

Sekvenovanie DNA v laboratórnej diagnostike bakteriálnych patogénov

Barbora Markusková, Aneta Lichvariková, Hana Drahovská

Katedra molekulárnej biológie, Prírodovedecká fakulta, Univerzita Komenského v Bratislave

V klinickom mikrobiologickom laboratóriu sa v súčasnosti čoraz častejšie využívajú metódy založené na sekvenovaní DNA. Tradičným postupom používaným pri identifikácii a typizácii patogénnych baktérií je sekvenovanie samostatných lokusov, najčastejšie génu pre 16S rRNA alebo siedmich génov kódujúcich proteíny pomocou metódy MLST. V poslednom období sa však rýchlo rozvíjajú aj metódy celogenómového sekvenovania, ktoré poskytujú komplexnú informáciu o pôvodcovi ochorenia, ale vyžadujú náročné bioinformatické vyhodnocovanie získaných dát. Na porovnanie genómov baktérií možno využiť viaceré aplikácie dostupné na internete, napríklad celogenómovú MLST alebo nástroj PATRIC. Viaceré špecializované databázy umožňujú aj predikciu prítomnosti génov rezistencie proti antibiotikám.

Kľúčové slová: patogén, celogenómové sekvenovanie, identifikácia, antibiotická rezistencia

DNA sequencing in laboratory diagnostics of bacterial pathogens

Presently, methods based on DNA sequencing are frequently used in clinical microbiology laboratory. Single locus sequencing is a traditional method used for bacterial identification and typing, 16S rRNA gene and MLST are the most frequently used ones. Recently the whole genome sequencing methods were applied in the clinical microbiology. This approach provides complex information about infectious agents. However, they demand advanced bioinformatic data evaluation. Bacterial genomes could be compared by several web-based applications; the examples are the whole genome MLST approach and PATRIC server. Several specialized databases are also devoted to detection of antibiotic resistance genes in sequenced bacterial genomes.

Keywords: pathogen, whole genome sequencing, identification, antibiotic resistance

NewsLab, 2019; roč. 10 (1): 25 – 27

V klinickom mikrobiologickom laboratóriu sa na identifikáciu a typizáciu pôvodcov infekčných ochorení tradične používajú fenotypové diagnostické metódy založené na kultivácii mikroorganizmov, mikroskopických technikách a špecializovaných postupoch, ako je biotypizácia, sérotypizácia alebo stanovenie citlivosti na antibiotiká. Zároveň v súčasnosti dochádza k významnému rozvoju technológií sekvenovania DNA, čo umožňuje v praktických mikrobiologických aplikáciách čoraz viac využívať priame genotypové metódy.

Sekvenovanie jednotlivých génov

V taxonómii baktérií sa tradične využíva úplné alebo čiastočné sekvenovanie génu pre 16S rRNA zodpovedného za syntézu RNA malej podjednotky ribozómu. Ako prvý tento prístup zaviedol Carl Woese, ktorý na základe porovnania sekvencií rRNA v roku 1977 definoval *Archea* ako samostatnú taxonomickú skupinu prokaryotov líšiacu sa od klasických baktérií⁽¹⁾. 16S rRNA gén má dĺžku približne 1,5 kbp a nachádzajú sa na ňom konzervatívnejšie a variabilnejšie úseky. To umožňuje navrhnuť univerzálne primery (pre všetky baktérie alebo skupinovo špecifické), ktoré sa používajú na PCR a sekvenovanie. Pri analýze je dôležité, že získané sekvencie možno porovnávať s databázami sekvencií, a tým vyhodnotiť príbuznosť kmeňov získaných v rozličných štúdiách. Existuje niekoľko relevantných databáz ribozómových sekvencií: Ribosomal Database Project, Silva, GreenGenes. Databáza Living Tree Project (<http://www.arb-silva.de/projects/living-tree/>) obsahuje sekvencie 16S and 23S rRNA génov typových kmeňov validne opísaných druhov baktérií a archeí dôležitých na taxonomické účely⁽²⁾.

Okrem nesporných výhod má však identifikácia baktérií na základe 16S rRNA génu svoje limity. Prvou nevýhodou je vysoká konzervatívnosť 16S rRNA sekvencií, preto v niektorých prípadoch nemožno jednoznačne odlíšiť kmene patriace do príbuzných druhov, ako sú niektoré streptokoky a enterobaktérie, a tiež nie je možná diskriminácia medzi rôznymi kmeňmi rovnakého druhu. Zároveň genómy baktérií obsahujú viacero kópií rRNA operónu (napríklad genóm *E. coli* má sedem rRNA operónov), ktoré nie sú sekvenčne identické, čo znižuje presnosť pri stanovení príbuznosti kmeňov⁽³⁾.

Okrem ribozómových génov sa pri identifikácii a typizácii baktérií často využíva sekvenovanie génov kódujúcich proteíny. Takéto sekvencie možno porovnávať na úrovni nukleotidov (vysoká variabilita spôsobená degeneratívnosťou genetického kódu) alebo na úrovni aminokyselín (vyššia konzervovanosť vhodná pri porovnávaní vzdialenejších taxónov). Spa typizácia, založená na sekvenovaní variabilných oblastí *spa* génu kódujúceho stafylokokový proteín A, je často používaná metóda na stanovenie príbuznosti kmeňov *Staphylococcus aureus*⁽⁴⁾.

V roku 1998 bola na stanovenie príbuznosti kmeňov *Neisseria meningitidis* zavedená metóda multilokusovej sekvenčnej typizácie (Multi-locus sequence typing, **MLST**) a neskôr sa táto metóda začala používať aj pri typizácii iných klinicky významných baktérií⁽⁵⁾. Metóda je založená na sekvenovaní (najčastejšie) siedmich génov rozložených v rôznych častiach bakteriálneho genómu. Cieľové gény sú odlišné pre rôzne bakteriálne druhy, ale platí zásada, že sa vyberajú gény základného metabolizmu (tzv. housekeeping gény), ktoré predstavujú najstabilnejšiu časť bakteriálneho genómu.

Pri baktériách je častým javom horizontálny transfer génov medzi nepríbuznými baktériami. Preto pri sekvenovaní iba jedného lokusu, ktorý bol v predchádzajúcom období predmetom génového transferu, sa môžu dva kmene javiť ako identické, ale sekvenovaním iných častí ich genómov sa zistí, že tieto kmene nie sú príbuzné. Sekvenovaním v rámci MLST sa preto určí/eliminuje vplyv horizontálneho transferu génov medzi nepríbuznými baktériami.

Metóda MLST má oproti klasickým typizačným technikám (ako napr. PFGE) výhodu v digitálnom charaktere sekvenčných dát. Pri vyhodnocovaní príbuznosti kmeňov sa postupuje tak, že sekvencie z jedného lokusu (génu) sa navzájom porovnávajú a jednotlivé alely sa očísľujú v takom poradí, v akom sa získali (t. j. sekvencia lokusu z prvého sekvenovaného kmeňa má číslo 1, sekvencia nasledujúceho kmeňa líšiaca sa aspoň o jeden nukleotid má číslo 2 atď.). Na základe kombinácie alel zo všetkých lokusov (alelického profilu) sa kmeňu priradí výsledné číslo nazývané sekvenčný typ (ST). Vyhodnocovanie a uchovávanie MLST dát sa uskutočňuje v databázach špecifických pre jednotlivé druhy baktérií. Väčšinu z týchto databáz možno nájsť na stránke <http://pubmlst.org/databases/>, na konci roku 2018 tam bolo deponovaných 105 schém pre baktérie a 10 schém pre eukaryotické mikroorganizmy.

Pri vyhodnocovaní podobnosti medzi izolátmi sa vychádza z alelických profilov. Príbuzné kmene majú rovnaké ST alebo ST, ktoré sa líšia len v niekoľkých alelách, nepríbuzné kmene majú sekvenciu odlišnú vo väčšine lokusov. Týmto spôsobom sa príbuzné ST zhlukujú do tzv. klonálnych komplexov, ktoré zahrňujú súbor geneticky príbuzných, ale nie úplne identických baktérií⁽⁵⁾.

Metóda MLST umožňuje dobre rozlíšiť kmene bakteriálnych druhov s vysokou vnútrodruhovou variabilitou, napr. *Neisseria meningitidis*, *Campylobacter* sp., *Streptococcus pneumoniae*, *Cronobacter* sp. Na druhej strane existujú patogény napr. *Bacillus anthracis* alebo *Yersinia pestis*, ktoré sa vyznačujú vysokou genetickou homogenitou a pre ich typizáciu je metóda MLST málo diskriminatívna^(5,6).

Celogenómové sekvenovanie

Prvým bakteriálnym druhom so známym genómom bol *Haemophilus influenzae*, publikovaný v roku 1995⁽⁷⁾. Na jeho sekvenovanie bol použitý klasický prístup prípravy genómovej knižnice v plazmidových vektoroch, sekvenovanie náhodných fragmentov pomocou Sangerovej metódy a bioinformatické skladanie získaných sekvencií do výslednej molekuly, preto bol tento postup prácny, pomalý a cenovo náročný. V súčasnosti ho nahradili viaceré metódy sekvenovania novej generácie, ktoré sú založené na vysokokapacitnom paralelnom sekvenovaní DNA molekúl. Tieto metódy sú schopné generovať obrovské množstvá základných sekvenčných dát v krátkom čase a pri nízkych nákladoch na jednu sekvenovanú bázu. Cena samotného sekvenovania, ktorá sa v súčasnosti pohybuje pod sumou 100 € za genóm, už nie je limitujúcim faktorom, a tak sa celogenómové sekvenovanie stáva dostupným aj pre mnohé rutinné laboratóriá. Napríklad v USA sú od roku 2013 sekvenované genómy všetkých izolátov z humánnych klinických prípadov *L. monocytogenes*^(8,9).

Dôležitou súčasťou sekvenovania bakteriálnych genómov je bioinformatické vyhodnocovanie získaných dát. Vo

väčšine súčasných next-gen technológií je výstupom veľké množstvo krátkych navzájom sa prekrývajúcich sekvenčných čítaní. Preto prvým krokom pri analýze dát je skladať (assembly) sekvencií do čo najdlhších kontinuálnych úsekov (contigov). Zložené sekvencie, najčastejšie vo forme viacerých nespojených lineárnych sekvencií (whole genome contigs), sa v klinickom mikrobiologickom laboratóriu môžu ďalej využiť na presnú identifikáciu pôvodcu ochorenia, stanovenie génov virulencie a rezistencie proti antibiotikám a na určenie príbuznosti kmeňov pri epidemiologických štúdiách.

Pre spracovanie údajov z next-gen sekvenovania je potrebné mať k dispozícii bioinformatické nástroje, ktoré sú schopné spracovať veľké množstvo dát vo veľkých súboroch kmeňov. Je žiaduce, aby tieto nástroje boli jednoduché na obsluhu a dostupné pre širokú odbornú verejnosť bez hlbších znalostí z informatiky a programovania. Jednou z možností pri vyhodnocovaní je využitie dostupných webových programov, napríklad pomocou aplikácií prístupných na stránke Center of genomic epidemiology (<http://www.genomicepidemiology.org/>). Program SpeciesFinder vyhľadáva v čiastočne sekvenovaných genómoch 16S rRNA gény a na základe najpríbuznejšej sekvencie zaradí kmeň do bakteriálneho druhu. Iný spôsob na druhovú identifikáciu je K-mer-Finder, ktorý porovnáva v genómových sekvenciách prítomnosť navzájom sa prekrývajúcich k-merov s dĺžkou 16 bp bez zohľadnenia ich lokalizácie na chromozóme a kmeň zaradí na základe najpríbuznejšieho kmeňa z databázy. Tento prístup je veľmi rýchly a dostatočne presný. Ďalšie programy umožňujú zistiť z celogenómových sekvencií niektoré dôležité vlastnosti kmeňov: napr. MLST, spa typ, sérotyp, prítomnosť plazmidov, génov virulencie a génov rezistencie proti antibiotikám⁽¹⁰⁾.

Iným spôsobom porovnávanie celogenómových sekvencií je využitie prístupu MLST, tzv. celogenómové MLST (wgMLST) pomocou softvérového nástroja BIGSdb (Bacterial Isolate Genome Sequence Database)⁽⁶⁾. Tento program extrahuje z celogenómových sekvencií bakteriálnych izolátov sekvencie jednotlivých génov, sekvenciám priradí čísla alel podobným spôsobom ako v klasickej MLST, a potom porovnáva kmene na základe zhody alebo rozdielov v alelách jednotlivých génov. Program je vytvorený hierarchicky a možno v ňom ľubovoľne zvoliť súbor génov, ktoré sa použijú pri analýze. Rovnakú databázu preto možno využiť na druhovú identifikáciu pôvodcu ochorenia pomocou analýzy ribozómových génov (rMLST), na zaradenie kmeňov do sekvenčných typov a klonálnych komplexov pomocou klasickej MLST aj na typizáciu kmeňov s vysokou diskrimináciou (napr. pri sledovaní vývoja epidémií v rámci jednej nemocnice), keď sa porovnávajú sekvencie všetkých spoločných génov (wgMLST)^(6,11).

Komplexným nástrojom na analýzu dát z celogenómového sekvenovania prokaryotických organizmov je webová aplikácia PATRIC (The Pathosystems Resource Integration Center)⁽¹²⁾. PATRIC umožňuje online analýzu bakteriálnych genómov jednotlivých používateľov a ich porovnanie s verejne dostupnými údajmi. Program je primárne zameraný na patogény, preto má veľkú perspektívu využitia v klinickom výskume. Jednou z významných vlastností je vyhľadávač génov rezistencie proti antibiotikám, ktorý využíva databázy so známymi mechanizmami rezistencie ako CARD (Comprehensive Antibiotic Resistance Database) a NDARO (National Database

of Antibiotic Resistant Organisms). PATRIC ďalej umožňuje identifikovať gény kódujúce virulénne faktory a transportéry, modelovať metabolické dráhy a porovnávať proteómy jednotlivých organizmov. Na identifikáciu baktérií možno využiť najbližšieho príbuzného organizmu v databáze a vyhodnotiť fylogenetické vzťahy v rámci vlastnej zbierky mikroorganizmov. Komplexnosť systému PATRIC dopĺňajú aj aplikácie umožňujúce základné spracovanie sekvenčných dát, t. j. skladanie (assembly) krátkych čítaní do contigov a anotácia sekvencií.

Stanovenie citlivosti na antibiotiká

Stanovenie citlivosti pôvodcu ochorenia k antimikrobiálnym látkam je jednou zo základných úloh klinického mikrobiologického laboratória. V súčasnosti sa testovanie antibiotickej citlivosti určuje takmer výhradne fenotypovými metódami na základe rastu baktérií v prítomnosti antibakteriálnych látok. Zároveň však súčasné technológie umožňujú odvodiť rezistenciu kmeňov pomocou vyhľadávania génov rezistencie v bakteriálnej DNA. Identifikácia génov rezistencie je vhodná aj na overenie citlivosti kmeňov, ktorá bola určená fenotypovo, a tiež na identifikáciu kmeňov so slabou expresiou v podmienkach *in vitro*. Detekcia týchto génov sa tradične robí pomocou metód PCR alebo DNA microarray⁽¹³⁾. Výhodou analýzy celogenómových sekvencií je, že v princípe poskytujú všetky údaje na určenie molekulárnych mechanizmov antimikrobiálnej rezistencie, a to aj bez predchádzajúcej znalosti fenotypu. Na predikciu génov rezistencie existujú viaceré špecializované databázy. Jednou z najpoužívanejších je databáza ARG-ANOT (Antibiotic Resistance Gene-ANNOTation), ktorú možno stiahnuť a využívať pomocou lokálneho softvéru na analýzu sekvencií⁽¹⁴⁾. Druhou možnosťou je využitie webových aplikácií,

ako je ResFiner⁽¹⁵⁾ alebo CARD⁽¹⁶⁾, ktoré sú pravidelné aktualizované. Stanovenie rezistencie priamo z bakteriálnych celogenómových sekvencií zatiaľ stále komplikujú nedostatočné vedomosti o všetkých genetických zmenách vedúcich k zníženej citlivosti bakteriálnych kmeňov na antibiotiká. Najväčším problémom pri rutinnej implementácii WGS je aj nedostatok automatizovaných, používateľsky prívetivých interpretačných nástrojov, pomocou ktorých by bolo možné rýchlo a jednoducho poskytnúť klinicky významné informácie⁽¹⁷⁾. Rýchly vývoj technológií v tejto oblasti je predpokladom efektívneho využitia v blízkej budúcnosti.

Záver

Techniky celogenómového sekvenovania majú veľký potenciál stať sa pevnou súčasťou rutinných vyšetrení v klinickej mikrobiológii, pretože umožňujú detailnú analýzu pôvodcov infekčných ochorení pomocou univerzálneho prístupu. Ich výhodou je aj možnosť identifikácie patogénov v zmesových vzorkách bez predchádzajúcej kultivácie. Okrem identifikácie a typizácie mikroorganizmov sú tieto techniky principiálne vhodné na určenie rezistencie proti antibiotikám, aj keď v súčasnosti ešte nemožno jednoznačne odvodiť fenotyp mikroorganizmu zo známej genómovej sekvencie. V budúcnosti je predpokladom ďalšie zlepšovanie technológií sekvenovania DNA, skrátenie času a zníženie ceny analýz, takže tieto postupy budú čoraz viac nahrádzať tradičné metódy závislé od kultivácie, používané v klinických laboratóriách.

Grantová podpora: Táto publikácia vznikla s podporou grantov APVV-16-0168, APVV-16-0119 a projektu Biomakro 2 (ITMS 26240120027).

LITERATÚRA

1. Woese CR. Bacterial evolution. *Microbiol Rev* 1987; 51(2): 221-271.
2. Rossello-Mora R, Amann R. Past and future species definitions for Bacteria and Archaea. *Syst Appl Microbiol* 2015; 38(4): 209-16.
3. Turcovsky I, Kunikova K, Drahovska H, Kaclikova E. Biochemical and molecular characterization of *Cronobacter* spp. (formerly *Enterobacter sakazakii*) isolated from foods. *Antonie Van Leeuwenhoek* 2011; 99(2): 257-69.
4. Sabat AJ. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* 2013; 18(4): 20380.
5. Urwin R, Maiden MCJ. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* 2003; 11(10): 479-487.
6. Maiden MC, Jansen van Rensburg MJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013; 11(10): 728-36.
7. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269(5223): 496.
8. Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin Microbiol Inf* 2018; 24(4): 355-360.
9. Stasiewicz MJ, den Bakker HC, Wiedmann M. Genomics tools in microbial food safety. *Current Opinion in Food Science* 2015; 4: 105-110.
10. Larsen MV, Cosentino S, Lukjancenko O, et al. Benchmarking of Methods for Genomic Taxonomy. *J Clin Microbiol* 2014; 52: 1529.
11. Kadlicekova V, Kajsik M, Soltys K, et al. Characterisation of *Cronobacter* strains isolated from hospitalised adult patients. *Antonie van Leeuwenhoek* 2018; 111(7): 1073-1085.
12. Wattam AR, Davis JJ, Assaf R, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 2017; 45(D1): D535-D542.
13. Batchelor M, Hopkins KL, Liebana E, et al. Development of a miniaturised microarray-based assay for the rapid identification of antimicrobial resistance genes in Gram-negative bacteria. *Int. J. Antimicrob. Agents* 2008; 31(5): 440-451.
14. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014; 58(1): 212-220.
15. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; 67(11): 2640-4.
16. McArthur AG, Waglechner N, Nizam F, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013; 57(7): 3348-57.
17. Ellington MJ, Ekelund O, Aarestrup FM, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect* 2017; 23(1): 2-22.



doc. RNDr. Hana Drahovská, PhD.

Katedra molekulárnej biológie
Prírodovedecká fakulta
Univerzita Komenského v Bratislave
Ilkovičova 6, 842 15 Bratislava
e-mail: hana.drahovska@uniba.sk