as South Corea, Japan, Canada, Denmark, and Sweden were obtained from GenBank. All sequences were aligned using MEGA 6 and BioEdit version 7.0.9.0 and were used for phylogenetic analyses. Neighbour-joining (NJ) tree was created in MEGA 6 and PAUP 4.0b10 with 1000 bootstrap replicates. Maximum likelihood (ML) analysis was performed in RAXMLGUI 1.0 using GTRGAMMA model with 1000 bootstrap replicates. BEAST v1.8.0 was used for Bayesian analysis.

Our phylogenetic analyses indicated that the Japanese, South Corean and American *C. sylvestris* were different from European and Turkish members. Turkish members of *C. sylvestris* were closely related to European ones according to our Bayesian, NJ and ML analyses. In Turkish members, *C. sylvestris* collected from Hazar and Çıldır Lake was more ancient than those from Marmara, Sapanca, Çıldır, Aygır, Beyşehir, Eğirdir and Sıhke Lakes.

In conclusion, our results clearly suggest that several transoceanic dispersal events among the continents may have occurred and that the entrance of Turkish *C. sylvestris* to Turkey may have been from southeast and northeast of the country.

P-03.03.2-023
Metagenomic analysis of microbial community in fresh water lake

S. Selim Abdel Salam
Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Aljouf University, Sakaka, Saudi Arabia

Metagenomics is providing great help to explore world of unculturable microorganisms in the natural samples to enhance our knowledge about microbial diversity. Here, we have performed metagenomic analysis of fresh water lake bacterial community using 454 pyrosequencing techniques. We have observed a wide array of bacteria from phylum proteobacteria and family Enterobacteriaceae as well as very few viruses from Podoviridae, Siphoviridae and unclassified phages. We have conducted a metagenomics analysis with the primary focus on the examination of the community of bacteria in a fresh water lake ecosystem. Roche GS FLX Software gave us total 156 253 reads (with an average read length of 795.274 bp). There were 15 226 contigs having > 100 bp sequence length whereas 10 481 contigs with > 500 bp sequence length. For further analysis we have taken contigs with > 500 bp only. Further, we have analyzed the microbial community composition using BLASTN/BLASTX against NT/NR databases with *E*-value cutoff of 10^{-5}. ≥70% of total contigs were mapped to the reference with ≥60% contig match coverage. The community analysis revealed that domain bacteria is predominantly present (99.8%) in the water sample, followed by Eukaryota (0.02%), viruses (0.08%) and other sequences (0.02%). Most abundant phyla was proteobacteria (99.8%) and the most dominant family was Enterobacteriaceae (89%) followed by Xanthomonadaceae (10%), Vibrionaceae (0.4%), Pasteurellaceae (0.1%), Shewanellaceae (0.07%). We performed functional analysis of all 5974 contigs using Rapid Annotation using Subsystems Technology (RAST)[4] which detected 15 319 coding sequences and 197 RNAs in 619 subsystems. Among the classified CDS from RAST showed major CDS hits for enzymes involved in the subsystems amino acids and derivatives and the carbohydrate metabolism. The great diversity of microorganisms present in the Lake may reflect the human activity in the area.

P-03.03.2-024
Characterization of unknown proteins: Additional information available from MALDI-TOF mass spectrometry data

O. Markelov[1], A. Kayumov[2], M. Bogachev[3]
[1]*Saint-Petersburg Electrotechnical University, Saint-Petersburg, Russia, Kazan (Volga Region) Federal University, Kazan, Russia,* [3]*Saint-Petersburg Electrotechincal University, Saint-Petersburg, Russia*

MALDI-TOF mass spectrometry is a ubiquitous and widespread tool for protein identification. Once the protein sequence is unavailable, unambiguous identification cannot be performed, and predictability is limited by the presence of sequenced homologous proteins. We present a statistical approach to predict a number of structural, localization and functional properties of unknown proteins by direct analysis of mass distribution shapes of their post-cleavage fragments obtained from MALDI-TOF mass spectrometry data. Secondary structure of proteins is best predicted by their specific cleavage at the inertial hydropathy group amino acid residues (FILMV), with Thermolysin (AFILMV) being the closest commercially available reagent, leading to distinguishing between proteins with presumably α-helixes or β-sheets with 90% accuracy. Cellular localization of proteins is best predicted by their specific cleavage at the external hydropathy group amino acid residues (DEHKNQR), exemplified by GluC(phosphate)+LysC(DEK) cleavage. Protein location in the cell membrane and its localization character (monotopic/transmembrane, single-pass/multi-pass transmembrane) are predictable with ~ 75% accuracy by this single cleavage, with optimal combination of 3–4 cleavages improving the accuracy to ~ 80%. Functional prediction of proteins is the best among membrane-associated proteins with characteristic structural conformations. We attribute the differences in the mass distribution shapes to the characteristic clustering of amino acids residues with respective hydropathy properties that are involved in the formation of 3d structural conformations of proteins. The suggested approach allows for a non-parametric statistical prediction of uncharacterized proteins from their MALDI-TOF mass spectrometry data without knowledge or reconstruction of their primary sequence. Potential applications include proteomic studies of organisms with unavailable genomic sequences and highly variable proteins analysis.

P-03.03.2-025
CrossHub update 2016: identification of functionally important transcription factor targets based on the mutational analysis and associations with clinical characteristics

G. Krasnov, A. Kudryavtseva, N. Melnikova, A. Dmitriev
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

The Cancer Genome Atlas (TCGA) represents a comprehensive database of genomic, transcriptomic and epigenetic alterations across more than 20 tumor types. Earlier we developed CrossHub tool aimed at multi-way analysis of TCGA data in the context of gene expression regulation. In the present work, the software was updated with new features that are described below.

CrossHub is a Python-based application. One of the features of CrossHub is the combining TCGA RNA-Seq co-expression analysis to ENCODE ChIP-Seq data in order to reveal most possible transcription factor (TF) targets and coupling miRNA-

The FEBS Journal **283** (Suppl. 1) (2016) 129–417 DOI: 10.1111/febs.13808
© 2016 The Authors. The FEBS Journal © 2016 FEBS