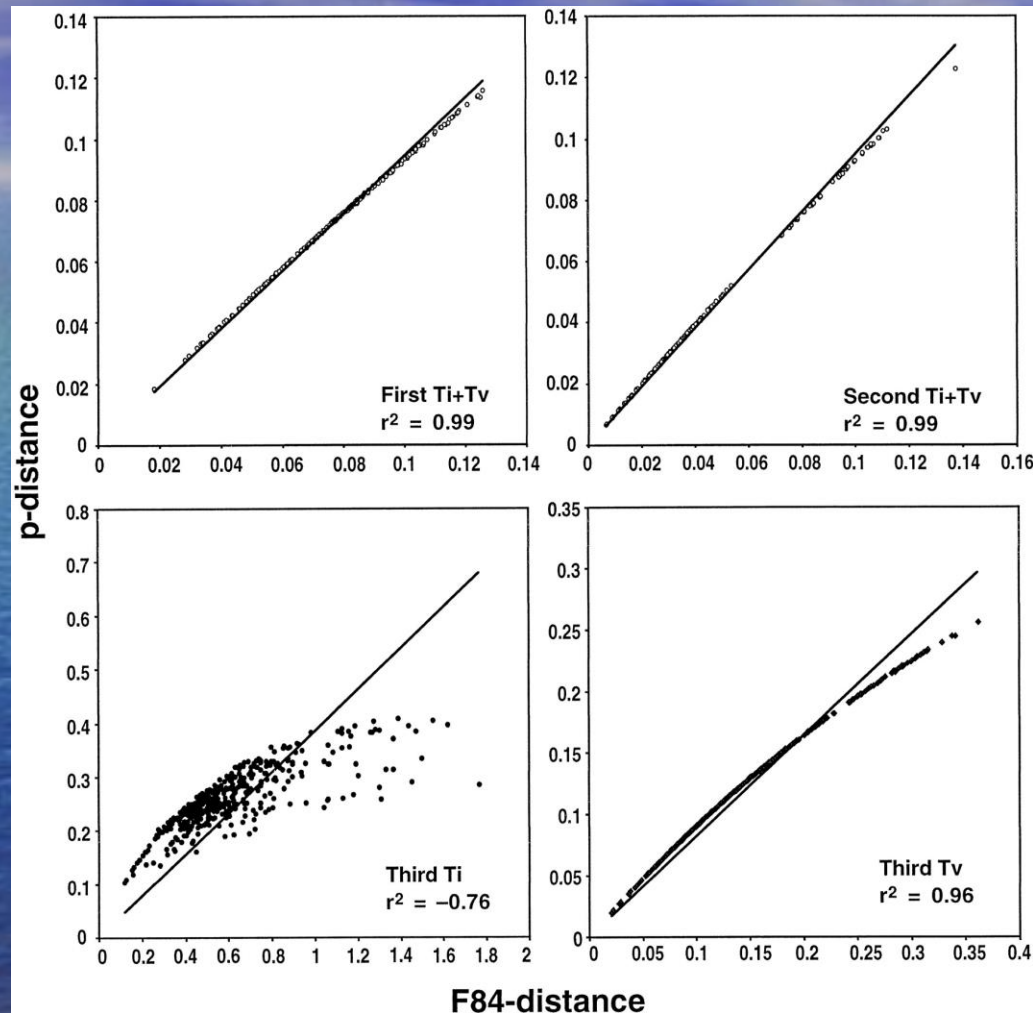
First, second, and third codon positions in rbcL of land plants
(Nickrent et al. 2000 Mol. Biol. Evol. 17:1885-1895)

A

B
y = 0.6818x
R² = 0.762

C

D
y = 0.589x
R² = 0.830

E

F
y = 0.4596x
R² = 0.605

C (=convergence) index

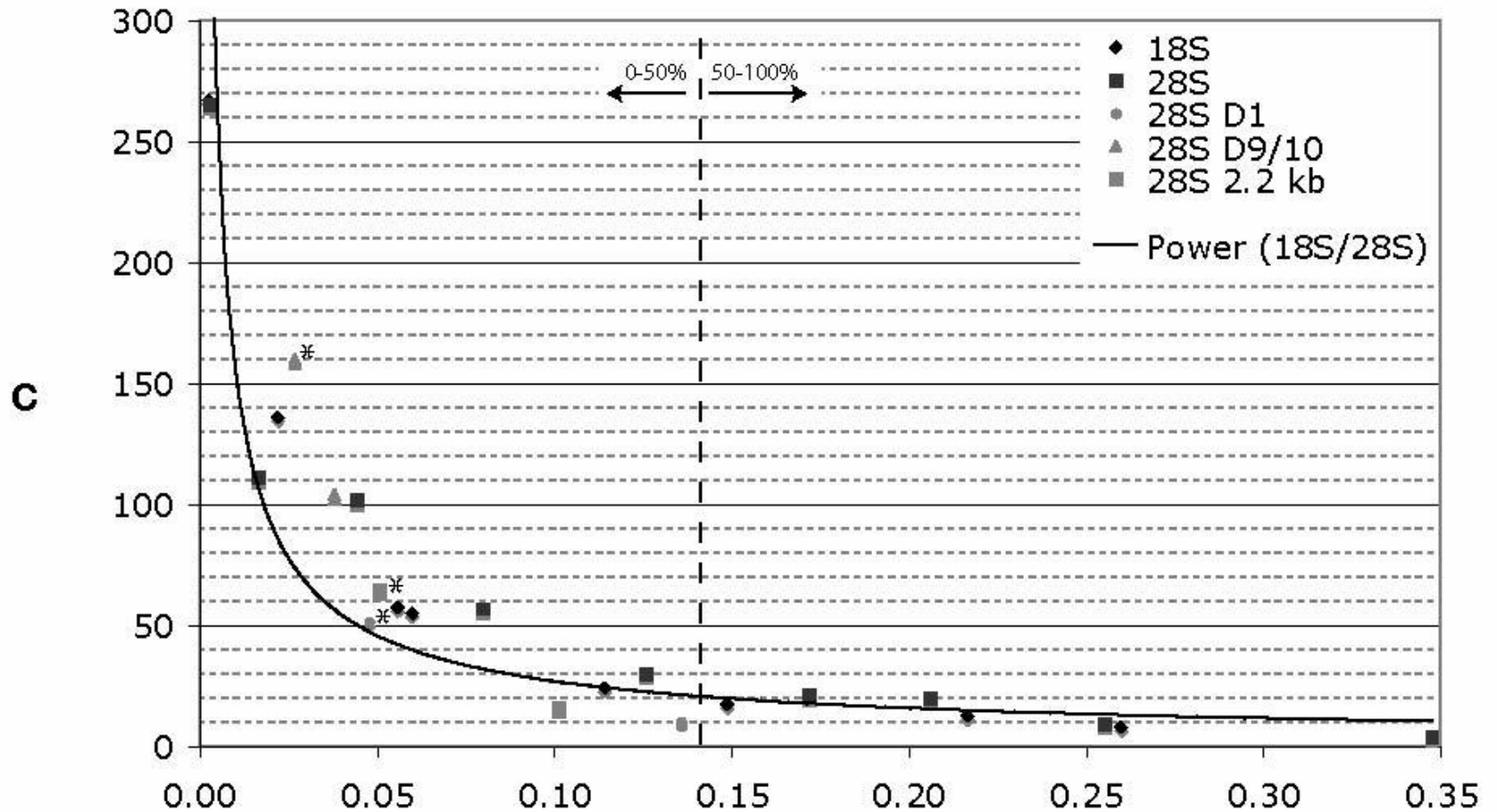**What are the properties of convergence?**

One factor is increasingly spread out, whereas the other becomes more and more condensed around a single value. Therefore, the standard deviation should increase as well as decrease, respectively.

$$C = \frac{\sigma(\frac{ti}{tv})}{\sigma(p)}$$
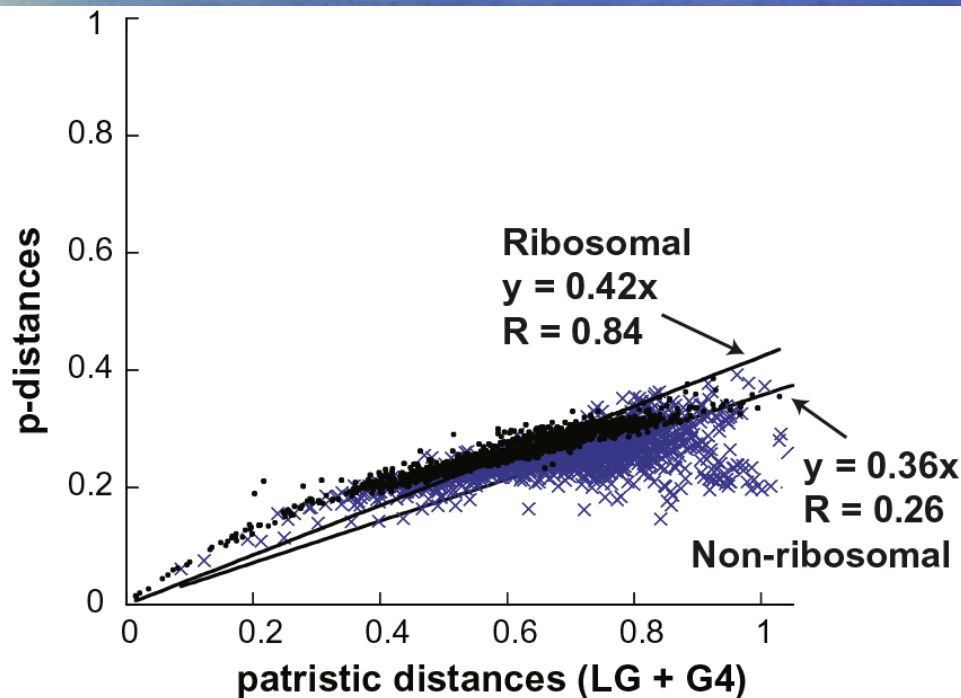
Struck et al. (2008) MPE

C (=convergence) index

R² and slope

Deep metazoan phylogeny: When different genes tell different stories

Tetyana Nosenko [a], Fabian Schreiber [b], Maja Adamska [c], Marcin Adamski [c], Michael Eitel [d,1], Jörg Hammel [e], Manuel Maldonado [f], Werner E.G. Müller [g], Michael Nickel [e], Bernd Schierwater [d], Jean Vacelet [h], Matthias Wiens [g], Gert Wörheide [a,i,j,*]

Fig. 3. Saturation analysis. The relative saturation levels were estimated for the ribosomal and non-ribosomal gene matrices containing 50 taxa by computing the Pearson correlation coefficient *R* and slope of the regression line of patristic vs. *p*-distances. The patristic distances between pairs of taxa were inferred from the branch lengths of ML trees constructed under the LG + Γ8 + I model.

Nosenko et al. (2013) MPE

I_{ss}

An index of substitution saturation and its application

Xuhua Xia,[a,d,*] Zheng Xie,[b] Marco Salemi,[c] Lu Chen,[d] and Yong Wang[d]

- ➢ ancestral state reconstruction of DNA sequences and comparison with simulated data
- ➢ estimation of substitution numbers of neighboring nodes along the tree
- ➢ I_{ss} is the ratio of the average information entropy of positions to the expected entropy of the alignment
- ➢ the smaller I_{ss}, the less substitution saturation has occurred

I_{SS}

Fig. 1. Two extreme topologies used in simulation: (a) symmetrical; (b) asymmetrical.
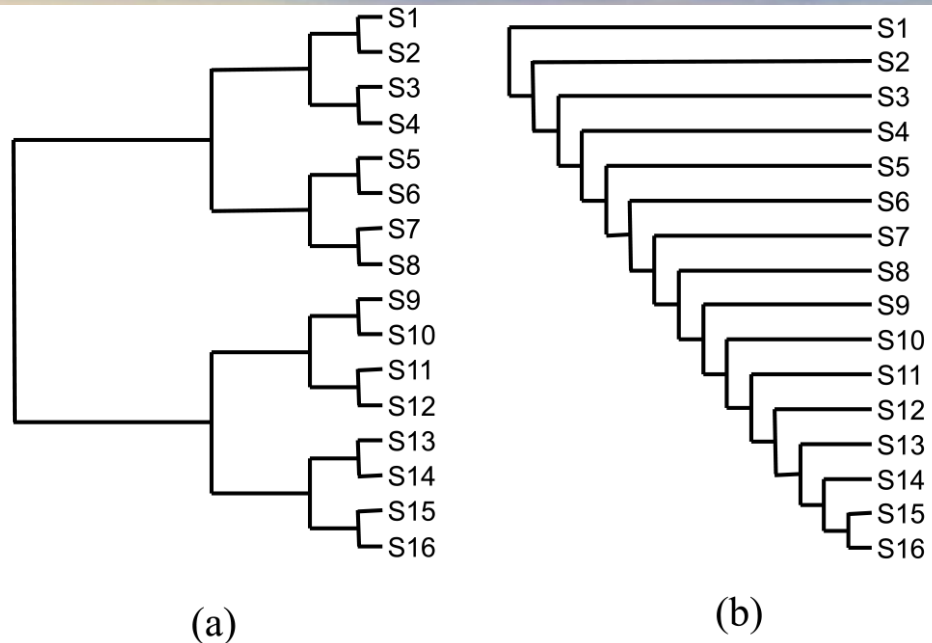
For simulation, only two possibilities of tree topology are considered. The two extremes:
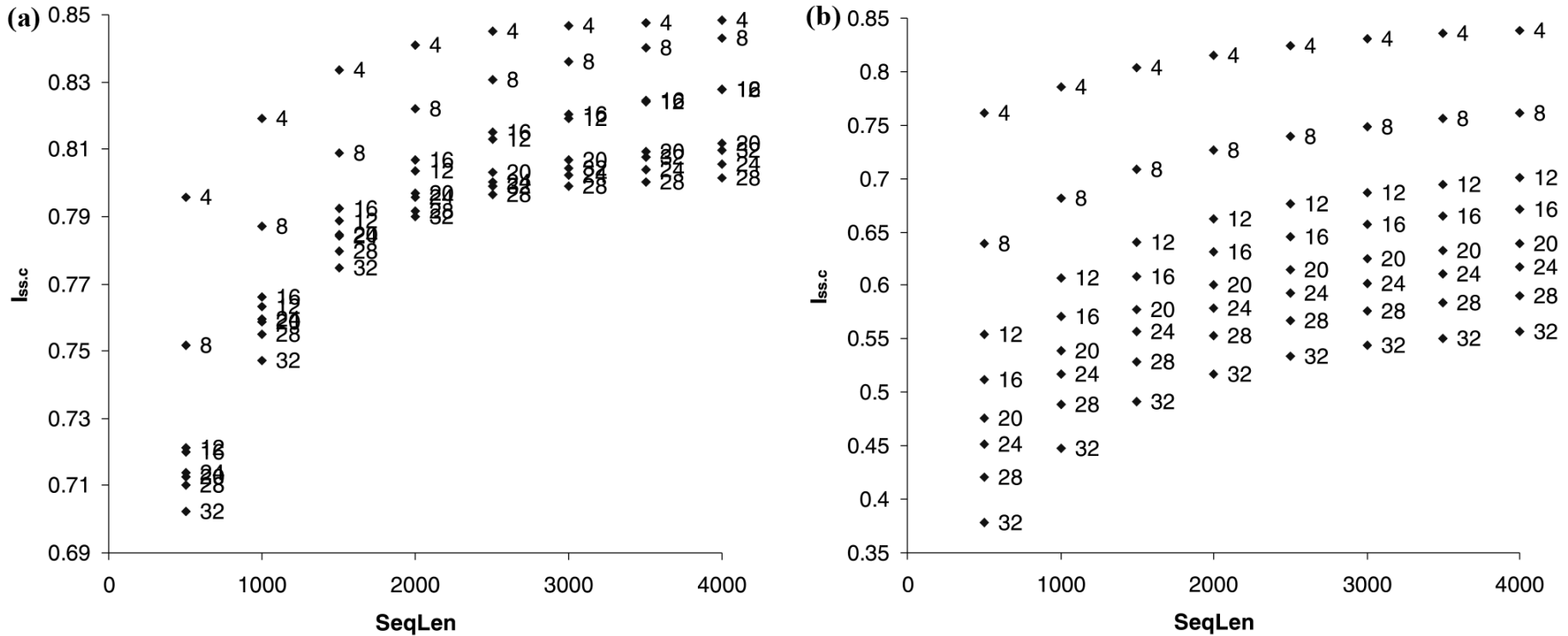- Absolutely symmetrical
- Absolutely asymmetrical

Xia et al. (2003) MPE

Fig. 3. The critical index of substitution saturation ($I_{ss.c}$) depends on the sequence length (SeqLen) and the number of OTUs (N_{OTU}). Data labels are N_{OTU} values: (a) with a symmetrical topology; (b) with an asymmetrical topology.

These are then used for all similar datasets to get the critical I_{ss} value indicating saturation.

Xia et al. (2003) MPE