

GENOTYPING OF *KLEBSIELLA PNEUMONIAE* ISOLATES

Markéta Nykrýnová

Master (2.), FEEC BUT

E-mail: xnykry01@stud.feec.vutbr.cz

Supervised by: Denisa Maděránková

E-mail: maderankova@feec.vutbr.cz

Abstract: Typing methods capable of distinguishing bacterial strain are essential for epidemiologist because they help to prevent and control infections. Unfortunately, many typing methods are time-consuming, expensive and reproducibility of their results is questionable. Therefore, we propose a new genotyping method, which is capable of identifying genes with high rate of variability to distinguish different bacterial strains.

Keywords: genotyping, *Klebsiella pneumoniae*, bacterial strains, mini-MLST

1 ÚVOD

Typizace bakterií je nezbytná pro identifikaci vztahů mezi jednotlivými bakteriálními liniemi. Nalezení těchto propojení představuje klíčový krok pro pochopení epidemií a zároveň nám pomáhá vystopovat zdroj infekce a způsob jejího šíření. Typizace je též důležitá pro detekci přenosu nosokomiálních patogenů, jejich diagnostiku a léčbu. [1]

V dnešní době se využívají genotypové metody, které dokáží rozlišit blízce příbuzné linie. Současné typizační metody lze rozdělit do třech skupin. První skupinu tvoří DNA metody, které jsou založené na analýze elektroforeogramu získaného pomocí elektroforézy, kde jsme schopni rozlišit jednotlivé linie na základě velikosti DNA proužků. Do druhé části spadají metody založené na sekvenování, které jsou schopny rozlišit jednotlivé bakteriální linie z polymorfismů v DNA. Do poslední skupiny lze zařadit metody založené na hybridizaci, které spoléhají na hybridizaci DNA se sondami o známé sekvenci. [2]

Metoda typizace, která byla použita pro analyzované genomy *Klebsiella pneumoniae* je mini-multilokusová sekvenční typizace neboli mini-MLST, která je odvozena od multilokusové sekvenční typizace a skládá se ze dvou částí. Nejprve dochází k amplifikaci vybraných genů (provozní geny) pomocí PCR. Následně probíhá vysokorozlišovací analýza křivek tání, což je citlivá a rychlá metoda, která dokáže zachytit jedno- i vícenukleotidové změny. DNA je obarvena pomocí barviva a během analýzy dochází ke zvyšování teploty, což má za následek tání dvoušroubovice DNA a uvolňování barviva, tudíž dochází k poklesu fluorescence, kterou sledujeme. Ze získaných dat je za pomoci převodního klíče určen melt-typ. Výhodou metody je nízká cena, která je zhruba 10 - 20 % multilokusové sekvenční typizace. [3], [4]

2 TESTOVANÝ ORGANISMUS

Klebsiella pneumoniae je gram-negativní bakterie obalena polysacharidovou kapsulou, jež poskytuje buňce resistenci vůči mnohým hostitelským obranným mechanismům. Tuto bakterii lze nalézt v půdě, vodě nebo na povrchu rostlin. Některé linie způsobují v nemocnicích infekce, obzvláště pacientům se sníženou imunitou a jelikož se šíří velmi rychle, mohou vést k vypuknutí nozokomiálních epidemií. Bakterie způsobuje pneumonie, infekce močového ústrojí, po vstupu do cévního oběhu sepsi a může vést k životu ohrožujícím infekcím. [5]

3 NAVRŽENÝ ALGORITMUS

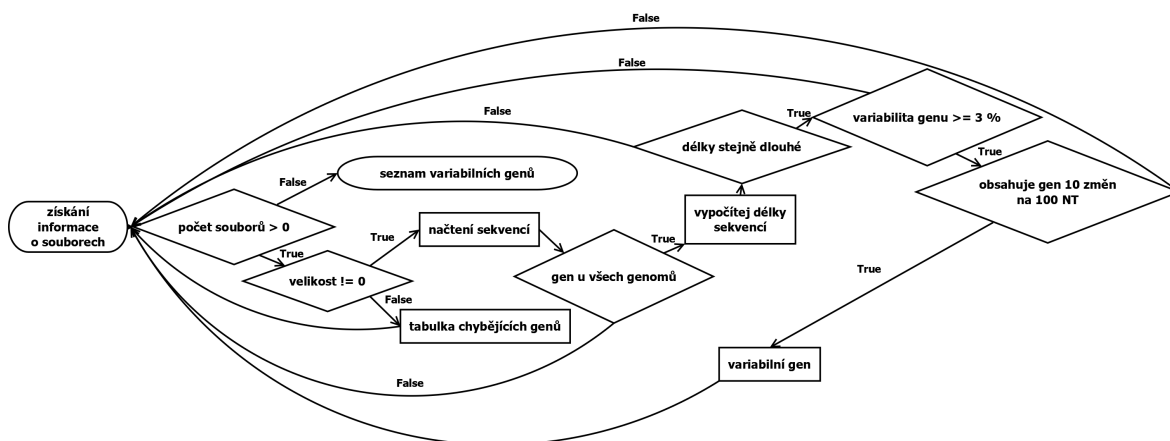
Představený algoritmus slouží pro nalezení genů s vysoce variabilními úseky v sekvenčních datech, přičemž nalezené geny budou použity pro typizaci metodou mini-MLST.

V prvním kroku je potřeba otestovat data získaná z DNA sekvenátoru. Jednotlivé genomy bakterie *Klebsiella pneumoniae* byly osekvenovány pomocí sekvenátoru Illumina a bylo použito párové-koncové sekvenování, což znamená, že pro jeden fragment máme sekvenované oba konce. Pro otestování kvality čtení byl použit program MultiQC [6]. Testování bylo provedeno pro všechny testované genomy.

Následně dochází ke složení sekvenčních dat. Pro skládání genomů byl použit program Burrows-Wheeler Aligner [7], přičemž skládání bylo provedeno oproti referenčnímu genomu získaného z databáze NCBI. Poté byl použit program Samtools [8], za jehož pomoci byla odstraněna nenamapovaná čtení. Nakonec byla z poskládaných dat postupně vytvořena konsenzuální sekvence pro všechny genomy.

Z referenčního genomu získaného z NCBI byly vyextrahovány jednotlivé geny, které byly následně vyhledány v našich analyzovaných genomech pomocí algoritmu BLAST, jež slouží právě k nalezení podobných oblastí v biologických sekvencích. Jako výstup dostaneme soubory ve formátu FASTA, které obsahují gen, pokud byl nalezen, a to vždy pro všechny sekvence.

Pro nalezení genů, které vykazují větší míru variability, a tudíž jsou vhodné pro genotypizaci byl navrhnout algoritmus, jehož vývojový diagram je uveden na obrázku 1.



Obrázek 1: Vývojový diagram navrhnutého algoritmu.

Nejprve dojde k načtení všech souborů a podle informací o jednotlivých souborech vytřídíme ty, které mají velikost rovnou 0 bitů, což znamená, že gen z reference nebyl obsažen v analyzovaných genomech. Pro takovéto geny je vytvořena tabulka obsahující sekvenci chybějícího genu, jeho pozici v referenčním genomu a genový produkt.

Následně pracujeme již se soubory, které obsahují nalezené geny. Po načtení sekvencí vždy z jednoho souboru, dojde ke kontrole, zda byl hledaný gen přítomen ve všech genomech. Pokud ne, je takovýto soubor vyloučen z analýzy.

Poté jsou genomové sekvence v rámci jednoho genu vícenásobně zarovnané. U zarovnaných sekvencí zjišťujeme, zda jsou stejně dlouhé, popřípadě zda je nalezený gen u některého genomu výrazně zkrácen či naopak. Pokud jsou v rámci jednoho genu variabilní délky genomových sekvencí je gen vyřazen.

Pro geny, které se nachází u všech genomů a mají téměř shodnou délku je spočtena variabilita jako počet změněných pozic podělený počtem všech pozic a převeden na procentuální hodnotu. Pokud je takto vypočtená hodnota větší nebo rovna 3 %, je gen považován za dostatečně variabilní a postupuje v analýze do další části.

V posledním kroku algoritmu zkoumáme v posuvném okně s překryvem, zda je ve vybraných genech dosaženo alespoň 10 nukleotidových změn na 100 bází. Pokud ano, je takovýto gen označen jako variabilní a může být použit pro genotypizaci.

4 ZÁVĚR

Některé linie bakterie *Klebsiella pneumoniae* jsou hypervirulentní a šíří se napříč jednotlivými odděleními nemocnice. Proto je nezbytné mít k dispozici přesné a rychlé metody k typizaci a odlišení jednotlivých druhů bakterií, které v ideálním případě nevyžadují značné finanční prostředky. Zároveň je důležité zjistit, jakým způsobem se skrz nemocnici bakterie šíří a zahájit úkony pro zastavení přenosu infekce a pro předejití případné epidemie. K řešení uvedeného problému může pomoci námi navržený algoritmus, který umožňuje vybrat z genomů vysoce variabilní geny, jež bude možné přidat k provozním genům a provést typizaci za použití mini-MLST. Pro každý genom bude určen melt-typ, což povede k rozlišení mezi jednotlivými bakteriálními liniemi.

Navržený algoritmus byl realizován pomocí programového prostředí Matlab R2017a a pro sekvenční data byly použity nástroje MultiQC [6], Burrows-Wheeler Aligner [7] a Samtools [8].

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory grantu GAČR 17-01821S.

REFERENCE

- [1] OLIVE, D. Michael a Pamela BEAN. Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. *Journal of Clinical Microbiology*. 1999, 37(6), 1661-1669.
- [2] LI, Wenjun, Didier RAOULT et al. Bacterial strain typing in the genomic era. *FEMS Microbiol Reviews*. 2009, 33(5), 892-916. DOI: 10.1111/j.1574-6976.2009.00182.x.
- [3] BRHELOVA, Eva, Iva KOCMANOVA, et al. Validation of Minim typing for fast and accurate discrimination of extended-spectrum, beta-lactamase-producing *Klebsiella pneumoniae* isolates in tertiary care hospital. 2016, 1999. DOI: 10.1016/j.diagmicrobio.2016.03.010.
- [4] ANDERSSON, Patiyan, Steven Y. C. TONG, et al. Minim Typing – A Rapid and Low Cost MLST Based Typing Tool for *Klebsiella pneumoniae*. 2012. DOI: 10.1371/journal.pone.0033530.
- [5] *Klebsiella Infections: Background, Pathophysiology, Epidemiology of Klebsiellae*. Medscape. 2017.
- [6] EWELS, Philip, Mans MAGNUSSON, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016, 32(19), 3047-3048. DOI: 10.1093/bioinformatics/btw354.
- [7] LI, H. a R. DURBIN. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009, 25(14), 1754-1760. DOI: 10.1093/bioinformatics/btp324.
- [8] LI, H., B. HANDSAKER, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009, 25(16), 2078-2079. DOI: 10.1093/bioinformatics/btp352.