



VYSOKÁ ŠKOLA
CHEMICKO-TECHNOLOGICKÁ
V PRAZE

Konferenční sborník ENBIK2014

Editoři:

Petr Čech, Daniel Svozil

Praha 2014

Publikace neprošla jazykovou ani odbornou úpravou.
Za obsah příspěvků odpovídají autoři.

© Petr Čech, Daniel Svozil, 2014
Cover Design © Petr Čech, 2014

ISBN 978-80-7080-883-2

Obsah	3
Organizační a vědecký výbor	4
Abstrakty	7
Sekce 1	7
<i>Dobřej hardvér. A co softvér?</i>	
Sekce 2	15
<i>Sekvence – nukleové kyseliny</i>	
Sekce 3	25
<i>Představení vědeckých skupin</i>	
Sekce 4	31
<i>Metody</i>	
Sekce 5	37
<i>Chemoinformatika</i>	
Sekce 6	45
<i>Struktury</i>	
Sekce 7	55
<i>Sekvence – proteiny</i>	
Sekce 8	61
<i>Aplikace</i>	
Poster Session	69
Seznam přednášek	105
Seznam posterů	109
Autorský index	113
Seznam účastníků	117

Organizační výbor

Doc. Daniel Svozil, Ph.D.

RNDr. David Hoksza, Ph.D.

Ing. Petr Čech

Vědecký výbor

Doc. Daniel Svozil, Ph.D.

RNDr. Petr Bartůněk, CSc.

RNDr. David Hoksza, Ph.D.

Doc. Jan Holub, Ph.D.

RNDr. Jiří Vondrášek, Ph.D.



SEKCE 1

Dobřej hardvér. A co softvér?

L1-01

Rozvoj Bioinformatiky v prostředí IT4Innovations národního superpočítačového centra

Jansík B.¹*¹ IT4Innovations, 17. listopadu 15/2172, 708 33 Ostrava*

Kapacitní a výpočetní náročnost zpracování genomických dat přináší unikátní spojení Bioinformatiky a HPC, domény superpočítačů. Náročnost genomické analýzy, zejména zarovnání sekvencí a identifikace variant překračuje možnosti běžných počítačů. S klesající cenou sekvencování, kdy je dnes možné přečíst lidský genom v řádu několika hodin je výskumný potenciál využití tohoto ohromného množství informací limitován zejména dostupnou výpočetní kapacitou.

V této prezentaci se zaměříme na možnosti zpracování genomických dat na superpočítači IT4Innovations národního superpočítačového centra. Budou diskutovány nástroje, bioinformatické algoritmy a výpočetní zdroje, dostupné v centru IT4Innovations pro rychlou a bezpečnou analýzu genomických dat.

L1-02

OMICS – High Performace Computing řešení pro genomiku

Hlavsa T.¹, Vojtěch J.^{1,2}

¹ *Bull s.r.o., Lazarská 6, 120 00 Praha 2*

Genomika je na vzestupu a to především díky razantnímu poklesu ceny sekvencování, kdy je dnes možné sekvencovat lidský genom v řádu několika hodin a to za rozumnou cenu. Když tedy sekvencování není oním úzkým hrdlem je zde zásadní výzva v podobě zpracování, ukládání, interpretace a korelace dat.

Řešení společnosti BULL s názvem OMICS (Omics & Medical Integrated Compute Systems) cílí právě na tyto oblasti a překlenuje svým technologickým řešením prostor mezi genomikou a klinickým prostředím s cílem realizovat tzv. P4 (prediktivní, preventivní, personalizovaná a aktivní) medicínu.

V této prezentaci se zaměříme na řešení OMICS-MASTER založeném na High Performance Computing architektuře, pokročilých bioinformatických algoritmech a nástrojích pro bezpečné a rychlé transfery dat.

L1-03

**ChemAxon – chem a bioinformatický software pro Váš
výzkum**Hamerník P.¹¹ *ChemAxon, s.r.o, Karlovo náměstí 290/16, 120 00 Praha 2*

ChemAxon patří mezi hlavní světové dodavatele softwarové platformy a aplikací pro chemický, farmaceutický a biochemický průmysl. Mezi velké priority patří podpora akademické sféry a výzkumu. V této prezentaci budou představeny hlavní produkty firmy s důrazem na novinky od minulé konference Enbik 2012.

L1-04

Biowes - řešení pro správu a sdílení experimentálních dat a metadat

Císař P.¹, Štys D.¹, Bárta A.¹

¹ *Jihočeská univerzita, Ústav komplexních systémů, Nové Hradky, Zámek 136, 373 33*

Experimentální metadata jsou data o datech, která zachycují všechny informace potřebné k reprodukovatelnosti experimentu. Z pohledu vlastníka metadat je automatické vyhledávání informací experimentálních dat či sdílení dat nedílnou součástí budoucí reprodukovatelnosti experimentu.

Ještě významnější je pohled kognitivní, kdy metadata a jejich projekce do návrhu následného experimentálního postupu odrážejí skutečnou podobu pracovní hypotézy. To pregnantně definoval šéfredaktor časopis *Molecular Cell Biology* Thomas Lemberger ve svém úvodníku z února 2014, který zavádí novou strukturu časopisu [1]. Systém BioWes [2] odráží a formalizuje strukturu kognitivního procesu a archivuje jej v elektronické, strojově zpracovatelné podobě.

Práce se systémem BioWes je organizována jako zápis o provedení experimentu. Tím je myšlena, jak přípravná fáze (definování experimentálních podmínek a popis experimentu), tak samotné měření i popis experimentu a závěrů. Schválené popisy návrhu experimentu jsou v současné době vesměs buď archivovány v papírové podobě, nebo v elektronické podobě, kdy není jednoznačně zajištěn: vztah návrhu experimentálních dat a popisu experimentu. Problémem při návrhu experimentu může být i nedostupnost informací o experimentech prováděných v rámci daného pracoviště. Tyto informace by mohly být využity jednak jako inspirace pro samotný návrh, tak i pro ověření tohoto návrhu. Zkušenost jednotlivých pracovišť dokládá nedostatečný popis experimentů v článkách, kdy z daného popisu není možné dosáhnout reprodukovatelnosti experimentu. Poměrně málo pracovišť využívá pro návrh experimentu statistické metody návrhu experimentu.

V případě předávání zpráv o experimentu pak není zajištěno provázání experimentálních dat, popisu experimentu a poznatků o experimentu. Znalosti jsou centrálně soustředěny pouze u jedné osoby a izolovaně u jednotlivých experimentátorů. Též poskytovatelé veřejných prostředků stále častěji požadují ukládání a zpřístupňování dat.

Standardizace je jedním ze stěžejních bodů provádění biologických experimentů. Přínos standardizace je především v možnosti přesného pojmenování všech významných pojmů a tím zamezení chybné interpretace výsledků. Standard je také základem pro objektivní srovnání experimentů a analýzy dat. V současné době jsou dostupné mnohé databáze

terminologických nebo ontologických standardů. Systém BioWes řadu z nich podporuje a umožňuje z nich jednoduše čerpat. To vše Biowes systém zajišťuje.

Literatura

- [1] Tools of discovery, Thomas Lemberger, Mol. Sys. Biol. 10: 715 (2014)
- [2] <http://www.biowes.org/>



SEKCE 2

Sekvence – nukleové kyseliny

L2-01

Zpracování a analýza HTS dat

Daněček P.¹

¹ *Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, UK*

Rozvoj High-Throughput Sequencing (HTS) technologii v uplynulých deseti letech přinesl dramatický nárůst objemu sekvencovaných dat. Zvyšují se nároky na hardwarovou infrastrukturu (úložiště a sdílení dat) a vznikají nové výpočetní metody umožňující jejich efektivní zpracování. Velké projekty (1000 Genomes Project, UK10k) stimulovaly vývoj nových datových formátů (BAM, VCF) a algoritmů, ale bylo již řečeno poslední slovo?

Představím nedávno ještě experimentální, dnes rutinní, ale zítra již pravděpodobně zastaralou metodiku zpracování a analýzy dat ve Wellcome Trust Sanger Institute.

L2-02

Assembly komplikovaných genomů

Pačes J.¹

¹ *Ústav molekulární genetiky AVČR, v.v.i. Václavská 1083, 142 20 Praha*

V genomech eukaryotických organismů, ale i v genomech některých prokaryot se vyskytují sekvence, které jsou i za současného stavu vysokokapacitních sekvenačních technologií komplikovaným problémem. Jedná se zejména o repetitivní oblasti, u diploidních organismů pak silně polymorfní oblasti, ale i například regiony tvořící kvadruplexní DNA. V krátkém přehledu přinášíme několik problematických vlastností a přístupy k jejich řešení, včetně softwaru.

L2-03

Globální pohled na výstupy z 454 pyrosekvenací a jejich problémy

Mokrejš M.¹¹ *Bioinformatics.cz, Donovalská 1658, 149 00 Praha*

Sláva pyrosekvenování již prý pohasíná a tak by bylo zajímavé podívat se trochu do historie jejího vývoje. Ačkoli byla technologie veřejně zpřístupněna asi v roce 2005-2006 zůstala v mnoha ohledech černou magickou skřínkou do které nikdo nevidí. Délky čtení se prodloužily po některých vylepšeních přístroje méně než slibovali obchodníci ale vysvětlení jsme se nedočkali. Existují desítky postupů na primární zpracování vzorku (typicky převod RNA do cDNA) a nepochybně desítky způsobů zpracování pro samotné sekvenování.

Vývoj univerzálního nástroje na analýzu a čištění hrubých dat z pyrosekvenátoru Roche 454 se stal nečekaným oříškem a naznačuje že podobně komplikované to musí být i u jiných, konkurenčních technologií. Analýzou asi 1500 datasetů z celého světa vznikla zajímavá paleta grafických srovnání sekvenací, distribucí sekvenačních chyb a databáze artefaktů. V některých případech lze monitorovat chování softwaru od výrobce při generování sekvencí v různých fázích procesu a jeho chybovost. Lze najít i sekvenační soubory nad jejichž obsahem zůstává rozum stát. Snad jediným vysvětlením by mohlo být že ani samotní vědci nebo laboranti tehdy nevěděli co se vzorkem provádí (a nebo již zapomněli co s ním již provedli). Je také s podivem jak málo se lidé věnovali efektivitě investic ve smyslu počtů získaných vyčištěných bází vzorku v poměru vůči množství bází promarněných sekvenováním všelijakých adaptérů a značek vzorků. Některé laboratorní postupy totiž vedou k obzvláště dlouhým sekvencím adaptérů, v sekvenační kapacitě pak nezbyvá mnoho místa pro samotný vzorek a bylo by lepší se jim vyhnout.

Vyvinutý software obsahuje asi 15-ti stupňový rozhodovací algoritmus na jehož základě se rozhoduje další chod programu, tedy vybírají základní moduly sekvencí adaptérů, značek vzorků a artefaktů které se vzájemně mezi sebou kombinují tak jako v experimentu. Zvláštní kód je věnován přebírání různých podobných výsledků a vybírání toho jednoho, správného. Další kód je věnován problematickým homopolymerním sekvencím a snaží se správně určit odkud kam až dosahuje homopolymerický úsek rozbitý sekvenačními chybami na několik kusů. Kromě grafických a tabulkových výstupů jsou výstupem i opravené soubory se sekvencemi zbavenými adaptérů, značek vzorků a artefaktů. Software umí analyzovat veškeré typy dat z přístrojů Roche 454, alespoň některé „shotgun“ sekvenace z přístrojů IonTorrentPGM a nezávisle na

sekvenační technologii zpracovat sekvence cDNA získaných podle protokolů firem Evrogen a Clontech nebo některých dalších, méně obvyklých postupů.

L2-04

Vliv sekvenační délky na rekonstrukci transkriptomu s využitím RNA-Seq

Oppelt J.¹, Blake J.², Beneš V.², Koča J.¹

¹ Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova univerzita, Kamenice 5, 625 00 Brno-Bohunice, Česká republika

² Genomics Core Facility, EMBL Heidelberg, Meyerhofstr. 1, D-69117 Heidelberg, Německo

Nástupem sekvenování nové generace (NGS) se vědě otevřely nové možnosti výzkumu transkriptomu - RNA-Seq [1]. Tato metoda má na rozdíl od microarrays tu výhodu, že dokáže odhalit i předem neznámé transkripty [2] či provést rekonstrukci celého transkriptomu. S postupem času se samotné sekvenátory a sekvenační metody vyvíjí. Jedním z možných zlepšení je i prodloužení maximální sekvenační délky. Otázkou ale je, jaký vliv má délka těchto sekvencí na samotnou analýzu a zdali se vyplatí investovat do delších a nákladnějších sekvencí či postačí dnes již standartních 100 bází.

V rámci naší práce jsme provedli analýzu, jenž vycházela ze dvou vzorků myších tkání, které byly osekvenovány pomocí párové knihovny a s délkou 250 bází. Z této knihovny pak byly ještě vytvořeny další dvě knihovny o různých délkách – párová knihovna s délkou 100 bází a nepárová knihovna se sekvencemi, které měly překryv a bylo je možné spojit a vytvořit tím knihovnu s délkami až 500 bází. Ke zjištění vlivu délky sekvencí jsme použili *ab initio* rekonstrukci transkriptomu [3].

Analýza odhalila, že délka sekvencí nemá výrazný vliv na rekonstrukci transkriptomu pomocí zvolené metody. Výsledky z knihoven s delšími sekvencemi překonaly výsledky z knihoven se 100 bází dlouhými sekvencemi, avšak rozdíly nebyly nijak výrazné a mohly být způsobeny rozdílným pokrytím u testovaných knihoven. V souvislosti s různými délkami sekvencí jsme se ještě pokusili zanalyzovat jejich vliv na relativní expresi samotných transkriptů. Zajímavé zjištění bylo, že i přes použití FPKM normalizace [4] byly velké rozdíly mezi relativními expresemi transkriptů mezi jednotlivými knihovnami. Rozdíly byly též zjištěny při detailnějším pohledu na expresi různě dlouhých transkriptů.

Při celkovém hodnocení však nebyly zjištěny tak markantní rozdíly mezi různými délkami sekvencí, aby musel být přehodnocen celý design experimentů sekvenování transkriptomu. Delší sekvence jsou nákladnější a složitější na získání a čas strávený optimalizací experimentu na větší délky by v současné době nepřinesl o tolik lepší výsledky, kolik by do tohoto procesu bylo investováno času. Musíme však mít na

paměti, že experimenty s rozdílnými sekvenačními délkami nejsou přímo porovnatelné a mohou mít odlišné výsledky.

Literatura

- [1] Wang Z, Gerstein M, Snyder M., RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009 Jan;10(1):57-63. doi: 10.1038/nrg2484
- [2] Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X., Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One.* 2014 Jan 16;9(1):e78644. doi: 10.1371/journal.pone.0078644. eCollection 2014
- [3] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010 May;28(5):511-5. doi: 10.1038/nbt.1621. Epub 2010 May 2
- [4] Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L., Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 2011;12(3):R22. doi: 10.1186/gb-2011-12-3-r22. Epub 2011 Mar 16

L2-05

RepeatExplorer: Analýza repetitivních sekvencí na Galaxy serveru

Novák P.¹

¹ *Biologické centrum AV ČR, Ústav molekulární biologie rostlin, Branišovská 31, České Budějovice*

RepeatExplorer je soubor programů na analýzu repetitivních sekvencí v eukaryotních genomech vyvinutý v Laboratoři molekulární cytogenetiky BC AVČR. Tyto programy jsou k dispozici na veřejném webovém serveru [1], který je provozován na Galaxy platformě [2]. *RepeatExplorer* je vhodný především na *de-novo* charakterizaci repetit z krátkých shotgunových sekvencí získaných pomocí technologie NGS. *RepeatExplorer* poskytuje metody pro kontrolu kvality a přípravu sekvenačních dat, identifikaci, anotaci, kvantifikaci a vizualizaci repetit pomocí metod založených na teorii grafů. Hlavní výhodou RepeatExploreru je, že jej lze použít též pro charakterizaci repetit u druhů, u nichž dosud neznáme kompletní genomickou sekvenci. I když byl *RepeatExplorer* vyvinut pro studium rostlinných genomů, je vhodný také ke studiu repetit u jiných taxonomických skupin a pro srovnávací studie repetit z více druhů a identifikaci repetit nových.

Literatura

- [1] <http://www.repeatexplorer.org>
- [2] <http://galaxyproject.org/>



SEKCE 3

Představení vědeckých skupin

L3-01

Bioinformatická analýza mitochondriálnych genómov kvasiniek

Brejová B.¹

¹ *Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave, Mlynská dolina, 842 48 Bratislava, Slovensko*

Bioinformatická skupina na našej fakulte spolupracuje s Laboratóriom Komparatívnej a funkčnej genomiky eukaryotických organel na Prírodovedeckej fakulte UK na sekvenovaní a komparatívnej analýze mitochondriálnych genómov kvasiniek, pričom doteraz sme sekvenovali 40 mitochondriálnych genómov. Z informatickej stránky táto oblasť inšpirovala nový algoritmus na predikciu ancestrálnych poradí génov, ktorý sa obzvlášť hodí na spracovanie väčšieho počtu pomerne krátkych genómov. Z biologickej stránky naša spolupráca viedla k objaveniu ribozomálneho skákania v jednom z genómov, pričom tento neobvyklý jav bol dovtedy pozorovaný iba v genóme bakteriofága T4. V príspevku tiež spomeniem ďalšie výskumné a pedagogické aktivity našej skupiny.

L3-02

Od sekvencí k chromozómům: výzkum repetitivní DNA rostlin v Laboratoři molekulární cytogenetiky BC AVČR

Macas J.¹

¹ *Biologické centrum AVČR, Ústav molekulární biologie rostlin, Branišovská 31, České Budějovice*

U většiny vyšších rostlin tvoří geny jen malou část genomu. Většina jejich jaderné DNA je totiž tvořena mnohonásobně se opakujícími sekvencemi, tzv. repetitivní DNA, která zahrnuje především transpozóny (mobilní elementy) a tandemově uspořádané sekvence satelitní DNA. Rozdílný stupeň akumulace repetitivní DNA je jednou z hlavních příčin obrovské variability genomů vyšších rostlin, které se mezi různými taxony liší svou velikostí až 2000x. Repetitivní DNA má navíc významnou roli jako zdroj strukturálních a regulačních sekvencí a jako nástroj pro rychlé přestavby genomu. Vzhledem k jejich značné komplexitě je však detailní složení populací repetitivní DNA známo jen u malého počtu rostlinných druhů a ještě méně jsou objasněny otázky evoluční dynamiky, regulace amplifikace a interakce repetitivních sekvencí s ostatními složkami genomu. V naší laboratoři se snažíme přispět k řešení těchto otázek komplexním přístupem kombinujícím nejnovější technologie sekvenování (“next-generation sequencing”, NGS) s nově vyvíjenými bioinformatickými nástroji a experimentálními postupy molekulární cytogenetiky. Zabýváme se zejména: (i) vývojem sady programů a provozováním serveru RepeatExplorer [1], které slouží k identifikaci, kvantifikaci a klasifikaci repetitivních elementů v genomických sekvencích rostlin a živočichů získaných pomocí technologií NGS [2]; (ii) srovnávací analýzou složení repetitivní DNA u vybraných skupin rostlinných druhů; (iii) studiem role repetitivní DNA při determinaci centromer a evoluci chromozómů [3].

Literatura

- [1] <http://www.repeatexplorer.org> ; <http://w3lmc.umbr.cas.cz/repeatexplorer/>
- [2] Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. (2013) *Bioinformatics* 29: 792-793
- [3] Seznam recentních publikací viz <http://w3lmc.umbr.cas.cz/lmc/publ.php>

L3-03

Bioinformatický a chemoinformatický výzkum v Loschmidtových laboratořích

Brezovský J.¹, Bendl J.^{1,2,3}, Daniel L.¹, Bednář D.^{1,2}, Štourač J.^{1,2}, Marques S.¹,
Damborský J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Centrum biomolekulárního a buněčného inženýrství, Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, Brno 612 66*

Loschmidtovy laboratoře se zabývají základním i aplikovaným výzkumem. Centrálním tématem výzkumu je proteinové inženýrství enzymů halogenalkandehalogenas. Řada výzkumných projektů je řešena na rozhraní disciplín sekvenční a strukturní bioinformatiky, chemoinformatiky a výpočetní chemie. Výstupy výzkumných projektů jsou validovány ve spolupráci s experimentálními týmy Loschmidtových laboratořích. V rámci přednášky budou představeny následující vybrané výzkumné projekty: i) vývoj softwarových nástrojů pro analýzu a design proteinů – Caver Analyst, HotSpot Wizard a PredictSNP, ii) vývoj *in silico* platformy pro stabilizaci proteinů, iii) vývoj *in silico* platformy pro identifikaci zajímavých biokatalyzátorů, iv) optimalizace metabolických drah a v) inženýrství transportních tunelů v enzymech.

L3-04

Hľadanie genómov RNA vírusov po sekvenovaní celkovej RNA tkanív hlodavca

Szemes T.¹, Hýblová M.¹, Šoltys K.¹, Bokorová S.¹, Klempa B.¹, Budiš J.², Turňa J.¹

¹ *Prírodovedecká fakulta, Katedra molekulárnej biológie, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika*

² *Fakulta matematiky, fyziky a informatiky, Katedra Informatiky, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika*

Na našom pracovisku sme v roku 2013 rozbehli viacero výskumným projektov z oblasti analýzy genómov, pre ktoré je spoločným menovateľom použitie technológie sekvenovania novej generácie. Špeciálne sa venujeme analýze minoritne zastúpených genómov v zmesných vzorkách. V historickej zbierke tkanív myších mozgov, sa nám po prepise RNA, sekvenovaní a použití de novo skladania, a teda bez a priori hypotézy o tom aký genóm hľadáme, podarilo najst' a rekonštruovať takmer všetky virálne genómy vo vzorke.



SEKCE 4

Metody

L4-01

Klasifikační metody pro genetická data: regularizace a robustnost

Kalina J.¹¹ Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 2, 182 07 Praha 8

V posledních letech byly navrženy různé metody pro extrakci informace z vysoce rozměrných dat pro situaci, kdy se měří desítky tisíc proměnných jen na desítkách či stovkách pozorování [1]. Jedná se zejména o metody pro klasifikační analýzu, shlukovou analýzu, lineární regresi či redukci dimenzionality. Jsou založeny na vhodné regularizaci standardních statistických či data miningových metod.

Pro analýzu vysoce rozměrných dat jsme navrhli novou verzi lineární diskriminační analýzy, která je založena na deformaci Mahalanobisovy vzdálenosti, a také několik rychlých algoritmů na její výpočet. Střední hodnoty jednotlivých skupin jsou odhadnuty jako lineární kombinace průměru dané skupiny a celkového průměru. Současně je populační varianční matice odhadnuta jako lineární kombinace empirické varianční matice a vhodné (např. diagonální) regulární matice. Tato metoda je však stejně jako většina dosavadních postupů citlivá vůči přítomnosti odlehlých hodnot v pozorovaných datech. V oblasti robustní statistiky jsme navrhli novou metodu pro redukci dimenzionality, která je vhodná i pro silně korelované proměnné (geny). Vybírá nejvíce relevantní geny tak, aby množina vybraných genů vykazovala minimální redundanci. Přitom se měří relevance genů i jejich redundance robustním způsobem.

Teoretické postupy jsme použili při analýze studie genových expresí u pacientů s akutním infarktem myokardu nebo cévní mozkovou příhodou [2]. Zde nové postupy umožňují spolehlivější určení diagnózy oproti stávajícím metodám. Vytipovali jsme i sadu 10 genů, které indikují zvýšené riziko úmrtí v souvislosti s kardio-vaskulárními příčinami. Další výzkum bude zaměřen na robustifikaci regularizovaných metod pro klasifikační analýzu a regresi, a to v rámci projektu Robustní analýza vysoce rozměrných dat od Nadačního fondu Neuron.

Literatura

- [1] Kalina J.: Classification methods for high-dimensional genetic data. *Biocybernetics and Biomedical Engineering*, **34**, 2014. In press
- [2] Kalina J., Seidl L., Zvára K., Grünfeldová H., Slovák D., Zvárová J.: Selecting relevant information for medical decision support with application to cardiology. *European Journal for Biomedical Informatics*, **9** (1), 2-6, 2013

L4-02

Autoindex nad DNA sekvencemi

Procházka P.¹, Holub J.¹

¹ Katedra teoretické informatiky, Fakulta informačních technologií, České vysoké učení technické v Praze, Thákurova 2700/9, 160 00 Praha 6, Česká republika

Práce představuje datovou strukturu pro množinu DNA sekvencí jednoho druhu, která umožňuje sekvence nejen efektivně ukládat, ale v nich i rychle vyhledávat. Struktura je založena na metodě FM-index a na vlnkovém stromu. Struktura je velmi efektivní pro vyhledávání krátkých sekvencí. Dosahuje kompresního poměru až 2,26 % a času vyhledání 2,56 μ s.

L4-03

Duální selekce genetických obvodů v bakteriích

Beneš D.¹, Sosík P.^{1,2}, Rodríguez-Patón A.²

¹ *Výzkumný ústav centra excelence IT4Innovation, Filozoficko-přírodovědecká fakulta, Slezská univerzita v Opavě, 74601 Opava, Česká republika*

² *Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte, 28660 Madrid, Spain*

Úspěch syntetické biologie závisí na možnosti efektivně sestavovat robustní genetické obvody. Sestavení byť i nejjednodušších regulačních komponent (přepínače, logická hradla) je však stále ještě poměrně komplikované a vyžaduje intenzivní laboratorní práci. S nárůstem složitosti biologických obvodů se tento přístup stává neefektivním, jelikož závisí na mnohačetném opakování experimentů stylem pokus-omyl. Jedním z důvodů, proč je nutné experimenty mnohokrát opakovat, je, že jednotlivé komponenty v komplexním prostředí vzájemně interagují nepředvídatelným způsobem. Proto také narůstá popularita řízené evoluce založené na screeningu / selekci náhodných rekombinací knihoven genetických obvodů. V tomto příspěvku využíváme techniku zvanou duální selekce. Pomocí této techniky, s využitím párů toxin-antitoxin a konjugační bílkoviny relaxázy, navrhujeme autonomní in vivo evoluční protokol pro selekci jednoduchých booleovských obvodů v bakteriích E. Coli. Oproti ostatním evolučním protokolům tento umožňuje paralelní realizaci více selekčních kroků v jediném prostředí bez lidského zásahu. Naše metoda řízené evoluce kombinuje novou metodu duální selekce s fluorescenčním screeningem pro zvýšení odolnosti proti nežádoucím mutacím. Simulačně byly potvrzeny vysoké selekční i vyhledávací schopnosti tohoto protokolu.

L4-04

Metody kombinování shlukovacích algoritmů

Bartoň T.^{1,2}, Kordík P.²

¹ *Ústav molekulární genetiky AV ČR, Vídeňská 1083, Praha 4*

² *Fakulta informačních technologií, Thákurova 9, 160 00 Praha 6*

Shlukovací algoritmy typicky generují velmi rozdílné výsledky, u některých algoritmů je výsledek závislý na inicializaci algoritmu. Cílem kombinování výsledků různých algoritmů je dosáhnout stabilního shlukování.

Představíme metody kombinování algoritmů a zároveň se zaměříme na integraci externích znalostí v průběhu procesu shlukování.



SEKCE 5

Chemoinformatika

L5-01

ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat

Jindřich J.¹, Müller T.¹, Škuta C.^{1,2}, Sedlák D.¹, Pombinho A.¹, Svozil D.^{1,2}, Bartůněk P.¹

¹ CZ-OPENSOURCE: Národní infrastruktura chemické biologie, Ústav molekulární genetiky AV ČR, v.v.i, Vídeňská 1083, 142 20 Praha 4

² VŠCHT v Praze, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6

ChemGenDB je databázový/LIMS systém původně vyvinutý pro Centrum chemické genetiky. Je to webová aplikace využívající převážně programovací jazyk Python. Dále využívá následující open source a volně dostupné softwarové komponenty: webový/databázový framework Django, databázový systém MySQL pro úschovu dat a cheminformatický framework RDKit, jenž slouží ke konverzi chemických strukturních formátů, substrukturní a podobnostní vyhledávání a generování 2D struktur pro zobrazování ve webovém prohlížeči. Webové prohlížeče používané pro přístup k LIMS musí podporovat JavaScript a HTML 5. Pro GUI funkce jsou využívány JavaScript knihovny jQuery, jQueryUI a KineticJS. Pro editaci chemických struktur je také používán Marvin for JavaScript.

Doposud byly implementovány následující funkce: import informací o chemických vzorcích (včetně struktur), substrukturní a podobnostní vyhledávání, generování destiček pro chemické vzorky (96, 384 a 1536 jamkové destičky), reformáty destiček (kopírování, ředění, Z-reformát), tvorba a tisk čárových kódů destiček pro chemické vzorky a vzorky pro biologické testy, sofistikované zadávání dat pro screening včetně vizualizačních a analytických nástrojů, tvorba destiček pro testování a automatický upload dat s výsledky testů na server.

L5-02

Nová metoda pro generování 2D farmakoforového modelu

Hoksza D.^{1,2}, Svozil D.²

¹ *Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Malostranské nám. 25, 118 00 Praha*

² *VŠCHT Praha, Laboratoř informatiky a chemie, Technická, 166 28 Praha*

Využití farmakoforů při virtuálním screeningu se ukázalo jako metoda vhodná především v raných fázích vývoje léčiv. Standardně je virtuální screening s pomocí farmakoforů prováděn ve 3D prostoru, kde jsou využívány známé nebo predikované struktury bioaktivních ligandů. Tyto jsou zarovnány a následně jsou identifikovány farmakoforové vlastnosti, které jsou agregovány do farmakoforového modelu aplikovaného při screeningu. V mnoha případech ovšem 3D struktury známy nejsou, což vede k nutnosti použití 2D farmakoforového modelu. V tomto příspěvku představíme novou metodu pro generování 2D farmakoforového modelu založenou na znalosti struktur aktivních a neaktivních na daném biologickém cíli. Z množin aktivních a neaktivních sloučenin jsou nejdříve vytvořeny 2D farmakoforové deskriptory. Následně je na ně aplikována statistická metoda pro identifikaci farmakoforových vlastností separujících množinu aktivních molekul od množiny neaktivních. Tyto vlastnosti jsou použity k vytvoření výsledného 2D farmakoforového modelu, který je následně využit ve virtuálním screeningu. Metoda byla testována na MUV (Maximum Unbiased Validation) datových množinách [1] a porovnána s několika dalšími přístupy k virtuálnímu screeningu (molekulární deskriptory MOE, MACCS, SESP [1]). Na 10 ze 17 datových množin vykazuje naše metoda nejlepší výkon z porovnávaných metod (měřený vzhledem k ROC), přičemž na zbylých datových sadách je její výkon srovnatelný. Prezentovaná metoda tak představuje užitečný nástroj obohacující portfolio postupů pro virtuální screening.

Literatura

- [1] Rohrer, S. G., Baumann, K., Jan. 2009. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* 49 (2), 169-184

L5-03

Jaký vliv na přesnost QSPR modelů má metodika přípravy 3D struktury? Případová studie pro predikci pK_a .

Geidl S.¹, Svobodová Vařeková R.¹, Petrušek L.¹, Sehnal D.¹, Koča J.¹

¹ *National Centre for Biomolecular Research, Faculty of Science and CEITEC - Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic*

Disociační konstanta pK_a je velmi důležitou vlastností molekul a její hodnoty jsou významné pro chemický, biologický, environmentální a farmaceutický výzkum. Z tohoto důvodu je velká snaha o vývoj přesných a rychlých metod pro predikci pK_a . Velmi úspěšným přístupem pro pK_a predikci je použití QSPR (Quantitative Structure-Property Relationship) modelů [1]. Tyto QSPR modely mohou využívat různé deskriptory, přičemž tyto deskriptory jsou většinou založeny na 3D struktuře molekul (tzv. 3D deskriptory). Velmi důležitou otázkou je, jak metodologie přípravy 3D struktury ovlivňuje kvalitu QSPR modelů.

Cílem naší práce je zodpovědět tuto otázku. Pro naše analýzy jsme si vybrali QSPR modely založené na atomových nábojích, protože nábojové deskriptory jsou často používány pro pK_a predikci a poskytují velmi přesné výsledky [2, 3]. Konkrétně jsme se zaměřili na dva typy atomových nábojů – kvantově mechanické (QM) náboje a empirické náboje. QM náboje počítáme pomocí čtyř různých metod a empirické náboje pomocí metody EEM (Electronegativity Equalization Method) [4] parametrizované pro stejné přístupy. Provedli jsme analýzy 3D struktur generovaných šesti nástroji (Baloon, CORINA, Frog, Omega, OpenBabel a RDKit) a optimalizovaných třemi přístupy (bez optimalizace, MMFF94 optimalizace a B3LYP/6-31G* optimalizace). Analýza byla realizována pro tři skupiny molekul – fenoly, aniliny a karboxylové kyseliny. V průběhu práce jsme připravili více než 450 QSPR models.

Zjistili jsme, že zdroj 3D struktury je velmi důležitý pro přesnost QSPR modelů. Pro kvalitní zdroje 3D struktury je pK_a velmi přesné a optimalizace molekul není nezbytná. I pro méně kvalitní 3D struktury poskytuje predikce pK_a akceptovatelné výsledky. V případě nízké kvality 3D struktur není predikce pK_a přesná a optimalizace struktur je nezbytná.

Literatura

- [1] Lee AC, Crippen GM: Predicting pK_a , J Chem Inf Model 2009, 49, 2013–2033
- [2] Svobodová Vařeková R, Geidl S, Ionescu CM, Skřehota O, Kudera M, Sehnal D, Bouchal T, Abagyan R, Huber HJ, Koča J: Predicting pK_a Values of

- Substituted Phenols from Atomic Charges: Comparison of Different Quantum Mechanical Methods and Charge Distribution Schemes, *J Chem Inf Model* 2011, 51, 1795–1806
- [3] Svobodová Vařeková R, Geidl S, Ionescu CM, Skřehota O, Bouchal T, Sehnal D, Abagyan R, Koča J: Predicting pK_a values from EEM atomic charges, *J Cheminf* 2013, 5, 1-15
- [4] Svobodová Vařeková R, Jiroušková Z, Vaněk J, Suchomel Š, Koča J: Electronegativity equalization method: Parameterization and validation for large sets of organic, organohalogene and organometal molecule, *Int J Mol Sci* 2007, 8, 572-582

L5-04

Radosti a strasti dokování

Bazgier V.¹, Berka K.²

¹ *Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra fyzikální chemie, CR Haná, 17. listopadu 12, 771 46 Olomouc*

² *Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra fyzikální chemie, RCPTM, 17. listopadu 12, 771 46 Olomouc*

Oblíbeným způsobem, jak studovat interakce mezi proteiny a malými organickými molekulami, je molekulární dokování. V tomto přístupu se hledá především vhodná vazebná póza mezi interagujícími partnery a podle nějakého modelu se stanovuje její skóre, pomocí kterého je pak možné dokování vyhodnocovat. Zní to jednoduše, ale celý tento proces málokdy sám o sobě stačí a tak se pak z dokování do jednotlivých proteinových systémů stává do značné míry expertní ruční práce, která se jen málokdy dá provádět skutečně masově, protože expertní vhléd do proteinu a do struktur testovaných ligandů mnohdy zachrání jinak zcela selhávající proces dokování. Na několika příkladech z naší praxe z poslední doby ukážeme, jak se mohou lišit jednotlivé případy dokování u systémů tak odlišných jako cytochromy P450,¹ CDK2 kinázy,^{2,3} sodno-draselná pumpa,^{4,5} a nakonec i receptorů pro histamin⁶ a brassinosteroidy⁷ a jak z toho (možná) vede cesta ven.

Tato práce byla podpořena OP VaVPI (CZ.1.05/2.1.00/03.0058). K.B. byl podpořen také post-doc grantem GAČR (P303/12/P019). V.B. byl podpořen studentským grantem UP (IGA_PrF_2014023).

Literatura

- [1] Berka K et al *J. Inorg. Biochem.* 110, 46, 2012
- [2] Gucký T et al *J. Med. Chem.* 56, 6234, 2013
- [3] Mojzych M et al *Eur.J. Med. Chem.* 78, 217, 2014
- [4] Havlíková M et al *BBA - Biomem* 1828, 568, 2013
- [5] Huličiak M et al *submitted*
- [6] Pozharitskaya O et al *Planta Med.* 79, 1698, 2013
- [7] Kvasnica M et al *submitted*

L5-05

Konzistence křížových odkazů v databázích malých molekul

Galgonek J.¹

¹ *Ústav organické chemie a biochemie AV ČR, v.v.i., Flemingovo nám. 2, 166 10 Praha*

Databáze malých molekul hrají důležitou roli v mnoha oblastech chemoinformatiky. Jejich zaměření bývá různé - od všeobecného až po úzce specializované. Pro maximální využití těchto databází je důležité vzájemné propojení jejich obsahu pomocí křížových odkazů. To platí zejména v případech, kdy se požadované informace nacházejí v různých databázích. Vytvoření křížových odkazů mezi databázemi však není jednoduchý úkol. Různé databáze totiž mohou používat rozdílná jména sloučenin, nebo nemusí stejnou sloučeninu reprezentovat pomocí stejné struktury (stejně tautomerní formy). Křížové odkazy mohou být buď spravovány ručně odborníky, anebo mohou být generovány automaticky. Při automatickém generování odkazů je důležité přiřadit sloučenině jedinečný identifikátor. Tento identifikátor je typicky generován na základě její struktury tak, aby nezávisel na použité tautomerní formě. Takovýmto identifikátorem může být například InChI vyvinutý International Union of Pure and Applied Chemistry (IUPAC). V naší práci jsme se zaměřili na srovnání konzistence a úplnosti automaticky generovaných odkazů založených na InChI s ručně spravovanými křížovými odkazy, které jsou uloženy přímo v databázích. Experimenty jsme prováděli na pěti databázích - ChEBI, DrugBank, PDBeChem, Human Metabolome Database (HMDB) a NCGC Pharmaceutical Collection (NPC). Ukázalo se, že odkazy jsou značně nekonzistentní (v nejhorším případě až 28.9% ručně spravovaných odkazů má nekonzistentní InChI identifikátory). Taktéž úplnost ručně spravovaných odkazů se ukázala jako nepříliš dobrá (pouze 93.8% v nejlepším případě).



SEKCE 6

Struktury

L6-01

Vyhledávání, porovnávání, validace a charakterizace strukturních motivů v rámci biomakromolekul

Svobodová Vařeková R.¹, Sehnal D.¹, Pravda L.¹, Ionescu C.-M.¹, Wimmerová M.¹, Koča J.¹

¹ *National Centre for Biomolecular Research and CEITEC - Central European Institute of Technology, Masaryk University, 625 00 Brno, Czech Republic*

V současné době máme k dispozici velké množství informací o 3D struktuře biomakromolekul. Toto množství dat poskytuje příležitost k analýze strukturních motivů (fragmentů) uvedených látek – např. vazebných míst, elementů sekundární struktury, ligandů, kavit a tunelů. Dané analýzy lze využít k identifikaci vzorů pro vývoj léků, k porozumění vztahů mezi strukturou proteinu a jeho funkcí, pro klasifikaci proteinů apod. [1-2]. Proto byly pro vyhledávání strukturních motivů a jejich analýzu vytvořeny nové metodiky a softwarové nástroje.

V prezentaci představíme nově vyvinuté nástroje pro zpracování strukturních motivů biomakromolekul. Motivy mohou být popsány pomocí jednoduchého a robustního jazyka MotiveQuery a na základě tohoto popisu pak automaticky vyhledány v Protein Data Bank pomocí MotiveQuery Serveru. Další nástroj, SiteBinder [3], je schopen přiložit tisíce proteinových strukturních motivů ve velmi krátkém čase. Softwarový nástroj Mole [4, 5], je určen pro vyhledávání tunelů. MotiveValidator je pak schopen určit, zda se testovaný ligand shoduje s referenční molekulou (má stejnou topologii i chiralitu). Poslední nástroj, ChargeCalculator, poskytuje hodnoty empirických nábojů vypočítaných pro vstupní molekulu. Využití nástrojů je ukázáno na případových studiích, zaměřených na biologicky významné proteinové strukturní motivy.

Literatura

- [1] Baran I, Svobodová Vařeková R, Parthasarathi L, Suchomel S, Casey F, Shields DC: Identification of potential small molecule peptidomimetics similar to motifs in proteins, *J Chem Inf Model* 2007, 47, 464-474
- [2] Watson JD, Laskowski RA, Thornton JM: Predicting protein function from sequence and structural data, *Curr Opin Struct Biol* 2005, 15, 275-284
- [3] Sehnal D, Svobodová Vařeková R, Huber HJ, Geidl S, Ionescu CM, Wimmerová M, Koča J: SiteBinder – an improved approach for comparing multiple protein structural motifs, *J Chem Inf Model* 2012, 52, 2, 343–359
- [4] Berka K, Hanák O, Sehnal D, Banás P, Navrátilová V, Jaiswal D, Ionescu CM, Svobodová Vařeková R, Koca J and Otyepka M: MOLEonline 2.0: interactive

- web-based analysis of biomacromolecular channels. *Nucleic Acids Res* 2012, 40, W222–7
- [5] Sehnal D, Svobodová Vařeková R, Berka K, Pravda L, Navrátilová V, Banáš P, Ionescu CM, Otyepka M and Koča J: MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminform* 2007, 5, 39

L6-02

Dolce: Databáze lokálních konformací DNA

Čech P.¹, Schneider B.², Svozil D.¹

¹ *VŠCHT Praha, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6*

² *Biotechnologický ústav AV ČR, v. v. i., Vídeňská 1083, 142 00 Praha 4*

Vývoj databází je ve strukturní bioinformatice zcela zásadní a přirozenou záležitostí. Jednou z nejvýznamnějších databází, které spravují strukturní data biologických makromolekul, zejména proteinů, nukleových kyselin a jejich komplexů, je Protein Data Bank (PDB) [1]. Základní strukturní data, která PDB poskytuje, se obvykle dále využívají a zpracovávají v projektech, které se zaměřují na specifickou oblast vědeckého výzkumu. Již existující data jsou tak díky četným a rozsáhlým analýzám obohacována o další informace, které je potřeba systematicky třídit a uchovávat. V tomto příspěvku představíme Dolce – Databázi lokálních konformací DNA. Vývoj databáze Dolce bezprostředně navazuje na náš předchozí projekt [2], ve kterém jsme pomocí metod strojového učení vyvinuli automatický postup pro klasifikaci dinukleotidových stepů v molekulách DNA. Dolce zahrnuje a udržuje stále aktuální výsledky klasifikace pro všechny struktury DNA uvolněné z PDB. K databázi jsme vytvořili uživatelské rozhraní s vestavěným systémem pro tvorbu dotazů, jehož pomocí lze prohledávat strukturní data DNA včetně klasifikace dinukleotidových konformací. Veškeré výsledky dotazů jsou vizualizovány přímo v prostředí webovské aplikace nebo mohou být uloženy ve formátu CSV pro další zpracování. Dolce je volně dostupná na <http://ich.vscht.cz/projects/dolce>.

Literatura

- [1] Berman H. M., et al., The Protein Data Bank. *Nucleic Acids Research*, 2000, 28(1): 235-242
- [2] Čech P., et al., Automatic workflow for the classification of local DNA conformations. *BMC Bioinformatics*, 2013, 14(1)

L6-03

MOLE: Pokročilý nástroj pro vyhledávání a charakterizaci tunelů

Pravda L.¹, Svobodová Vařeková R.¹, Sehnal D.¹, Ionescu C.-M.¹, Berka K.², Otyepka M.², Koča J.¹

¹ *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*

² *Regionální centrum pokročilých technologií a materiálů, Katedra fyzikální chemie, Přírodovědecká fakulta, Univerzita Palackého, Olomouc, tř. 17. listopadu 12, 771 46 Olomouc*

Tunely a póry mají nezastupitelnou úlohu v různých životně důležitých procesech, jelikož umožňují enzymatickou aktivitu v mnoha proteinech s aktivním místem umístěným hluboko pod povrchem proteinu. Jejich geometrické vlastnosti, jako délka nebo průměr, ale i chemické vlastnosti (polarita, hydropatie nebo potenciál) mají velký vliv na specifitu a selektivitu celé řady enzymatických reakcí. Nejen z těchto důvodů je identifikace a charakterizace tunelů předmětem zájmu pro mnoho různých oblastí – např. pro vývoj léků nebo proteinové inženýrství.

MOLE je nástroj pro snadnou, rychlou a intuitivní analýzu tunelů a pórů v biomakromolekulárních systémech. Díky svým vlastnostem jako jsou interaktivita, rychlost nebo propojení s externími systémy pro vizualizaci molekul nachází uplatnění např. při hloubkovém studiu jednotlivých biologických systémů od malých proteinů až po ribozomy. Navíc je schopen dávkově zpracovávat velké množství proteinů paralelně. Funkcionalita MOLE bude ukázána prostřednictvím analýzy tunelů v rámci více než 4000 proteinů s katalytickou aktivitou uvedenou v CSA databázi.

MOLE je dostupný zdarma jako online služba (<http://mole.upol.cz/>) nebo jako samostatný softwarový balík s vestavěnou vizualizací (<http://mole.chemi.muni.cz>).

Literatura

- [1] Sehnal,D., Svobodová Vařeková,R., Berka,K., Pravda,L., Navrátilová,V., Banáš,P., Ionescu,C.-M., Otyepka,M. and Koča,J. (2013) MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform.*, **5**, 39.doi:10.1186/1758-2946-5-39.
- [2] Berka,K., Hanák,O., Sehnal,D., Banáš,P., Navrátilová,V., Jaiswal,D., Ionescu,C.-M., Svobodová Vařeková,R., Koca,J. and Otyepka,M. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular

- channels. *Nucleic Acids Res.*, **40**, W222–
7. doi:10.1093/nar/gks363. PMID:22553366.
- [3] Otyepka, M., Berka, K. and Anzenbacher, P. (2012) Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Curr. Drug Metab.*, **13**, 130–42. PMID:22208528.
- [4] Pravda, L., Berka, K., Svobodová Vařeková, R., Sehnal, D., Banáš, P., Laskowski, R.A., Koča, J., Otyepka, M. (2014) Anatomy of Enzyme Channels. *Bioinformatics*, submitted.
- [5] Furnham, N., Holliday, G.L., de Beer, T. a P., Jacobsen, J.O.B., Pearson, W.R. and Thornton, J.M. (2013) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* doi:10.1093/nar/gkt1243. PMID:24319146.

L6-04

Vývoj nových metod a nástrojů pro metadynamické simulace

Hošek P.¹, Spiwok V.¹

¹ *VŠCHT Praha, Ústav biochemie a mikrobiologie, Technická 5, 166 28 Praha*

Molekulární dynamika se stala oblíbeným nástrojem pro studium proteinů a jiných biomakromolekul díky své možnosti zobrazit časový průběh ke statickým strukturám získaných krystalovou rentgenografií či NMR. Avšak pro studium pomalejších dějů (sbalování proteinů, výraznější konformační změny) molekulární dynamikou je potřeba delší simulace, které mohou být pro současné výpočetní kapacity příliš náročné. Proto byla velká pozornost věnována rozvoji metod s lepším vzorkováním, které vyžadují mnohem kratší délky simulací. Jedna z těchto metod je Metadynamika, která využívá uměle dodávaný potenciál pro znevýhodnění již navštívených konformačních stavů. Díky tomuto znevýhodňování snadněji překonává energetické bariéry a prozkoumává další lokální minima. V naší laboratoři jsme vyvinuli dvě nové metody využívající Metadynamiku jako základ. První z nich, Moving Hills, nepoužívá postupné ukládání umělého potenciálu jako Metadynamika, ale potenciál se přemísťuje podle okamžitého stavu v mnoha paralelních simulacích. Rozmístění umělého potenciálu díky vzájemně odpuzivému účinku probíhá mnohem rychleji než v případě klasické metadynamiky. Pro ověření metody bylo provedeno srovnání na alanin-dipeptidu a bylo dosaženo dobré shody. Druhá metoda, Phantom MTD, je implementací alchemických simulací využívajících Hessova zákona do metadynamického programu PLUMED. Metoda byla použita pro odhad solvatační energie několika malých molekul a tento odhad byl v dobré shodě s experimentálními výsledky. Navíc byla tato metoda testována na několika modelových případech (tvorba katenanů, průchod molekuly membránou). Kromě výše zmíněných metod byl vytvořen v naší laboratoři nástroj pro vizualizaci dat získaných z Metadynamiky či výše zmíněných metod ve formě webové stránky. Tento nástroj kombinuje nové technologie jako HTML5, WebGL a JavaScriptové typované pole pro větší výkon při prohlížení dat jako videa.

L6-05

Simulace oscilační molekulární dynamiky allosterických proteinů v programu GROMACS

Šerá L.¹, Spiwok V.¹

¹ *Vysoká škola chemicko-technologická Praha, Ústav biochemie a mikrobiologie, Technická 5, 166 28, Praha 6, Česká republika*

Alosterie zahrnuje široké spektrum jevů, při kterých je akce v jedné části molekuly následována reakcí v jiné části. Je za určitých podmínek vlastní všem strukturovaným proteinům [1]. Jelikož zahrnuje změny odehrávající se v řádech mikrosekund, je studium alosterie pomocí klasické molekulové dynamiky velice náročné. Oscilační molekulová dynamika (OMD) problém zjednodušuje zavedením oscilační síly o vhodně zvolené frekvenci, amplitudě a směru na vybrané atomy nebo residua. Tento signál se poté šíří strukturou proteinu pomocí řídké sítě alostericky propojených residuí. Šíření takového impulzu může být detekováno v různých místech za použití Fourierovy transformace atomových souřadnic. Následně lze použitím vhodných metod vícerozměrné statistiky [2] získat profil spřažení a tudíž kvantifikovat míru interakce mezi spřaženými residuí [3].

OMD byla implementována do programu *mdrun* programového balíku GROMACS. Dále byl do tohoto balíku přidán program pro FFT analýzu získaných trajektorií. Bylo provedeno několik sérií testů, z nichž byly nakonec získány odpovídající data. V současné době testujeme vliv různých nastavení a zvolených parametrů jako frekvence, amplituda, směr, silové pole, ošetření solventu. Dále máme v plánu zmapovat alosterickou propojenost v receptorech vázaných na G-proteiny, které jsou cílem více než jedné třetiny všech na trhu dostupných léků. V neposlední řadě také plánujeme implementaci metody ve formě webového serveru dostupného širší akademické veřejnosti.

Literatura

- [1] K. Gunasekaran; B.Y. Ma, R. Nussinov. *Proteins* 57 (2004) 433-443
- [2] N. Prabhu; A. Lee; J. W. Wand; K.A. Sharp. *Biochemistry* 42 (2002) 562-570
- [3] K. Sharp; J. J. Skinner. *Proteins* 65 (2006) 347-361



SEKCE 7

Sekvence – proteiny

L7-01

PredictSNP: prediktor vlivu mutací na funkci proteinů

Bendl J.^{1,2,3}, Štourač J.^{1,3}, Šalanda O.², Pavelka A.¹, Wieben E. D.⁴, Zendulka J.², Damborský J.^{1,3}, Brezovský J.¹

¹ *Loschmidovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, 612 66 Brno*

³ *Centrum biomolekulárního a buněčného inženýrství, Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

⁴ *Department of Biochemistry and Molecular Biology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA*

Pokrok v oblasti sekvenčních technik způsobil obrovský nárůst počtu identifikovaných jednobodových mutací. Vliv nalezených mutací na životaschopnost a zdraví organismu však zpravidla není známa. K predikci míry škodlivosti konkrétních mutací v kontextu lidského zdraví byla dosud vyvinuta řada výpočetních metod, které se lze využít pro prvotní ohodnocení efektu mutací a jejich následnou prioritizaci pro experimentální charakterizaci. Ačkoliv je výběr metody zcela zásadní, jejich přehledné a věrohodné srovnání dosud přesto chybělo. Hlavní limitace existujících komparativních analýz spočívá ve zkrácení výsledků predikcí v důsledku překryvů trénovací a testovací datové sady. V rámci této studie vznikly tři datové sady nezávislé na trénovacích sadách osmi známých predikčních nástrojů. Tyto datové sady byly použity k objektivní evaluaci nástrojů. Následně bylo šest nejvýkonnějších z nich integrováno do podoby konsenzuálního klasifikátoru, který významně zvyšuje úspěšnost binární predikce škodlivosti mutace i přesnost odhadu predikční spolehlivosti. Kromě váhovaného konsenzuálního výsledku nástroj poskytuje uživatelské rozhraní s přístupem k výsledkům všech integrovaných nástrojů. Analýza je doplněna experimentálními anotacemi mutací z databází PMD a UniProt. Webový server a datové sady jsou volně dostupné na adrese <http://loschmidt.chemi.muni.cz/predictsnp>.

L7-02

Systém pre identifikáciu mutácií v proteomických dátach

Hruška M.^{1,2}, Voller J.¹, Džubák P.¹, Hajdúch M.¹

¹ Ústav Molekulární a Translační Medicíny Lékářské Fakulty Univerzity Palackého v Olomouci, Hněvotínska 5, 779 00 Olomouc

² Univerzita Palackého v Olomouci, Přírodovědecká fakulta, 17. listopadu 1192/12, 771 46 Olomouc

Typické nádorové bunky nesú mutácie v desiatkach, či dokonca stovkách génov. Znalosť mutačných profilov umožňuje porozumieť procesom, ktoré sú pozmenené v nádorových bunkách u jednotlivých pacientov a podľa toho zvoliť terapiu. Pre skríning mutácií sú bežne používané moderné vysokoformátové metódy ako SNP čipy a moderné celogenomové sekvenovanie (NGS). Hmotnostná spektrometria proteómu, ďalšia vysoko perspektívna metóda pre personalizovanú medicínu sa pre tento účel v dostatočnej miere nevyužíva. Limitujúcim faktorom je nedostatočné zastúpenie mutantných foriem v proteomických databázach a nemožnosť rekonštrukcie peptidov pomocou de-novo prístupov v komplexných vzorkách.

Pre dosiahnutie pokroku v tejto oblasti sme vytvorili systém pre identifikáciu mutantných, polymorfných a alternatívne zostríhaných génov na úrovni proteínov. V jadre systému sa nachádza databáza, systematicky skonštruovaná z unikátneho množstva zdrojov (mutácie, integrované informačné zdroje). Navyiac, užívateľ môže poskytnúť vlastný zdroj mutácií a na jeho základe vygenerovať peptidóm, ktorý môže použiť pri vyhľadávaní. Systém taktiež umožňuje užívateľom prehľadávať MS/MS spektrá a vykonáva potrebnú štatistickú evaluáciu kandidátnych výsledkov. Identifikácia prebieha pomocou nami vyvinutého systému Dymka, ktorý integruje viacero vyhľadávacích systémov a je paralelizovaný na výpočtovom klastri. Štatisticky významné výsledky následne podstupujú dôkladnú automatickú analýzu, kde sú oddelené relevantné identifikácie od potenciálnych artefaktov, napr., zámena leucin/isoleucin. Výsledky sú ďalej anotované z viacerých informačných zdrojov a prezentované vo forme vhodnej pre ďalšie spracovanie.

Systém umožňuje používať hmotnostnú spektrometriu ako nástroj pre identifikáciu mutácií, polymorfizmov a zostríhových variant. Informácia daného typu je kľúčová pre výskum nádorových ochorení a diagnostiku somatických mutácií. Aj vďaka možnosti generovať peptidóm na základe mutácií z vlastných zdrojov, môže byť systém použitý pre iné genomicko-proteomické aplikácie.

L7-03

Bioinformatická analýza náhodných proteinových sekvencí

Vymětal J.¹, Tretyachenko V.², Hloučová K.^{1,2}, Vondrášek J.¹

¹ Ústav Organické Chemie a Biochemie AV ČR, Flemingovo nám. 2, 166 10 Praha 6

² Karlova Univerzita v Praze, Přírodovědecká fakulta, Katedra Biochemie, Hlavova 2030, 128 00 Praha 2

Díky masivnímu rozvoji sekvenování obsahuje proteinová databáze Uniprot sekvence více než 50 milionů předpovězených proteinů. V proteinové strukturní databázi (PDB) bylo již deponováno více než 90 tisíc proteinových struktur, které reprezentují přibližně 1300 známých proteinových strukturních rodin (foldů). Tato čísla jsou však naprosto zanedbatelná v porovnání s astronomickým množstvím nevyskytujících se polypeptidů, které lze teoreticky získat kombinací 20 proteinogenních aminokyselin. Tyto úvahy vedou k představě, že strukturované proteiny musí být v celém sekvenčním prostoru neobyčejně vzácné a prakticky nerealizovatelné prostřednictvím náhodné kombinace aminokyselin.

Abychom otestovali tento předpoklad, provedli jsme analýzu náhodných proteinových sekvencí s cílem sestavit protokol, který by odhalil sekvence s potenciálem vytvořit stabilní sekundární a terciální struktury. Tyto vybrané sekvence plánujeme charakterizovat experimentálně a validovat tak protokol samotný i jednotlivé predikční metody, které využívá.

Námi navržený protokol je založen na predikci obsahu sekundárních struktur (gor4, jnet, predator, simpa, psipred), obsahu nestrukturovaných elementů (disopred, disembl, vs12, iupred), agregačních tendencí (zyggregator) a dalších fyzikálně-chemických vlastností. Prezentovány budou výsledky získané analýzou 1 milionu náhodných sekvencí s různými distribucemi aminokyselin, které odrážejí statistiky Uniprotu, PDB a databáze nestrukturovaných proteinů (DisProt). Další distribuce využívají omezenou aminokyselinovou abecedu či arteficiální poměry mezi aminokyselinami.



SEKCE 8

Aplikace

L8-01

Horizontální přenos genů v komplexní struktuře genomických signatur

Martínková N.^{1,2}, Moravec J. C.¹, Jaron K. S.²

¹ *Akademie věd ČR, Ústav biologie obratlovců, v.v.i., Květná 8, 603 65 Brno*

² *Masarykova univerzita, Institut biostatistiky a analýz, Kamenice 3, 625 00 Brno*

Horizontální přenos genů boří zavedené dogma o přenosu vlastností z jedince na potomka. Adaptivní vlastnosti organismů získány horizontálním přenosem zahrnují např. patogenicitu, rezistenci vůči lékům, schopnosti přežít v toxickém, nebo teplotně extrémním prostředí. Horizontálně přenesený genomický úsek se odlišuje od domovského genomu signály evoluční historie. Mezi ně patří kromě fylogenetického signálu i genomická signatura, která představuje vektor frekvencí oligonukleotidů a vzniká zřejmě systematickou chybovostí molekulárního aparátu replikujícího DNA. Místa genomu s jiným než specifickým poměrem frekvencí oligonukleotidů mohou být produkty horizontálního přenosu. U vyšších organismů metody založené na genomické signatuře selhávají, protože složitější struktura genomu znemožňuje popsání organismu jedinou statistikou hlavně z důvodu heterogenity sekvence DNA podél chromozomů. Pomocí klouzavé hustoty prvků genomické signatury demonstrujeme, že rozdělení frekvence oligonukleotidů se mění v závislosti na pozici sekvence na chromozomu. Lokálně se v některých oblastech vyskytují multimodální hustoty signatury. Vznik multimodálního rozdělení signatury může reflektovat kombinaci dvou a více úseků se specifickou signaturou, například při horizontálním přenosu. Tyto úseky ovšem mohou představovat i strukturální nebo konzervativní části původní v domovském genomu. Popsání těchto struktur je dalším krokem pro úspěšnou detekci genomických ostrovů ve vyšších organismech, a tak k nahrazení konzervativních fylogenetických metod, jejichž efektivita je silně závislá na známých sekvencích a schopnosti algoritmů odhalit homologii.

L8-02

Využití signálového zpracování ve fylogenetice a fylogenomice

Sedlář K.¹, Škutková H.¹, Provazník I.¹

¹ *Ústav biomedicínského inženýrství, Fakulta elektrotechniky a komunikačních technologií, VUT v Brně, Technická 12, 616 00 Brno*

Vyhodnocení strukturální podobnosti nukleotidových sekvencí je základním krokem při analýze příbuznosti nebo funkční podobnosti organismů. Stěžejní problém dříve spočíval v nalezení vhodných fylogenetických markerů reprezentovaných konkrétními geny, které byly dostatečně dlouhé pro spolehlivé odlišení nejbližších i nejvzdálenějších organismů a zároveň dostatečně krátké pro zajištění přijatelné výpočetní náročnosti. S rostoucími výpočetními výkony a zejména s klesající cenou sekvenačních technik, kdy se stále více do popředí dostávají multigenové analýzy nebo dokonce fylogenomické studie, nastává problém nový. Jedná se o nalezení dostatečně efektivního nástroje pro zpracování velkého objemu genomických dat. Vzhledem k tomu, že i pouhý průměrný bakteriální genom má 2 – 3 milióny párů bází, vyhodnocení vzájemné podobnosti více takových záznamů obvykle ztroskotává na nemožnosti realizace jejich mnohonásobného zarovnání. Řešení prezentované v příspěvku spočívá v konverzi genetické informace do signálové podoby (genomický signál). Bylo prokázáno, že různé organismy se projevují taxonomicky specifickým průběhem genomického signálu, a to jak na úrovni celogenomové reprezentace, tak i na úrovni jednotlivých genů. Signálová forma navíc umožňuje decimaci objemných dat s přijatelnou odchylkou ve výsledku při současném exponenciálním zrychlení výpočtu. Tím se zarovnání dlouhých sekvencí, které je ve znakové podobě nemožné, stává snadno realizovatelným. Navíc nedochází k zarovnání jednotlivých pozic v závislosti na jejich podobnosti, ale přiřazují se navzájem celé lokální homologie projevující se podobným trendem v signálu. Díky větší míře obsažené fylogenetické informace v kompletních genomech bakterií umožňuje fylogenomická analýza s využitím signálového zpracování dosažení přesnějšího výsledku a odlišení i velmi blízkých druhů, které klasická klasifikace na základě jednoho genu (např. standard 16S rRNA) neumožňuje.

L8-03

Evolučně konzervované vzory sekvenční komplementarity mezi eukaryotickými mRNA UTR a rRNA a implikace jejich existence pro regulaci genové translace

Pánek J.¹, Valášek L. S.^{1,2}

¹ *Laboratoř bioinformatiky, Mikrobiologický ústav Akademie věd České republiky*

² *Laboratoř regulace genové exprese, Mikrobiologický ústav Akademie věd České republiky*

Existuje několik klíčových mechanismů regulujících eukaryotickou genovou expresi na úrovni syntézy proteinů. Kupodivu, nejméně prozkoumané mechanismy translační kontroly jsou ty, které zahrnují centrální jednotku syntézy proteinů, tj. překládající ribozom, a které by mohly být zprostředkované např. interakcemi mezi ribozomálními RNA (rRNA) and mRNA.

Využili jsme rostoucí objemy dat sekvencí mRNA četných organismů, nedávno vyřešené struktury ribozomů a dostupný vysoký výpočetní výkon pro bioinformatickou identifikaci statisticky významné komplementarity mezi sekvencemi mRNA a rRNA četných biologických druhů s potenciálem k vytváření interakcí mezi mRNA a překládajícím ribozomem.

Zjistili jsme, že komplementarita mezi sekvencemi 18S RNA a oběma 5' a 3' UTR formuje evolučně konzervované, přesně definované a specifické vzory na strukturách ribozomální 40S podjednotky. Jejich evoluční konzervovanost a specifická, v kontextu už publikovaných experimentálních výsledků, napovídá, že představují molekulární podstatu nového regulačního mechanismu genové translace.

L8-04

Lokalizace potenciálních G4-kvadruplexů v eukaryotických genomech

Lexa M.¹, Kejnovský E.², Martínek T.³

¹ *Fakulta informatiky Masarykovy univerzity v Brně, Botanická 68a, 602 00 Brno*

² *Biofyzikální ústav Akademie věd České republiky, Královopolská 135, 612 65 Brno*

³ *Fakulta informačních technologií Vysokého učení technického v Brně, Božetěchova 2, 612 66 Brno*

V posledních letech dochází k obnovenému zájmu biologů o guaninové kvadruplexy. Bylo zjištěno, že jsou v mnohých eukaryotických genomech přítomny častěji než by napovídalo nukleotidové složení DNA. Vyskytují se obzvláště často v telomerech, promotorech a počátcích replikace. Nedávno jsme popsali i jejich výskyt v LTR sekvencích rostlinných retrotranspozonů a jejich asociaci s lidskými repetitivními sekvencemi jako je např. SVA nebo LINE-1. V souvislosti s touto asociací spekulujeme, že kvadruplexy plní dvojí biologickou funkci, jednu v rámci mobilních elementů a druhou vůči hostitelskému genomu, kde jejich disperzi napomáhá právě mobilní DNA. G4 kvadruplexy jsou zpravidla tvořeny čtyřmi blízkými shluky guaninů v DNA nebo RNA, které vzájemně interagují a vytváří strukturu ze čtyř paralelních nebo antiparalelních vláken, kde vždy čtyři guaniny ze čtyř vláken tvoří kruh, ve kterém guaniny vytváří pomocí Hoogsteenových vazeb guaninovou tetradu. V přednášce popíšeme výpočetní postup založen na prostředí R-Bioconductor, pomocí kterého lze sekvence tvořící takové struktury efektivně vyhledávat a výsledky udržovat ve formátu, který umožňuje jejich další zpracování a vizualizaci. Ukážeme několik způsobů vizualizace rozmístění potenciálních kvadruplexů ve vybraných genomech a jejich asociaci s některými jejich regiony. Zvláště zajímavá je asociace s repetitivními elementy, u člověka například SVA, LINE-1 a HERV, u rostlin pak retrotransposony, jako Copia nebo Gypsy, které mají takové sekvence v LTR. Ukážeme některé pravidelnosti v rozmístění kvadruplexů ve studovaných genomech a budeme spekulovat o jejich biologické funkci.

L8-05

Využití genomických metod pro studium transkripčních faktorů u modelové kvasinky *Schizosaccharomyces pombe*

Převorovský M.^{1,2}, Oravcová M.¹, Rodríguez-López M.², Zach R.¹, Folk P.¹, Půta F.¹, Bähler J.²

¹ *Univerzita Karlova v Praze, Přírodovědecká fakulta, Viničná 5, 128 43 Praha 2*

² *University College London, Department of Genetics, Evolution & Environment and UCL Cancer Institute, Gower Street, London WC1E 6BT, UK*

Poltivá kvasinka *Schizosaccharomyces pombe* představuje oblíbený model pro studium regulace genové exprese a struktury chromatinu. Má kompaktní haploidní genom čítající kolem 5000 protein-kódujících genů a 1500 nekódujících RNA, který byl osekvenován již v roce 2002 a k dipozici je velmi kvalitní anotace. Díky tomu je *S. pombe* velmi vhodným organismem pro aplikaci genomických metod, které přinesly zásadní poznatky např. o regulaci globální buněčné odpovědi na stresové podmínky nebo o biologii heterochromatinu.

Naše laboratoř se věnuje studiu regulace genové exprese v rámci buněčného cyklu a při oxidativním, nutričním a genotoxickém stresu a jejímu vztahu ke struktuře chromatinu. S využitím expresních microarrays a technik CHIP-chip/CHIP-seq jsme identifikovali cílové geny transkripčních faktorů Cbf11 a Cbf12 z evolučně konzervované rodiny CSL (CBF1/Su(H)/Lag-1), která u živočichů řídí vývoj a diferenciaci v rámci signální dráhy Notch. Proteiny CSL u kvasinky *S. pombe* hrají významnou úlohu při správném načasování průchodu buněk buněčným cyklem a zásadním způsobem se podílejí na udržování integrity kvasinkového genomu, pravděpodobně díky jejich vlivu na globální strukturu chromatinu.

L8-06

Bioinformatika v environmentální mikrobiologii

Uhlík O.¹, Strejček M.¹, Musilová L.¹, Šuman J.¹, Wald J.¹, Řídl J.², Hroudová M.², Macek T.¹

¹ *Vysoká škola chemicko-technologická v Praze, Fakulta potravinářské a biochemické technologie, Ústav biochemie a mikrobiologie, Technická 3, 166 28 Praha 6*

² *Ústav molekulární genetiky v.v.i., Akademie věd České republiky, Oddělení genomiky a bioinformatiky, Vídeňská 1083, 142 20 Praha 4*

Jedním z hlavních úkolů environmentální mikrobiologie je porozumět mikrobiální diverzitě, fylogenetické i metabolické, a faktorům, které ji ovlivňují. Donedávna předcházela charakterizaci mikroorganismů vždy jejich izolace z životního prostředí a následná kultivace. Rutinně kultivovat za laboratorních podmínek však lze pouze okolo 1 % mikroorganismů. Mikrobiální diverzita tak začala být v druhé polovině 80. let studována pomocí molekulárně biologických technik založených na studiu buď taxonomicky významných genů (v drtivé většině 16S rRNA geny pro studium prokaryot a 18S rRNA geny pro studium eukaryot) nebo funkčních genů, které byly amplifikovány pomocí PCR a amplikony byly využity pro tvorbu genových knihoven. Do příchodu "next-generation" sekvenování ale nebylo z kapacitních důvodů možné studovat mikrobiální diverzitu v dostatečné hloubce. Při využití pyrosekvenace ampliconů taxonomicky významných či funkčních genů je tento problém odbourán. I tento přístup má ale určitá úskalí – zejména se jedná o nutnost eliminace chyb způsobených metodikou vedoucích k nadhodnocení diverzity. Ke vzniku chyb dochází při polymerasové řetězové reakci jednak chybovostí enzymu DNA-polymerasy a vznikem tzv. chimér. Četnost chybného zařazení nukleotidu do syntetizovaného řetězce může být snížena výběrem DNA-polymerasy s opravnou aktivitou. Eliminace chimér vznikajících během vlastního experimentu ale možná není. Této problematice je věnována velká pozornost, protože chiméry mohou představovat 5 – 45 % z celkového počtu sekvencí. Pokud by nedošlo k eliminaci chimerních sekvencí při analýze, diverzita by byla uměle nadhodnocována. Další komplikovanou částí může být vlastní pyrosekvenace, kde je problematické především čtení homopolymerů. Cílem příspěvku je představit tyto problémy a zároveň prezentovat bioinformatické nástroje, které je nutno použít za účelem získání adekvátního obrazu mikrobiální diverzity.

Poděkování: Autoři studie děkují za finanční podporu poskytnutou MŠMT ČR (projekt LH14004) a GAČR (projekt 13-20414P).

Poster Session

Ceny pro nejlepší postery jsou věnovány
Středoevropským technologickým institutem Ceitec.



CEITEC

středoevropský technologický institut

BRNO | ČESKÁ REPUBLIKA

P-01

Biowes - řešení pro správu a sdílení experimentálních dat a metadat

Císař P.¹, Štys D.¹, Bárta A.¹

¹ *Jihočeská univerzita, Ústav komplexních systémů, Nové Hradky, Zámek 136, 373 33*

Experimentální metadata jsou data o datech, která zachycují všechny informace potřebné k reprodukovatelnosti experimentu. Z pohledu vlastníka metadat je automatické vyhledávání informací experimentálních dat či sdílení dat nedílnou součástí budoucí reprodukovatelnosti experimentu.

Ještě významnější je pohled kognitivní, kdy metadata a jejich projekce do návrhu následného experimentálního postupu odrážejí skutečnou podobu pracovní hypotézy. To pregnančně definoval šéfredaktor časopis *Molecular Cell Biology* Thomas Lemberger ve svém úvodníku z února 2014, který zavádí novou strukturu časopisu [1]. Systém BioWes [2] odráží a formalizuje strukturu kognitivního procesu a archivuje jej v elektronické, strojově zpracovatelné podobě.

Práce se systémem BioWes je organizována jako zápis o provedení experimentu. Tím je myšlena, jak přípravná fáze (definování experimentálních podmínek a popis experimentu), tak samotné měření i popis experimentu a závěrů. Schválené popisy návrhu experimentu jsou v současné době vesměs buď archivovány v papírové podobě, nebo v elektronické podobě, kdy není jednoznačně zajištěn: vztah návrhu experimentálních dat a popisu experimentu. Problémem při návrhu experimentu může být i nedostupnost informací o experimentech prováděných v rámci daného pracoviště. Tyto informace by mohly být využity jednak jako inspirace pro samotný návrh, tak i pro ověření tohoto návrhu. Zkušenost jednotlivých pracovišť dokládá nedostatečný popis experimentů v člancích, kdy z daného popisu není možné dosáhnout reprodukovatelnosti experimentu. Poměrně málo pracovišť využívá pro návrh experimentu statistické metody návrhu experimentu.

V případě předávání zpráv o experimentu pak není zajištěno provázání experimentálních dat, popisu experimentu a poznatků o experimentu. Znalosti jsou centrálně soustředěny pouze u jedné osoby a izolovaně u jednotlivých experimentátorů. Též poskytovatelé veřejných prostředků stále častěji požadují ukládání a zpřístupňování dat.

Standardizace je jedním ze stěžejních bodů provádění biologických experimentů. Přínos standardizace je především v možnosti přesného pojmenování všech významných pojmů a tím zamezení chybné interpretace výsledků. Standard je také základem pro objektivní srovnání experimentů a analýzy dat. V současné době jsou dostupné mnohé databáze

terminologických nebo ontologických standardů. Systém BioWes řadu z nich podporuje a umožňuje z nich jednoduše čerpat. To vše Biowes systém zajišťuje.

Literatura

- [1] Tools of discovery, Thomas Lemberger, Mol. Sys. Biol. 10: 715 (2014)
- [2] <http://www.biowes.org/>

P-02

Dva príklady využitia bioinformatiky v potravinárskom výskume

Brežná B.¹

¹ Národné poľnohospodárske a potravinárske centrum Lužianky, pobočka Výskumný Ústav Potravinársky Bratislava, Priemyselná 4, 82475 Bratislava, Slovenská republika

Využitie bioinformatiky na našom pracovisku bude demonštrované na dvoch problémoch: A) predpovedanie falošnej pozitivity dôkazu mandlí kvôli podobnosti s príbuznými rastlinami, B) vyhodnocovanie nekultivačných stanovení mikroflóry korenia.

Metódy na dôkaz mandlí v potravinárskych výrobkoch slúžia na kontrolu falšovania potravín ako aj na ochranu alergických jedincov, ktorí sa musia spoľchať na pravdivé označovanie výrobkov, vrátane stôp alergénu. V poslednej dekáde bolo vyvinutých viacero takýchto metód, založených na princípe dôkazu DNA, hlavne pomocou PCR. Mandle sú však geneticky veľmi príbuzné kôstkoviciam, ako napríklad broskyňa, marhuľa, slivka, čerešňa alebo višňa. Skúmaním databáz DNA pomocou bioinformatických nástrojov bolo možné predpovedať, ktoré z publikovaných metód vykazujú riziko falošnej pozitivity v prítomnosti kôstkovíc.

Iný set bioinformatických nástrojov je potrebný pri analýze výstupov z novogeneračného sekvenovania DNA (next-generation sequencing). Táto technika bola použitá pri nekultivačnom skúmaní mikrobiálnych komunit vo vzorkách čierneho korenia a petržlenu.

Tento príspevok bol financovaný z kolaboratívneho projektu SPICED č. 312631.

P-03

Aplikácia a webová služba pre funkčnú anotáciu variantov

Budiš J.¹, Hekel R.¹, Minárik G.², Szemes T.²

¹ Fakulta matematiky, fyziky a informatiky, Katedra Informatiky, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika

² Prírodovedecká fakulta, Katedra molekulárnej biológie, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika

Štandardné postupy identifikácie genetických variantov z NGS dát vedú k identifikácii veľkého množstva kandidátnych variantov, z ktorých len malá časť súvisí so skúmaným javom. Nutným krokom pri interpretácii analýz je preto filtrácia a prioritizácia na základe atribútov opisujúcich ich funkčný dopad.

Navrhli sme grafickú aplikáciu Variant Annotation Analyser (VAA), ktorá umožňuje automatizované anotovanie variantov atribútmi z webových databáz. Základným poskytovateľom atribútov je webová služba Variant Annotation Service (VAS), ktorá združuje informácie zo štandardných webových databáz a predikčných nástrojov. Architektúra programu založená na zásuvných moduloch navyše umožňuje jednoduché rozšírenie aplikácie, čo vedie ku komplexnej anotácii umožňujúcej bohaté možnosti na výber kandidátnych pozícií.

P-04

Automatický postup pro klasifikaci lokálních konformací DNA

Automatic workflow for the classification of local DNA conformations

Čech P.¹, Kukul J.², Schneider B.³, Černý J.³, Svozil D.¹

¹ *VŠCHT Praha, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6*

² *ČVUT v Praze, Fakulta jaderná a fyzikálně inženýrská, Trojanova 13, 122 00 Praha 2*

³ *Biotechnologický ústav AV ČR, v. v. i., Vídeňská 1083, 142 00 Praha 4*

Rostoucí množství krystalových a NMR struktur odhalilo významný strukturní polymorfismus DNA na lokální úrovni. V závislosti na sekvenčním okolí jsou dinukleotidové kroky v DNA nezanedbatelně flexibilní a DNA je tak překvapivě variabilní strukturou. Existující klasifikace DNA dinukleotidů [1] je založena z větší části na manuální anotaci, která je jak časově náročná, tak náchylná k chybám. Abychom překonali tato omezení, vyvinuli jsme automatický postup pro klasifikaci konformací DNA dinukleotidů [2] založený na metodách strojového učení. Navržený postup využívá metodu *k*-NN, pomocí níž jsou dinukleotidy buď klasifikovány do jedné z 24 tříd, nebo jsou označeny jako neklasifikované. V takto označených datech lze pak pomocí našeho algoritmu shlukové analýzy najít případné nové třídy zcela automaticky. Uvedený postup jsme použili k analýze 816 X-ray a 664 NMR struktur DNA, a jeho pomocí jsme identifikovali šest nových konformačních tříd. Výsledky ilustrují využití metod strojového učení v klasifikaci lokálních konformací DNA.

Literatura

- [1] Svozil D., et al., DNA conformations and their sequence preferences. *Nucleic Acids Research*, 2008, 36(11):3690-3706
- [2] Čech P., et al., Automatic workflow for the classification of local DNA conformations. *BMC Bioinformatics*, 2013, 14(1)

P-05

Sekvence a anotace genomu půdní bakterie *Streptomyces* sp. W6 indikuje potenciál produkce annimycinu

Genome sequence and annotation of *Streptomyces* sp. W6 natural putative producer of annimycin antibiotic

Chrudimský T.¹, Chroňáková A.¹, Petříčková K.², Petříček M.², Křišťůfek V.¹

¹ *Biologické Centrum AV ČR, v. v. i. - Ústav půdní biologie, Na Sádkách 7, České Budějovice*

² *Mikrobiologický ústav AV ČR, v. v. i. - Laboratoř molekulární biologie aktinomycet, Vídeňská 1083, Praha 4*

Streptomycety reprezentují důležitý a slibný zdroj přirozených látek, které mají uplatnění v léčbě onemocnění. V této práci předkládáme sekvenci genomu půdní bakterie *Streptomyces* sp. W6, která byla izolována z rekultivované plochy po těžbě černého uhlí ve státě Wyoming (USA) pomocí standardních mikrobiologických technik. *Streptomyces* sp. W6 roste na kultivačních médiích M2, GYM a Mueller-Hilton-a je dlouhodobě uložen v depositáři Sbírký půdních aktinomycetů v Českých Budějovicích (<http://www.actinomycetes.cz>). Bakterie vykazuje určité antibiotické účinky proti gram pozitivním bakteriím a pomocí cíleného genového skríninku byl odhalen biosyntetický klastr nesoucí jednotku C5N odvozenou od 5-aminolevulinátové kyseliny. Tento genetický skrínink sloužil pro odhalení nových pravděpodobných producentů sekundárních metabolitů s protizánětlivými aktivitami z přirozeného prostředí. Tyto látky jsou dnes velmi žádané pro využití ve farmacii. Výsledkem bylo potvrzení přítomnosti genů zodpovědných za cyklizaci C5N jednotky, které se objevují v biosyntetických klastrech zodpovědných za syntézu manumycinových antibiotik, známých protizánětlivých metabolitů. Pilotní průzkum sekvence genomu v okolí genového shluku naznačil syntézu látek podobných moenomycinovým antibiotikům. Sekvence genomu byla získána pomocí 454-pyrosequencingu na platformě GS Junior pomocí Titanium chemie a jednotlivá čtení byla složena pomocí software Newbler (Roche/454). Výsledné kontigy byly prozkoumány na přítomnost biosyntetických genových klastrů a jejich anotace indikuje přítomnost annimycinového genového klastru, který byl nedávno identifikován v jiném kmenu *Streptomyces clavus*. Dle sekvence podobnosti 16S rRNA genu, je kmen W6 příbuzný *Streptomyces sioyaensis*.

P-06

Predikce nových agonistů glukokortikoidního receptoru explorací chemického prostoru

Chemical space mining for new glucocorticoid receptor agonists

Čmelo I.^{1,2}, Sedlák D.², Hoksza D.^{1,3}, Bartůněk P.², Svozil D.^{1,2}

¹ *Laboratoř informatiky a chemie, VŠCHT Praha, Technická 5, 166 28 Praha 6*

² *CZ-OPENSSCREEN, Ústav molekulární genetiky AV ČR, v.v.i., Vídeňská 1083, 142 20 Praha 4*

³ *Výzkumná skupina SIRET, Katedra softwarového inženýrství, MFF UK Praha, Malostranské nám. 25, 118 00 Praha 1*

Glukokortikoidní receptor (GR) je transkripční faktor hrající významnou roli při regulaci metabolismu a odezvy imunitního systému. GR patří do rodiny steroidních receptorů, které představují jeden z důležitých molekulárních cílů při léčbě rakoviny. V tomto příspěvku je popsán výpočetní přístup pro vývoj nových ligandů GR založený na exploraci chemického prostoru, která je prováděna našim algoritmem Molpher [1]. Molpher slouží k tvorbě virtuální chemické knihovny pomocí procesu zvaného “molekulové morfování”, který sestává z iterativní aplikace malých strukturních změn na množinu vybraných sloučenin, zde konkrétně známých agonistů GR. Množinu GR agonistů jsme sestavili z aktivních sloučenin pocházejících z databáze ChEMBL a z našich experimentálních HTS dat. Molpherem vytvořenou virtuální chemickou knihovnu potenciálních agonistů GR následně podobnostním vyhledáváním promítneme do databáze komerčně dostupných molekul ZINC [2]. Nejslibnější látky zakoupíme a jejich aktivitu experimentálně ověříme.

Literatura

- [1] Hoksza D., Skoda P., Vorsilak M., Svozil D.: Molpher: a software framework for systematic chemical space exploration. *Journal of Cheminformatics* **6** : 7, 2014
- [2] Irwin J. J., Sterling T., Mysinger M. M., Bolstad E. S. and Coleman R. G., ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **52**(7), 1757-1768, 2012

P-07

Sekvenování chromosomálního ramene 3DS pšenice

Cviková K.¹

¹ *Centrum strukturální a funkční genomiky rostlin, Ústav experimentální botaniky AV ČR, Šlechtitelů 31, 78371 Olomouc*

Pšenice setá je základní potravinou pro přibližně 40 % světové populace, avšak její výnos v závislosti na vnějších faktorech stále klesá. Vyšlechtění nových kultivarů, které by lépe odolávaly biotickým a abiotickým stresům je přitom časově velmi náročné. Znalost sekvence celého pšeničného genomu tuto práci časově zefektivní. Kvalita získané referenční sekvence je ovlivněna složitostí genomu, vybráním vhodné sekvenační strategie a optimálního assembleru. Pšenice setá je allohexaploid a její genom je složen ze 42 chromosomů. Celková velikost genomu je 17 Gbp, což je 6x více než u člověka. Pouze 1-2 % z tohoto genomu je tvořeno samotnými geny, zbytek zahrnuje převážně repetitivní sekvence, které významně komplikují přečtení genomu. Sekvenování genomu pšenice jako celku je velmi obtížné z důvodu nutnosti pracovat s velkým množstvím dat, které by žádný software nebyl schopen kvalitně poskládat. Proto byla pro čtení genomu pšenice zvolena nová strategie, kdy jsou jednotlivé chromosomy pšenice tříděny pomocí průtokové cytometrie. Z každého chromosomu je poté vytvořena DNA knihovna (BAC knihovna), která uchovává v nezměněné formě části chromosomu (BAC klon) o velikosti ~ 100 000 párů bází. Ty jsou na základě fragmentační analýzy uspořádány do větších celků a následně sekvenovány. Sekvence BAC klonů ve formě statisíců 100-250 bp čtení jsou poté skládány do větších celků. Jednotlivé assembly se liší svými postupy, algoritmy a parametry, které je nutné optimalizovat podle analyzovaného druhu a kvality sekvenačních dat. V naší studii jsme použili sekvence z krátkého ramene chromosomu 3DS. Sekvence jednotlivých BAC klonů jsme skládali pomocí tří assemblerů: Edena, ABySS, Masurca.

Tato práce je podporována Grantovou agenturou České republiky (13-8786S) a IGA (PrF-2014-001).

P-08

Štatistický model pre neinvazívnu prenatalnú diagnostiku aneuploidii

Ďuriš F.¹, Budiš J.², Szemes T.¹, Minárik G.¹, Turňa J.¹

¹ *Prírodovedecká fakulta, Katedra molekulárnej biológie, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika*

² *Fakulta matematiky, fyziky a informatiky, Katedra Informatiky, Univerzita Komenského, Mlynská dolina, Bratislava, Slovenská republika*

Recentné štúdie poukázali na využitie NGS technológií pre prenatalnú diagnostiku aneuploidii neinvazívnym spôsobom. Zavedené metodiky dosahujú vysokú sezitivitu a špecificitu pri predikcii trizómie chromozómu 21, stále ale nedosahujú optimálne výsledky pri predikcii trizómií T18 a T13. Štúdie navyše využívajú vysoké pokrytie genómu, čo bráni masovému využitiu diagnostiky v klinickom prostredí.

V našej štúdií navrhujeme štatistický model, na ktorom skúmame vplyv faktorov ako kvalita mapovania, dĺžka fragmentov a GC zloženie na schopnosť diagnostikovať trizómiu. Cieľom našej práce je navrhnúť spoľahlivú metodiku predikcie pri nízkom pokrytí, ktorá by umožnila cenovo dostupnú diagnostiku namiesto zavedených invazívnych metód.

P-09

Studium vlivu polysubstituovaných pyrimidinů na expresi genů spojených se zánětlivou reakcí

Hájek M.¹, Kužmová E.¹, Šmídková M.¹, Kolman V.¹, Zídek Z.², Jansa P.³

¹ Ústav organické chemie a biochemie AV ČR, v.v.i., Flemingovo nám. 2, 166 10 Praha

² Ústav experimentální medicíny AV ČR, v.v.i., Vídeňská 1083, 142 20 Praha

³ Gilead Sciences, Inc, 333 Lakeside Drive, Foster City, CA 94404

Některé nové látky ze skupiny polysubstituovaných pyrimidinů zmírňují průběh experimentálního střevního zánětu vyvolaného dextransulfátem sodným (DSS). V systému *in vitro* bylo prokázáno, že zástupce této skupiny látek, PJ-32A, významně snižuje expresi genů spojených se zánětlivou reakcí prostřednictvím inhibice NF- κ B a JAK2/STAT1 signálních drah. Cílem této práce bylo pomocí microarray analýzy stanovit vliv PJ-32A na expresi klíčových cytokinů a enzymů metabolických drah vedoucích k zánětlivé odpovědi v modelové buněčné linii myších makrofágů J774A.1. Celogenomové expresní profilování bylo provedeno mikročipovou technologií ve firmě Geneti Biotech v Hradci Králové. Vzorky byly hybridizovány na 2 skla Agilent 4x44K v duplikátech s vzájemně opačným barvením. Tímto způsobem byly porovnány expresní profily nestimulovaných makrofágů, makrofágů 4 hodiny po aktivaci LPS/IFN- γ , aktivovaných makrofágů po 4 hodinovém působení 10 μ M PJ-32A a nestimulovaných makrofágů po 4 hodinovém působení 10 μ M PJ-32A. Výsledky byly zpracovány v programovacím prostředí R s využitím softwarového balíčku LIMMA. Diferenciálně exprimované geny byly analyzovány v programu MultiExperiment Viewer 4.9 a dále vyhodnoceny pomocí webové aplikace DAVID Bioinformatics Resources 6.7. Analýzou genové ontologie bylo zjištěno, že v souboru genů, jejichž exprese byla u aktivovaných makrofágů testovanou látkou snížena, jsou významně zastoupeny geny, jejichž produkty se podílí na imunitní odpovědi (GO:0006955), aktivitě cytokinů (GO:0005125), aktivitě chemokinů (GO:0008009) a chemotaxi (GO:0006935). Z hlediska lokalizace jsou mezi nimi výrazně zastoupeny geny, jejichž produkty jsou součástí extracelulárního prostoru (GO:0005615). Z analýzy zapojení produktů inhibovaných genů do buněčných drah vyplynulo, že nejvíce zastoupená je signální dráha spojená s interakcí cytokinů a jejich receptorů. Ze získaných dat budou vybrány zajímavé geny, u kterých bude studován biologický význam jejich regulace zvláště ve vztahu k ovlivnění migrace prozánětlivých leukocytů do místa zánětu. Protizánětlivý účinek PJ-32A bude následně ověřen na experimentálním modelu myši ulcerózní kolitidy imunohistochemickým stanovením míry infiltrace střevní stěny zánětlivými buňkami.

P-10

Zjednodušená implementace umělých neuronových sítí v prognostice přežití pacientů s kolorektálním karcinomem na základě klinických a molekulárně-biologických dat

Hošek P.^{1,2,3}, Pitule P.^{1,2}, Ostašov O.^{1,2}

¹ *Univerzita Karlova v Praze, Lékařská fakulta v Plzni, Biomedicínské centrum, Husova 3, 306 05 Plzeň*

² *Univerzita Karlova v Praze, Lékařská fakulta v Plzni, Ústav histologie a embryologie, Karlovarská 48, 306 05 Plzeň*

³ *České vysoké učení technické v Praze, Fakulta biomedicínského inženýrství, nám. Sítná 3105, 272 01 Kladno 2*

Umělé neuronové sítě jsou typickým příkladem aplikace strojového učení na řadu klasifikačních a regresních úloh. Analýza přežití a z ní vyplývající prognostické modely však vyžadují specifické postupy kvůli přítomnosti cenzorovaných pozorování. Existuje několik zavedených metod trénování neuronových sítí pro predikci přežití na cenzorovaných datech. Většina z nich spočívá v použití neuronové sítě jako nelineárního rozšíření často uplatňovaného Coxova modelu proporcionálních rizik. Přímočarou alternativou k tomuto postupu je zohlednění intervalového charakteru cenzorovaného pozorování při trénování sítě základním algoritmem backpropagation. Neuronová síť tak může být použita k predikci doby přežití přímo, bez použití rizikové funkce.

Na několika sadách dat pacientů s kolorektálním karcinomem z dřívějších studií jsme vytvořili prognostické modely využívající umělé neuronové sítě pro predikci celkového a bezpříznakového přežití. Tyto modely využívají klinická data dostupná pro onkologické pacienty (věk, pohlaví, grading a staging tumoru, aplikace chemoterapie, pooperační komplikace apod.) i cíleně získaná molekulárně-biologická data (expresi vybraných genů a imunohistochemické stanovení markerů nádorových kmenových buněk). Vlastnosti vytvořených modelů jsme srovnali s tradičními způsoby analýzy přežití a Coxovým modelem proporcionálních rizik.

Publikace je spolufinancovaná Evropským sociálním fondem a státním rozpočtem České republiky. Registrační číslo projektu: CZ.1.07/2.3.00/30.0061, GAUK 1268213.

P-11

Skryté retroviry: Hledání nových rodin endogenních retrovirů

Hron T.¹, Elleder D.¹, Pačes J.¹

¹ *Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4*

Endogenní retroviry (ERVs) jsou genetické elementy vyskytující se u většiny obratlovců. K jejich vzniku dochází při integraci exogenního viru do zárodečné buněčné linie, což vede k vertikálnímu přenosu virové DNA do dalších generací jako součást genomu hostitele. Většina ERVs je fixována v populaci po mnoho miliónů let a jejich sekvence je poškozena v důsledku rekombinace a akumulace mutací. Některé, převážně mladší, ERVs si však zachovaly intaktní genetickou informaci a hrají významnou roli v některých buněčných procesech a rozvoji patologií. Všechny retroviry mají teoreticky schopnost endogenizace, nicméně u některých významných virových rodin se doposud nepodařilo najít jejich zástupce v endogenní formě. V nedávné době bylo popsáno hned několik ERVs patřících do rodiny lentivirů, jež byla doposud spojována pouze s exogenními viry. Otázkou tedy je, zda se některé retrovirové rodiny v evolučně fixované endogenní formě skutečně nevyskytují, anebo jsou pouze skryty v záplavě genomických dat a čekají na své odhalení.

Abychom identifikovali nové zástupce ERVs, navrhli jsme algoritmus, pomocí něhož lze plošně prohledávat genomové sekvence obratlovců a kategorizovat nalezené kandidáty podle podobnosti se známými exogenními viry. Takto jsme analyzovali více než 100 veřejně přístupných genomů obratlovců, kde jsme hledali nové zástupce virových rodin lentiviry a deltaretroviry. Podařilo se nám potvrdit výskyt všech dříve popsaných endogenních lentivirů, což dokazuje funkčnost navrhované metody. Tento přístup umožňuje rychlé a efektivní prohledávání objemných genomických dat, jejichž počet v posledních letech stále narůstá.

P-12

Interakčné preferencie v komplexoch proteín - nukleová kyselina

Jakubec D.¹, Vondrášek J.¹

¹ *Ústav organické chemie a biochemie AV ČR, Flemingovo nám. 2, 166 10 Praha 6, Česká Republika*

Cieľom práce je hľadanie výberových pravidiel riadiacich špecifické interakcie medzi bázou nukleovej kyseliny a aminokyselinou. To je realizované vytvorením energetického profilu pre vybrané štatisticky významné reprezentatívne interakčné páry. Ďalej je uvážený význam tohto reprezentatívneho kontaktu jeho lokalizáciou na profile interakčných energií získanom analýzou zhľuku kontaktov s podobnou priestorovou distribúciou. Tieto štatisticky významné zoskupenia sú získané rozborom neredundantnej sady štruktúr z Protein Data Bank. Sada empirických a *ab initio* metód je použitá k výpočtu interakčných energií a ich presnosť je porovnaná. Zhromaždené data sú integrované do jednotnej databázy zahŕňajúcej tak komplexy proteín - nukleová kyselina, ako aj výsledky skoršej práce na komplexoch proteín - proteín.

P-13

Klasifikace metagenomických vzorků pomocí diskriminativních nadřetězců DNA

Jalovec K.¹, Železný F.¹

¹ *České vysoké učení technické, Katedra počítačů, Karlovo nám. 13, 121 35 Praha 2*

S rostoucím množstvím dat získaných metodami NGS roste i potřeba tato data efektivně vyhodnocovat. Tato práce se zaměřuje na binární klasifikaci metagenomických vzorků získaných metodami NGS.

Metagenomické vzorky sestávají z velkého množství krátkých řetězců DNA, které mohou patřit rozličným organismům vyskytujícím se v prostředí, ze kterého byl daný vzorek odebrán. Funkce prostředí může být ovlivněna kontaminací organismy, které se běžně v tomto prostředí nevyskytují. Pak vzorek obsahuje i genomické řetězce těchto kontaminantů. Cílem této práce je navrhnout metodu, která je schopna sestavit klasifikátor či klasifikátory ve formě množiny nadřetězců DNA. Tyto klasifikátory mohou být použity pro určení, zda jsou nově odebrané vzorky kontaminované či nikoliv bez nutnosti identifikace jednotlivých organismů ve vzorku. Sestavení takovýchto klasifikátorů koresponduje s úlohou nalezení konzistentního nadřetězce.

Námi navržený postup sestává v sestrojení grafu překryvů z řetězců DNA obsažených v odebraných vzorcích. Klasifikátory (diskriminativní nadřetězce) je potom možno vyhledávat jako cesty v grafu. Klasifikátor vznikne zřetěžením uzlů (řetězců DNA) vyskytujících se na cestě v grafu. Cílem je nalézt takové cesty v grafu, které nejlépe rozlišují pozitivní (kontaminované) vzorky od negativních.

P-14

SigHunt: Nová metoda pro hledání kandidátních genomických ostrovů v DNA sekvencích

Jaron K. S.¹, Moravec J. C.², Martínková N.^{1,2}

¹ Masarykova univerzita, Institut biostatistiky a analýz, Kamenice 3, 625 00 Brno

² Akademie věd ČR, Ústav biologie obratlovců, v.v.i., Květná 8, 603 65 Brno

Genomické ostrovy vznikají horizontálním přenosem genů a jsou v současnosti jedněmi z nejčastěji zkoumaných genomických úseku. Význam horizontálního přenosu genů na evoluci prokaryot byl v minulosti popsán, teprve však nedávno byli objeveny genomické ostrovy nesoucí klíčové geny pro životní styl eukaryotních organismů, jako například u plísní *Aspergillus* nebo *Pyrenophora*. Představujeme zde první metodu optimalizovanou pro detekci genomických ostrovů v eukaryotických genomech a nekompletních genomických sekvencích. SigHunt je založený na variabilitě tetranukleotidů mezi organismy, ale zároveň bere v úvahu variabilitu uvnitř sekvence. Lokální odchylky od domovské signatury SigHunt hodnotí podle toho, nakolik se frekvence tetranukleotidu v krátké sekvenci DNA odlišuje od lokální hustoty jeho frekvence. Metoda byla testována na sekvencích s uměle vloženými genomickými ostrovy v rámci široké škály organismů. SigHunt prokázal vysokou sensitivitu a specifitu určenou pomocí ROC (průměrná plocha pod křivkou byla 0.77). Při testování na biologických datech, sestávající z publikovaných a fylogeneticky verifikovaných ostrovů u rodů *Aspergillus*, *Saccharomyces*, *Pyrenophora*, *Galdieria* a *Cryptosporidium*, byl SigHunt schopen detekovat bezmála 70% popsáných genomických ostrovů v krátkém výpočetním čase. Tato metoda byla vytvořena s ohledem na velikost a heterogenitu eukaryotických genomů. Díky své výpočetní nenáročnosti umožňuje zpracovávat celé eukaryotické genomy na osobních počítačích a v relativně krátké době naleznout kandidátní genomické ostrovy pro následnou verifikaci.

P-15

Predikcia mitochondriálnych adresných štiepných sekvencií v *Euglena gracilis* a Trypanosomatidách

Krnáčová K.¹, Vesteg M.^{2,3}, Horváth A.¹

¹ Univerzita Komenského v Bratislave, Prírodovedecká Fakulta, Katedra biochémie, Mlynská dolina CH1, 842 15 Bratislava 4, Slovenská republika

² Ostravská univerzita v Ostravě, Katedra biologie a ekologie, Přírodovědecká fakulta, Chittussiho 10, 710 00 Ostrava

³ Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra genetiky, Mlynská dolina B1, 842 15 Bratislava 4, Slovenská republika

Kmeň Euglenozoa spájajúci tak odlišné druhy ako sú *Euglena gracilis*, *Trypanosoma brucei*, *Trypanosoma cruzi* a *Leishmania major* tvorí jednu z najstarších eukaryotických vetiev. Väčšina mitochondriálnych proteínov tohto taxónu je rovnako ako v ostatných eukaryotoch kódovaná jadrom, no ich mitochondriálny proteínový importný systém je pravdepodobne odlišný. V prezentovanej práci sme sa zamerali na mitochondriálne adresné štiepne sekvencie (CTS) jadrom kódovaných preproteínov participujúcich na procesoch oxidačnej fosforylácie, Krebsovho cyklu, RNA editigu, syntézy Fe-S klastrov a syntézy tetrapyrolov a ich *in silico* analýze. Práca, na základe dvoch počítačových programov, predikuje CTS a v ďalších krokoch analyzuje ich biochemické vlastnosti. Súčasťou projektu je aj hľadanie spoločných motívov vyskytujúcich sa v CTS spomínaných organizmov. Naše *in silico* predikcie sú v súlade s niektorými experimentálnymi výsledkami prác zaoberajúcich sa CTS v týchto organizmoch, no nie je ich veľa a tak naša práca dopĺňa obraz o CTS Euglenozoa.

P-16

Integrace klinických a biomedicínských dat v projektu THALAMOSS

Lexa M.¹, Štefanič F.¹¹ *Masarykova univerzita, Fakulta informatiky; Botanická 68a; 602 00 Brno*

Od roku 2012 je UVT a FI Masarykovy univerzity zapojena do mezinárodního projektu 7.RP Thalamoss [1]. Tento projekt shromáždí velké množství klinických a biomedicínských dat o pacientech (převážně z Evropy) trpících beta-thalassemií, genetickou poruchou tvorby funkčního hemoglobinu beta (gen HBB). Jedním z cílů projektu je zlepšit medicínské postupy s využitím molekulárních přístupů a shromážděných dat, například lepší diferenciací pacientů pro různá doporučení (stratifikace pacientů), či genovou terapií. Ve svém výzkumu se zaměřujeme na bioinformatické zpracování těchto dat, zejména pak problematiku integrace informací z heterogenních zdrojů. V tomto okruhu problémů jsme se dosud věnovali nebo se začínáme věnovat:

- a) využití anotace lidského genomu za účelem filtrace polymorfismů pro GWAS studie [2,3],
- b) využití statistických, pravděpodobnostních modelů a strojového učení za účelem stratifikace pacientů s beta-thalassemií podle volitelných kritérií [4],
- c) integrace klinických a biomedicínských dat za účelem vizualizace struktury populace pacientů a jejich biomedicínských charakteristik.

V naší prezentaci budeme informovat o dosažených výsledcích a studovaných postupech formou posteru.

Literatura

- [1] Internetová prezentace projektu Thalamoss: <http://www.thalamoss.eu>
- [2] Štefanič S, Lexa M (2013). Generovanie simulovaných testovacích dát pre genómové asociačné štúdie. ITAT 2013, Donovaly, Slovensko.
- [3] Lexa M, Štefanič S (2014). The possibilities of using biological knowledge for filtering pairs of SNPs in GWAS studies: an exploratory study on public protein-interaction and pathway data. Bioinformatics 2014, Angers, France.
- [4] WWW stránka se statistickými modely pro stratifikaci: <http://helix.fi.muni.cz/thalamoss>

P-17

ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat

Jindřich J.¹, Müller T.¹, Škuta C.^{1,2}, Sedlák D.¹, Pombinho A.¹, Svozil D.^{1,2}, Bartůněk P.¹

¹ *CZ-OPENSOURCE: Národní infrastruktura chemické biologie, Ústav molekulární genetiky AV ČR, v.v.i, Vídeňská 1083, 142 20 Praha 4*

² *VŠCHT v Praze, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6*

ChemGenDB je databázový/LIMS systém původně vyvinutý pro Centrum chemické genetiky. Je to webová aplikace využívající převážně programovací jazyk Python. Dále využívá následující open source a volně dostupné softwarové komponenty: webový/databázový framework Django, databázový systém MySQL pro úschovu dat a cheminformatický framework RDKit, jenž slouží ke konverzi chemických strukturních formátů, substrukturní a podobnostní vyhledávání a generování 2D struktur pro zobrazení ve webovém prohlížeči. Webové prohlížeče používané pro přístup k LIMS musí podporovat JavaScript a HTML 5. Pro GUI funkce jsou využívány JavaScript knihovny jQuery, jQueryUI a KineticJS. Pro editaci chemických struktur je také používán Marvin for JavaScript.

Doposud byly implementovány následující funkce: import informací o chemických vzorcích (včetně struktur), substrukturní a podobnostní vyhledávání, generování destiček pro chemické vzorky (96, 384 a 1536 jamkové destičky), reformáty destiček (kopírování, ředění, Z-reformát), tvorba a tisk čárových kódů destiček pro chemické vzorky a vzorky pro biologické testy, sofistikované zadávání dat pro screening včetně vizualizačních a analytických nástrojů, tvorba destiček pro testování a automatický upload dat s výsledky testů na server.

P-18

Analýza taxonomicky jednoduché bakteriální komunity z prostředí hlubinného dolu s využitím metagenomiky a bioinformatiky

Rídl J.¹, Falteisek L.², Hroudová M.¹, Pačes J.¹, Strnad H.¹, Vlček Č.¹

¹ Ústav molekulární genetiky, AV ČR, v.v.i., Vídeňská 1083, Praha

² Katedra zoologie, PřF UK, Viničná 7, Praha

V extrémně kyselém prostředí bývalého dolu ve Zlatých Horách vznikají vlivem metabolické aktivity bakterií duté krápníkovité útvary s obsahem mikrobiálního biofilmu. Za účelem studia příslušného bakteriálního konsorcia byla ze vzorku biofilmu izolována DNA a RNA, následně sekvenovaná s využitím genomových sekvenátorů GS FLX, Roche a MiSeq, Illumina. Pro detailnější taxonomickou analýzu byl pomocí PCR amplifikován úsek genu 16S rRNA a výsledné amplikony byly sekvenovány na GS FLX. Ze sekvenačních dat vyplývá, že jde o velmi jednoduché konsorcium s dominantně naboženými třemi kmeny bakterií: *Ferroplasma myxofaciens*, *Acidithiobacillus ferrivorans* a dosud necharakterizované bakterie. V příspěvku budou diskutovány přístupy ke skládání, mapování a anotaci získaných metagenomických a metatranscriptomových dat a jejich využití pro popis této přirozené jednoduché bakteriální komunity.

P-19

Analýza genomu *Pseudomonas fluorescens* ICT v souvislosti s jejím bioremediačním potenciálem

Sakmaryová I.¹, Martinčová M.¹, Ševců A.¹

¹ *Technická Univerzita v Liberci, Ústav pro nanomateriály, pokročilé technologie a inovace, Bendlova 7, Liberec*

Dekontaminace životního prostředí pomocí částic nulmocného nanoželeza v souvislosti s použitím mikrobiálních společenstev patří mezi nadějně a stále častěji používané remediační technologie. Nicméně jejím omezením je nedostatek znalostí o bakteriálním společenstvu podílejícím se na degradaci polutantů. Tento fakt významně omezuje její daleko širší použití. Molekulárně genetická analýza může pomoci odkrýt bioremediační potenciál, tzn. enzymatické aparáty, jednotlivých bakteriálních kmenů. Z tohoto důvodu jsme provedli sekvenaci genomu bakterie *Pseudomonas fluorescens* ICT, která byla izolována z kontaminované lokality a vykazovala flexibilní katabolismus širokého spektra polutantů.

Genom *P. fluorescens* ICT byl sekvenován metodou 454 - pyrosekvenování. Celkově jsme získali 121 Mb ve 2 sekvenačních bězích na platformě GS Junior System. Získaná data vedla ke konstrukci 73 kontigů s průměrnou délkou 99,068 pb. Anotace získaných sekvencí byla provedena pomocí MG-RAST and Blast2Go serverů. Předpokládaná délka genomu byla stanovena na 7.2 MB. Z celkové genomické sekvence bylo 97.2% proteinů anotováno pomocí MG-RAST z čehož 91.2% bylo zařazeno do funkčních kategorií. Geneticky nejbližším organismem byl určen kmen *P. fluorescens* Pf-5 (Accession No. CP000076) s 96% shodou v genomické sekvenci s E hodnotou nižší než 10^{-30} . Mezi nejzajímavější výsledky vyplývající z analýzy genomu řadíme nálezy 18 genů zapojených do degradace bifenyly, 10 zapojených do degradace chlorovaných aromatických sloučenin a 9 zapojených do degradace naftalenu a antracenu.

Podrobnější analýza genomu testovaného bakteriálního kmene povede nejen k vysvětlení bioremediačního potenciálu, ale i nutričních požadavků, neboli k osvětlení faktorů nejvíce ovlivňujících jeho použitelnost pro bioremediace půdního ekosystému. Komplexní přístup při dekontaminaci životního prostředí zahrnující použití nanotechnologií ve spojení s přesně směřovanou mikrobiální degradací podpořenou znalostí genetických vlastností použitých kmenů může efektivně zrychlit remediační proces.

Výsledky tohoto projektu LO1201 byly získány za finančního přispění Ministerstva školství, mládeže a tělovýchovy v rámci účelové podpory programu "Národní program udržitelnosti I" a projektu OP VaVpl Centrum pro nanomateriály, pokročilé technologie

a inovace CZ.1.05/2.1.00/01.0005 a zároveň Ministerstva průmyslu a obchodu v rámci projektu FR-TI3/564.

P-20

Nové modulatory steroidních receptorů identifikované profilováním s vysokou propustností na panelu reportérových buněčných linií

Novel steroid receptor modulators identified by high-throughput profiling with panel of reporter cell lines

Sedlák D.¹, Škuta C.^{1,2}, Svozil D.^{1,2}, Bartůněk P.¹

¹ *CZ-OPENSREEN: National Infrastructure for Chemical Biology, Institute of Molecular Genetics AS CR, Vídeňská 1083, 14220 Prague, Czech Republic*

² *Laboratory of Informatics and Chemistry, Institute of Chemical Technology, Technická 5, 166 28 Prague, Czech Republic.*

Steroidní receptory jsou evolučně i funkčně příbuzné transkripční faktory regulující mnohé fyziologické funkce organismu. Jejich rodina sestává ze dvou receptorů pro estrogeny (ER α a ER β) a z androgenního (AR), glukokortikoidního (GR), mineralokortikoidního (MR) a progesteronového (PR) receptoru. Ačkoliv je regulace transkripce za pomoci steroidních receptorů velmi komplexní proces, vykazuje jeden společný rys. Aktivitu steroidních receptorů lze modulovat nízkomolekulárními ligandy, což z těchto receptorů činí významný molekulový cíl pro léčbu závažných lidských onemocnění, jako jsou např. různá onkologická, metabolická, neurodegenerativní nebo zánětlivá onemocnění.

Současné snahy ve výzkumu terapeutického využití steroidních receptorů se soustředí jak na identifikaci nových aktivních ligandů, tak, u ligandů stávajících, na vylepšení jejich funkce, selektivity či módu působení. V tomto příspěvku představíme práci, ve které jsme využili naši unikátní technologickou a informatickou platformu pro profilování steroidních receptorů. Nejdůležitější součástí této platformy je kolekce selektivních luciferasových reportérových buněčných linií vytvořených na unifikovaném buněčném pozadí v U2OS buňkách. Tato kolekce sestává ze dvou panelů. První panel je založen na expresi chimérického proteinu obsahujícího DNA vazebnou doménu přenesenou z kvasinkového transkripčního faktoru Gal4. Tento systém je obzvláště vhodný pro primární screeny, protože je méně senzitivní k aktivitám dalších buněčných signálních drah. Ve druhém panelu je aktivace steroidního responzivního elementu v promotoru reportérového vektoru zprostředkována plnodelkovým steroidním receptorem. Takový systém poskytuje, ve srovnání s prvním panelem, kvalitnější informace o ligandy indukované transaktivaci steroidními receptory.

V naší práci byla systematicky profilována chemická knihovna obsahující 3000+ unikátních steroidních derivátů získaných z Ústavu organické chemie a biochemie v

Praze. Každý ze steroidních receptorů byl testován ve třech koncentracích (1 μ M, 100 nM, 10 nM) v agonistickém i antagonistickém módu. Data byla normalizována a analyzována pomocí shlukové analýzy za použití InChlibu, interaktivní knihovny pro tvorbu “cluster heatmap”. Téměř třetina (cca 1000 sloučenin) celé steroidní knihovny je aktivní nejméně na jednom steroidním receptoru. Analýza vztahů mezi strukturami a aktivitami odhalila překvapivě komplexní aktivitní profily. Některé látky byly např. velmi selektivní pro daný receptor, tj. byly aktivní v celém koncentračním rozsahu pouze na tomto receptoru. V jiných případech však byly pozorovány významné aktivity rozprostřené přes několik receptorů. Jedním z takových příkladů jsou aktivitní profily sestávající z agonismu pro ER β a antagonismu pro ER α .

Tato studie demonstruje silný potenciál skrytý v datech obsahujících bohatou informaci o vztahu struktur a aktivit nejen na jednom vybraném receptoru, ale na kompletní receptorové rodině v různých módech působení (agonistický či antagonistický mód). Ligandy, vykazující specifické kombinace aktivit, mohou být pak dále využitelné např. při léčbě nádorových onemocnění, např. v karcinomu prsu nebo prostaty.

P-21

Generická reprezentace chemických sloučenin

Škoda P.¹, Hoksza D.¹

¹ *Univerzita Karlova v Praze, MFF, Katedra softwarového inženýrství, Malostranské nám. 25, 118 00 Praha, Česká republika*

Standardní součástí procesu vývoje léčiv je screening chemických knihoven. Vzhledem k velikosti stávajících knihoven může být klasický screening velice nákladný a nabízí se alternativa v podobě screeningu virtuálního. Dnes již standardní metodou je tzv. ligand-based virtuální screening. Základem tohoto přístupu je využití podobnosti screenovaných sloučenin k sloučeninám s již známou aktivitou vzhledem k danému cíli. Na výsledky virtuálního screeningu má vliv nejen použitá podobnostní metoda ale také zvolená reprezentace chemických sloučenin.

V našem příspěvku představujeme novou reprezentaci chemických sloučenin založenou na fingerprintech. Naše reprezentace využívá fragmentů chemické sloučeniny k její reprezentaci jakožto celku. Každý fragment je reprezentován fyzikálně-chemickými vlastnostmi. Reprezentace je vysoce parametrizovatelná a to zejména v oblasti výběru fyzikálně-chemických vlastností a jejich aplikace. Pro otestování naší reprezentace jsme využili existujícího frameworku pro benchmark virtuálních screeningu. Předběžné výsledky ukazují, že naše metoda je i se základní parametrizací srovnatelná s nejlepšími současnými přístupy k reprezentaci chemických sloučenin.

P-22

Mining chemogenomického prostoru za účelem nalezení vztahů mezi biologickými cíli

Discovering target relations through chemogenomic space mining

Škuta C.^{1,2}, Tetko I.³, Bartůněk P.², Svozil D.^{1,2}

¹ *Laboratoř informatiky a chemie, VŠCHT Praha, Technická 5, 166 28 Praha 6*

² *Oddělení buněčné diferenciaci, Ústav molekulární genetiky AV ČR, v. v. i., Videňská 1083, 142 20 Praha 4*

³ *Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 857 64 Neuherberg*

Chemogenomika se jako obor snaží pokrýt vliv všech možných ligandů na všechny biologické cíle. [1] Tento záměr je daleko ke svému naplnění a patrně ani nikdy naplněn nebude. Velké podмноžiny chemogenomického prostoru jsou však dostupné ve formě databází bioaktivních molekul jako jsou např. *ChEMBL*, *BindingDB* nebo *PubChem BioAssay*. Tento prostor si můžeme představit jako matici s ligandy v řádcích a biologickými cíli ve sloupcích, kde je každý průnik ligand/cíl reprezentován hodnotou vazné afinity (např. *IC50*, *EC50*, *Ki*).

Princip podobnosti molekul říká, že podobné molekuly mají sklon vykazovat podobné vlastnosti. [2] Z toho lze odvodit, že se také podobně chovají a váží se na podobné cíle. [1] Na tomto základě můžeme hodnoty v chemogenomické matici považovat za biologické deskriptory ligandů. Na rozdíl od klasických deskriptorů však netrpí nedostatky striktně strukturního přístupu molekulové podobnosti. Metodu afinitních fingerprintů, jak se popisuje ligandů na základě změřené vazné potence říká, představil *Kauvar a spol.* v roce 1995. [3,4] Afinitní fingerprinty demonstroval použitím experimentálně změřených hodnot pro referenční množinu proteinů.

Chemogenomická matice však zachycuje nejen biologické vztahy mezi ligandy v řádcích, ale také mezi biologickými cíli ve sloupcích, které tvoří afinitní fingerprinty cílů. [5] Dva cíle jsou tak považovány za podobné, pokud tvoří komplex se stejnými ligandy s podobnou vaznou afinitou. Velikost a malá hustota hodnot v chemogenomickém prostoru mohou však být důsledkem značných odchylek při určování podobnosti cílů. [6] Tomu se snažíme předcházet QSAR modelováním jednotlivých cílů a vyplňováním prostoru *in silico* získanými hodnotami afinit. Vytvořené *in vitro/in silico* afinitní fingerprinty biologických cílů mohou být použity např. při určování vedlejších účinků bioaktivních látek nebo hledání nových cílů pro známé ligandy.

Literatura

- [1] Klabunde, T., Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *British Journal of Pharmacology*, 2007. 152(1): p. 5-7
- [2] Maggiora, G.M. and M.A. Johnson, Introduction to Similarity in Chemistry. *Concepts and Applications of Molecular Similarity*, 1990: p. 1-13
- [3] Kauvar, L.M., et al., Predicting Ligand-Binding to Proteins by Affinity Fingerprinting. *Chemistry & Biology*, 1995. 2(2): p. 107-118
- [4] Kauvar, L.M., et al., Protein affinity map of chemical space. *Journal of Chromatography B*, 1998. 715(1): p. 93-102
- [5] Frye, S.V., Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chemistry & Biology*, 1999. 6(1): p. R3-R7
- [6] Vieth, M., et al., Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 2004. 1697(1-2): p. 243-257

P-23

HotSpot Wizard 2.0: automatický webový nástroj pro proteinové inženýrství

Štourač J.^{1,2}, Bendl J.^{1,2,3}, Šebestová E.¹, Pavelka A.⁴, Brezovský J.¹, Damborský J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Centrum biomolekulárního a buněčného inženýrství, Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, Brno 612 66*

⁴ *Fakulta informatiky, Masarykova univerzita, Botanická 68a, 602 00 Brno*

HotSpot Wizard je webový server poskytující anotaci důležitých aminokyselinových zbytků v proteinových strukturách, jejichž mutagenézí lze dosáhnout změny substrátové specifity, enzymové aktivity nebo enantioselektivity. Vhodné pozice pro mutagenézi jsou vyhodnocovány pomocí kombinace strukturních, funkčních a evolučních informací získaných z internetových databází a výpočetních nástrojů. Výsledky analýzy jsou seřazeny podle mutovatelnosti a vizualizovány v proteinové struktuře. HotSpot Wizard nachází využití v komunitě proteinových inženýrů, kteří modifikují vlastnosti proteinů s pomocí místně cílené mutagenéze nebo řízené evoluce. Aktuální verze nástroje je použitelná jen pro proteiny s dostupnou trojrozměrnou strukturou a návrh mutagenéze se omezuje na pozici, nikoliv konkrétní substituci. Nová verze řeší tyto limity: (i) integraci nástroje pro homologní modelování s automatizovaným výběrem vhodného templátu, (ii) predikci změny funkce v důsledku mutace s vyloučením škodlivých mutací a (iii) predikci stabilizujících mutací pomocí „back-to-consensus“ přístupů. Samozřejmostí je také implementace nových verzí nástrojů a databází, jako i nahrazení mnoha nástrojů jejich výkonnější a přesnější alternativou. V neposlední řadě je vyvinuto nové webové rozhraní, které díky větší interaktivitě nabízí uživateli pohodlnější přípravu analýzy a vyhodnocení výsledků. Cílem nové verze je poskytnout širokému spektru uživatelů plně automatickou analýzu proteinových sekvencí a struktur.

P-24

Strategie analýzy pyrotagů funkčních genů

Strejček M.¹, Macek T.¹, Uhlík O.¹

¹ *Vysoká škola chemicko-technologická v Praze, Ústav biochemie a mikrobiologie, Technická 3, Praha 6*

Vysokokapacitní sekvenace 16S rRNA genů se stala standardní technikou pro stanovení prokaryotní diversity. Existuje mnoho postupů zpracování těchto sekvenčních dat, které jsou často založeny na shlukování sekvencí do definovaných operačních taxonomických jednotek (*angl.* OTU). Tyto techniky dobře maskují chyby odvozené od DNA-polymerasy zajišťující amplifikaci markerového genu a jedná se především o substituce. V případě pyrosekvenace hrají důležitou roli i delece a inzerce (indely), které v protein-kódujících sekvencích způsobují kritický posun čtecího rámce.

Cíl práce bylo porovnání několika algoritmů navržených pro korekci amplikonových pyrotagů – genových markerů amplifikovaných 454 pyrosekvenací – a jejich schopnosti redukce posunů čtecích rámců. Jako modelové funkční geny byly zvoleny *bphA* a *benA* kódující bifenylyl- a benzoátdioxygenasy. Výsledkem práce je navržení optimálního postupu pro zpracování pyrotagů funkčních genů.

Poděkování: Autoři studie děkují za finanční podporu poskytnutou MŠMT ČR (projekt LH14004) a GAČR (projekt 14-32432S).

P-25

Genomový projekt *Rhodococcus erythropolis* CCM2595

Strnad H.¹, Kolář M.¹, Pátek M.², Fousek J.¹, Ulbrich P.³, Nešvera J.², Pačes V.¹, Vlček Č.¹

¹ Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Mikrobiologický ústav AV ČR, v.v.i, Vídeňská 1083, 142 20 Praha 4

³ Vysoká škola chemicko-technologická v Praze, Ústav biochemie a mikrobiologie, Technická 5, 166 28 Praha 6

Bakterie *Rhodococcus erythropolis* CCM2595 je studována pro své bioremediační schopnosti jako je například degradace fenolu, katecholu, resorcinolu, hydroxybenzoátu, hydrochinonu, p-chlorofenolu, p-nitrofenolu, pyrimidinů a sterolů.

Genom *R. erythropolis* CCM2595 byl sekvenován za použití 454 technologie. Párová i nepárová sekvenční čtení (500k) byla složena v programu Newbler (2.5) do 44 kontigů se střední délkou 375 kb (N_{50}), které byly uspořádány do dvou scaffoldů. Kompletní genomová sekvence byla připravena uzavřením sekvenčních mezer v programu Consed bez nutnosti další cílené sekvenace. Genom sestává z jednoho chromosomu (6,28 Mb, gb: CP003761) a plasmidu pRECF1 (90 kb, gb: CP003762). Oba replikony mají relativně vysoký obsah GC (62,5%).

V genomu byly identifikovány oblasti kódující proteiny (programy: Critica, Prodigal a Glimmer) a geny kódující transferové RNA (programy: Aragorn, tRNAscan). Program RNAmmer byl použit k nalezení genů pro ribosomální RNA a nekódující RNA (ncRNA). Anotace genů byla provedena různými metodami s preferencí pojmenování dle PGAAP (www.ncbi.nlm.nih.gov/genome/). Celkové zpracování anotace bylo provedeno v prostředí Artemis. Finální anotace se skládá z 5830 kódujících oblastí, 12 rRNA genů, 53 tRNA genů, jednoho tmRNA genu a 5 nekódujících RNA genů (ncRNA).

Financováno: GAČR: 13-28283S, MŠMT: RVO68378050, 2B08062, RVO 61388971.

P-26

Transkriptomická analýza oddenků pýru plazivého v rané fázi dehydratačního stresu

Svoboda P.¹, Janská A.¹, Spiwok V.², Kučera L.¹, Ovesná J.¹

¹ *Tým molekulární genetiky, Výzkumný ústav rostlinné výroby, v.v.i., Drnovská 507, 166 00 Praha*

² *Oddělení biochemie a mikrobiologie, Fakulta potravinářské a biochemické technologie, Vysoká škola chemicko-technologická, Technická 5, 166 28 Praha*

Sucho je závažným problémem současného zemědělství s prognózou eskalace do budoucna. Jedním z předpokladů zajištění potravinové bezpečnosti při předpokládaném nárůstu lidské populace bude vývin kultivarů hospodářských plodin, jež budou schopny zachovat si životaschopnost i v nepříznivých podmínkách při minimální výnosové penalizaci. Cílená konstrukce odolných genotypů vyžaduje poznání mechanismů, jež rostlinám umožňují se na stres adaptovat. Výborným prostředkem k dosažení tohoto cíle je expresní profilování s využitím DNA čipů. Tato technologie se v minulosti ukázala jako nanejvýš vhodná pro studium konsekvencí dehydratačního stresu u rozličných kulturních rostlin. Méně početné jsou experimenty, jež hodnotily dopad sucha na plané druhy rostlin. Tyto plané genotypy se však často vyznačují vyšší odolností vůči nepříznivým podmínkám a představují tak cenný zdroj genetické variability pro kulturní druhy. Jedním z takovýchto genotypů je i pýr plazivý (*Elytrigia repens* (L.) Nevski), jenž je závažným plevelným druhem jak v národním, tak i globálním měřítku. V půdě pýr setrvává ve formě článkovaných oddenků, jež mají obrovský regenerační potenciál a jsou rovněž odolné vůči vysychání, což z nich činí výborný objekt pro studium adaptačních mechanismů rostlin na stres suchem. V tomto experimentu jsme se proto zaměřili na transkriptomickou analýzu oddenků pýru plazivého v průběhu raného stádia dehydratačního stresu. S využitím DNA čipů byla porovnáována odezva nestresovaných oddenků a oddenků po expozici 1, 2 a 3 hodinám stresu suchem. Jelikož dosud neexistuje specializovaný čip pro pýr plazivý, byla použita vhodná substitute, a to ve formě DNA čipu ječmene (Affymetrix 22 K Barley1 GeneChip Genome Array), který byl vybrán s ohledem na sestavu genomu (StStH) *E. repens*. Genom H tohoto allohexaploidního druhu (2n=42) má nejbliže k r. *Hordeum*, což umožňuje využití molekulárních postupů uplatňovaných při studiu ječmene i pro studium pýru plazivého.

P-27

Vyhledávání, porovnávání, validace a charakterizace strukturních motivů v rámci biomakromolekul

Svobodová Vařeková R.¹, Sehnal D.¹, Pravda L.¹, Ionescu C.-M.¹, Wimmerová M.¹, Koča J.¹

¹ *National Centre for Biomolecular Research and CEITEC - Central European Institute of Technology, Masaryk University, 625 00 Brno, Czech Republic*

V současné době máme k dispozici velké množství informací o 3D struktuře biomakromolekul. Toto množství dat poskytuje příležitost k analýze strukturních motivů (fragmentů) uvedených látek – např. vazebných míst, elementů sekundární struktury, ligandů, kavit a tunelů. Dané analýzy lze využít k identifikaci vzorů pro vývoj léků, k porozumění vztahů mezi strukturou proteinu a jeho funkcí, pro klasifikaci proteinů apod. [1-2]. Proto byly pro vyhledávání strukturních motivů a jejich analýzu vytvořeny nové metodiky a softwarové nástroje.

V prezentaci představíme nově vyvinuté nástroje pro zpracování strukturních motivů biomakromolekul. Motivy mohou být popsány pomocí jednoduchého a robustního jazyka MotiveQuery a na základě tohoto popisu pak automaticky vyhledány v Protein Data Bank pomocí MotiveQuery Serveru. Další nástroj, SiteBinder [3], je schopen přiložit tisíce proteinových strukturních motivů ve velmi krátkém čase. Softwarový nástroj Mole [4, 5], je určen pro vyhledávání tunelů. MotiveValidator je pak schopen určit, zda se testovaný ligand shoduje s referenční molekulou (má stejnou topologii i chiralitu). Poslední nástroj, ChargeCalculator, poskytuje hodnoty empirických nábojů vypočítaných pro vstupní molekulu. Využití nástrojů je ukázáno na případových studiích, zaměřených na biologicky významné proteinové strukturní motivy.

Literatura

- [1] Baran I, Svobodová Vařeková R, Parthasarathi L, Suchomel S, Casey F, Shields DC: Identification of potential small molecule peptidomimetics similar to motifs in proteins, *J Chem Inf Model* 2007, 47, 464-474
- [2] Watson JD, Laskowski RA, Thornton JM: Predicting protein function from sequence and structural data, *Curr Opin Struct Biol* 2005, 15, 275-284
- [3] Sehnal D, Svobodová Vařeková R, Huber HJ, Geidl S, Ionescu CM, Wimmerová M, Koča J: SiteBinder – an improved approach for comparing multiple protein structural motifs, *J Chem Inf Model* 2012, 52, 2, 343–359
- [4] Berka K, Hanák O, Sehnal D, Banás P, Navrátilová V, Jaiswal D, Ionescu CM, Svobodová Vařeková R, Koca J and Otyepka M: MOLEonline 2.0: interactive

- web-based analysis of biomacromolecular channels. *Nucleic Acids Res* 2012, 40, W222–7
- [5] Sehnal D, Svobodová Vařeková R, Berka K, Pravda L, Navrátilová V, Banáš P, Ionescu CM, Otyepka M and Koča J: MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminform* 2007, 5, 39

P-28

InCHlib – interaktivní shluková heatmapa pro webové aplikace

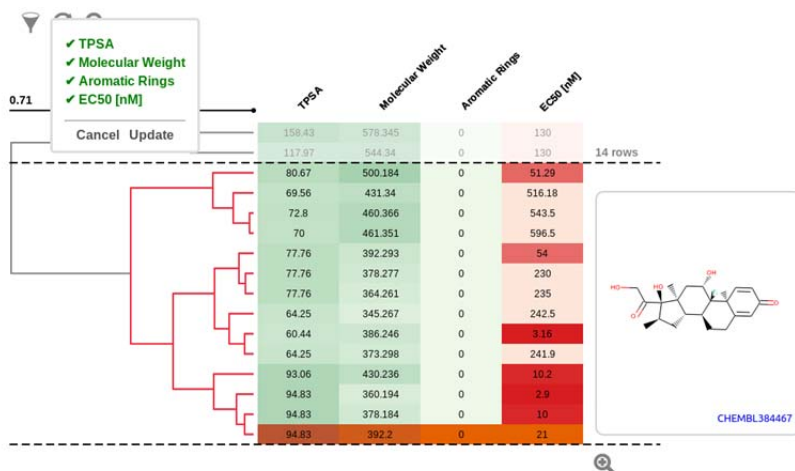
InCHlib – Interactive Cluster Heatmap for web applications

Škuta C.^{1,2}, Bartůněk P.², Svozil D.^{1,2}

¹ *Laboratoř informatiky a chemie, Fakulta chemické technologie, Vysoká škola chemicko-technologická, Technická 5, CZ-166 28 Praha, Česká republika*

² *CZ-OPENSREEN, Ústav molekulární genetiky AV ČR., v. v. i., Videňská 1083, CZ-142 20 Praha, Česká republika*

Shluková heatmapa (v angličtině “cluster heatmap”) je vizualizační metoda, ve které se zobrazí datová matice a současně hierarchická struktura jejích řádků či sloupců. Datová matice je zobrazena jako obdélníková mřížka, v níž jsou pro reprezentaci hodnot jednotlivých prvků použity různé barvy. Hierarchická struktura řádků či sloupců je vizualizována pomocí tzv. dendrogramu. InCHlib (Interactive Cluster Heatmap Library) je javascriptová knihovna pro vizualizaci a interaktivní exploraci shlukových heatmap. InCHlib lze snadno propojit s externími nástroji pro datovou analýzu a obohatit tak vizualizaci o další prvky. InCHlib je určena k integraci do stávajících webovských řešení a již se úspěšně používá např. v HTS informačním systému vyvíjeném Národní infrastrukturou pro chemickou biologii CZ-OPENSREEN. Ačkoliv je InCHlib primárně cílena na chemická a biologická data, knihovna je dostatečně obecná a její aplikační doména není omezena pouze na přírodní vědy. InCHlib je pouze vizualizační metoda, vstupem do ní jsou data již dopředu shluknutá a uložená ve správném formátu. Proto byl v Pythonu vyvinut také skript `inchlib_clust`, který obstarává oba dva výše zmíněné kroky. Jak InCHlib, tak `inchlib_clust` jsou k dispozici ke stažení zdarma [1]. InCHlib je také součástí [2] sady komponent pro zpracování biologických informací BioJS [3].



Literatura

- [1] <http://openscreen.cz/software/inchlib>
- [2] <https://www.ebi.ac.uk/Tools/biojs/registry/Biojs.InChIlib.html>
- [3] Gomez, J., *et al.* (2013) BioJS: an open source JavaScript framework for biological data visualization, *Bioinformatics*, **29**, 1103-1104

P-29

Scaffoldová analýza environmentální a farmaceutických databází

Scaffold analysis of environmental and drug discovery databases

Voršilák M.^{1,2}, Svozil D.¹, Tetko I.²

¹ *Laboratory of Informatics and Chemistry, ICT Prague, Technická 5, Prague 6, 166 28 Prague, Czech Republic*

² *Institute of Structural Biology, Helmholtz Zentrum Munich, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany*

Molekulární scaffold tvoří strukturální jádro sloučeniny, propojené kruhy [1]. Některé scaffoldy, jako např. kondenzované heterocyklické sloučeniny, jsou spojeny s toxicitou a karcinogenicitou sloučenin. Jiné, např. benzodiazepinové jádro, jsou naopak přednostňované při vývoji léčiv. Výběr neredundantní reprezentativní množiny pokrývající různé scaffoldy je důležitý při biologickém screeningu a při sestavování kombinatoriálních knihoven. V této práci jsme provedli scaffoldovou analýzu více než 60 miliónů komerčně dostupných sloučenin z databáze ChemNavigator. Při analýze jsme získali 10 miliónů scaffoldů do velikosti 5 kruhů, pro které byly vypočítány a normalizovány E-state descriptor. Pro náhodně vybranou sadu scaffoldů byla natrénovaná Kohonenova mapa [2]. Z volně dostupných environmentální a farmaceutických databází DrugBank [3], ChEMBL DrugStore [4] a EINECS (European Inventory of Existing Commercial chemical Substances) byly získány scaffoldy, které byly projektovány do Kohonenovy mapy. Množiny scaffoldů nepokrývají stejný prostor, proto jsou rozdíly mezi nimi dále zkoumány. Kohonenova mapa je užitečný nástroj pro prohledávání chemického prostoru a pro navrhování a porovnávání chemických knihoven.

Literatura

- [1] Bemis, G.W.; Murcko, M.A. *J. Med. Chem.* 39 (1996) 2887-93
- [2] Kohonen, T., *Self-Organizing Maps*. 3 ed.; Springer Verlag: New York, 2000
- [3] Knox, C.; Law, V.; et al. *Nucleic Acids Res.* 39 (2011) D1035-1041
- [4] Gaulton A.; Bellis, L.J.; et al. *Nucl. Acids Res.* 40 (2012) D1100-D1107

	<i>strana</i>
L1-01 Rozvoj Bioinformatiky v prostředí IT4Innovations národního superpočítačového centra <i>Branislav Jansík, IT4Innovations, Ostrava</i>	7
L1-02 OMICS – High Performace Computing řešení pro genomiku <i>Tomáš Hlavsa, BULL s.r.o., Praha 2</i>	8
L1-03 ChemAxon – chem a bioinformatický software pro Váš výzkum <i>Petr Hamerník, ChemAxon s.r.o., Praha 2</i>	9
L1-04 Biowes - řešení pro správu a sdílení experimentálních dat a metadat <i>Petr Císař, Jihočeská univerzita v Českých Budějovicích, Nové Hradky</i>	10
L2-01 Zpracování a analýza HTS dat <i>Petr Daněček, Wellcome Trust Sanger Institute, Hinxton, UK</i>	15
L2-02 Assembly komplexních genomů <i>Jan Pačes, Ústav molekulární genetiky AV ČR, v.v.i., Praha</i>	16
L2-03 Globální pohled na výstupy z 454 pyrosekvenací a jejich problémy <i>Martin Mokrejš, Bioinformatics.cz, Praha</i>	17
L2-04 Vliv sekvenační délky na rekonstrukci transkriptomu s využitím RNA-Seq <i>Jan Opett, Přírodovědecká fakulta, Masarykova univerzita, Brno</i>	19
L2-05 RepeatExplorer: Analýza repetitivních sekvencí na Galaxy serveru <i>Petr Novák, Biologické centrum AV ČR, v. v. i. , Ceske Budejovice</i>	21
L3-01 Bioinformatická analýza mitochondriálních genomů kvasiniek <i>Bronislava Brejová, KI FMFI UK, Bratislava</i>	25

	<i>strana</i>
L3-02 Od sekvencí k chromozómům: výzkum repetitivní DNA rostlin v Laboratoři molekulární cytogenetiky BC AVČR <i>Jiří Macas, Biologické centrum AVČR, České Budějovice</i>	26
L3-03 Bioinformatický a chemoinformatický výzkum v Loschmidtových laboratořích <i>Jan Brezovský, Přírodovědecká fakulta, Masarykova univerzita, Brno</i>	27
L3-04 Hľadanie genómov RNA vírusov po sekvenovaní celkovej RNA tkanív hľadavca <i>Tomáš Szemes, Prírodovedecká fakulta Univerzity Komenského, Bratislava 4</i>	28
L4-01 Klasifikační metody pro genetická data: regularizace a robustnost <i>Jan Kalina, Jan Kalina, Praha 6</i>	31
L4-02 Autoindex nad DNA sekvencemi <i>Jan Holub, České vysoké učení technické v Praze, FIT, Praha</i>	32
L4-03 Duální selekce genetických obvodů v bakteriích <i>Petr Sosík, Slezská univerzita v Opavě, Opava</i>	33
L4-04 Metody kombinování shlukovacích algoritmů <i>Tomáš Bartoň, Ústav molekulární genetiky AV ČR, v. v. i., Praha</i>	34
L5-01 ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat <i>Jindřich Jindřich, Ústav molekulární genetiky AV ČR, v. v. i., Praha 4</i>	37
L5-02 Nová metoda pro generování 2D farmakoforového modelu <i>David Hoksza, Univerzita Karlova v Praze, Praha</i>	38
L5-03 Jaký vliv na přesnost QSPR modelů má metodika přípravy 3D struktury? Případová studie pro predikci pKa. <i>Stanislav Geidl, Přírodovědecká fakulta, Masarykova univerzita, Brno</i>	39

	<i>strana</i>
L5-04 Radosti a strasti dokování <i>Karel Berka, Univerzita Palackého v Olomouci, Olomouc</i>	41
L5-05 Konzistence křížových odkazů v databázích malých molekul <i>Jakub Galgonek, Ústav organické chemie a biochemie AV ČR, v.v.i., Praha 6</i>	42
L6-01 Vyhledávání, porovnávání, validace a charakterizace strukturálních motivů v rámci biomakromolekul <i>Radka Svobodová Vařeková, CEITEC, Masarykova univerzita, Brno</i>	45
L6-02 Dolce: Databáze lokálních konformací DNA <i>Petr Čech, VŠCHT Praha, Praha</i>	47
L6-03 MOLE: Pokročilý nástroj pro vyhledávání a charakterizaci tunelů <i>Lukáš Pravda, Přírodovědecká fakulta MU, Brno</i>	48
L6-04 Vývoj nových metod a nástrojů pro metadynamické simulace <i>Petr Hošek, Vysoká škola chemicko-technologická, Praha</i>	50
L6-05 Simulace oscilační molekulární dynamiky allosterických proteinů v programu GROMACS <i>Leona Šerá, VŠCHT Praha, Praha</i>	51
L7-01 PredictSNP: prediktor vlivu mutací na funkci proteinů <i>Jaroslav Bendl, Fakulta Informačních Technologií, Vysoké Učení Technické v Brně, Brno</i>	55
L7-02 Systém pre identifikáciu mutácií v proteomických dátach <i>Miroslav Hruška, Univerzita Palackého v Olomouci, Olomouc</i>	56
L7-03 Bioinformatická analýza náhodných proteinových sekvencí <i>Jiří Vymětal, Ústav Organické Chemie a Biochemie AV ČR, Praha 6</i>	57

	<i>strana</i>
L8-01 Horizontální přenos genů v komplexní struktuře genomických signatur <i>Natália Martínková, Ústav biologie obratlovců AV ČR, v.v.i., Brno</i>	61
L8-02 Využití signálového zpracování ve fylogenetice a fylogenomice <i>Karel Sedlář, Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Brno</i>	62
L8-03 Evolučně konzervované vzory sekvenční komplementarity mezi eukaryotickými mRNA UTR a rRNA a implikace jejich existence pro regulaci genové translace <i>Josef Pánek, MBÚ AV ČR, Praha</i>	63
L8-04 Lokalizace potenciálních G4-kvadruplexů v eukaryotických genomech <i>Tomáš Martínek, Vysoké učení technické v Brně, Fakulta informačních technologií, Brno</i>	64
L8-05 Využití genomických metod pro studium transkripčních faktorů u modelové kvasinky <i>Schizosaccharomyces pombe</i> <i>Martin Převorovský, Univerzita Karlova v Praze, Přírodovědecká fakulta, Praha 2</i>	65
L8-06 Bioinformatika v environmentální mikrobiologii <i>Ondřej Uhlík, Vysoká škola chemicko-technologická v Praze, Praha</i>	66

	<i>strana</i>
P-01 Biowes - řešení pro správu a sdílení experimentálních dat a metadat <i>Bárta Antonín, University of South Bohemia, České Budějovice</i>	69
P-02 Dva příklady využití bioinformatiky v potravinářském výskume <i>Brežná Barbara, Národní zemědělské a potravinářské centrum, Lužianky</i>	71
P-03 Aplikácia a webová služba pre funkčnú anotáciu variantov <i>Budiš Jaroslav, Lambda Life a.s., Bratislava 5</i>	72
P-04 Automatický postup pro klasifikaci lokálních konformací DNA <i>Čech Petr, VŠCHT Praha, Praha 6</i>	73
P-05 Sekvence a anotace genomu půdní bakterie <i>Streptomyces</i> sp. W6 indikuje potenciál produkce antimycinu <i>Chrudimský Tomáš, BC AVČR v.v.i, Ústav půdní biologie, České Budějovice</i>	74
P-06 Predikce nových agonistů glukokortikoidního receptoru explorační chemického prostoru <i>Čmelo Ivan, VŠCHT Praha, Praha 6</i>	75
P-07 Sekvenování chromosomálního ramene 3DS pšenice <i>Cviková Kateřina, Centrum strukturní a funkční genomiky rostlin ÚEB AV ČR v.v.i., Olomouc - Holice</i>	76
P-08 Štatistický model pre neinvazívnu prenatalnú diagnostiku aneuploidii <i>Đuriš František, Prírodovedecká fakulta Univerzity Komenského, Bratislava 4</i>	77
P-09 Studium vlivu polysubstituovaných pyrimidinů na expresi genů spojených se zánětlivou reakcí <i>Hájek Miroslav, Ústav organické chemie a biochemie AV ČR, Praha 6</i>	78
P-10 Zjednodušená implementace umělých neuronových sítí v prognostice přežití pacientů s kolorektálním karcinomem na základě klinických a molekulárně-biologických dat <i>Hošek Petr, Plzeň</i>	79

	<i>strana</i>
P-11 Skryté retroviry: Hledání nových rodin endogenních retrovirů <i>Hron Tomáš, Ústav molekulární genetiky AV ČR, v. v. i., Praha 4</i>	80
P-12 Interakčné preferencie v komplexoch proteín - nukleová kyselina <i>Jakubec David, ÚOCHB AV ČR, Praha 6</i>	81
P-13 Klasifikace metagenomických vzorků pomocí diskriminativních nadřetězců DNA <i>Jalovec Karel, ČVUT Fakulta Elektrotechnická, Praha</i>	82
P-14 SigHunt: Nová metoda pro hledání kandidátních genomických ostrovů v DNA sekvencích <i>Jaroň Kamil, Institut biostatistiky a analýz, Brno</i>	83
P-15 Predikcia mitochondriálnych adresných štiepných sekvencií v Euglena gracilis a Trypanosomatidách <i>Krnáčková Katarína, Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra biochémie, Bratislava</i>	84
P-16 Integrace klinických a biomedicínských dat v projektu THALAMOSS. <i>Lexa Matěj, Masarykova univerzita, Fakulta informatiky, Brno</i>	85
P-17 ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat <i>Müller Tomáš, Ústav molekulární genetiky AV ČR, v. v. i., Praha 4</i>	86
P-18 Analýza taxonomicky jednoduché bakteriální komunity z prostředí hlubinného dolu s využitím metagenomiky a bioinformatiky <i>Rídl Jakub, Ústav molekulární genetiky AV ČR, Praha</i>	87
P-19 Analýza genomu Pseudomonas fluorescens ICT v souvislosti s jejím bioremediačním potenciálem <i>Sakmaryová Iva, Technická Univerzita v Liberci, Liberec</i>	88

	<i>strana</i>
P-20 Nové modulatory steroidních receptorů identifikované profilováním s vysokou propustností na panelu reportérových buněčných linií <i>Sedlák David, Ústav Molekulární Genetiky, AV CR, v.v.i, Praha</i>	90
P-21 Generická reprezentace chemických sloučenin <i>Škoda Petr, Univerzita Karlova v Praze - MFF, Praha 1</i>	92
P-22 Mining chemogenomického prostoru za účelem nalezení vztahů mezi biologickými cíli <i>Škuta Ctibor, VŠCHT Praha, Praha</i>	93
P-23 HotSpot Wizard 2.0: automatický webový nástroj pro proteinové inženýrství <i>Štourač Jan, Brno</i>	95
P-24 Strategie analýzy pyrotagů funkčních genů <i>Strejček Michal, VŠCHT Praha, Praha</i>	96
P-25 Genomový projekt <i>Rhodococcus erythropolis</i> CCM2595 <i>Strnad Hynek, Ústav molekulární genetiky, AV ČR, Praha 4</i>	97
P-26 Transkriptomická analýza oddenků pýru plazivého v rané fázi dehydratačního stresu <i>Svoboda Pavel, Výzkumný ústav rostlinné výroby, Praha</i>	98
P-27 Vyhledávání, porovnávání, validace a charakterizace strukturálních motivů v rámci biomakromolekul <i>Svobodová Vařeková Radka, CEITEC, Masarykova univerzita, Brno</i>	99
P-28 InChIlib – interaktivní shluková heatmapa pro webové aplikace <i>Svožil Daniel, Vysoká škola chemicko-technologická, Praha</i>	101
P-29 Scaffoldová analýza environmentálních a farmaceutických databází <i>Voršílák Milan, VŠCHT Praha, Praha</i>	103

	<i>strana</i>
Bartoň Tomáš	34
Bartůněk Petr	37, 75, 86, 90, 93, 101
Bárta Antonín	10, 69
Bendl Jaroslav	27, 55, 95
Berka Karel	41, 48
Brejová Bronislava	25
Brežná Barbara	71
Brezovský Jan	27, 55, 95
Budiš Jaroslav	28, 72, 77
Císař Petr	10, 69
Cviková Kateřina	76
Čech Petr	47, 73
Černý Jiří	73
Čmelo Ivan	75
Daněček Petr	15
Đuriš František	77
Galgonek Jakub	42
Geidl Stanislav	39
Hamerník Petr	9
Hájek Miroslav	78
Hlavsa Tomáš	8
Hoksza David	38, 75, 92
Holub Jan	32
Hošek Petr	50
Hošek Petr	79
Hron Tomáš	80
Hroudová Miluše	66, 87
Hruška Miroslav	56
Chroňáková Alica	74
Chrudimský Tomáš	74
Jakubec David	81
Jalovec Karel	82
Jansík Branislav	7
Jaroň Kamil	61, 83
Jindřich Jindřich	37, 86
Kalina Jan	31
Kolář Michal	97

Krnáčová Katarína	84
Kužmová Erika	78
Lexa Matěj	64, 85
Macas Jiří	26
Martincová Marie	88
Martínek Tomáš	64
Martínková Natália	61, 83
Mokrejš Martin	17
Müller Tomáš	37, 86
Novák Petr	21
Oppelt Jan	19
Ostašov Pavel	79
Pačes Jan	16, 80, 87
Pánek Josef	63
Pitule Pavel	79
Pravda Lukáš	45, 48, 99
Převorovský Martin	65
Rídl Jakub	66, 87
Sakmaryová Iva	88
Sedlák David	37, 75, 86, 90
Sedlář Karel	62
Sosík Petr	33
Strejček Michal	66, 96
Strnad Hynek	87, 97
Svoboda Pavel	98
Svobodová Vařeková Radka	39, 45, 48, 99
Svozil Daniel	37, 38, 47, 73, 75, 86, 90, 93, 101, 103
Szemes Tomáš	28, 72, 77
Šerá Leona	51
Škoda Petr	92
Škuta Ctibor	37, 86, 90, 93, 101
Šmídková Markéta	78
Štourač Jan	27, 55, 95
Uhlík Ondřej	66, 96
Vojtěch Jaroslav	8
Voller Jiří	56
Vondrášek Jiří	57, 81

	<i>strana</i>
Voršilák Milan	103
Vymětal Jiří	57

Bartoň Tomáš (*tomas.barton@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha

Bartůněk Petr (*bartunek@img.cas.cz*)

Ústav molekulární genetiky AV ČR, Praha 4

Bárta Antonín (*abarta@frov.jcu.cz*)

University of South Bohemia, České Budějovice

Bendl Jaroslav (*ibendl@fit.vutbr.cz*)

Fakulta Informačních Technologií, Vysoké Učení Technické v Brně, Brno

Berka Karel (*karel.berka@upol.cz*)

Univerzita Palackého v Olomouci, Olomouc

Brejová Bronislava (*brejova@dcs.fmph.uniba.sk*)

KI FMFI UK, Bratislava

Brežná Barbara (*brezna@vup.sk*)

Národné poľnohospodárske a potravinárske centrum, Lužianky

Brezovský Jan (*brezovsky@mail.muni.cz*)

Přírodovědecká fakulta, Masarykova universita, Brno

Budiš Jaroslav (*budis@dcs.fmph.uniba.sk*)

Lambda Life a.s., Bratislava 5

Císař Petr (*cisar@frov.jcu.cz*)

Jihočeská univerzita v Českých Budějovicích, Nové Hradky

Cviková Kateřina (*cvikova@ueb.cas.cz*)

Centrum strukturální a funkční genomiky rostlin | ÚEB AV ČR v.v.i., Olomouc - Holice

Čech Petr (*cechp@vscht.cz*)

VŠCHT Praha, Praha

Černý Jiří (*jiri.cerny@img.cas.cz*)

Biotechnologický ústav AV ČR, Praha

Čmelo Ivan (*cmelo.ivan@gmail.com*)

VŠCHT Praha, Praha 6

Daněček Petr (*petr.danecek@gmail.com*)

Wellcome Trust Sanger Institute, Hinxton, UK

Dvořák Pavel (*pavel.dvorak@uochb.cas.cz*)

Ústav organické chemie a biochemie AV ČR, v.v.i., Praha 6

Đuriš František (*fduris@dcs.fmph.uniba.sk*)

Prirodovedecká fakulta Univerzity Komenského, Bratislava 4

Galgonek Jakub (*jakub.galgonek@marge.uochb.cas.cz*)

Ústav organické chemie a biochemie AV ČR, v.v.i., Praha 6

Geidl Stanislav (*standag@chemi.muni.cz*)

Přírodovědecká fakulta, Masarykova univerzita, Brno

Hamerník Petr (*kkolska@chemaxon.com*)

ChemAxon s.r.o., Praha 2

Hájek Miroslav (*hajek@uochb.cas.cz*)

Ústav organické chemie a biochemie AV ČR, Praha 6

Hlavsa Tomáš (*tomas.hlavsa@bull.cz*)

BULL s.r.o., Praha 2

Hokszá David (*david.hokszá@gmail.com*)

Univerzita Karlova v Praze, Praha

Holub Jan (*Jan.Holub@fit.cvut.cz*)

České vysoké učení technické v Praze, FIT, Praha

Hošek Petr (*petr.hosek@lfp.cuni.cz*)

Plzeň

Hošek Petr (*hosekp@vscht.cz*)

Vysoká škola chemicko-technologická, Praha

Hron Tomáš (*tomas.hron@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Hroudová Miluše (*hroudova@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v.v.i., Praha 4

Hruška Miroslav (*hruska.miro@gmail.com*)

Univerzita Palackého v Olomouci, Olomouc

Hřibová Eva (*hribova@ueb.cas.cz*)

Centrum strukturální a funkční genomiky rostlin | ÚEB AV ČR v.v.i., Olomouc - Holice

Chroňáková Alica (*alicach@upb.cas.cz*)

Biologické Centrum AV ČR, Ústav půdní biologie, České Budějovice

Chrudimský Tomáš (*tomas.chrudimsky@gmail.com*)

BC AV ČR v.v.i., Ústav půdní biologie, České Budějovice

Jakubec David (*jakubecd@natur.cuni.cz*)

ÚOCHB AV ČR, Praha 6

Jalovec Karel (*jalovec.karel@gmail.com*)

ČVUT Fakulta Elektrotechnická, Praha

Janský Branislav (*branislav.jansk@vsb.cz*)

IT4Innovations, Ostrava

Jaroň Kamil (*376090@mail.muni.cz*)

Institut biostatistiky a analýz, Brno

Jindřich Jindřich (*jindrich@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Kalina Jan (*kalina@cs.cas.cz*)

Jan Kalina, Praha 6

Kolář Michal (*kolarmi@img.cas.cz*)

Ústav molekulární genetiky AV ČR, Praha 4

Krnáčová Katarína (*katkasysova@gmail.com*)

Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra biochémie, Bratislava

Kužmová Erika (*kuzmovae@gmail.com*)

Ústav organické chemie a biochemie AV ČR, Prague 6

Lexa Matěj (*lexa@fi.muni.cz*)

Masarykova univerzita, Fakulta informatiky, Brno

Macas Jiří (*macas@umbr.cas.cz*)

Biologické centrum AVČR, České Budějovice

Martincová Marie (*marie.martincova@tul.cz*)

Technická univerzita v Liberci, Liberec

Martínek Tomáš (*martinto@fit.vutbr.cz*)

Vysoké učení technické v Brně, Fakulta informačních technologií, Brno

Martínková Natálie (*martinkova@ivb.cz*)

Ústav biologie obratlovců AV ČR, v.v.i., Brno

Mokrejš Martin (*mmokrejs@bioinformatics.cz*)

Bioinformatics.cz, Praha

Müller Tomáš (*tomas.muller@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Novák Petr (*petr@umbr.cas.cz*)

Biologické centrum AV ČR, v. v. i., Ceske Budejovice

Oppelt Jan (*323639@mail.muni.cz*)

Přírodovědecká fakulta, Masarykova univerzita, Brno

Ostašov Pavel (*pavel.ostasov@lfp.cuni.cz*)

Lékařská fakulta v Plzni, Univerzita Karlova v Praze, Plzeň

Pačes Jan (*hpaces@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v.v.i., Praha

Pánek Josef (*panek@biomed.cas.cz*)

MBÚ AV ČR, Praha

Pitule Pavel (*pitulep@seznam.cz*)

Lékařská fakulta UK v Plzni, Plzeň

Pravda Lukáš (*xpravda@ncbr.muni.cz*)

Přírodovědecká fakulta MU, Brno

Provazník Jan (*provaz@entu.cas.cz*)

Biologické Centrum AVČR v.v.i., Entomologický ústav, České Budějovice

Převorovský Martin (*prevorov@natur.cuni.cz*)

Univerzita Karlova v Praze, Přírodovědecká fakulta, Praha 2

Přistoupilová Anna (*anna.pristoupilova@lf1.cuni.cz*)

ÚDMP, 1. lékařská fakulta UK, Praha

Rídl Jakub (*ridlj@img.cas.cz*)

Ústav molekulární genetiky AV ČR, Praha

Sakmaryová Iva (*iva.sakmaryova@tul.cz*)

Technická Univerzita v Liberci, Liberec

Sedlák David (*sedlak@img.cas.cz*)

Ústav Molekulární Genetiky, AV ČR, v.v.i, Praha

Sedlář Karel (*sedlar@feec.vutbr.cz*)

Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Brno

Skalický Tomáš (*skalicky@paru.cas.cz*)

Přírodovědecká fakulta Jihočeské Univerzity v Českých Budějovicích, České Budějovice

Sosík Petr (*petr.sosik@fpf.slu.cz*)

Slezská univerzita v Opavě, Opava

Strejček Michal (*strejcem@vscht.cz*)

VŠCHT Praha, Praha

Strnad Hynek (*strnad@img.cas.cz*)

Ústav molekulární genetiky, AV ČR, Praha 4

Svoboda Pavel (*pavel.svoboda@vurv.cz*)

Výzkumný ústav rostlinné výroby, Praha

Svobodová Vařeková Radka (*radka.svobodova@ceitec.muni.cz*)

CEITEC, Masarykova univerzita, Brno

Svozil Daniel (*daniel.svozil@gmail.com*)

Vysoká škola chemicko-technologická, Praha

Szemes Tomáš (*szemes@fns.uniba.sk*)

Přírodovědecká fakulta Univerzity Komenského, Bratislava 4

Šauer Max (*kkolska@chemaxon.com*)

ChemAxon s.r.o., Praha 2

Šerá Leona (*seral@vscht.cz*)

VŠCHT Praha, Praha

Škoda Petr (*skodapetr@gmail.com*)

Univerzita Karlova v Praze - MFF, Praha 1

Škuta Ctibor (*skuta.ctibor@gmail.com*)

VŠCHT Praha, Praha

Šmídková Markéta (*smidkova@uochb.cas.cz*)

ÚOCHB AVČR, Praha 6

Štefanič Stanislav (*stefanic@mail.muni.cz*)

Masarykova univerzita, Fakulta informatiky, Brno

Štourač Jan (*stourac.jan@gmail.com*)

Jan Štourač, Brno

Uhlík Ondřej (*uhliko@vscht.cz*)

Vysoká škola chemicko-technologická v Praze, Praha

Vojtěch Jaroslav (*jaroslav.vojtech@bull.cz*)

BULL s.r.o., Praha 2

Voller Jiří (*jirivoller@gmail.com*)

IMTM, Univerzita Palackého, Olomouc

Vondrášek Jiří (*jiri.vondrasek@uochb.cas.cz*)

ÚOCHB AV ČR, Praha 6

Voršilák Milan (*vorsilam@vscht.cz*)

VŠCHT Praha, Praha

Vymětal Jiří (*jiri.vymetal@uochb.cas.cz*)

Ústav Organické Chemie a Biochemie AV ČR, Praha 6

Konferenční sborník ENBIK2014

Vydala: Vysoká škola chemicko-technologická v Praze
Technická 5, 166 28 Praha 6

Editoři: **Petr Čech, Daniel Svozil**

Tisk: powerprint s.r.o.
Kamýcká 751/60, 165 00 Praha – Suchdol

Rok vydání: 2014

Počet stran: 124

Náklad: 100 ks