

# Biological Applications of Mutual Information

Peter Grassberger<sup>1,2</sup>

<sup>1</sup>*Max Planck Institute for the Physics of Complex Systems, D-01187, Dresden, Germany;*

<sup>2</sup>*Jülich Supercomputing Center, Forschungszentrum Jülich, D-52425, Jülich, Germany;*

*E-mail address: p.grassberger@fz-juelich.de*

Previous applications of information theory to biological problems have met with mixed success, to say the least, because it is not clear what one learns from an estimate of any quantity of information (as opposed to its meaning), and that is -- it seems -- all which information theory is about. I will show that this criticism does not apply to mutual information, which basically provides a universally applicable distance measure between any two objects -- or between an object and its environment, in which case it is related to meaning.

I will first recall main features of the mutual information (MI) as a measure for similarity or statistical dependence. In particular, I will discuss its embodiments in two versions of information theory: Probabilistic (Shannon) versus algorithmic (Kolmogorov). I will compare two different strategies for estimating the algorithmic complexity of "texts", one involving sequence alignment and file compression ("zipping"), the other just zipping alone. When applied to mitochondrial DNA, both versions will lead to distance measures which outperform other distance currently in use. The last part of the talk will be devoted to estimating Shannon MI from real-valued data, and an application to microarray gene expression measurements. In particular, I will show that large differences between dependencies estimated from MI and from linear correlation measures may hint to interesting structures which can then be further explored.