

Analýza výsledku sekvenování (cvičení č. 3)

Úvodní slovo

Toto cvičení je přímým pokračováním cvičení č. 1. Pro rodové zařazení může postačovat sekvence části genu pro 16S rRNA. Ověření a přesnější klasifikace (na druh či kmen) lze dosáhnout porovnáním sekvencí více genů v databázi. Pro účely klasifikace použitého mikroorganismu jsme zvolili kromě genu pro 16S rRNA i druhý gen, jehož sekvence je specifická pro daný bakteriální kmen.

Sekvenování se provádí z jednoho z primerů pro amplifikaci genu nebo lze navrhnout vhodnou sekvenci uvnitř naamplifikovaného genu.

Pro porovnávání námi získaných sekvencí s celosvětovou databází použijeme algoritmus BLAST (Basic Local Alignment Search Tool, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), který je v současnosti asi nejrozšířenějším heuristickým algoritmem, jenž umožňuje prohledávat rozsáhlé databáze. Pro účely srovnávání nukleotidových (DNA) sekvencí bakterií využijeme nástroj BLASTN. Po seznámení se se vzhledem příslušných webových stránek zvolíme nastavení default, je ale možné zvolit vlastní parametry prohledávání.

Cíl cvičení

Analyzovat výsledky sekvenování

Seznam přístrojů

- Notebooky s připojením k internetu a staženými volně dostupnými programy -
- **WinZip** (<https://www.winzip.com/win/en/winzip-10.html>)
- **BioEdit** (<https://bioedit.software.informer.com/7.2/>)
- Výsledky ze sekvenování (od české firmy SeqMe)

Pracovní postup

- 1) Stáhněte si soubory s výsledky sekvenací (formát .zip nebo .rar), soubory rozbalte pomocí programu WinZip a jednotlivé sekvence si prohlédněte v programu BioEdit.**
- 2) Při kontrole kvality sekvencí postupujte podle pokynů vyučujícího a dělejte si poznámky.**
- 3) Zkontrolujte a opravte sekvence:**
 - A) Prohlédněte si záznamy sekvencí jednotlivých genů a vyhledejte ty nukleotidy, které nejsou dobře čitelné.
 - B) Opravte v záznamu software neurčené nukleotidy, případně vyřadte ty, které jsou nečitelné.
 - C) Lze ze záznamu zjistit, kde se vázal sekvenační primer?
 - D) Zkontrolujte, jak dlouhý je získaný sekvenační záznam oproti délce amplikonu.
 - E) Vyznačte v sekvenci celistvý úsek, který neobsahuje žádné nečitelné nebo špatně čitelné nukleotidy.

4) Prohledejte databázi a zařadte výsledky

- F) Vyhledejte prohlédavač označovaný jako BLAST (Basic Local Alignment Search Tool, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Celý proces přirovnání sekvencí Vám nejprve popíše a názorně ukáže vyučující. Postup následně zopakujete individuálně s vlastními sekvencemi 16S rRNA genu.
- G) Vložte sekvenci celistvého úseku do prohlédavače.
- H) Vyčkejte, až program provede porovnání vaší sekvence s databází.
- I) Odečtěte výsledek a zaznamenejte nejpravděpodobnější zařazení rodu a druhu (případně i kmene) mikroorganismu.
- J) Klikněte na odkaz s Accession kódem sekvence s nejvyšší shodou, který propojuje BLAST službu s NCBI nukleotidovou databází. Vyexportujte sekvenci genu pro celý gen 16S rRNA ve formátu FASTA a v programu BioEdit ji srovnajte pomocí lokálního a globálního přirovnání (Local vs. Global Alignment) s Vaší sekvencí ze sekvenace. Postupujte podle pokynů vyučujícího. Postup si zapisujte.
- K) Odpovězte na následující otázky: Jak velký je celý gen 16S rRNA a kterou jeho část jsme amplifikovali pomocí PCR? Jak velká je shoda (%) Vaší sekvence s porovnávaným úsekem genu 16S rRNA? Jak velká je shoda (%) Vaší sekvence s celým genem 16S rRNA? Pozorujete rozdíly mezi výsledky lokálního a globálního přirovnání? Popište je. Jaký typ přirovnání je v tomto případě vhodnější a proč?
- L) Pro přesnou identifikaci druhu a kmene bakterie použijeme výsledky ze sekvenace neznámého genu. Sekvenci přirovnáme pomocí algoritmu nBLAST ve speciální databázi, jejíž fungování Vám popíše vyučující.
- M) Na základě přirovnání sekvence neznámého genu popište nejpravděpodobnější druh a kmen bakterie, ze které gen pochází. Odpovězte na následující otázky: Jakému genu patří amplifikovaná sekvence? Jaké jsou jeho koordináty v genomu identifikované bakterie? Jaká je funkce produktu genu a kde v bakterii se nachází? Zjištění si zapište do protokolu ze cvičení.

Další informace k této problematice najdete v následující literatuře

Cvrčková, F. 2006: Úvod do praktické bioinformatiky. Academia, Praha.

Šmarda et al. 2005: Metody molekulární biologie, Masarykova Universita, Brno

Simmon K.E., Croft A.C., Petti C.A. 2006: Application of SmartGene IDNS Software to Partial 16S rRNA Gene Sequences for a Diverse Group of Bacteria in a Clinical Laboratory. Journal of Clinical Microbiology 44(12), 4400-4406.