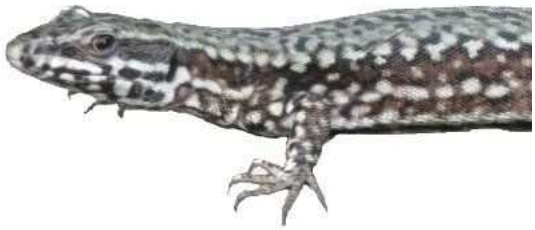


FOBIA CONFERENCE

Zlín, November 6th 2009

Book of Abstracts

Univerzita Tomáše Bati ve Zlíně

budova U5, 6p. místnost A612
Fakulta aplikované informatiky
Univerzita Tomáše Bati ve Zlíně
Nad Stráněmi 4511
Zlín 76001

49° 13' 50.282" N 17° 39' 27.863" E

http://fobia.img.cas.cz
http://web.vscht.cz/spiwokv/fobia/

10:00 *Opening*

10:00 **V. Spiwok**: Evolution of Protein Dynamics

10:15 **L. Kouřil**: Možnosti predikce aminokyselinových sekvencí v proteinech

10:30 **V. Bystrý**: Segmentace proteinových sekvencí pomocí HMM

10:45 *Coffee break*

11:00 **M. Lexa**: Stochastický model tvorby sekundarných struktúr v DNA

11:15 **B. Brejová**: Searching for Genes in Novel Genomes

11:30 **T. Vinař**: Evolučné histórie zhlukov génov

11:45 *Lunch*

12:45 *Excursion*

13:30 **J. Pačes**: New Sequencing Technologies: Current Computational Challenges

13:45 **M. Kolář**: Transcription Profiling of Head and Neck Squamous Cell Carcinoma

14:00 *Coffee break*

14:15 **J. Vohradský**: Inference of Active Genetic Networks from Microarray and ChIP-on-Chip Experiments by Evolutionary Modeling

14:30 **J. Pánek**: Wrong Structures Find Right RNAs: Homology Search for Bacterial Non-coding RNAs in Evolutionary Distant Species Using Structures with Higher than Minimum Free Energy

14:45 ELIXIR Workshop

Evolution of protein dynamics

V. Spiwok
Institute of Chemical Technology, Prague

The relationships between amino acid sequence of a protein, its structure and its dynamics still remains unclear. A complicated protein dynamics (for example a simulation trajectory or an experimental set of structures) can be "dissected" to few collective atomic motions for example by principal component analysis. Evolution of proteins, on the other hand, can be examined by comparing their amino acid sequences. Alternatively, instead of pair-wise comparison of sequences, it is possible to analyse certain protein family by testing how the presence of certain amino acid residue in one position influences the presence of another residue in a second position. This approach is known as statistical coupling analysis (SCA). Here we present an analysis of a model protein in terms of i. coupled molecular motions, and ii. in terms of co-evolving amino acid residues and comparison of observed networks of coupled residues.

Možnosti predikce aminokyselinových sekvencí v proteinech

L. Kouřil, I. Zelinka
Univerzita Tomáše Bati ve Zlíně

Proteiny, tvořící stavební prvky života a živých organismů, jsou z pohledu své primární struktury tvořeny aminokyselinovými řetězci. Pokud vezmeme v úvahu, že veškeré proteiny, kterých je v lidských buňkách až 84 000, jsou tvořeny pouze z 20 aminokyselin, přičemž každý protein je sekvenčně odlišný, musejí se v jednotlivých proteinech nacházet oblasti, které jsou shodné u více proteinů. Variabilitu proteinových sekvencí potom tedy zajišťují pouze minoritní alternace aminokyselin. Jako důsledek této úvahy vyvstává, že je teoreticky možné provádět predikci aminokyselinových řetězců tvořících ucelené oblasti v sekvencích primárních struktur proteinů.

Cílem této práce je ověřit možnosti predikce aminokyselinových sekvencí v proteinech za použití umělé inteligence ve formě neuronových sítí a statistických dat získaných analýzou proteinů z databáze RCSD Protein Data Bank.

Segmentace proteinovych sekvenci pomoci HMM

V. Bystry, M. Lexa
Masaryk University, Brno

Prediction of cruciform structure formation in topologically constrained DNA by a probabilistic model

M. Lexa, M. Brázdová
Masaryk University, Brno,
Institute of Biophysics, CAS, Brno

Sequence-dependent secondary DNA structures, such as cruciform or triplex DNA, are implicated in regulation of gene transcription and other important biological processes at the molecular level. Sequences capable of forming these structures can readily be identified in entire genomes by appropriate searching techniques. However, not every DNA segment containing the proper sequence has equal probability of forming an alternative structure. Calculating the free energy of the potential structures provides an estimate of their stability *in vivo*, but there are other structural factors, both local and non-local, not taken into account by such simplistic approach. We present the procedure we currently use to identify potential cruciform structures in DNA sequences. The procedure relies on identification of palindromes (or inverted repeats) and evaluation of their sequences by a nucleic acid folding program (UNAFold). We further extended the procedure by adding a modelling step to filter the predicted cruciforms. The model takes into account superhelical density of the analyzed segments of DNA and calculates the probability of cruciforms forming at several locations of the analyzed DNA, based on the sequences in the stem and loop areas of the structures and competition among them.

This research has been supported by grant No. 204/08/1560 from the Czech Grant Agency.

Searching for Genes in Novel Genomes

B. Brejová
Univerzita Komenského, Bratislava

New rapid sequencing methods now allow affordable sequencing of previously unexplored genomes. The gene prediction in these novel genomes is difficult due to the lack of reliable training data necessary for adjusting parameters of models used for this task.

We have developed a novel method for estimating the parameters of hidden Markov models for gene finding in newly sequenced species. Our approach does not rely on curated training data sets, but instead uses extrinsic evidence (including paired-end ditags that have not been used in gene finding previously) and iterative training. This new method is particularly suitable for annotation of species with large evolutionary distance to the closest annotated species. We have used our approach to produce an initial annotation of the newly sequenced *Schistosoma japonicum* draft genome. Our new gene set provides a first glimpse at a gene complement of a flatworm (phylum platyhelmintes).

This work was published as: B. Brejová, T. Vinař, Y. Chen, S. Wang, G. Zhao, D.G. Brown, M. Li, Y. Zhou: Finding genes in *Schistosoma japonicum*: annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Research*, **37**(7):e52. April 2009.

Evolučné histórie zhlukov génov

T. Vinař, B. Brejová, A. Siepel
Univerzita Komenského v Bratislave

Zhruba 5% ľudského genómu je pokrytých zhlukmi génov, ktoré vznikli opakovanými segmentálnymi duplikáciami. V týchto regiónoch často nájdeme gény s rýchlo sa vyvíjajúcimi funkciami, či gény dôležité z medicínskeho hľadiska. Zhluky génov je najlepšie analyzovať v kontexte ich duplikačných histórií, keďže tieto nám umožňujú zostaviť akurátne génové stromy potrebné pre ďalšiu komparatívnu analýzu. V prednáške predstavím nové metódy na rekonštrukciu takýchto duplikačných histórií.

New Sequencing Technologies: Current Computational Challenges

J. Pačes
IMG, ASCR, Prague

Transcription profiling of head and neck squamous cell carcinoma

M. Kolář, J. Šáchová, L. Lacina, J. Pačes, M. Urbanová, V. Pačes, J. Betka, Č. Vlček, K. Smetana, J. Plzák, M. Chovanec, Z. Čada, H. Strnad
IMG, ASCR, Prague, Charles University in Prague, Faculty of Medicine and Faculty Hospital Motol, Prague, Czech Republic

Head and neck squamous cell carcinoma (HNSCC) is one of the most prevalent tumour types which is characterised by a high mortality. The poor prognosis results from a lack of markers of early stages of the disease. In search of these markers we profiled transcriptoms of tumoural and normal tissues of a cohort of 37 patients with conventional HNSCC. The data has been collected over last two years and here we present our data analysis pipeline and preliminary results.

The presented data consisting of patient-matched samples reveal several marker genes distinguishing tumour and normal tissues. The results are interpreted in the broad scale of metabolic, regulatory and signalling pathways. Further, we describe differences between three classes of peritumoural tissues potentially allowing to assess tumour resection accuracy.

Inference of active genetic networks from microarray and ChIP-on-chip experiments by evolutionary modeling

C.C. To, J. Vohradský
Institute of Microbiology, ASCR, Prague

Genetic networks and their dynamic properties are inferred from the analysis temporal microarray experiments of yeast cell cycle using differential equation model of gene expression. Genome-wide location data

for a small genetic network (ChIP-on-chip experiments) are used as a constrained in reconstruction of gene regulatory interactions. Evolutionary computing approach is used to identify the network structure in an unbiased way. Results show several principles of regulation which can be considered as common for other genetic networks. Analysis presented here shows that with currently available experiments and due to experimental error we are not able to predict one genetic network participating in given cellular process, but we have to work with a hypothesis of multiple equivalent networks which. It is also shown that ChIP-on-chip experiments are, not sufficient to predict functional networks which are active during an investigated process. Such predictions have to be considered as potential and their actual realization during particular cellular process has to be identified by incorporating dynamic data and consequently additional *in vivo* experiments.

Wrong structures find right RNAs: Homology search for bacterial non-coding RNAs in evolutionary distant species using structures with higher than minimum free energy

J. Pánek, J. Vohradský.
Institute of Microbiology, ASCR, Prague

Non coding RNAs (ncRNAs) are small RNAs with potential regulatory role which are coded in intergenic or intragenic region of genome. Identification of their sequence and location on the chromosome in prokaryotes remain a challenge for bioinformatics. The reason is a week conservation of sequences, synteny and genome locus. ncRNA secondary structure is relatively better conserved, but so far it did not allow for efficient ncRNA identification by structural homology search. Difficulty of the search by homology search increases dramatically with increasing evolutionary divergence.

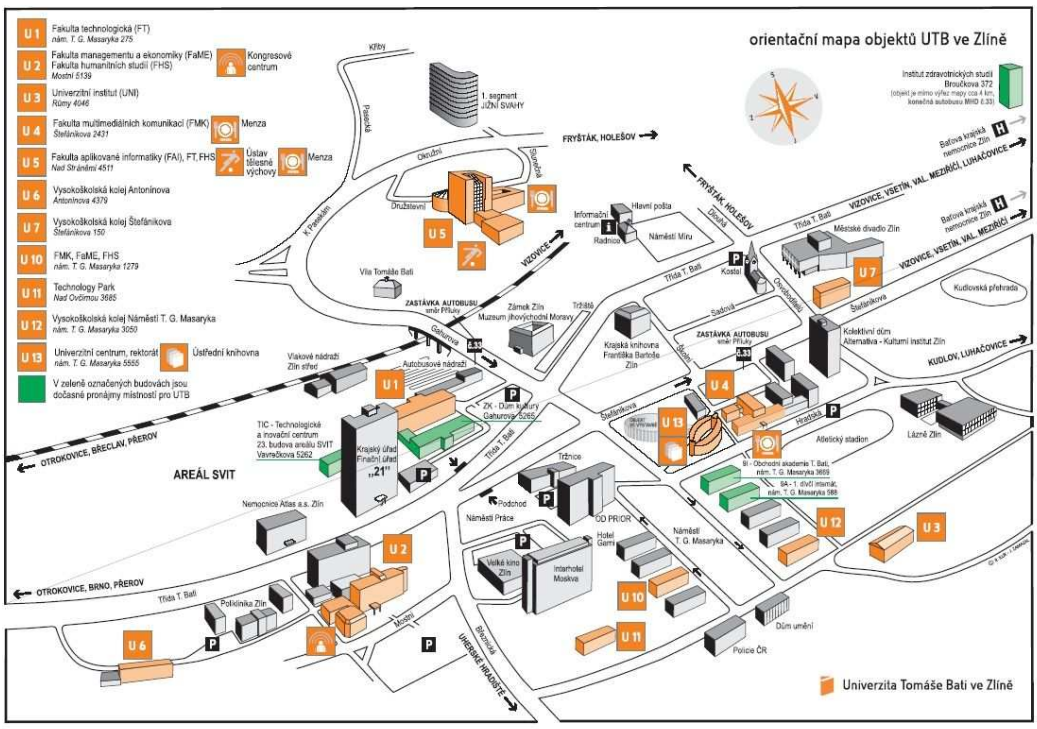
Presented study focusing on the sequence and structural homology of ncRNAs was done on 172 known 6S RNAs from X bacterial species. It shows that biologically relevant ncRNAs in different bacterial species do not have week sequence homology and the minimum free energy structures are not homologous among species. It shows that the structures with higher than minimum free energy can be homologous despite of the level of sequence similarity.

Strategy for bacterial ncRNAs search, suggested here, is based on non-minimum free energy structures and it avoids the use of sequence similarity at all, overcoming the observed week sequence conservation. The method was applied to identification of 6S RNA in *Streptomyces coelicolor* which was consequently experimentally verified. Search in related species of *Streptomyces*, *Bacillus* and *Haemofilus* identified novel 6S RNA candidates. Results show, that the sequence and structural similarity, as we understand and use it nowadays for ncRNA identification, is not biologically relevant for evolutionary distant species and its use becomes a limitation for their bioinformatic identification.

Workshop

ELIXIR – European infrastructure for the management and integration of information in the live science

ELIXIR is a pan-European initiative for creation of bioinformatics infrastructure in Europe. Current state of the structure and form of integration of bioinformatics resources suggested by the consortium will be presented. Involvement of the Czech Republic and the local structure should be a topic of discussion.


orientační mapa objektů UTB ve Zlíně
U 1 Fakulta technologická (FT) nám. T. G. Masaryka 275
U 2 Fakulta managementu a ekonomiky (FaME) Fakulta humanitních studií (FHS) Mostní 5139
U 3 Univerzitní institut (UNI) Rámy 4046
U 4 Fakulta multimediálních komunikací (FMK) Štefánikova 2431
U 5 Fakulta aplikované informatiky (FAI), FT, FHS Nad Stráněmi 4511
U 6 Vysokoškolská kolej Antonínova Antonínova 4379
U 7 Vysokoškolská kolej Štefánikova Štefánikova 150
U 10 FMK, FaME, FHS nám. T. G. Masaryka 1279
U 11 Technology Park Nad Ovčírnou 3685
U 12 Vysokoškolská kolej Náměstí T. G. Masaryka nám. T. G. Masaryka 3050
U 13 Univerzitní centrum, rektorát nám. T. G. Masaryka 5555
Kongresové centrum
Menza
Ústav tělesné výchovy
Ústřední knihovna
V zeleně označených budovách jsou dočasně pronájmy místností pro UTB
AREÁL SVIT
OTROKOVICE, BŘECLAV, PŘEROV
OTROKOVICE, BRNO, PŘEROV
FRYŠTÁK, HOLEŠOV
VIZOVICE, VSETÍN, VAL. MEZIŘÍČÍ, LUHAČOVICE
KUDLOV, LUHAČOVICE
Univerzita Tomáše Bati ve Zlíně