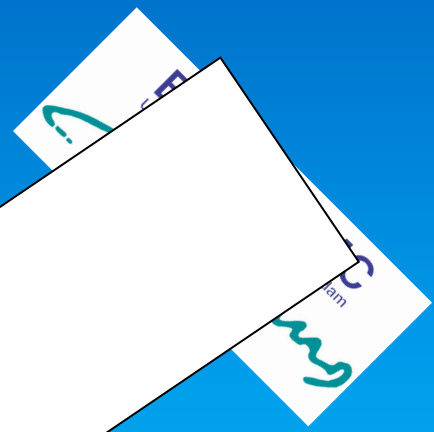




DNA Sequence Patterns a Successful Example of Grid Computing in Genome Research



**(c) 1996 - 2011
All Rights Reserved!**

Tobias A. Knoch

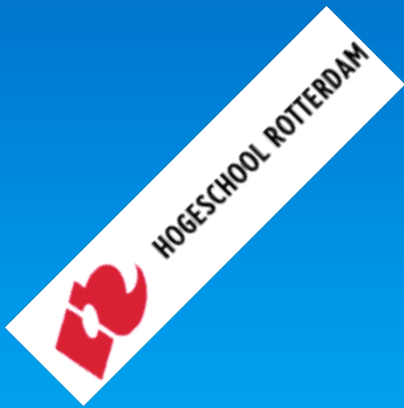
with

A. A. Graaf, M. Lesnussa, N. Kepper, Ad Emmen, & F. G. Grosveld

Biophysical Genomics & Erasmus Computing Grid

Dept. Cell Biology & Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.

BioQuant Center & German Cancer Research Center (DKFZ), 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

with

A. Abuseiris, R. de Graaf, M. Lesnussa, N. Kepper, Ad Emmen, & F. G. Grosveld

Biophysical Genomics & Erasmus Computing Grid

Dept. Cell Biology & Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.

BioQuant Center & German Cancer Research Center (DKFZ), Heidelberg, Germany.

Abstract & Literature



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.

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.

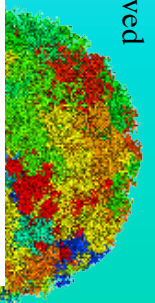
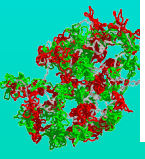
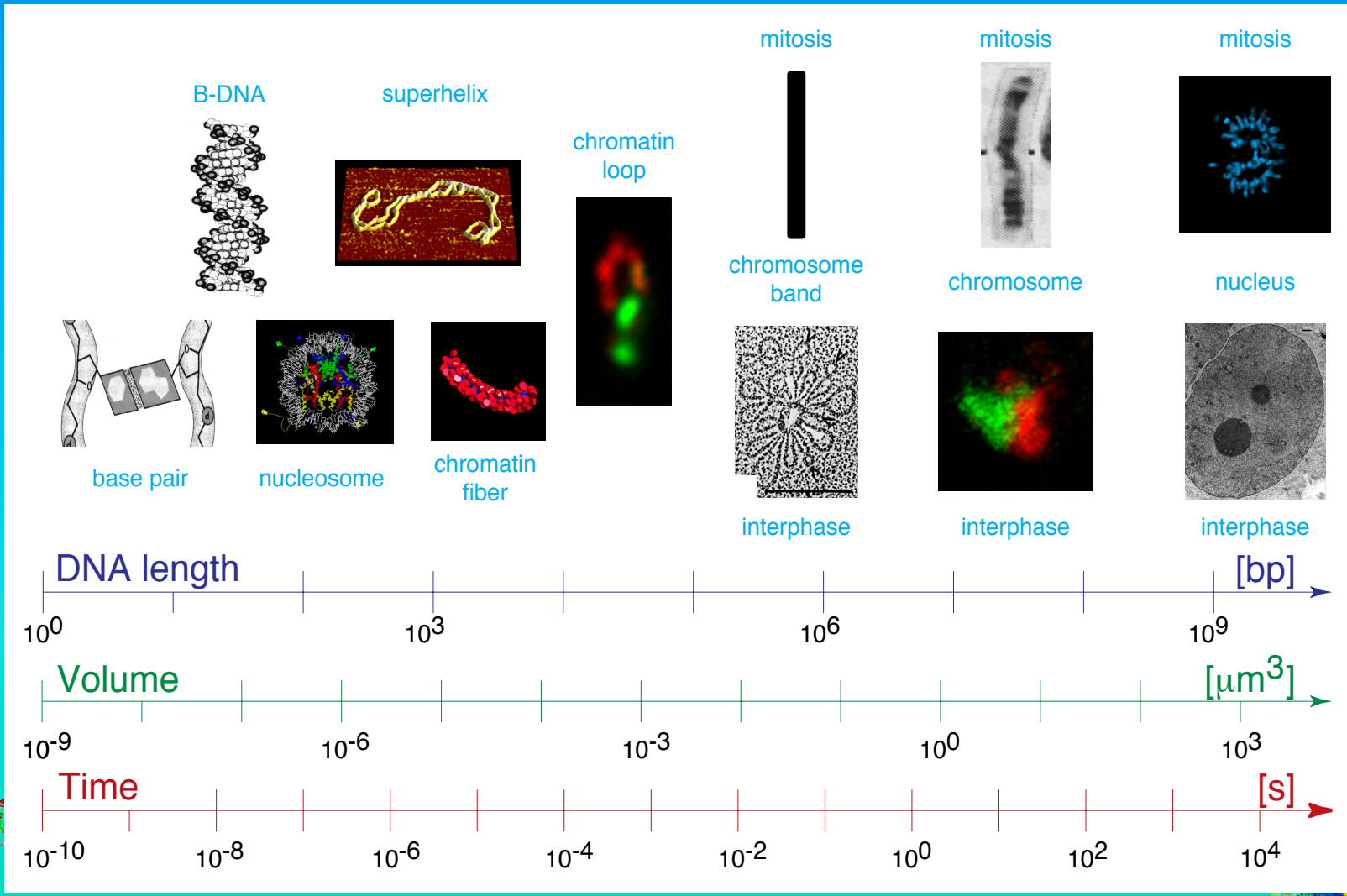
Knoch, T. A., Göcker, M., Lohner, R., Abuseiris, A. & Grosveld, F. G. Fine-structured multi-scaling long-range correlations in completely sequenced genomes - features, origin and classification. *Eur. Biophys. J.* 38(6), 757-779, 2009.

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.

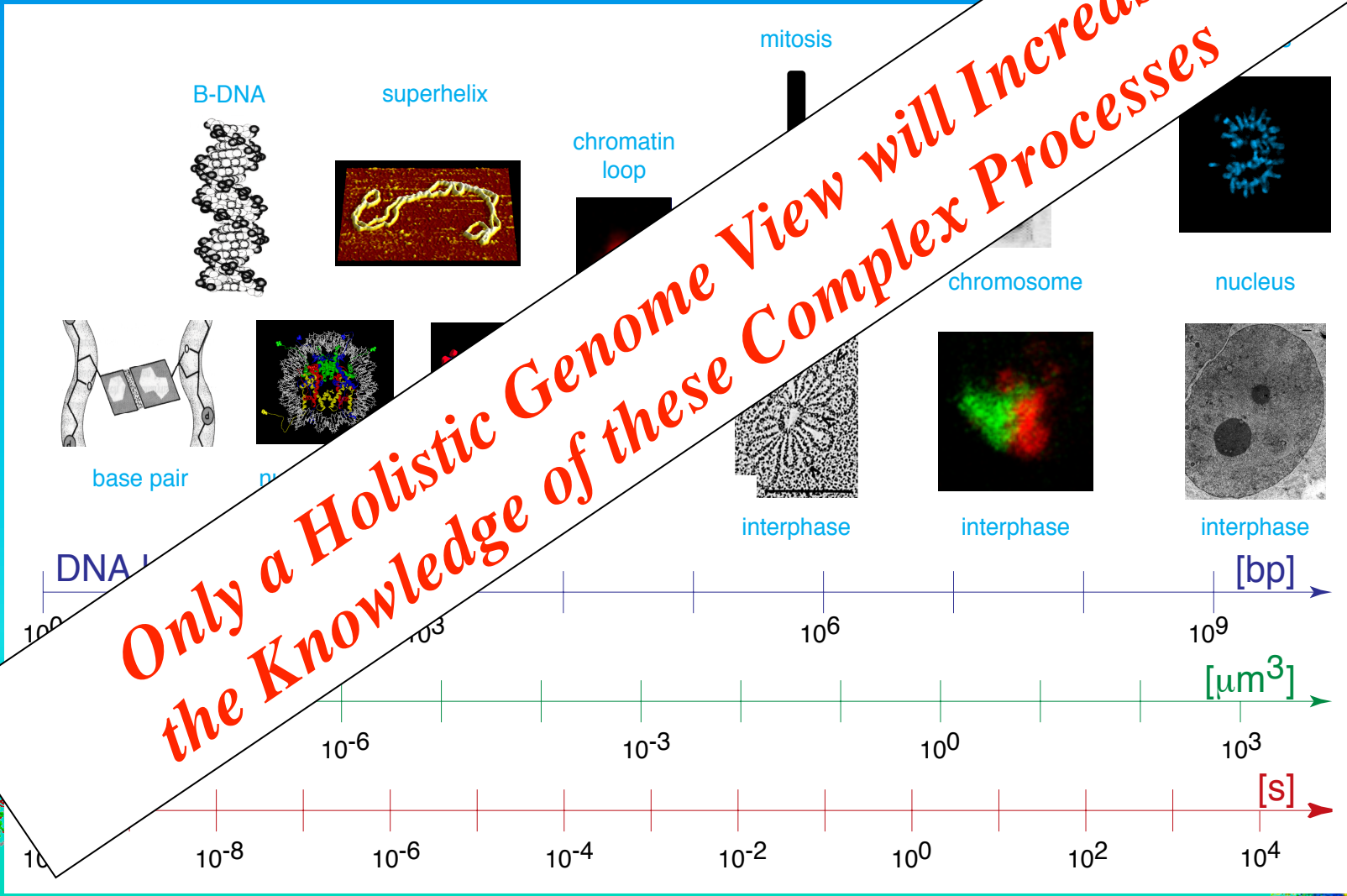
Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogenesis, or cancer development is still unclear?



Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogenesis, or cancer development is still unclear?

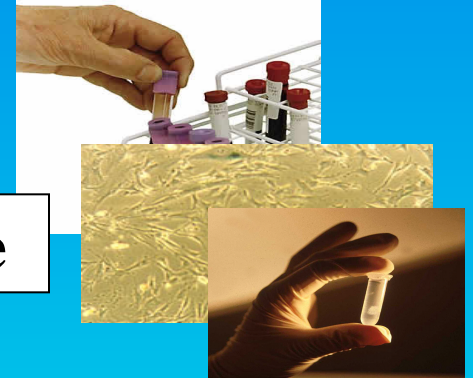
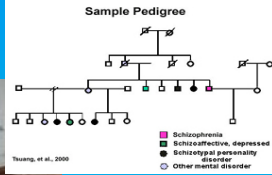


Only a Holistic Genome View will Increase the Knowledge of these Complex Processes

The Complexity of Cytogenetic Diagnostics

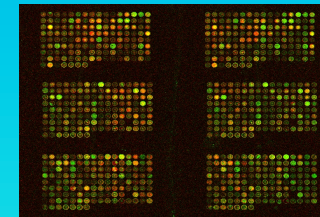
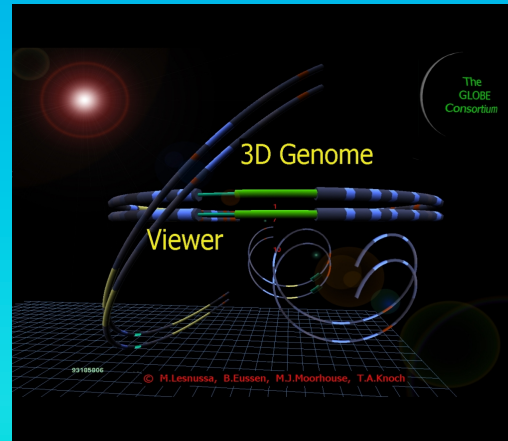
The process of cytogenetic analysis requires proper patient and sample analysis as well as a comprehensive evaluation of the results.

Patient



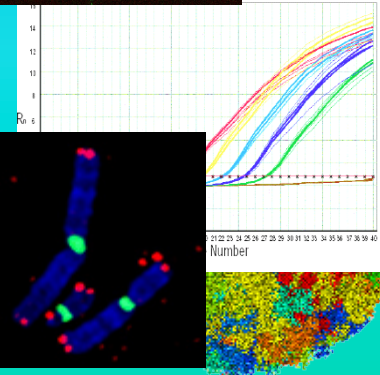
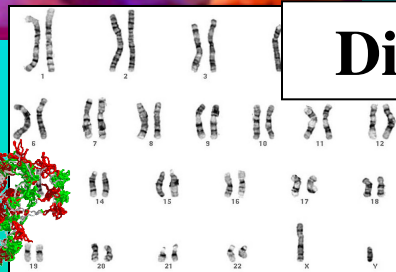
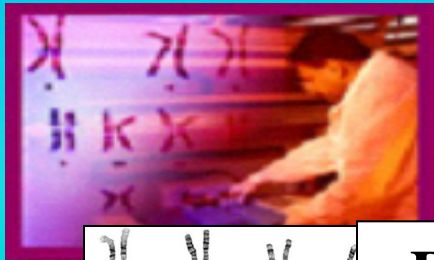
Sample

Treatment



Analysis

Diagnosis

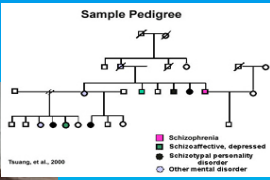


The Complexity of Cytogenetic Diagnostics

The process of cytogenetic analysis requires proper patient and sample analysis as well as a comprehensive evaluation of the results.



Patient



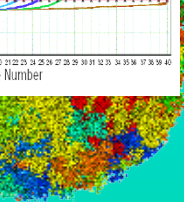
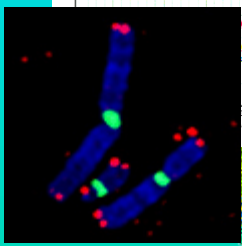
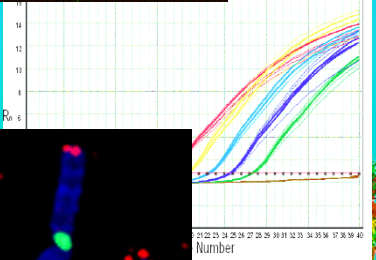
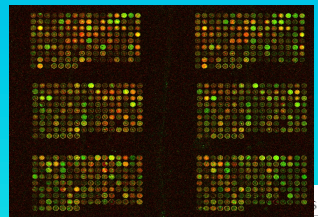
Treatment

Sample



Diagnosis

Analysis



Better Cytogenetic Diagnostics is Based on Advanced Genome Understanding!

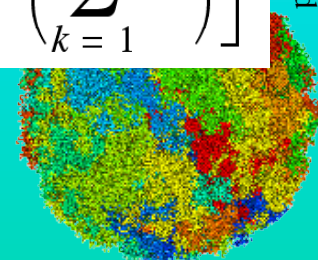
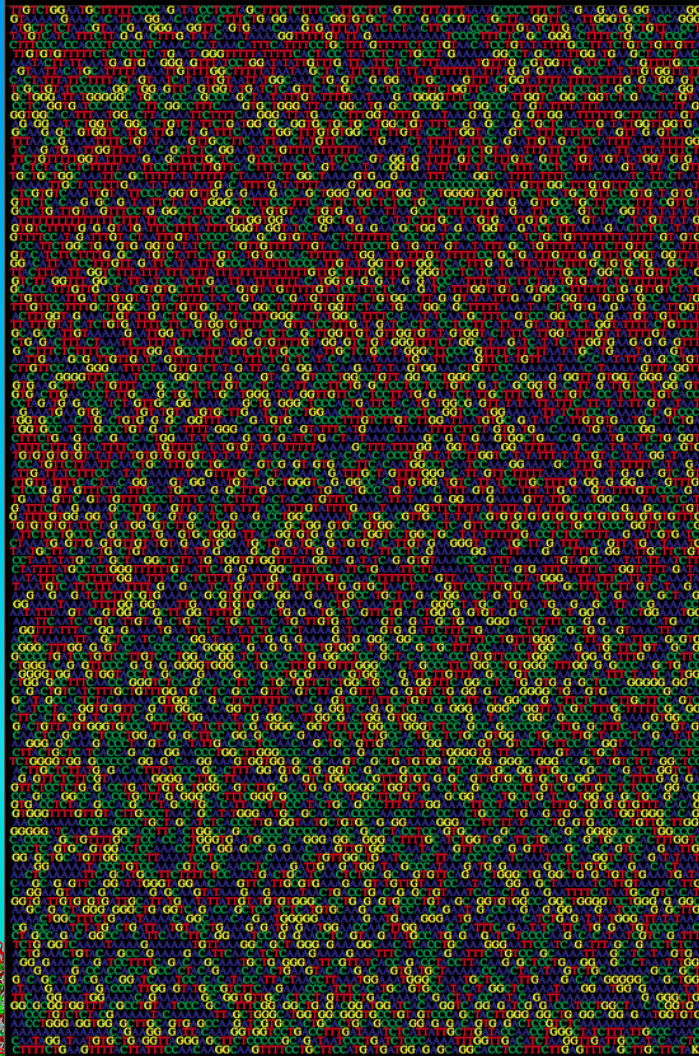
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ are an indication for the i) degree of long-rang scaling behaviour, ii) general multi-scaling, and iii) fine-structure features, which all are connected to all levels of genome organization and especially also to the three-dimensional genome architecture.

$$C(l) = \sqrt{\langle (c_l - \bar{c}_L)^2 \rangle_s}$$

$$C(l) = \sqrt{\frac{1}{L-l+1} \sum_{s=1}^{L-l} \left(\frac{1}{l} \sum_{k=1}^l n - \frac{1}{L} \sum_{k=1}^L N \right)^2}$$

$$C(l) = \frac{1}{Ll} \sqrt{\frac{1}{L-l} \sum_{s=1}^{L-l} \left[\left(\sum_{k=1}^l Ln \right) - \left(\sum_{k=1}^L lN \right) \right]^2}$$



The Erasmus Computing Grid



The largest desktop grid for the biomedical research and care sectors with now ~13 Tera FLOPS and a potential ~50 Tera FLOPS and ~20,000 desktops, at two city wide institutions: the Hogeschool Rotterdam and the Erasmus Medical Center.

~ 10 BioMedical
User Groups

ECG - Centralized Office

Two Donor Organizations



~20,000 PC "Owners",
i.e. Local PC Donors.

Research:

- ❖ genomic and proteomic analysis
- ❖ epidemiology
- ❖ image analysis, e.g. Applied Molecular Imaging (AMI)

Education:

- ❖ training of the coming grid generation of IT specialists
- ❖ developing new concepts for grid computing

Diagnostics:

- ❖ clinical image and data analysis
- ❖ operation planning and operation support

Industry:

- ❖ brokerage of computing resources

Dedicated and Secured!

ECG Projects and Users

The projects and users currently run on the ECG are mainly basic research and application development in the biological and medical areas. Currently there are ~15 projects and ~10 users of the ECG with large expansion already in sight.



The ECG is one of the few grids on this level of complexity !

<i>Image Analysis</i>	->	Dept. Medical Informatics
<i>DNA Sequence Correlation</i>	->	Dept. Cell Biology & Genetics
<i>Genome Duplication Analysis</i>	->	Dept. Cell Biology & Genetics
<i>Epidemiology studies</i>	->	Dept. Public Health
<i>CT Image Analysis</i>	->	Dept. Internal Medicine
<i>Structural Genome Simulations</i>	->	Dept. Cell Biology & Genetics
<i>Protein Motive Patterns</i>	->	Dept. Urology
<i>Mass Spectrometry Analysis</i>	->	Center for Biomics
<i>Chip & Sequencing Analysis</i>	->	Center for Genomics
<i>Genome Wide Association Studies</i>	->	Dept. Internal Medicine
<i>Protein Structure Prediction</i>	->	Dept. Cell Biology & Genetics
<i>Diffusion of Molecules</i>	->	Dept. Pathology
<i>Microscopic Image Analysis</i>	->	Dept. Microscopy & Cell Biology

ECG Basic Grid Structure

Erasmus Computing Grid



The ECG consists of the computer pools of the Erasmus MC and the Hogeschool Rotterdam. The client computers are controlled by the general ECG Emperor of the grid at EMC and a local master at HR. "Condor" is used as middleware since it is very well established and open-source.

Erasmus Computing Grid

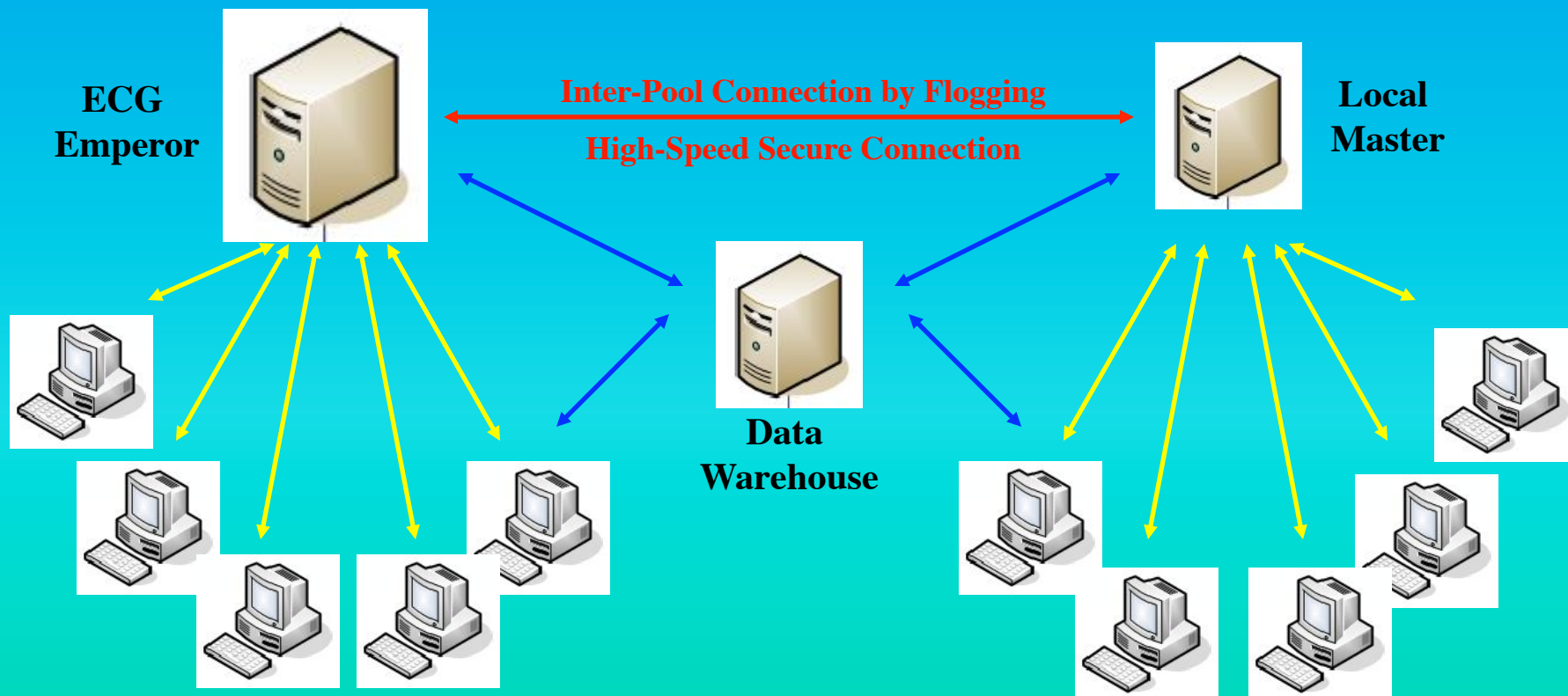


~16000 CPUs or ~50000 vHosts



HOGESCHOOL ROTTERDAM

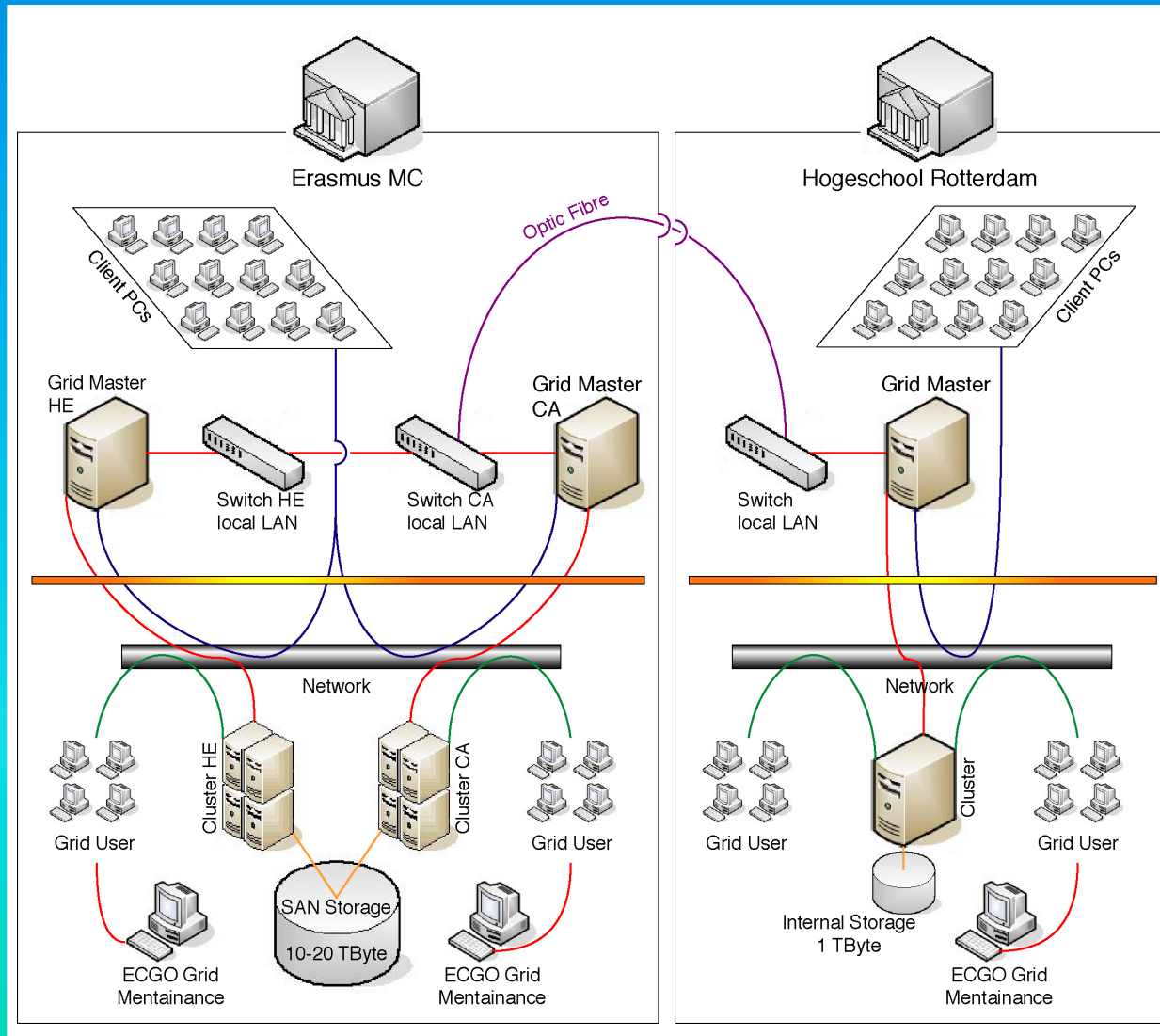
~4500 CPUs or ~15000 vHosts



ECG - Infrastructure

The infrastructure of the ECG consist of a dedicated architecture respecting client, user and management security and efficiency in respect to the fact that the ECG is one of the largest grids in the world with the highest possible degree of complexity.

The new ECG infrastructure guaranties the exploitation of the capabilities to the maximum !



MediGRID and Services@MediGRID



MediGRID and Services@MediGRID operate the national biomedical research and care cluster-grid within the national German D-Grid initiative and integrate various disciplines, institutions, and states throughout Germany.



Module Coordination:

- ❖ coordination of the distributed office

Module Resource Integration:

- ❖ sharing of the integrated resources

Module Middleware:

- ❖ grid technical virtualization

Module Ontology Tools:

- ❖ ontology development for grid user projects

Module BioMedical Informatics:

- ❖ user projects in biomedical research

Module Clinical Imaging:

- ❖ user projects in clinical imaging

Module Clinical Research:

- ❖ user projects in general clinical research

Module e-Science:

- ❖ general research on e-grid science

Services@MediGRID:

- ❖ services towards MediGRID

Large-Scale Resource Sharing in IT: The *Inverse* Tragedy of the Commons



The grid phenomenon and its implications are similar complicated to the ecology/climate/environmental challenge!

The Tragedy of the Commons:

=> A resource belonging to all and being on limited demand is overexploited / destroyed by the users due to responsibility diffusion!

<=> transforms into =>

:The *Inverse* Tragedy of the Commons

A Resource belonging to all and being in affluent availability on limited demand is <=> underexploited by potential users due to responsibility diffusion !!!!!

Similarity: Renewable Energy Resource!

The grid challenge lies in the e-Social embedding of grid phenomena:

- Micro-Social: the sharing attitude and socialization of the individual.
- Macro-Social: the organization culture of the embedding institution.

Autopoietic Social Sub-Systems: The Grid Challenge of Integration



The social systems theory by Niklas Luhmann (1927-1998) based on the autopoietic concept of Humberto Maturana and Francisco Varela (1946-2001) is so far the most advanced social systems theory existing to describe the complexity of grid implementation.

The Social Sub-Systems Involved:

❖ Religion

❖ Education

❖ Science

=> currently grid involves only considerably => SCIENCE

❖ Art

❖ Economy

❖ Jurisdiction

❖ Policy

:The Autopoietic Tragedy of Social Sub-Systems

The subsystems have their own code of communication and are separated from each other in a way blocking in principle a consistent integration although they form a society with all their contradictions !!!!!

The e-Social challenge lies in the integration of sub-systems towards a working grid society:

- **Micro-Sub-Systems:** the sub-system stickiness of individuals.
- **Macro-Sub-Systems:** the integration of institutionalized sub-systems via soft interfaces.

From Individual to Cultural Risk Management

Grid implementation and social spread is carried first by individuals as with any meme introduced into society: in the focus of the transformation of society to e-Society stands - as always - the balance between potential risk and opportunity.

The Risk Psychology Matrix:

- ❖ Individual Security Perception & Risk Acceptance
- ❖ Knowledge-Based Security & Risk Acceptance
- ❖ Incidental Security Reaction Behaviour
- ❖ Legal and Political Security Scenarios
- ❖ Religious & Cultural Security Archetypi

:The Autopoietic Link

Genetics & Deep Psychology (C.G. Jung)

Education & Science

Economics & Realities

Jurisdiction & Politics

Religion, Art & Culture

linking grid psychology with autopoietic social sub-systems

The grid challenge lies in a unified concept addressing the psychology of grid:

- **Micro-Risk-Management:** the micro-risk in the perception the individual and its emotional well-being.
- **Macro-Risk-Management:** the macro-risks in the procedural and institutionalization in organizations.

Security of the ECG

One of the major issues concerning grids is to create a security level sufficient for all involved parties as we did for the ECG. The technology, culture and even the psychology of security have been and are treated very seriously by the ECG and are to the highest achievable level guarantying the integrity and privacy of the system and data !



Security Technology

- ❖ system hardening
- ❖ firewalling
- ❖ encrypted network communications
- ❖ intrusion detection monitoring
- ❖ logging of session
- ❖ auditing and testing of applications
- ❖ virus/trojan checking of in-/output
- ❖ authentication

Security Culture

- ❖ trained employees
- ❖ secure programming
- ❖ change management system
- ❖ constant check of procedure
- ❖ constant test and training of employees
- ❖ transparent management structure
- ❖ transparent infra structure
- ❖ NAN and ISO certification

**The highest security level guarantying privacy is just good enough for us.
&
We still would like to have it better.**

The success of grid is based on a sustainable grid ecology within the e-Society, i.e. the e-Human Ecology of Grid reaches a equilibrated space within the integration of grid psychology with autopoietic e-Social sub-systems. Human Ecology first evolved in Chicago in the 1920's in the area of city development by Robert Park (1864-1944) and Ernest Burgess (1886-1966).

The Definition of e-Human "Grid" Ecology:

"Under e-Human "Grid" Ecology we understand the complete science of the relationships of grid to the surrounding environment to which we can count all conditions of existence in the widest sense."¹

¹ Haeckel, E., Generelle Morphology der Organismen, Berlin, Band 2, Allgemeine Entwicklungsgeschichte, p. 286, 1866.

² Haeckel, E., Natürliche Schöpfungsgeschichte, 9. Auflage, Berlin, p. 793, 1898

(e-Human "Grid" Ecology "is) ...the relationship between grid and all other e-Social systems."²

The solutions of the grid challenge on the operational layer are addressed by:

- **Micro-Operationality:** the participative integration of fundamental IT applications of major individual users complying with the psychology of grid in an e-Human Ecology manner.
- **Macro-Operationality:** the set-up of an open and sustainable management structure complying to all the autopoietic e-Social sub-systems in an e-Human Ecology manner.

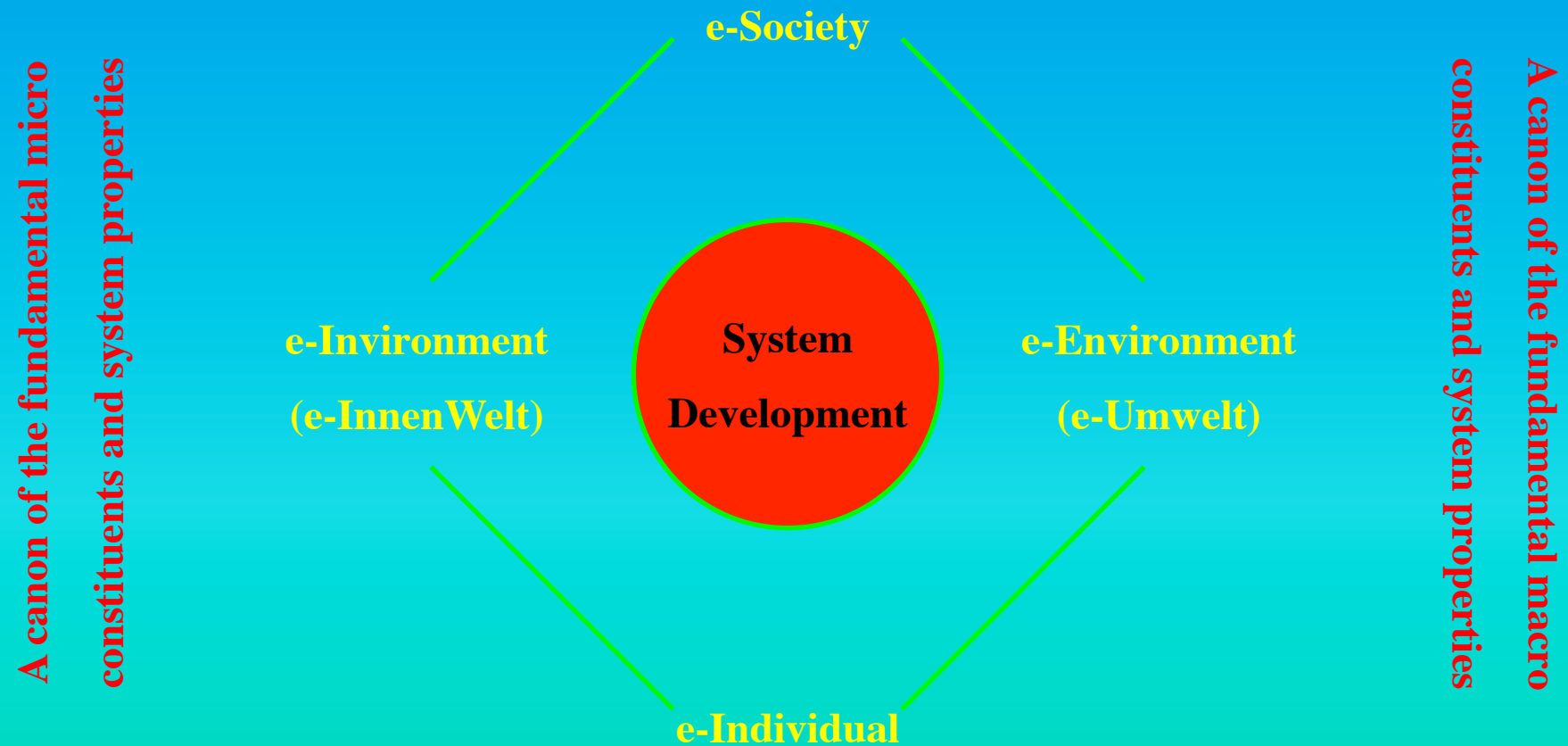
⇒ Solution to the Inverse Tragedy of the Commons!

The Curriculum of e-Human Ecology

Curricula of the canon and grammar of e-Human Ecology need to contain the further developed classic contents of the human ecology rectangle as well as the knowledge and understanding of the special aspects of e-Life and a-Life.

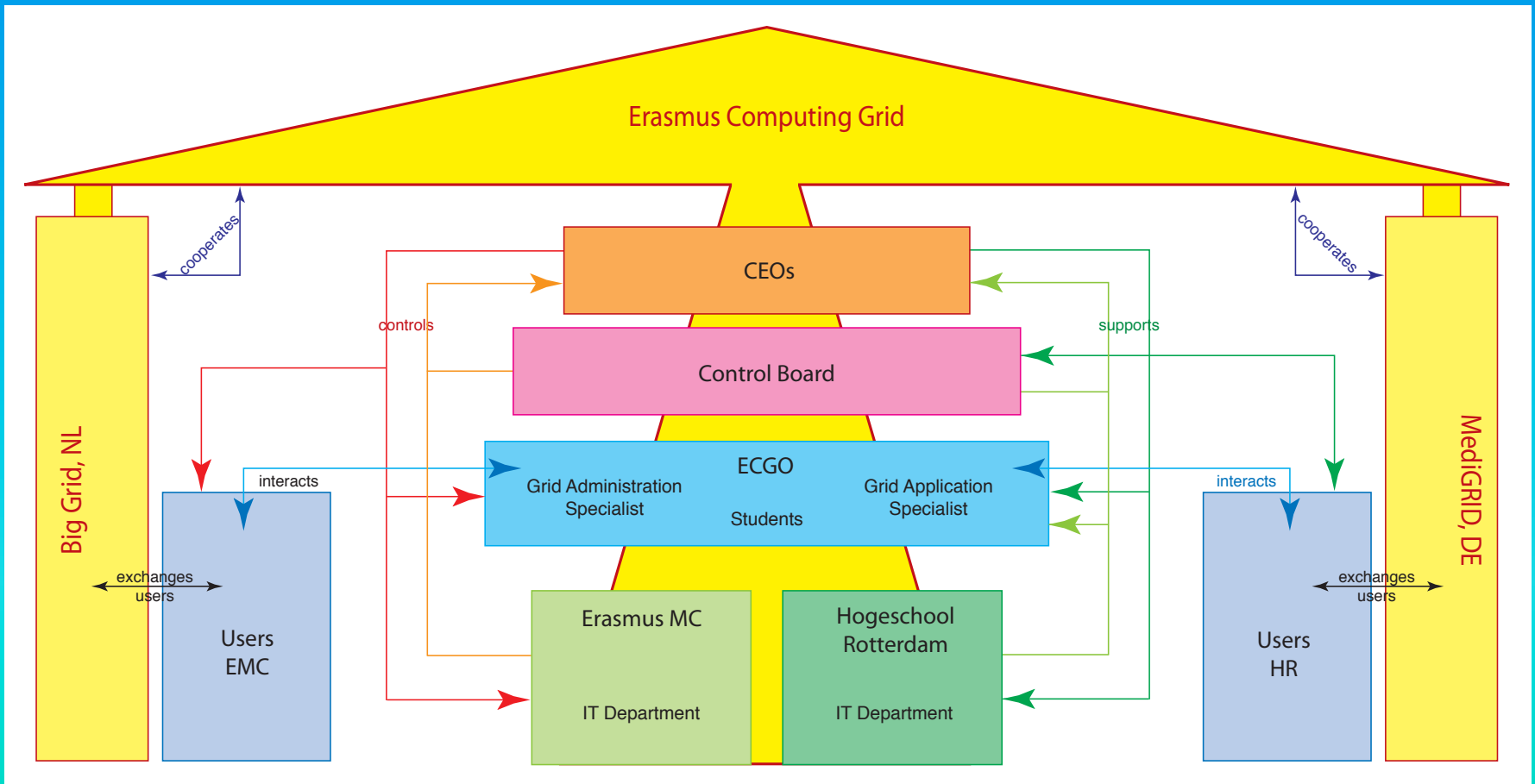


The classic human ecology triangle is transformed into the *human ecology rectangle*, to complete the circle for creation and dependencies for a curriculum canon and grammar!



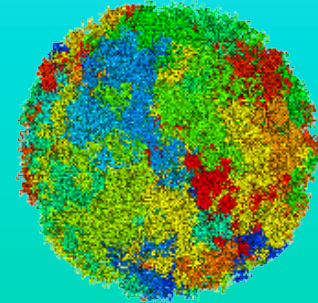
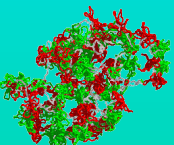
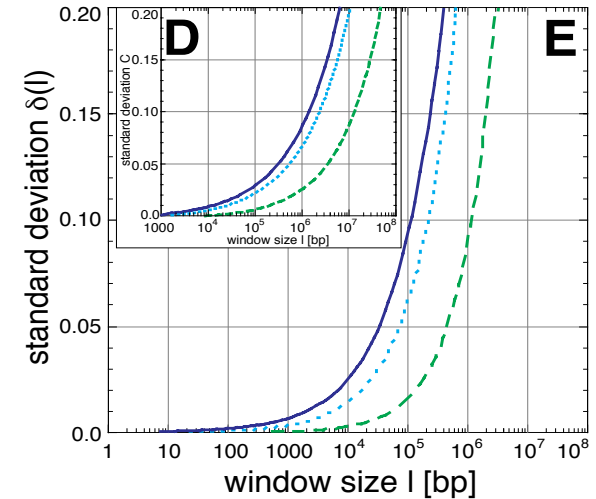
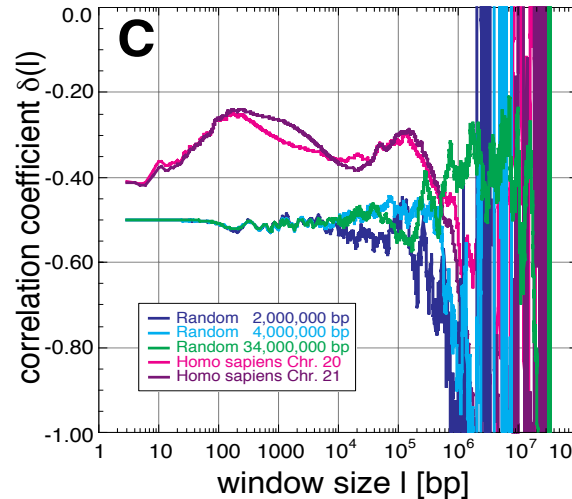
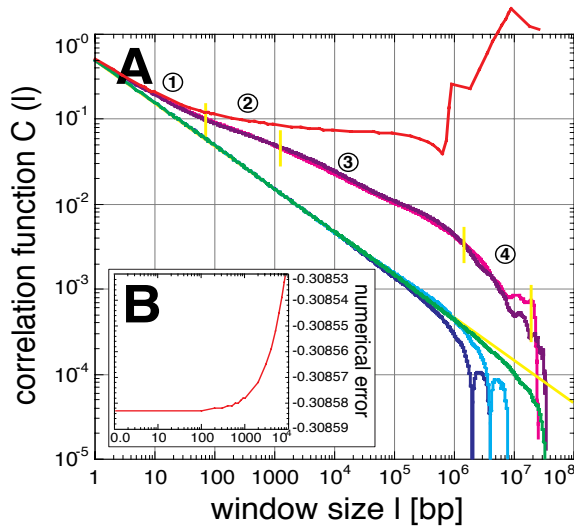
ECG - Management

To guaranty the integrity and efficiency of the ECG, a dedicated management structure has been put in place with corresponding checks and balances. The organization is made such, that also collaborations with other grids as e.g. the Dutch Big-Grid, the German nationwide MediGRID / D-Grid, or other European Grids as EGEE, EDGES can be exploited most efficiently.



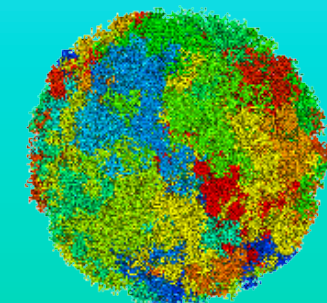
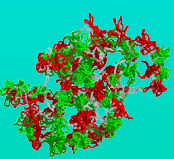
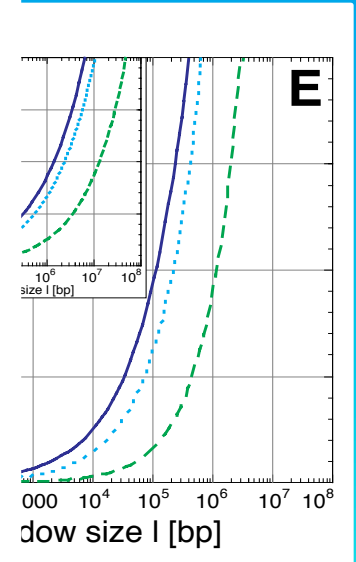
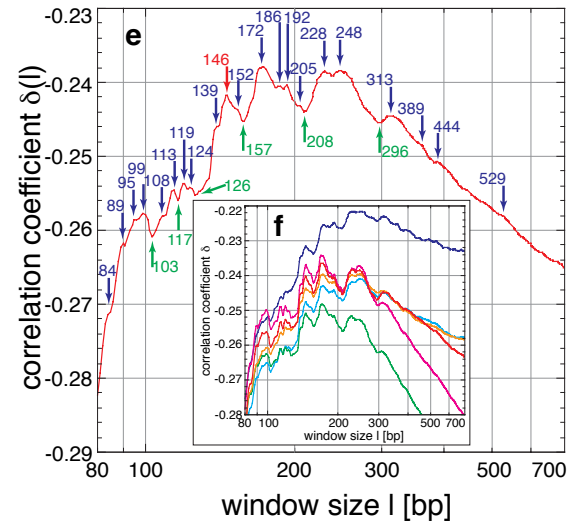
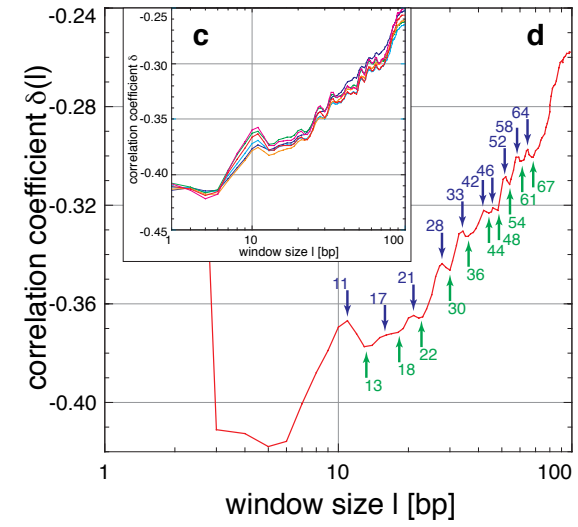
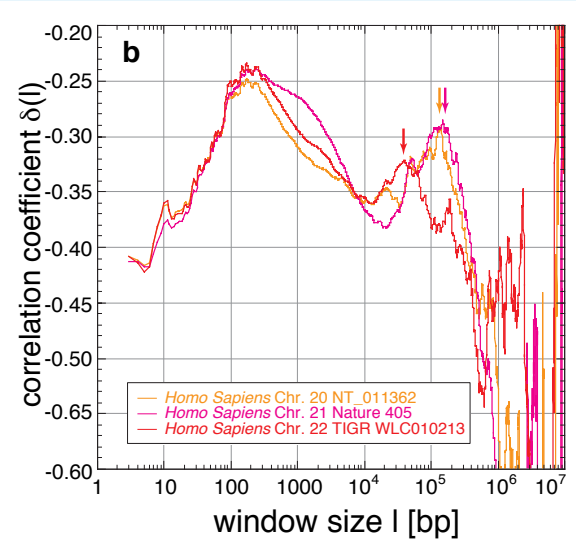
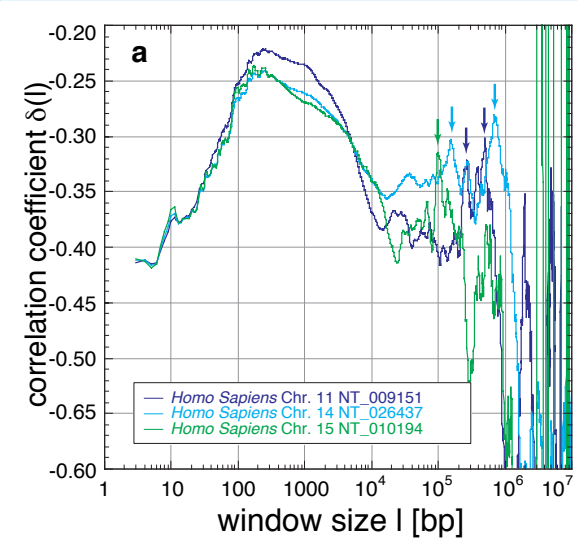
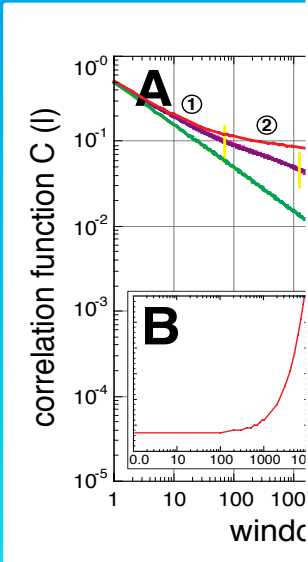
Long-Range Correlations in DNA Sequences

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ show that genomes show characteristic long-range correlations up to 10^8 bp, i.e. in principle over their complete length. Beyond, they show a specific multi-scaling behaviour, as well as a characteristic fine-structure. All correlates with the 3D-architecture of genomes.



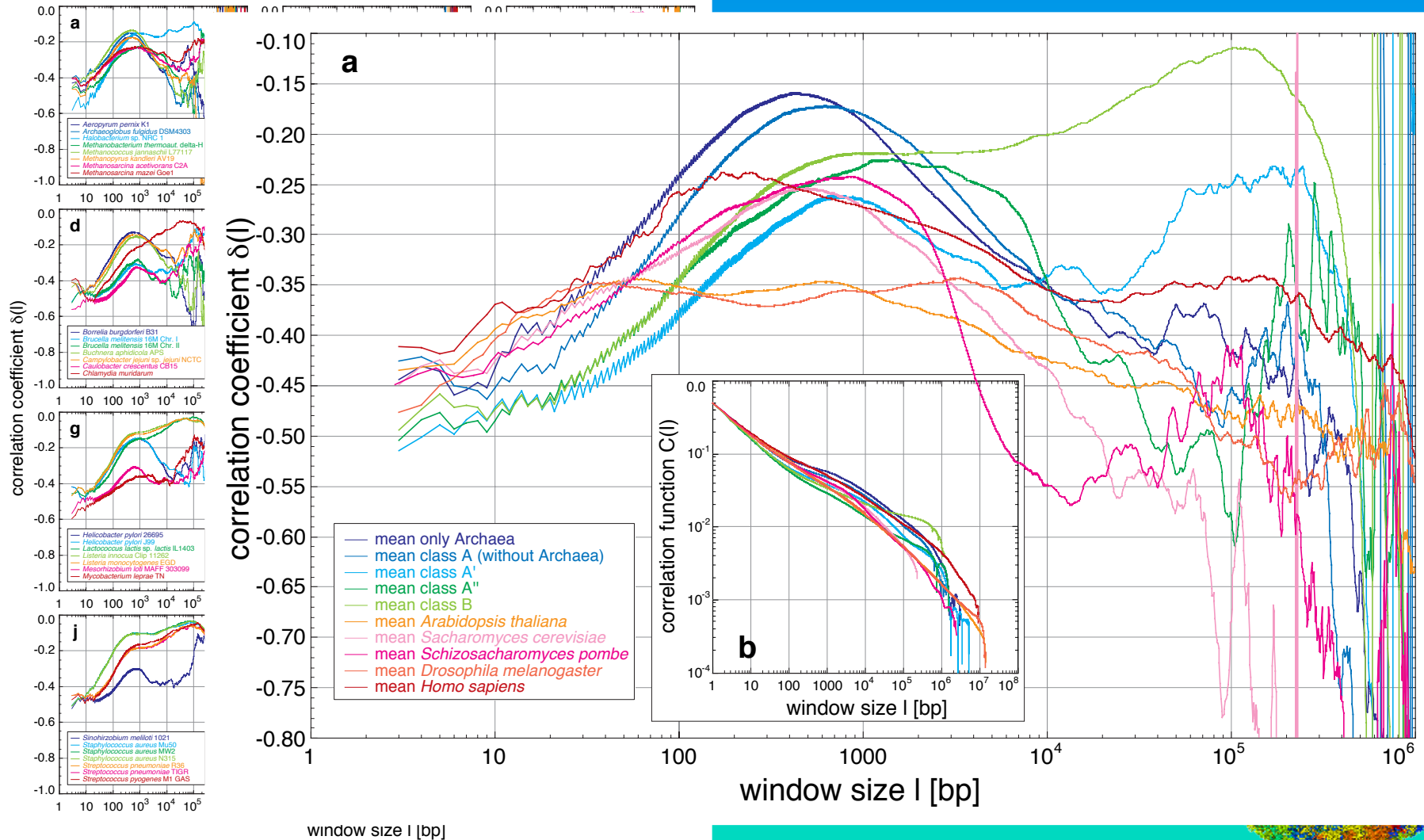
Long-Range Correlations in DNA Sequences

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ show that genomes show characteristic long-range correlations up to 10^8 bp, i.e. in principle over their complete length. Beyond, they show a specific multi-scaling behaviour, as well as a characteristic fine-structure. All correlates with the 3D-architecture of genomes.



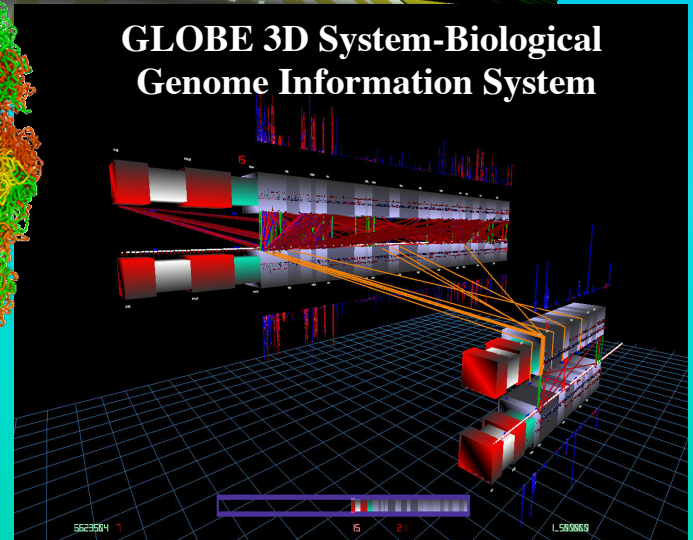
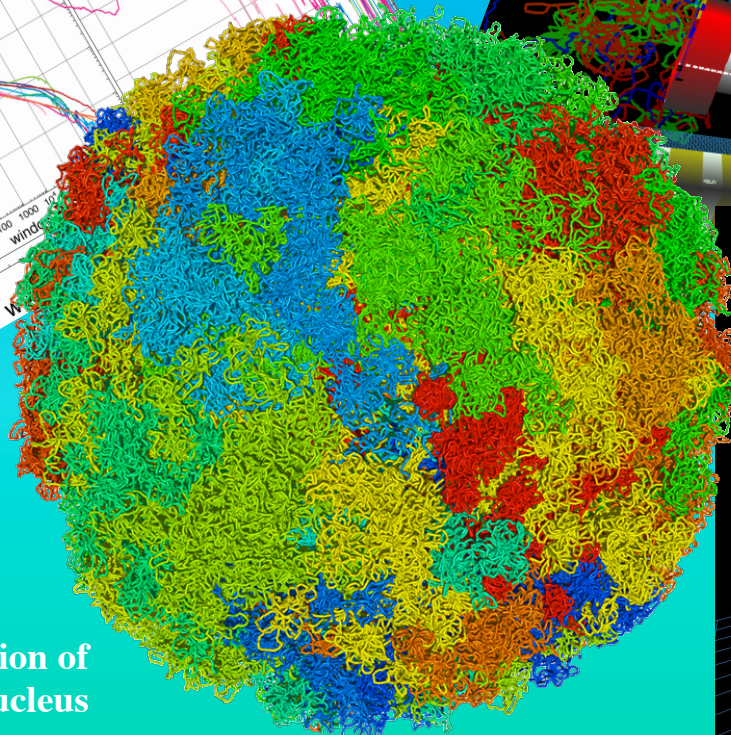
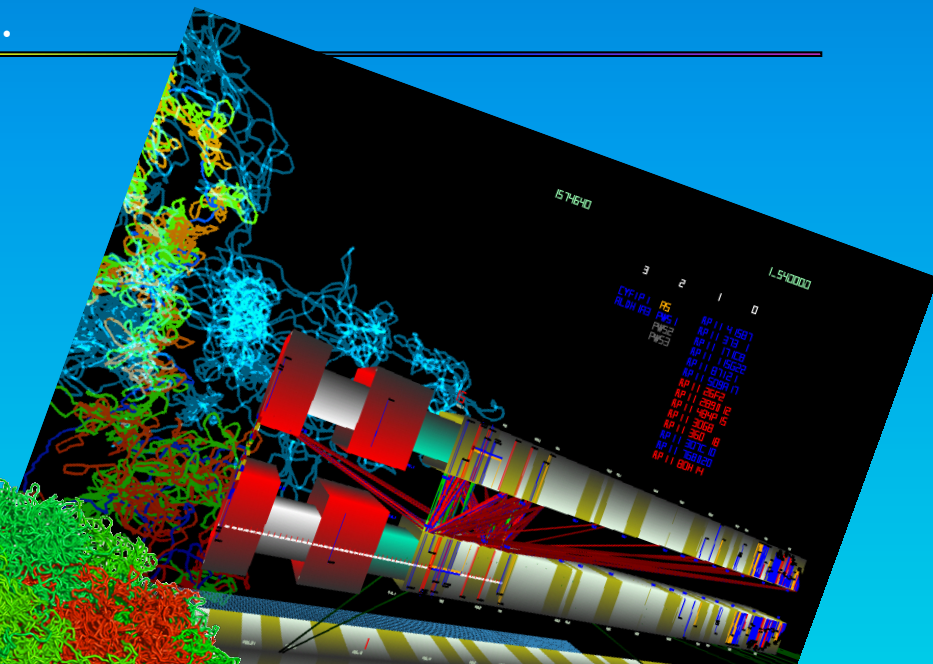
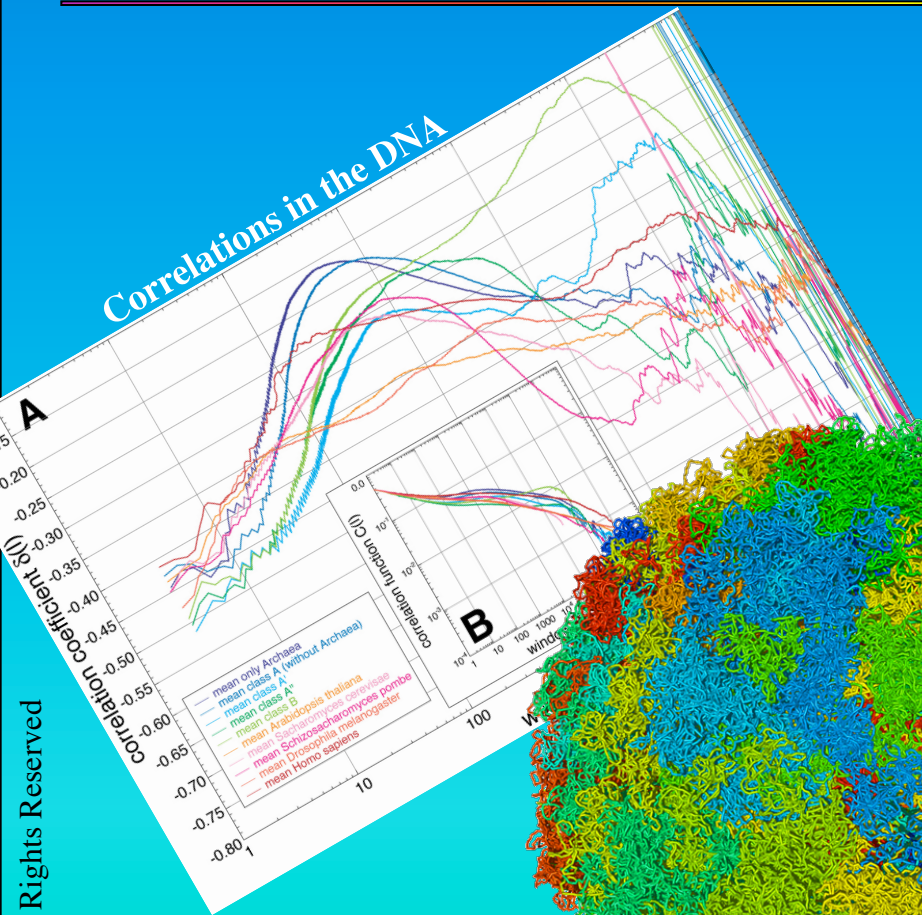
Sequence Organization and Phylogeny

The concentration fluctuation function $C(l)$ and the correlation coefficient $\delta(l)$ show a characteristic long-range fine-structured multi-scaling behaviour, which is species specific. Tree construction thus leads to characteristic groups, which are similar to those suggested by classic phylogenetic trees. Thus, the sequence organization represents evolutionary lines.



The Happy End - Profits Sharing the Commons

Both the Erasmus Computing Grid and the MediGRID/Services@MediGRID examples show that the IT challenges mankind faces in the biomedical research and care sectors can be successfully approached by exploitation of the commons by e-Human "grid" Ecology means.



© 1996-2011 All Rights Reserved

Simulation of the Organization of an entire Human Cell Nucleus

Acknowledgements



Gridders:

Anis Abuseiris (ECG R&D, systems, applications)

Michael Lesnussa (ECG visualization)

Nick Kepper (ECG parallel/MPI/CUDA)

Rob de Graaf (ECG build/administration)

Social Science:

Volkmar Baumgärtner

Julia Hampl

Kurt Egger

Steering Committee:

Frank G. Grosveld, ErasmusMC

Arthur van der Kamp, Erasmus MC

Menno Kok, Erasmus MC

Science:

Markus Göker

(phylogenies, DSMZ, Braunschweig)

Rudolf Lohner

(numerics, KIT, Karlsruhe)

Collaborators:

Ad Emmen (BOINC)

Alexander Beck-Ratzka (BOINC)

Martin Kühn (gaming infrastructures)

Institutions and/or Funders:

Erasmus Medical Center

Hogeschool Rotterdam

The Fraunhofer Society

The German MediGRID and Services@MediGRID

The German D-Grid Initiatives

The German Ministry for Science and Technology

The Dutch Science Organization (NWO)

The European EGEE Initiative

The European EDGES Consortium

The German Society for Human Ecology

The International Society for Human Ecology

The European Commission

DNA Sequence Patterns

—

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

Knoch, T. A., Abuseiris, A., Lesnussa, M., Kepper, N.,
de Graaf, R. & Grosveld, F. G.

8th International Desktop Grid Foundation (IDGF) Workshop, Max Planck Institute
for Gravitational Physics, Hannover, Germany, 17th August, 2011.

Abstract

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.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome statistical mechanics, genomic uncertainty principle, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, chromatin cross-linking, chromosome conformation capture (3C), selective high-resolution high-throughput chromosome interaction capture (T2C), confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool.

Literature References

Knoch, T. A. Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.) *Diploma Thesis*, Faculty for Physics and Astronomy, Ruperto-Carola University, Heidelberg, Germany, 1998, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5 and ISBN 978-3-00-010685-9 (soft cover, 2rd ed.), ISBN 3-00-035857-9 and ISBN 978-3-00-0358857-0 (hard cover, 2rd ed.), ISBN 3-00-035858-7, and ISBN 978-3-00-035858-6 (DVD, 2rd ed.), 1998.

Knoch, T. A., Münkkel, C. & Langowski, J. Three-dimensional organization of chromosome territories and the human cell nucleus - about the structure of a self replicating nano fabrication site. *Foresight Institute - Article Archive*, Foresight Institute, Palo Alto, CA, USA, <http://www.foresight.org>, 1- 6, 1998.

Knoch, T. A., Münkkel, C. & Langowski, J. Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, editor Wilfried Juling, Scientific Supercomputing Center (SSC) Karlsruhe, University of Karlsruhe (TH), 27- 29, 1999.

- Knoch, T. A.**, Münkkel, C. & Langowski, J. Three-dimensional organization of chromosome territories in the human interphase nucleus. *High Performance Computing in Science and Engineering 1999*, editors Krause, E. & Jäger, W., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.
- Bestvater, F., **Knoch, T. A.**, Langowski, J. & Spiess, E. GFP-Walking: Artificial construct conversions caused by simultaneous cotransfection. *BioTechniques* 32(4), 844-854, 2002.
- Knoch, T. A. (editor)**, Backes, M., Baumgärtner, V., Eysel, G., Fehrenbach, H., Göker, M., Hampl, J., Hampl, U., Hartmann, D., Hitzelberger, H., Nambena, J., Rehberg, U., Schmidt, S., Weber, A., & Weidemann, T. Humanökologische Perspektiven Wechsel - Festschrift zu Ehren des 70. Geburtstags von Prof. Dr. Kurt Egger. Human Ecology Working Group, Ruperto-Carola University of Heidelberg, Heidelberg, Germany, 2002.
- 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 3-00-009959-X and ISBN 978-3-00-009959-5 (soft cover, 3rd ed.), ISBN 3-00-009960-3 and ISBN 978-3-00-009960-1 (hard cover, 3rd ed.), ISBN 3-00-035856-9 and ISBN 978-3-00-010685-9 (DVD, 3rd ed.) 2002.
- Knoch, T. A.** Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. *High Performance Computing in Science and Engineering 2003*, editors Krause, E., Jäger, W. & Resch, M., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3- 540-40850-9, 421-440, 2003.
- Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, **Knoch, T. A.**, Waldeck, W. & Langowski, J. Analyzing intracellular binding and diffusion with continuous fluorescence photobleaching. *Biophys. J.* 84(5), 3353-3363, 2003.
- Weidemann, T., Wachsmuth, M., **Knoch, T. A.**, Müller, G., Waldeck, W. & Langowski, J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J. Mol. Biol.* 334(2), 229-240, 2003.
- Fejes Tóth, K., **Knoch, T. A.**, Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C. P., Müller, G. & Rippe, K. Trichostatin A induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Science* 117, 4277-4287, 2004.
- Ermler, S., Kronic, D., **Knoch, T. A.**, Moshir, S., Mai, S., Greulich-Bode, K. M. & Boukamp, P. Cell cycle-dependent 3D distribution of telomeres and telomere repeat-binding factor 2 (TRF2) in HaCaT and HaCaT-myc cells. *Europ. J. Cell Biol.* 83(11-12), 681-690, 2004.
- Kost, C., Gama de Oliveira, E., **Knoch, T. A.** & Wirth, R. Spatio-temporal permanence and plasticity of foraging trails in young and mature leaf-cutting ant colonies (*Atta spp.*). *J. Trop. Ecol.* 21(6), 677- 688, 2005.
- Winnefeld, M., Grewenig, A., Schnölzer, M., Spring, H., **Knoch, T. A.**, Gan, E. C., Rommelaere, J. & Cziepluch, C. Human SGT interacts with BAG-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* 312, 2500-2514, 2006.
- Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Yassene, M., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. Grid-basierte Services für die elektronische Patientenakte der Zukunft. *E- HEALTH-COM - Magazin für Gesundheitstelematik und Telemedizin*, 4(2), 61-63, 2007.
- de Zeeuw, L. V., **Knoch, T. A.**, van den Berg, J. & Grosveld, F. G. Erasmus Computing Grid - Het bouwen van een 20 TeraFLOP virtuele supercomputer. *NIOC proceedings 2007 - het perspective of lange termijn.* editor Frederik, H. NIOC, Amsterdam, The Netherlands, 52-59, 2007.
- Rauch, J., **Knoch, T. A.**, Solovei, I., Teller, K. Stein, S., Buiting, K., Horsthemke, B., Langowski, J., Cremer, T., Hausmann, M. & Cremer, C. Lightoptical precision measurements of the Prader- Willi/Angelman

Syndrome imprinting locus in human cell nuclei indicate maximum condensation changes in the few hundred nanometer range. *Differentiation* 76(1), 66-82, 2008.

Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Mohammed, Y., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. C. Auf dem Weg zur individualisierten Medizin - Grid-basierte Services für die EPA der Zukunft. *Telemedizinführer Deutschland 2008*, editor Jäckel, A. Deutsches Medizinforum, Minerva KG, Darmstadt, ISBN 3-937948-06-6, ISBN-13 9783937948065, 47-51, 2008.

Drägestein, K. A., van Capellen, W. A., van Haren, J. Tsibidis, G. D., Akhmanova, A., **Knoch, T. A.**, Grosveld, F. G. & Galjart, N. Dynamic behavior of GFP-CLIP-170 reveals fast protein turnover on microtubule plus ends. *J. Cell Biol.* 180(4), 729-737, 2008.

Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J. M., Grosveld, F. G., **Knoch, T. A.**⁺ & Murre, C.⁺ The 3D-structure of the Immunoglobulin Heavy Chain Locus: implications for long-range genomic interactions. *Cell* 133(2), 265-279, 2008.

Krefting, D., Bart, J., Beronov, K., Dzhimova, O., Falkner, J., Hartung, M., Hoheisel, A., **Knoch, T. A.**, Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolxdorff, T., Vossberg, M., Viezens, F. & Weisbecker, A. MediGRID - Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems* 25(3), 326-336, 2008.

Knoch, T. A., Lesnussa, M., Kepper, F. N., Eussen, H. B., & Grosveld, F. G. The GLOBE 3D Genome Platform - Towards a novel system-biological paper tool to integrate the huge complexity of genome organization and function. *Stud. Health. Technol. Inform.* 147, 105-116, 2009.

Knoch, T. A., Baumgärtner, V., de Zeeuw, L. V., Grosveld, F. G., & Egger, K. e-Human Grid Ecology: Understanding and approaching the Inverse Tragedy of the Commons in the e-Grid Society. *Stud. Health. Technol. Inform.* 147, 269-276, 2009.

Dickmann, F., Kaspar, M., Löhnhardt, B., **Knoch, T. A.**, & Sax, U. Perspectives of MediGRID. *Stud. Health. Technol. Inform.* 147, 173-182, 2009.

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.

Dickmann, F., Kaspar, M., Löhnhardt, B., Kepper, N., Viezens, F., Hertel, F., Lesnussa, M., Mohammed, Y., Thiel, A., Steinke, T., Bernarding, J., Krefting, D., **Knoch, T. A.** & Sax, U. Visualization in health-grid environments: a novel service and business approach. *LNCS 5745*, 150-159, 2009.

Dickmann, F., Kaspar, M., Löhnhardt, B., Kepper, N., Viezens, F., Hertel, F., Lesnussa, M., Mohammed, Y., Thiel, A., Steinke, T., Bernarding, J., Krefting, D., **Knoch, T. A.** & Sax, U. Visualization in health-grid environments: a novel service and business approach. *Grid economics and business models - GECON 2009 Proceedings, 6th international workshop, Delft, The Netherlands*. editors Altmann, J., Buyya, R. & Rana, O. F., GECON 2009, LNCS 5745, Springer-Verlag Berlin Heidelberg, ISBN 978-3-642-03863-1, 150-159, 2009.

Estrada, K.* , Abuseiris, A.* , Grosveld, F. G., Uitterlinden, A. G., **Knoch, T. A.**⁺ & Rivadeneira, F.⁺ GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* 25(20), 2750-2752, 2009.

Kepper, N., Schmitt, E., Lesnussa, M., Weiland, Y., Eussen, H. B., Grosveld, F. G., Hausmann, M. & **Knoch T. A.**, Visualization, Analysis, and Design of COMBO-FISH Probes in the Grid-Based GLOBE 3D Genome Platform. *Stud. Health Technol. Inform.* 159, 171-180, 2010.

Kepper, N., Ettig, R., Dickmann, F., Stehr, R., Grosveld, F. G., Wedemann, G. & **Knoch, T. A.** Parallel high-performance grid computing: capabilities and opportunities of a novel demanding service and business class allowing highest resource efficiency. *Stud. Health Technol. Inform.* 159, 264-271, 2010.

Skrowny, D., Dickmann, F., Löhnhardt, B., **Knoch, T. A.** & Sax, U. Development of an information platform for new grid users in the biomedical field. *Stud. Health Technol. Inform.* 159, 277-282, 2010.

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.

- Dickmann, F., Brodhun, M., Falkner, J., **Knoch, T. A.** & Sax, U. Technology transfer of dynamic IT outsourcing requires security measures in SLAs. *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, 1-115, 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, 21-48, 2011.
- Kolovos, P., **Knoch, T. A.**, F. G. Grosveld, P. R. Cook, & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics and Chromatin* 5(1), 1-8, 2012.
- Dickmann, F., Falkner, J., Gunia, W., Hampe, J., Hausmann, M., Herrmann, A., Kepper, N., **Knoch, T. A.**, Lauterbach, S., Lippert, J., Peter, K., Schmitt, E., Schwardmann, U., Solodenko, J., Sommerfeld, D., Steinke, T., Weisbecker, A. & Sax, U. Solutions for Biomedical Grid Computing - Case Studies from the D-Grid Project Services@MediGRID. *JOCS* 3(5), 280-297, 2012.
- Estrada, K. *, Abuseiris, A. *, Grosveld, F. G., Uitterlinden, A. G., **Knoch, T. A.**⁺ & Rivadeneira, F.⁺ GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Dissection of the complex genetic architecture of human stature and osteoporosis*. cumulative dissertation, editor Estrada K., Erasmus Medical Center, Erasmus University Rotterdam, Rotterdam, The Netherlands, ISBN 978-94-6169-246-7, 25-30, 1st June 2012.
- van de Corput, M. P. C., de Boer, E., **Knoch, T. A.**, van Cappellen, W. A., Quintanilla, A., Ferrand, L., & Grosveld, F. G. Super-resolution imaging reveals 3D folding dynamics of the b-globin locus upon gene activation. *J. Cell Sci.* 125, 4630-4639, 2012.