



Záverečná karta projektu

Názov projektu

Evidenčné číslo projektu

APVV-18-0239

Nekonvenčné aplikácie nových sekvenačných technológií v komparatívnej a funkčnej genomike

Zodpovedný riešiteľ **prof. RNDr. Jozef Nosek, DrSc.**

Príjemca **Univerzita Komenského v Bratislave**

Názov pracoviska, na ktorom bol projekt riešený

Katedra biochémie, Prírodovedecká fakulta, Univerzita Komenského v Bratislave

Katedra genetiky, Prírodovedecká fakulta, Univerzita Komenského v Bratislave

Katedra informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave

Názov a štát zahraničného pracoviska, ktoré spolupracovalo pri riešení

Center for Computational Biology and Bioinformatics, Pennsylvania State University, USA

Centre for Genome Regulation, Španielsko

Department of Biochemistry, University of Montréal, Kanada

Department of Genetics and Applied Microbiology, University of Debrecen, Maďarsko

Department of Mathematics, Simon Fraser University, Kanada

Department of Microbiology, University of Szeged, Maďarsko

International Institute of Molecular and Cell Biology, Poľsko

Udelené patenty/podané patentové prihlášky, vynálezy alebo úžitkové vzory, ktoré sú výsledkami projektu

neboli

Najvýznamnejšie publikácie (knihy, články, prednášky, správy a pod.) zhrňujúce výsledky projektu – uveďte aj publikácie prijaté do tlače

Výsledky projektu boli prezentované formou 14 prednášok a posterov na medzinárodných vedeckých konferenciách (Belgicko, Francúzsko, Japonsko, Slovensko, Spojené kráľovstvo, Španielsko, Švajčiarsko, Švédsko, Turecko) a celkovo 26 vedeckých publikáciách. V nasledovnom zozname je uvedených 20 najvýznamnejších prác vo vedeckých časopisoch:

1. Boža, V., Perešíni, P., Brejová, B., Vinař, T. (2020) DeepNano-blitz: a fast base caller for MinION nanopore sequencers. *Bioinformatics* 36(14):4191-4192.

2. Boža, V., Perešíni, P., Brejová, B., Vinař, T. (2022) Dynamic Pooling Improves Nanopore Base Calling Accuracy. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19(6):3416-3424.

3. Brejová, B., Boršová, K., Hodorová, V., Čabanová, V., Gafurov, A., Fričová, D., Neboháčová, M., Vinař, T., Klempa, B., Nosek, J. (2021) Nanopore Sequencing of SARS-CoV-2: Comparison of Short and Long PCR-tiling Amplicon Protocols. *PLoS ONE*

16(10):e0259277.

4. Brejová, B., Boršová, K., Hodorová, V., Čabanová, V., Reizigová, L., Paul, E.D., Čekan, P., Klempa, B., Nosek, J., Vinař, T. (2021) A SARS-CoV-2 mutant from B.1.258 lineage with Δ H69/ Δ V70 deletion in the Spike protein circulating in Central Europe in the fall 2020. *Virus Genes* 57:556-560.
5. Brejová, B., Hodorová, V., Lichancová, H., Peričková, E., Šoucová, V.A., Sipiczki, M., Vinař, T., Nosek, J. (2023) Chromosome-level genome assembly of the yeast *Candida verbasci*. *Microbiology Resource Announcements* 12(3):e00005-23.
6. Brejová, B., Vozáriková, V., Agarský, I., Derková, H., Fedor, M., Harmanová, D., Kiss, L., Korman, A., Pašen, M., Brázdovič, F., Vinař, T., Nosek, J., Tomáška, L. (2023) Yeast Mitochondrial Post-Translational Modification Database. *Genetics* 224(3):iyad087.
7. Cillingová, A., Tóth, R., Mojáková, A., Zeman, I., Vrzoňová, R., Siváková, B., Baráth, P., Neboháčová, M., Klepcová, Z., Brázdovič, F., Lichancová, H., Hodorová, V., Brejová, B., Vinař, T., Tomáška, L., Gácser, A., Gabaldón, T., Nosek, J. (2022) Transcriptome and proteome profiling reveals complex adaptations of *Candida parapsilosis* cells assimilating hydroxyaromatic carbon sources. *PLoS Genetics* 18(3):e1009815.
8. Červenák, F., Sepšiová, R., Nosek, J., Tomáška, L. (2021) Step-by-step evolution of telomeres: lessons from yeasts. *Genome Biology and Evolution* 13(2):evaa268.
9. Frankovský, J., Keresztesová, B., Bellová, J., Kunová, N., Čanigová, N., Hanáková, K., Bauer, J.A., Ondrovičová, G., Lukáčová, V., Siváková, B., Zdráhal, Z., Pevala, V., Procházková, K., Nosek, J., Baráth, P., Kutejová, E., Tomáška, L. (2021) The yeast mitochondrial succinylome: Implications for regulation of mitochondrial nucleoids. *Journal of Biological Chemistry* 297(4):101155.
10. Frankovský, J., Vozáriková, V., Nosek, J., Tomáška, L. (2021) Mitochondrial protein phosphorylation in yeast revisited. *Mitochondrion* 57:148-162.
11. Gafurov, A., Brejová, B., Medvedev, P. (2022) Markov chains improve the significance computation of overlapping genome annotations. *Bioinformatics* 38:i203-i211.
12. Gafurov, A., Baláž, A., Amman, F., Boršová, K., Čabanová, V., Klempa, B., Bergthaler, A., Vinař, T., Brejová, B. (2022) VirPool: model-based estimation of SARS-CoV-2 variant proportions in wastewater samples. *BMC Bioinformatics* 23(1):1-18.
13. Hodorová, V., Lichancová, H., Zubenko, S., Sienkiewicz, K., Penir, S.M.U., Afanasyev, P., Bocek, D., Bonnin, S., Hakobyan, S., Smyczynska, U., Zhivkopias, E., Zlatohurska, M., Tralle, E., Frolova, A., Pryszcz, L.P., Brejová, B., Vinař, T., Nosek, J. (2019) Genome Sequence of the Yeast *Saprochaete ingens* CBS 517.90. *Microbiology Resource Announcements* 8(50):e01366-19.
14. Juríková, K., Gajarský, M., Hajikazemi, M., Nosek, J., Procházková, K., Paeschke, K., Trantírek, L., Tomáška, L. (2020) Role of folding kinetics of secondary structures in telomeric G-overhangs in the regulation of telomere maintenance in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 295(27):8958-8971.
15. Mixão, V., Hegedúsová, E., Saus, E., Pryszcz, L.P., Cillingová, A., Nosek, J., Gabaldón, T. (2021) Genome analysis of *Candida subhashii* reveals its hybrid nature and dual mitochondrial genome conformation. *DNA Research* 28(3):dsab006.
16. Mixão, V., del Olmo, V., Hegedúsová, E., Saus, E., Pryszcz, L.P., Cillingová, A., Nosek, J., Gabaldón, T. (2022) Genome analysis of five recently described species of the CUG-Ser clade uncovers *Candida theae* as a new hybrid lineage with pathogenic potential in the *Candida parapsilosis* species complex. *DNA Research* 29(2):dsac010.
17. Perešíni, P., Boža, V., Brejová, B., Vinař, T. (2021) Nanopore base calling on the edge. *Bioinformatics* 37(24):4661-4667.
18. Sitarčík, J., Vinař, T., Brejová, B., Krampl, W., Budiš, J., Radvánszky, J., Lucká, M. (2023) WarpSTR: Determining tandem repeat lengths using raw nanopore signals. *Bioinformatics* 39(6):btad388.
19. Tomáška, L., Nosek, J. (2020) Co-evolution in the jungle: From leafcutter ant colonies to chromosomal ends. *Journal of Molecular Evolution* 88(4):293-318.
20. Vrzoňová, R., Tóth, R., Siváková, B., Mořovská, A., Gaplovská, K., Baráth, P., Tomáška, L., Gácser, A., Gabaldón, T., Nosek, J., Neboháčová, M. (2021) OCT1 - a yeast mitochondrial thiolase involved in the 3-oxoadipate pathway. *FEMS Yeast Research* 21(5):foab034.

Softvérové produkty:

DeepNano-blitz: Rýchly nástroj na určovanie báz pre nanopórové sekvenátory MinION.

<https://github.com/fmfi-compbio/deepnano-blitz>

DeepNano-coral: Rýchly a energeticky úsporný nástroj na určovanie báz pre nanopórové sekvenátory MinION pomocou hardvérového akcelérátora Coral edge TPU.

<https://github.com/fmfi-compbio/coral-basecaller>

Heron: Vysoko presný nástroj na určovanie báz pre nanopórové sekvenátory MinION.

<https://github.com/fmfi-compbio/heron>

Osprey: presný nástroj na určovanie báz pre nanopórové sekvenátory MinION, ktorý pracuje v reálnom čase bez podpory GPU. <https://github.com/fmfi-compbio/osprey>

NaDaVCa: nástroj na zarovnávanie signálov nanopórového sekvenovania k referenčnému genómu <https://github.com/fmfi-compbio/nadavca>

WarpSTR: nástroj na určovanie dĺžky krátkych tandemových opakovaní z dát nanopórového sekvenovania <https://github.com/fmfi-compbio/warpstr>

pande-mic: nástroj na monitorovanie behu nanopórového sekvenovania pri sekvenovaní

SARS-CoV-2, prípadne iných vírusov <https://github.com/fmfi-compbio/pande-mic>

VirPool: nástroj na určovanie zastúpenia variantov vírusu SARS-CoV-2 pri sekvenovaní vzoriek z odpadovej vody <https://github.com/fmfi-compbio/virpool>

MCDP: výpočet štatistickej významnosti prekryvov genomických anotácií

<https://github.com/fmfi-compbio/mc-overlaps>

Genómové sekvencie:

3731 slovenských izolátov vírusu SARS-CoV-2 zverejnené v medzinárodnej databáze GISAID

49 druhov, resp. kmeňov kvasiniek, z ktorých bolo vedeckej komunite doteraz prístupných 9 v medzinárodnej databáze European Nucleotide Archive (ENA)

Verejné genómové prehliadače (Genome Browsers):

genóm kvasinky *Saprochaete ingens*

<http://genome.compbio.fmph.uniba.sk/cgi-bin/hgGateway?db=sapIngB1>

genóm kvasinky *Candida verbasci*

<http://genome.compbio.fmph.uniba.sk/cgi-bin/hgGateway?db=canVerA1>

Uplatnenie výsledkov projektu

Projekt mal charakter základného výskumu a bol zameraný na prípravu viacerých aplikácií pre nové sekvenačné technológie. Vytvorené laboratórne protokoly vhodné pre nanopórové sekvenovanie nukleových kyselín, bioinformatický softvér, genómové sekvencie a verejné genómové prehliadače môžu nájsť uplatnenie v biomedicínskych vedách a biotechnológiách. Časť našich výsledkov už našla uplatnenie pri analýze variantov vírusu SARS-CoV-2 v slovenskej populácii počas pandémie COVID-19 v rokoch 2020-2021, a sú tiež aplikovateľné pri ďalšom monitorovaní vírusu v odpadových vodách. Genómové sekvencie nekonvenčných druhov kvasiniek a charakterizované biochemické dráhy pre degradáciu hydroxyaromatických zlúčenín sú základom pre využitie týchto mikroorganizmov v biotechnológiách alebo bioremediáciách, v prípade patogénnych druhov sú potenciálne využiteľné aj v molekulárnej diagnostike alebo identifikácii molekulárných cieľov vhodných pre terapiu kvasinkových infekcií. Projekt mal aj edukačný prínos a prispel k príprave mladých vedeckých pracovníkov a študentov v programoch biochémie, genetiky a informatiky.

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v slovenskom jazyku (max. 20 riadkov)

K najvýznamnejším výsledkom riešenia projektu patria (1) protokoly pre (i) prípravu a analýzu vzoriek vysokomolekulárnej (HMW) DNA nanopórovou technológiou a (ii) biochemické obohacovanie telomerických fragmentov a analýzu dĺžky telomér individuálnych chromozómov; (2) sekvencie 49 genómov askomycétnych a bazidiomycétnych kvasiniek; (3) sekvencie 3731 izolátov vírusu SARS-CoV-2; (4) bioinformatický softvér vhodný pre (i) spracovanie surových dát z nanopórového sekvenátora (určovanie báz a ich zarovnávanie k referenčnému genómu), (ii) monitorovanie procesu sekvenovania vírusových genómov (napr. SARS-CoV-2) v reálnom čase, (iii) analýzu zastúpenia variantov vírusov vo vzorkách odpadových vôd, (iv) určovanie dĺžky krátkych tandemových repetícií (STR) v nanopórových dátach, a (v) výpočet štatistickej významnosti prekryvov genomických anotácií. Výsledky analýzy vírusových sekvencií viedli k identifikácii a charakterizácii nového variantu SARS-CoV-2 línie B.1.258 s deléciou v gène

pre Spike proteín. Komparatívne analýzy sekvencií kvasinkových genómov umožnili charakterizovať (i) biochemické dráhy a zodpovedajúce metabolické klastre génov zahrnuté v asimilácii hydroxyaromatických zlúčenín, (ii) nekanonické telomerické repetície, a (iii) rozdiely v zastúpení expanzných segmentov v ribozomálnych RNA kvasiniek rodu *Magnusiomyces/Saprochaete*.

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v anglickom jazyku (max. 20 riadkov)

The most important results of the project include (1) the protocols for (i) preparation and analysis of high molecular weight (HMW) DNA samples by nanopore sequencing technology and (ii) biochemical enrichment of telomeric fragments and telomere length analysis of individual chromosomes; (2) 49 genome sequences of ascomycete and basidiomycete yeasts; (3) sequences of 3731 isolates of the SARS-CoV-2 virus; (4) bioinformatics software suitable for (i) processing raw data from nanopore sequencers (basecalling and aligning the nanopore signal to the reference genome), (ii) real-time monitoring sequencing runs of viral genomes (e.g. SARS-CoV-2), (iii) analysis of the virus variants in wastewater samples, (iv) determination of the length of short tandem repeats (STR) in nanopore data, and (v) calculation of statistical significance of overlaps in genomic annotations. The results of the viral sequence analysis led to the identification and characterization of a new SARS-CoV-2 variant of the B.1.258 lineage carrying a deletion in the gene for Spike protein. Comparative analysis of the yeast genome sequences allowed characterization of (i) biochemical pathways and corresponding metabolic gene clusters involved in the assimilation of hydroxyaromatic compounds, (ii) non-canonical telomeric repeats, and (iii) differences of the expansion segments in the ribosomal RNAs in the yeast species of the genus *Magnusiomyces/Saprochaete*.