

# Pattern Avoidance and Genome Sorting

Edited by

Michael Albert<sup>1</sup>, Miklós Bóna<sup>2</sup>, István Miklós<sup>3</sup>, and  
Einar Steingrímsson<sup>4</sup>

1 University of Otago, NZ, [malbert@cs.otago.ac.nz](mailto:malbert@cs.otago.ac.nz)

2 University of Florida – Gainesville, US, [bona@ufl.edu](mailto:bona@ufl.edu)

3 Alfréd Rényi Institute of Mathematics – Budapest, HU, [Miklós@ramet.elte.hu](mailto:Miklós@ramet.elte.hu)

4 University of Strathclyde, GB, [ainer@alum.mit.edu](mailto:ainer@alum.mit.edu)

---

## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16071 “Pattern Avoidance and Genome Sorting”.

**Seminar** February 14–19, 2016 – <http://www.dagstuhl.de/16071>

**1998 ACM Subject Classification** D.3.1 Formal Definitions and Theory, Semantics, E.1 Data Structures, F.2 Analysis of Algorithms and Problem Complexity

**Keywords and phrases** evolutionary distance, lists, metrics, patterns, permutations, sorting

**Digital Object Identifier** 10.4230/DagRep.6.2.65

**Edited in cooperation with** Miklós Bóna

## 1 Executive Summary

*Miklós Bóna*

**License**  Creative Commons BY 3.0 Unported license  
© Miklós Bóna

The seminar took place from February 14, 2016, to February 19, 2016. It had 36 participants, who were researchers in theoretical computer science, combinatorics, and molecular biology. It was a geographically diverse group, with participants coming from the US, Canada, Brazil, Germany, Iceland, the United Kingdom, Sweden, France, Slovakia, Hungary and New Zealand. The seminar featured 18 talks, three of which were hourlong talks, and an open problem session.

Numerous collaborative research efforts have been started. Here is a sampling.

Megan Martinez and Manda Riehl worked on a bijection between LP matchings (one of the RNA matchings described in Vincent Vatter’s talk) and Klazar’s nesting equivalent matchings. They studied a paper by Klazar and Aziza Jefferson’s dissertation and made progress on the bijection.

István Miklós, Péter Erdős and Miklós Bóna worked on proving a log-convexity conjecture related to ordered degree sequences of bipartite graphs.

Brona Brejova and Manda Riehl discussed two potential future projects related to gene and species tree reconciliation. The most probable starting point is a project involving gene and species trees where a gene is allowed to duplicate a string inside itself. This situation was not allowed in previous models, however it seems that as long as the specific breakpoints are not reused from this insertion, a modification of the previous algorithms could still be effective.



Except where otherwise noted, content of this report is licensed  
under a Creative Commons BY 3.0 Unported license

Pattern Avoidance and Genome Sorting, *Dagstuhl Reports*, Vol. 6, Issue 2, pp. 65–77

Editors: Michael Albert, Miklós Bóna, István Miklós, and Einar Steingrímsson



DAGSTUHL  
REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Jay Pantone, David Bevan and Miklós Bóna collaborated on asymptotic enumeration of a balanced urns and balls model that was seen to be a step towards finding a better upper bound for a pattern avoidance enumeration problem.

We have all the reasons to believe that this, and many other joint research efforts that started during this seminar will lead to new results that would not have been possible without the seminar. Therefore, we strongly believe that the seminar was a success that we would like to repeat at some point in the future.

## 2 Table of Contents

### Executive Summary

<i>Miklós Bóna</i> . . . . .	65
------------------------------	----

### Overview of Talks

Decomposition trees of permutations, and how to use them for a (realistic?) study of perfect sorting by reversals <i>Mathilde Bouvel</i> . . . . .	69
Isometric Gene Tree Reconciliation <i>Brona Brejova</i> . . . . .	69
The SCJ small parsimony problem for weighted gene adjacencies <i>Cedric Chauve</i> . . . . .	70
Method of moments estimates for reversal and block transposition distances using symmetric group models <i>Niklas Eriksen</i> . . . . .	70
Gene orders, median of permutations and related combinatorial problems <i>Sylvie Hamel</i> . . . . .	71
Sorting with Forbidden Intermediates <i>Anthony Labarre</i> . . . . .	71
Efficient algorithms for permutation pattern matching <i>Marie-Louise Lackner</i> . . . . .	72
Social choice and permutation patterns <i>Martin Lackner</i> . . . . .	72
Counting and sampling genome rearrangement scenarios: a meeting-point of combinatorics and computer science <i>István Miklós</i> . . . . .	72
The method of differential approximants <i>Jay Pantone</i> . . . . .	73
Parametric Analysis of an SCFG-based model for RNA structure prediction <i>Svetlana Poznanovikj</i> . . . . .	73
Longest increasing subsequences and log concavity <i>Bruce Sagan, Miklós Bóna, and Marie-Louise Lackner</i> . . . . .	74
Consequences of the no-coincidence assumption in comparative gene order <i>David Sankoff</i> . . . . .	74
An Update on Gene Family-free Genome Comparison <i>Jens Stoye</i> . . . . .	75
Positional Constraints for Rearrangements through Noncrossing Colored Partitions and Cycle Packings <i>Krister Swenson</i> . . . . .	75
Breaking bad <i>Eric Tannier</i> . . . . .	75

The substitution decomposition of RNA secondary structures	
<i>Vincent Vatter</i> . . . . .	76
On sequence segmentation problems	
<i>Tomáš Vinař</i> . . . . .	76
<b>Participants</b> . . . . .	<b>77</b>

### 3 Overview of Talks

#### 3.1 Decomposition trees of permutations, and how to use them for a (realistic?) study of perfect sorting by reversals

*Mathilde Bouvel (Universität Zürich, CH)*

**License** © Creative Commons BY 3.0 Unported license  
© Mathilde Bouvel

**Joint work of** Mathilde Bouvel, Cedric Chauve, Marni Mishna, Cyril Nicaud, Carine Pivoteau, and Dominique Rossin

**Main reference** M. Bouvel, M. Mishna, C. Nicaud, “Some simple varieties of trees arising in permutation analysis”, in Proc. of the 25th Int’l Conf. on Formal Power Series and Algebraic Combinatorics (FPSAC 2013), DMTCS proceedings, 2013.

**URL** <http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/issue/view/130>

**URL** [http://user.math.uzh.ch/bouvel/presentations/Poster2013\\_BMN.pdf](http://user.math.uzh.ch/bouvel/presentations/Poster2013_BMN.pdf)

I will present the genome rearrangement problem of perfect sorting by reversals, and show its algorithmic solution by Bérard-Bergeron-Chauve-Paul. It uses the strong interval trees of permutations (whose definition will be recalled in the talk). Those trees are also known as (substitution) decomposition trees of permutations. I will present two results and a research project related to those trees.

First, I will show that the algorithm of Bérard-Bergeron-Chauve-Paul is polynomial on average (and with probability 1 as the size of the input goes to infinity). I will also describe average properties of commuting scenarios (a.k.a. separable permutations) for perfect sorting by reversals. These results are obtained using the tools of analytic combinatorics developed by Flajolet-Sedgewick.

Second, I will generalize these average properties (to some extent) to super-classes of the separable permutations. This demonstrates a phenomenon of convergence of a sequence of families of trees to the class of all permutations, whose analytic nature is essentially different. We only have a partial explanation of this phenomenon.

Finally, I will go back to the biological motivation, showing that all these models of trees do not represent well the strong interval trees obtained from the biological data. I will conclude by proposing a possibly better model, whose study is yet to be done.

#### 3.2 Isometric Gene Tree Reconciliation

*Brona Brejova (Comenius University in Bratislava, SK)*

**License** © Creative Commons BY 3.0 Unported license  
© Brona Brejova

**Joint work of** Brona Brejova, Michal Sabo

The infinite sites model, introduced by Jian Ma et al (PNAS 2008), formalizes the problem of recovering the evolutionary history of a set of related genomes allowing a large set of evolutionary operations including insertions, deletions, duplications, and rearrangements. One of the steps of their polynomial-time algorithm reconciles a gene tree with a species tree under the assumption that both trees have exact branch lengths known. This assumption simplifies the problem compared to the typical reconciliation scenario without branch lengths. We show several mistakes in the original algorithm and provide a corrected and simplified version. We also discuss related open problems.

### 3.3 The SCJ small parsimony problem for weighted gene adjacencies

*Cedric Chauve (Simon Fraser University – Burnaby, CA)*

**License** © Creative Commons BY 3.0 Unported license  
© Cedric Chauve

**Joint work of** Ashok Rajaraman, Yann Ponty, Cedric Chauve

**Main reference** A. Rajaraman, C. Chauve, Y. Ponty, “Assessing the Robustness of Parsimonious Predictions for Gene Neighborhoods from Reconciled Phylogenies: Supplementary Material”, *Bioinformatics Research and Applications, LNCS, Vol. 9096*, pp. 260–271, Springer, 2015.

**URL** [http://dx.doi.org/10.1007/978-3-319-19048-8\\_22](http://dx.doi.org/10.1007/978-3-319-19048-8_22)

Reconstructing ancestral gene orders in a given phylogeny is a classical problem in comparative genomics. Most existing methods compare conserved features in extant genomes in the phylogeny to define potential ancestral gene adjacencies, and either try to reconstruct all ancestral genomes under a global evolutionary parsimony criterion, or, focusing on a single ancestral genome, use a scaffolding approach to select a subset of ancestral gene adjacencies, generally aiming at reducing the fragmentation of the reconstructed ancestral genome. We describe an exact algorithm for the small parsimony problem that combines both approaches. We consider that gene adjacencies at internal nodes of the species phylogeny are weighted, and we introduce an objective function defined as a convex combination of these weights and the evolutionary cost under the Single-Cut-or-Join (SCJ) model. The weights of ancestral gene adjacencies can e.g. be obtained through the recent availability of ancient DNA sequencing data, which provide a direct hint at the genome structure of the considered ancestor, or through probabilistic analysis of gene adjacencies evolution. The algorithm we propose is Fixed-Parameter Tractable (FPT) based on the dynamic programming algorithm by (Sankoff and Rousseau, 1975) and allows to sample co-optimal solutions.

### 3.4 Method of moments estimates for reversal and block transposition distances using symmetric group models

*Niklas Eriksen (University of Örebro, SE)*

**License** © Creative Commons BY 3.0 Unported license  
© Niklas Eriksen

**Joint work of** Niklas Eriksen, Axel Hultman

**Main reference** N. Eriksen, A. Hultman, “Expected reflection distance in  $G(r,1,n)$  after a fixed number of reflections”, *Annals of Combinatorics*, 9(1):21–33, Springer, 2005.

**URL** <http://dx.doi.org/10.1007/s00026-005-0238-y>

The gene order of species change over time, and can hence be used to infer relationships between species. By estimating the number of reversals or block transpositions that separate two species, we get an idea of the time since speciation. Eriksen and Hultman introduced a permutation model with similar properties to the reversal model but much more tractable. In this model, method of moments estimates could be computed, corresponding to the reversal distance in a pure reversal model. We extend their results to obtain several estimates of the number of reversal and block transpositions that separate two species. We also give a method for estimating the proportion on reversals, which is a very important parameter.

### 3.5 Gene orders, median of permutations and related combinatorial problems

*Sylvie Hamel (University of Montréal, CA)*

**License** © Creative Commons BY 3.0 Unported license  
© Sylvie Hamel

**Joint work of** Sylvie Hamel and Robin Milosz

**Main reference** R. Milosz, S. Hamel, “Medians of Permutations: building constraints”, Algorithms and Discrete Applied Mathematics, LNCS, Vol. 9602, pp. 264–276, Springer, 2016.

**URL** [http://dx.doi.org/10.1007/978-3-319-29221-2\\_23](http://dx.doi.org/10.1007/978-3-319-29221-2_23)

Our aim is to discuss the general problem of finding a consensus ranking, given a set of different rankings of a set of objects. Here, one assumes that a set of different rankings are proposed for a given set of strictly ordered elements, and one is looking for a ranking that is in closest agreement to all of these. Thus, the problem becomes that of finding the median of a set of permutations under a distance  $d$ . In part this is motivated by the classical gene order problem of comparative genomics, where the difference in the order of appearance of genes in the genome of different species is used to evaluate the evolutionary distances between them.

We give close attention to this median problem for the “Kendall-tau” distance, which corresponds to counting the number of order disagreements between pairs of elements of two permutations. The problem has been shown to be NP-complete for sets of  $m$  permutations, where  $m > 3$ , and the complexity is still unknown for  $m = 3$ . From an algorithmic point of view, we present a deterministic heuristic for this median problem, and derive some theoretical properties of the starting set of permutations that drastically reduce the search space for the medians of this set. In a more combinatoric point of view, we consider the interesting automedian cases (when a set of permutations is equal to the set of its medians), deriving some of its properties under group actions, shuffle operation, etc. Finally, we generalize this problem to the problem of “aggregating ranking with ties”, with an application to the bioinformatic context of calculating such medians for biological data related to certain diseases.

### 3.6 Sorting with Forbidden Intermediates

*Anthony Labarre (University Paris-Est – Marne-la-Vallée, FR)*

**License** © Creative Commons BY 3.0 Unported license  
© Anthony Labarre

Most genome rearrangement problems on permutations can be recast as constrained sorting problems, where the goal is to compute of a shortest sorting sequence of operations for a given permutation under the restriction that the set of allowed operations is fixed beforehand. However, biologists have known for more than a century that some of these mutations at a given point in time can be lethal to a given organism. We revisit those problems by adding a new constraint on the sequences to be computed: they must avoid a given set of forbidden intermediates, which correspond to species that cannot exist because the mutations that would be involved in their creation are lethal. We initiate this study by focusing on the case where the only mutations that can occur are algebraic transpositions, and give a polynomial time algorithm for solving that problem when the permutation to sort is an involution.

### 3.7 Efficient algorithms for permutation pattern matching

*Marie-Louise Lackner (TU Wien, AT)*

**License**  Creative Commons BY 3.0 Unported license  
© Marie-Louise Lackner

**Joint work of** Michael Albert, Marie-Louise Lackner, Martin Lackner and Vincent Vatter

**Main reference** M. H. Albert, M.-L. Lackner, M. Lackner, V. Vatter, “The complexity of pattern matching for 321-avoiding and skew-merged permutations”, arXiv:1510.06051 [math.CO], 2015.

**URL** <http://arxiv.org/abs/1510.06051v1>

Given two permutations  $\tau$  and  $\pi$  where  $\pi$ , the pattern, is shorter than  $\tau$ , the Permutation Pattern Matching problem (PPM) asks whether  $\pi$  is contained in  $\tau$ . In general, this problem is known to be NP-complete, implying that we may not hope for efficient algorithms to solve PPM. Two directions have been pursued so far in order to circumvent this hardness result: First, one can look for special cases in which PPM can be solved efficiently, i.e., in polynomial time. Second, one can try to find a parameter that explains the computational hardness of this problem and confine the combinatorial explosion to this parameter. In this talk, I will give an overview of the algorithms known so far and present one algorithm following the first approach and one taking the second one in more detail.

### 3.8 Social choice and permutation patterns

*Martin Lackner (University of Oxford, GB)*

**License**  Creative Commons BY 3.0 Unported license  
© Martin Lackner

**Joint work of** Martin Lackner, Marie-Louise Lackner

**Main reference** M. Lackner and M.-L. Lackner, “On the Likelihood of Single-Peaked Preferences”, arXiv:1505.05852 [cs.GT], 2015.

**URL** <http://arxiv.org/abs/1505.05852v1>

In this talk I will discuss several combinatorial and algorithmic problems from Social Choice and how to they relate to questions about permutation patterns. In particular, some questions about structure in preferences can be answered by translating them to questions about permutation patterns. Also, I will present open problems about permutation patterns that arise from questions in Social Choice.

### 3.9 Counting and sampling genome rearrangement scenarios: a meeting-point of combinatorics and computer science

*István Miklós (Alfréd Rényi Institute of Mathematics – Budapest, HU)*

**License**  Creative Commons BY 3.0 Unported license  
© István Miklós

**Joint work of** István Miklós, Heather Smith

**Main reference** I. Miklós and H. Smith, “Sampling and counting genome rearrangement scenarios”, BMC Bioinformatics, 16(Suppl 14):S6, Springer, 2015.

**URL** <http://dx.doi.org/10.1186/1471-2105-16-S14-S6>

Even for moderate size inputs, there are a tremendous number of optimal rearrangement scenarios, regardless what the model is and which specific question is to be answered. Therefore giving one optimal solution might be misleading and cannot be used for statistical



inferring. Statistically well funded methods are necessary to sample uniformly from the solution space and then a small number of samples are sufficient for statistical inferring.

In this talk, we are going to give an overview of the state-of-the-art of sampling and counting rearrangement scenarios. The talk will focus on how combinatorial methods can be used in computational statistics.

### 3.10 The method of differential approximants

*Jay Pantone (Dartmouth College – Hanover, US)*

License © Creative Commons BY 3.0 Unported license  
© Jay Pantone

For decades, the method of differential approximants has been applied to the study of statistical mechanics to estimate the singularity structure of the generating function of a sequence of positive integers, using only a finite number of initial terms of the generating function. While all such approximations are of course only non-rigorous estimates, experience shows these estimates to be remarkably accurate.

Differential approximants can be extremely useful to combinatorialists. We provide several examples of combinatorial sequences for which no generating function is known or conjectured yet the method of differential approximants provides a very accurate approximation of the asymptotic behavior of the sequence. We then describe several extensions to the method that are in progress.

### 3.11 Parametric Analysis of an SCFG-based model for RNA structure prediction

*Svetlana Poznanovikj (Clemson University, US)*

License © Creative Commons BY 3.0 Unported license  
© Svetlana Poznanovikj

The function of the RNA molecule is often dependent on its structure and so understanding how the RNA nucleotide chain folds onto itself is an important problem. Language-based methods for RNA structure prediction use stochastic context-free grammars (SCFGs). The SCFG developed by Knudsen and Hein is relatively simple and yet has been shown to achieve good accuracy compared to other grammars. We performed an analysis of the probability distribution induced by this grammar and in this talk I'll present some interesting properties that we found.

### 3.12 Longest increasing subsequences and log concavity

Bruce Sagan (*Michigan State University – East Lansing, US*), Miklós Bóna (*University of Florida – Gainesville, US*), and Marie-Louise Lackner (*TU Wien, AT*)

**License** © Creative Commons BY 3.0 Unported license

© Bruce Sagan, Miklós Bóna, and Marie-Louise Lackner

**Joint work of** Miklós Bóna, Marie-Louise Lackner, Bruce Sagan

**Main reference** M. Bóna, M.-L. Lackner, B. Sagan, “Longest increasing subsequences and log concavity”, arXiv:1511.08653 [math.CO], 2015.

**URL** <http://arxiv.org/abs/1511.08653>

Let  $\mathfrak{S}_n$  be the set of all permutations of  $1, 2, \dots, n$  viewed as sequences. Let  $l_{n,k}$  be the number of  $\pi \in \mathfrak{S}_n$  having a longest increasing subsequence of length  $k$ . This length is closely related to the Ulam distance between permutations which is used to model evolutionary distance in DNA research. William Chen conjectured that the sequence  $l_{1,n}, l_{2,n}, \dots, l_{n,n}$  is log concave which means that

$$l_{k-1,n}l_{k+1,n} \leq l_{n,k}^2$$

for all  $k$ . We also conjecture that if  $i_{n,k}$  is the number of involutions in  $\mathfrak{S}_n$  with longest increasing subsequence length  $k$  then  $i_{1,n}, i_{2,n}, \dots, i_{n,n}$  is log concave. We show that these two conjectures are strongly related. We also present evidence to support the truth of both. Our main tool is the Robinson-Schensted correspondence. Many other associated conjectures will be discussed.

### 3.13 Consequences of the no-coincidence assumption in comparative gene order

David Sankoff (*University of Ottawa, CA*)

**License** © Creative Commons BY 3.0 Unported license

© David Sankoff

The number of genes in plant and animal genomes tends to exceed 25,000, so that the coincidental gene orders must be very rare unless the genomes have inherited common orders. We discuss two consequences of this. One is the number of chromosomes inferred in an ancestral genome construction, especially in the context of ancient whole genome duplication, where inferring an incorrect number of chromosomes may require unlikely coincidences. The other is the inference of the locus of genome rearrangement on a phylogenetic tree. In this case only certain branches may be inferred to contain a breakpoint, otherwise coincidental changes must have happened. Both of these principles are key to recent reconstructions of ancestral gene orders.

### 3.14 An Update on Gene Family-free Genome Comparison

*Jens Stoye (Universität Bielefeld, DE)*

**License** © Creative Commons BY 3.0 Unported license  
© Jens Stoye

**Joint work of** Daniel Doerr, Louis Antonio Kowada, Jens Stoye, Simone Dantas

**Main reference** L. A. B. Kowada, D. Doerr, S. Dantas, J. Stoye, “New Genome Similarity Measures based on Conserved Gene Adjacencies”, in Proc. of the 20th Annual Conf. on Research in Computational Molecular Biology (RECOMB’16), LNCS, Vol. 9649, pp. 204–224, Springer, 2016; pre-print available from author’s webpage.

**URL** [http://dx.doi.org/10.1007/978-3-319-31957-5\\_15](http://dx.doi.org/10.1007/978-3-319-31957-5_15)

**URL** <http://www.techfak.uni-bielefeld.de/~stoye/mypublications.html#KOW-DOE-DAN-STO-2016>

Many methods in computational comparative genomics require gene family assignments as a prerequisite. While the biological concept of gene families is well established, their computational prediction remains unreliable. In this talk I will present recent results in a new line of research, in which family assignments are not presumed. We study several family-free approaches in detecting conserved structures, genome rearrangements and in reconstructing ancestral gene orders. This leads to a number of interesting combinatorial optimization problems, some of which are easily polynomial-time solvable, while others turn out to be surprisingly hard.

### 3.15 Positional Constraints for Rearrangements through Noncrossing Colored Partitions and Cycle Packings

*Krister Swenson (University of Montpellier 2, FR)*

**License** © Creative Commons BY 3.0 Unported license  
© Krister Swenson

**Joint work of** Krister Swenson, Mathieu Blanchette, Pijus Simonaitis

**Main reference** K. M. Swenson, M. Blanchette, “Models and algorithms for genome rearrangement with positional constraints”, in Proc. of the 15th Int’l Workshop Algorithms in Bioinformatics (WABI’15), LNCS, Vol. 9289, pp. 243–256, Springer, 2015.

**URL** [http://dx.doi.org/10.1007/978-3-662-48221-6\\_18](http://dx.doi.org/10.1007/978-3-662-48221-6_18)

The number of moves is often the sole criterion used to measure the quality of a rearrangement scenario. A current challenge is to incorporate biological information into the gene-order evolutionary model in a manner that is computationally tractable. We present a model amenable towards positional constraints (as seen with chromatin conformation capture data), while elucidating connections between genome rearrangements and noncrossing colored partitions and cycle packings.

### 3.16 Breaking bad

*Eric Tannier (University Claude Bernard – Lyon, FR)*

**License** © Creative Commons BY 3.0 Unported license  
© Eric Tannier

The permutation as a model for gene order is flawed. It ignores the diversity of susceptibility to breakage across genomic regions, which is necessary even under a uniform random breakage model. I will propose a model of evolution of gene order by inversions where breakage probabilities vary across intergene regions and over time. It contains as a particular

case the uniform breakage model on the nucleotidic sequence, where breakage probabilities are proportional to intergene region lengths. This is very different from the frequently used pseudo-uniform model where all intergene regions have the same probability to break. Estimations of rearrangement distances based on the pseudo-uniform model completely fail on simulations with the truly uniform model. I will propose new combinatorial and statistical problems with this model.

### 3.17 The substitution decomposition of RNA secondary structures

*Vincent Vatter (University of Florida – Gainesville, US)*

**License** © Creative Commons BY 3.0 Unported license  
© Vincent Vatter

**Joint work of** Vincent Vatter and Aziza Jefferson

**Main reference** A. Jefferson, “The substitution decomposition of matchings and RNA secondary structures”, Doctoral dissertation, University of Florida, 2015.

**URL** <http://gradworks.umi.com/37/29/3729175.html>

The substitution decomposition has proved to be a powerful tool for analyzing classes of permutations and of graphs (where it is known as the modular decomposition). I will discuss applications of the substitution decomposition to RNA secondary structures, which can be modeled by matchings.

### 3.18 On sequence segmentation problems

*Tomáš Vinař (Comenius University in Bratislava, SK)*

**License** © Creative Commons BY 3.0 Unported license  
© Tomáš Vinař

**Joint work of** Martina Višňovská, Tomáš Vinař, Broňa Brejová

**Main reference** M. Višňovská, T. Vinař, and B. Brejová, “DNA Sequence Segmentation Based on Local Similarity”, in *Information Technologies – Applications and Theory (ITAT)*, CEUR Workshop Proceedings, Vol. 1003, pp. 36–43, 2013.

**URL** <http://ceur-ws.org/Vol-1003/>

Most rearrangement models work with short segments (called markers, genes, atoms, etc.) that are considered atomic, i.e. they are long enough so that they are identifiable in a genomic sequence and at the same time they are not broken by rearrangement operations. It turns out that precomputing such segments is a difficult and interesting problem, and it is a major obstacle in applying algorithms for rearrangement analyses to real data. In some cases, the problem is in fact equivalent to reconstructing evolutionary history in a given model. In the talk I would discuss several approaches to solving this problem and also include some notes on an approach to analysis of segmental duplications that we have previously introduced.

## Participants

- Michael Albert  
University of Otago, NZ
- David Bevan  
The Open University – Milton Keynes, GB
- Miklós Bóna  
University of Florida – Gainesville, US
- Mathilde Bouvel  
Universität Zürich, CH
- Marilia Braga  
Immetro – Duque de Caxias, BR
- Brona Brejova  
Comenius University in Bratislava, SK
- Robert Brignall  
The Open University – Milton Keynes, GB
- Cedric Chauve  
Simon Fraser University – Burnaby, CA
- Anders Claesson  
University of Strathclyde, GB
- Péter L. Erdős  
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Niklas Eriksen  
University of Örebro, SE
- Pedro Feijão  
Universität Bielefeld, DE
- Guillaume Fertin  
University of Nantes, FR
- Sylvie Hamel  
University of Montréal, CA
- Vít Jelínek  
Charles University – Prague, CZ
- Anthony Labarre  
University Paris-Est – Marne-la-Vallée, FR
- Marie-Louise Lackner  
TU Wien, AT
- Martin Lackner  
University of Oxford, GB
- Megan Martinez  
Ithaca College, US
- István Miklós  
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Jay Pantone  
Dartmouth College – Hanover, US
- Adeline Pierrot  
University of Paris South XI, FR
- Yann Ponty  
Ecole Polytechnique – Palaiseau, FR
- Svetlana Poznanovikj  
Clemson University, US
- Manda Riehl  
University of Wisconsin – Eau Claire, US
- Bruce Sagan  
Michigan State University – East Lansing, US
- David Sankoff  
University of Ottawa, CA
- Rebecca Smith  
The College at Brockport, US
- Einar Steingrímsson  
University of Strathclyde, GB
- Jens Stoye  
Universität Bielefeld, DE
- Krister Swenson  
University of Montpellier 2, FR
- Eric Tannier  
University Claude Bernard – Lyon, FR
- Vincent Vatter  
University of Florida – Gainesville, US
- Stéphane Vialette  
University Paris-Est – Marne-la-Vallée, FR
- Tomáš Vinař  
Comenius University in Bratislava, SK

