Program

- Monday, February 15

  - 9:00-9:10 Opening Remarks
  - 9:15 - 10:15 István Miklós
  - 10:30 – 11:00 Vincent Vatter
  - 12:15 Lunch
  - 3:00 – 3:30 Eric Tannier
  - 3:45 – 4:15 Anthony Labarre
  - 4:30 – 5:00 David Sankoff

- Tuesday, February 16

  - 9:00-10:00 Sylvie Hamel
  - 10:15 – 10:45 Tomáš Vinař
  - 12:15 Lunch
  - 3:00 – 3:30 Broňa Brejová
  - 3:45 – 4:15 Jens Stoye
  - 4:30 – 5:30 Open problem session

- Wednesday, February 17

  - 9:00 – 9:30 Niklas Ericksen
  - 9:45 – 10:15 Cedric Chauve
  - 12:15 Lunch
  - Afternoon hike

- Thursday, February 18

  - 9:00-10:00 Mathilde Bouvel
  - 10:15 – 10:45 Svetlana Poznanovikj
  - 12:15 Lunch
  - 3:00 – 3:30 Krister Swenson
  - 3:45 – 4:15 Marie-Louise Lackner
  - 4:30 – 5:00 Jay Pantone

- Friday, February 19

  - 9:00 – 9:30 Martin Lackner
  - 9:45 – 10:15 Bruce Sagan
  - 12:15 Lunch

Abstracts

Monday

István Miklós

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

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.

Vincent Vatter

The substitution decomposition of RNA secondary structures

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.

This is joint work with Aziza Jefferson.

Eric Tannier

Breaking bad

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 intergne 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.

Anthony Labarre
Sorting with Forbidden Intermediates

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.

David Sankoff
Consequences of the no-coincidence assumption in comparative gene order

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.

Tuesday

Sylvie Hamel
Gene orders, median of permutations and related combinatorial problems

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.

Jens Stoye
An Update on Gene Family-free Genome Comparison

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.

This is joint work with Daniel Doerr and many others.

Tomáš Vinař
On A Sequence Segmentation Problem

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.

Broňa Brejová
Isometric Gene Tree Reconciliation

Isometric Gene Tree Reconciliation 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.

Wednesday

Niklas Eriksen
Method of moments estimates for reversal and block transposition distances using symmetric group models

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.

Cedric Chauve
The SCJ small parsimony problem for weighted gene adjacencies

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.

Thursday

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

I will present the genome rearrangement problem of perfect sorting by reversals, and show its algorithmic solution by Brard-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 Brard-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.

This talk is based on joint works and ongoing projects with C. Chauve, M. Mishna, C. Nicaud, C. Pivoteau and D. Rossin.

Svetlana Poznanovikj
Parametric Analysis of an SCFG-based model for RNA structure prediction

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.

Krister Swenson
Positional Constraints for Rearrangements through Noncrossing Colored Partitions and Cycle Packings

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.

Marie-Louise Lackner
Efficient Algorithms for Permutation Pattern Matching

Given two permutations τ and π where π, the pattern, is shorter than τ, the Permutation Pattern Matching problem (PPM) asks whether π is contained in τ. 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.

This talk is based on joint work with Michael Albert, Martin Lackner and Vince Vatter.

Jay Pantone
The Method of Differential Approximants

For decades, the method of differential approximants has been applied to the study of statistical mechanics to estimate the singularity structure

7

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.

Friday

Martin Lackner
Social Choice and Permutation Patterns

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.

Bruce Sagan
Longest increasing subsequences and log concavity

Let $\mathcal{S}_n$ be the set of all permutations of $1, 2, \ldots, n$ viewed as sequences. Let $l_{n,k}$ be the number of $\pi \in \mathcal{S}_n$ 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 $l_{1,n}, l_{2,n}, \ldots, l_{n,n}$ 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 $\mathcal{S}_n$ with longest increasing subsequence length k then $i_{1,n}, i_{2,n}, \ldots, i_{n,n}$ 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. This is joint work with Miklós Bóna and Marie-Louise Lackner.