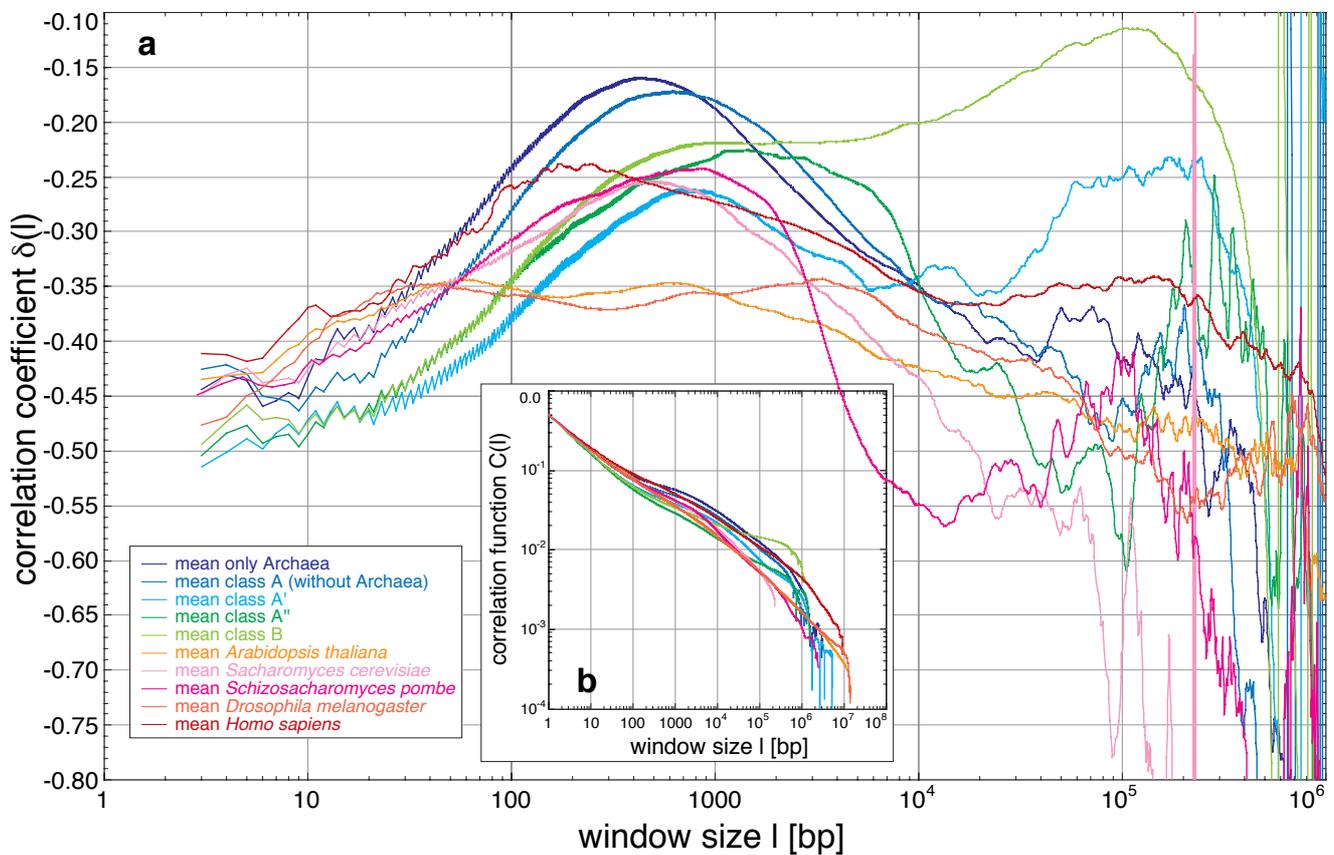


DNA Sequence Patterns

-

A Successful Example of Grid Computing in Genome Research and Building Virtual Super-Computers for the Research Commons of e-Societies



Tobias A. Knoch,

Anis Abuseiris, Rob de Graaf, Michael Lesnussa, Nick Kepper, & Frank G. Grosveld

TA.Knoch@taknoch.org, *Biophysical Genomics, Cell Biology & Genetics, Erasmus MC, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands, and Genome Organization and Function, BioQuant & German Cancer Research Center, Im Neuenheimer Feld 227, 69267 Heidelberg, Germany.*

DNA Sequence Patterns

A Successful Example of Grid Computing in Genome Research and Building Virtual Super-Computers for the Research Commons of e-Societies

Tobias A. Knoch,

Anis Abuseiris, Rob de Graaf, Michael Lesnussa, Nick Kepper, & Frank G. Grosveld

TA.Knoch@taknoch.org, *Biophysical Genomics, Cell Biology & Genetics, Erasmus MC, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands, and Genome Organization and Function, BioQuant & German Cancer Research Center, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany.*

The amount of information is growing exponentially with ever-new technologies emerging and is believed to be always at the limit. In contrast, huge resources are obviously available, which are underused in the IT sector, similar as e.g. in the renewable energy sector. Genome research is one of the boosting areas, which needs an extreme amount of IT resources to analyse the sequential organization of genomes, i.e. the relations between distant base pairs and regions within sequences, and its connection to the three-dimensional organization of genomes, which is still a largely unresolved problem.

The underusage of resources as those accessible by grid with its fast turnover rates is very astonishing considering the barriers for further development put forward by the inability to satisfy the need for such resources. The phenomenon is a typical example of the *Inverse Tragedy of the Commons*, i.e. resources are underexploited in contrast to the unsustainable and destructing overexploitation in the *Classic Tragedy of the Commons*. An analysis of IT and the grid sector which attempts to share resources for better usage efficiency, reveals two challenges, which lead to the heart of the paradox: i) From a macro perspective all grid infrastructures involve not only mere technical solutions but also dominantly all of the autopoietic social sub-systems ranging from religion to policy. ii) On the micro level the individual players and their psychology and risk behaviour are of major importance for acting within the macro autopoietic framework. Consequently, the challenges of grid implementation are similar to those of other pressing global issues as e.g. climate protection. This is well described by extending the *Human Ecology* triangle to a rectangle: environment-individual-society-environment. By applying this extension of this classical field of interdisciplinary basic and applied research to the grid sector, i.e. by further extension to an *e-Human Grid Ecology* rational, the *Grid Inverse Tragedy of the Commons* can be understood and approached regarding the internalization challenge into e-Society and e-Life, from which then guidelines for the day-to-day management can be derived. This is of general importance for many complex fields and thus with similar paradoxes and challenges.

By using grid Long-range power-law correlations were found using correlation analysis on almost the entire observable scale of 132 completely sequenced chromosomes of 0.5×10^6 to 3.0×10^7 bp from Archaea, Bacteria, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*. The local correlation coefficients show a species specific multi-scaling behaviour: close to random correlations on the scale of a few base pairs, a first maximum from 40 to 3400 bp (for *Arabidopsis thaliana* and *Drosophila melanogaster* divided in two submaxima), and often a region of one or more second maxima from 10^5 to 3×10^5 bp. Within this multi-scaling behaviour, an additional fine-structure is present and attributable to codon usage in all except the human sequences, where it is related to nucleosomal binding. Computer-generated random sequences assuming a block organization of genomes, the codon usage, and nucleosomal binding explain these results. Mutation by sequence reshuffling destroyed all correlations. Thus, the stability of correlations seems evolutionarily tightly controlled and connected to the spatial genome organization. In summary, genomes show a complex sequential organization related closely to their three-dimensional organization.

Consequently, grids can be established by solving the *Grid Inverse Tragedy of the Commons* using a *e-Human Grid Ecology* rational and indeed be used as e.g. in genome research for DNA sequence pattern analysis very successfully to determine for decades unresolved questions which demand very heavy IT support. Thus, indeed the solutions for the demand requirements in the research commons of e-Societies can be tackled successfully by such a systemic approach.

Literature:

Knoch, T. A. Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic properties in the simulation of interphase chromosomes and cell nuclei, long-range correlations in complete genomes, *in vivo* quantification of the chromatin distribution, construct conversions in simultaneous co-transfections. *Dissertation*, Ruperto-Carola University, Heidelberg, Germany, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN-10: 3-00-009959-X and 978-3-00-009959-5 (softcover), ISBN-10: 3-00-009960-3 and 978-3-00-009960-1 (hardcover), 2002.

Knoch, T. A., Göcker, M., Lohner, R., Abuseiris, A. & Grosveld, F. G. Fine-structured multi-scaling long-range correlations in completely sequenced genomes - features, origin and classification. *Eur. Biophys. J.* 38(6), 757-779, 2009.

Knoch, T. A., Baumgärtner, V., Grosveld, F. G. & Egger, K. Approaching the internalization challenge of grid technologies into e-Society by e-Human "Grid" Ecology. *Economics of Grids, Clouds, Systems, and Services – GECON 2010 Proceedings*, 7th International Workshop, Ischia, Italy, editors Altman, J., & Rana, O. F., Lecture Notes in Computer Science (LNCS) 6296, Springer Berlin Heidelberg New York, ISSN 0302-9743, ISBN-10 3-642-15680-0, ISBN-13 978-3-642-15680-9, 116-128, 2010.

Knoch, T. A. Sustained Renewability: approached by systems theory and human ecology. *Renewable Energy 2*, editors M. Nayeripour & M. Keshti, Intech, ISBN 978-953-307-573-0, in press, 2011.