

**Ze současné medicíny**

## **Genomika bakterií: možnosti, omezení a aplikace**

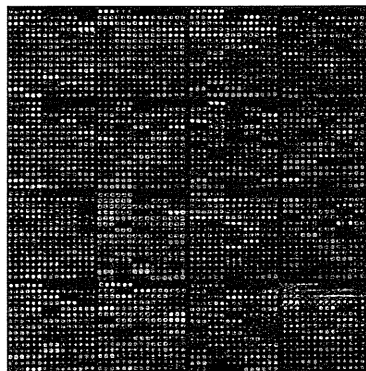
DAVID ŠMAJS

Bakteriální genomikou v nejširším slova smyslu rozumíme odvětví genetiky, které se zabývá studiem bakterií na úrovni celého jejich genomu, tedy na úrovni jejich kompletní sekvence DNA. Rozvoj bakteriální genomiky jako nástroje pro studium bakterií umožnily až pokroky v automatické sekvenaci DNA a v počítačovém zpracování sekvenčních dat. Bakteriální genomika tak – alespoň částečně – těží z projektu HUGO, z projektu sekvenace lidského genomu, který byl nastartován několik let před uveřejněním prvního bakteriálního genomu. Kompletní sekvence bakteriálního genomu (pořadí všech bází v bakteriální DNA) je jak prvním krokem genomického přístupu, tak i krokem, který nezbytně předchází použití metod dalších. Se znalostí kompletní sekvence DNA je možno využít metody funkční a komparativní genomiky. Funkční genomika se zajímá o to, jak zjistit funkci genů, o kterých zatím není nic známo, jak tyto geny fungují za změněných podmínek a jak navzájem interagují jejich genové produkty. Komparativní genomika pak srovnává genomy příbuzných bakterií, které se zpravidla liší v nějakém podstatném znaku, např. jeden kmen je virulentní a druhý nikoliv. Oba tyto základní přístupy se navzájem doplňují a překrývají: funkci genu lze odhalit na základě údajů získaných komparativní genomikou a naopak funkční genomika může upozornit na důležitý genetický rozdíl mezi kmeny. Oba přístupy přitom používají i řadu společných experimentálních přístupů.

Kromě genomové sekvenace jakožto základní metody používá genomika řadu postupů dalších. Je to analýza *in silico*, počítačové vyhodnocení získané sekvence, které se používá pro identifikaci genů, hledání homologií mezi genomy dalších bakterií, předpovědi genových funkcí, buněčné lokalizace genových produktů, identifikaci regulačních míst, promotorů, genetické „heterogenity“ na chromozomu, identifikace pohyblivých elementů v DNA, opakujících se sekvencí atp. Jak rostou naše znalosti v bakteriální genetice a genomice, zpřesňují se i počítačové algoritmy, které geny odhalují a předpovídají jejich funkce i to, jak jsou tyto geny regulovány.

Dalším významným nástrojem studia bakteriálních genomů je technika DNA čipů, která umožňuje sledování exprese (tedy funkční aktivity genů) všech genů daného organismu v jednom jediném experimentu. Podstatou metody je hybridizace DNA (Southernova hybridizace), která na rozdíl od

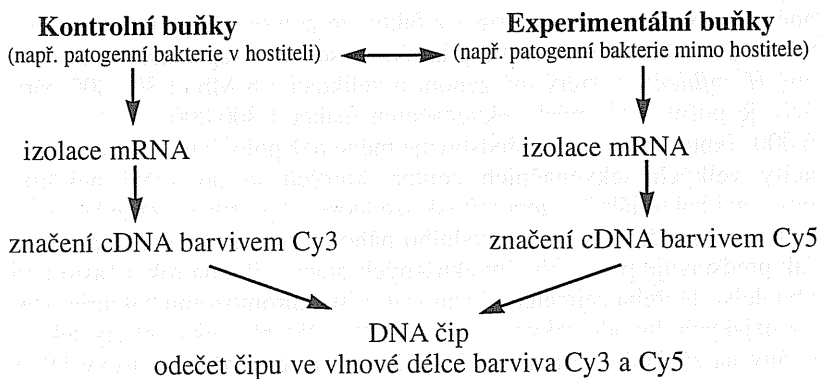
experimentů genetických neprobíhá pro jeden nebo několik genů, nýbrž pro všechny, ev. téměř všechny geny daného organismu. Tento pokrok umožnila miniaturizace a nově vyvinuté techniky povrchové modifikace skla, které umožňují efektivní vazbu DNA (hybridizace jednotlivých genů probíhá na kruhové ploše o průměru v řádu desetin milimetru). Celková plocha čipu (který je zpravidla „vytištěn“ na podložním mikroskopickém skle) je pak přibližně  $1 \text{ cm}^2$  a čip obsahuje až 40.000 hybridizačních plošek (obr.1).



Obr. 1. DNA čip obsahující 3.600 hybridizačních ploch reprezentujících všechny geny původce syfilis, *Treponema pallidum* subsp. *pallidum*.

Tato metodika umožňuje nejen sledování změn v genové expresi bakterií žijících za různých podmínek (obr. 2), ale i rychlé srovnávání genomů blízké příbuzných bakteriálních kmenů. Navíc je možné sledovat i odezvu lidských genů v populacích buněk zasažených bakteriální infekcí a v populacích buněk zneškodňujících infekci. Předběžné výsledky potvrzují, že se během bakteriální infekce mění nejen exprese genů patogenního mikroorganismu, ale že i hostitelský organismus reaguje na invadující bakterie „zapínáním a vypínáním“ některých svých genů.

Historie bakteriální genomiky je poměrně krátká, začíná v roce 1995, kdy byla uveřejněna kompletní sekvence prvního bakteriálního genomu. Touto bakterií byl *Haemophilus influenzae*, bakterie, která způsobuje celou řadu infekčních onemocnění u člověka, od onemocnění dýchacích cest u dospělých, které doprovází primární virové onemocnění, až po závažné a život ohrožující onemocnění u dětí. Dnes je dokončeno nebo těsně před dokončením více než 200 bakteriálních genomů a další sekvenované genomy je možno brzy očekávat.



Obr. 2. Schéma hybridizačního experimentu na DNA čipu. Informační RNA (mRNA) vzniká přepisem genu a počet mRNA molekul v buňce je tak mírou funkční aktivity genu. cDNA vzniká reverzním přepisem z mRNA v podmínkách *in vitro* a její množství přesně odpovídá množství původní mRNA. Při tomto přepisu se navíc cDNA „nabarví“ vhodným fluorescenčním barvivem. Pro každou hybridizaci se odečte signál z kanálu Cy3 a Cy5 a poměr intenzity fluorescenčního záření obou kanálů představuje poměr koncentrací mRNA v kontrolních a experimentálních buňkách (a tím relativní změnu v expresi daného genu).

Jak se vlastně sekvenace bakteriálního genomu provádí? Princip dnes používané genomové sekvenace spočívá ve fragmentaci genomu na poměrně malé kousky (o délce cca 500 – 2.000 nukleotidů) a jejich klonování (namnožení) do vektorů (vektory jsou modifikované plasmidy vhodné pro klonování DNA). Tento postup se označuje jako tzv. konstrukce DNA knihovny a jednotlivé klony DNA knihovny jsou pak sekvenovány z obou konců vložené (klonované) DNA. Protože klonování jednotlivých fragmentů DNA do vektoru je náhodné, je také náhodné složení DNA knihovny z hlediska fragmentů DNA, které obsahuje. Stejně tak je i náhodné, které fragmenty DNA bakterie budou sekvenovány. Protože se klony navzájem překrývají, je třeba sekvenovat mnohem více klonů než v případě, kdy bychom znali pozici jednotlivých klonů na chromozomu bakterie. Na začátku sekvenačního projektu se odhadne, kolik sekvenačních reakcí je třeba provést, a výsledná sekvence genomu se sestavuje až po dokončení potřebného počtu sekvenačních reakcí. Z metodických důvodů je délka čtení sekvenační reakce omezena; průměrná délka čtení jedné reakce je okolo 500 nukleotidů. (Ve stadiu experimentálních zkoušek jsou už sekvenační postupy, které toto omezení nemají a které – pokud budou úspěšné – způsobí revoluci v sekvenaci nejen genomů bakteriálních, ale i genomů eukaryontních organismů.) Množství sekvenačních reakcí, které je třeba provést k sekvenaci genomu bakterie, lze odvodit z předpoklá-

dané velikosti genomu bakterie a z faktu, že pomocí náhodné sekvenace je třeba přečíst každý nukleotid přibližně desetkrát. Například pro zmiňovaný *H. influenzae*, který má genom o velikosti 1,8 Mb (1.800.000 párů bází), je počet potřebných sekvenačních reakcí  $1.800.000 : 500 \times 10 = 36.000$ . Tento počet reakcí představuje méně než polovinu jednodenní kapacity velkých sekvenačních center, kterých je po světě několik. Sekvenaci bakteriálního genomu tak lze teoreticky provést za jeden jediný den. Dokončovací fáze původního náhodného sekvenačního přístupu však představuje práci pro tým zkušených pracovníků na rok a často i na dobu delší. Je třeba zejména sekvenovat místa chromozomu bakterie, která z nějakých důvodů sekvenována nebyla (některé úseky nebyly sekvenovány na základě prostého statistického rozložení, některé úseky DNA mohou vést po naklonování do vektoru k produkci proteinů toxických pro bakterie použité ke konstrukci knihovny atp.). Další komplikací je sestavení jednotlivých sekvencí do správného pořadí podél chromozomu bakterie, problémy způsobují zejména opakující se identické sekvence (repetitivní sekvence), podobné sekvence vmezežené do různých částí chromozomu (inzerční sekvence), duplikované geny a úseky DNA atd.

Více než tři čtvrtiny dnes sekvenovaných bakteriálních genomů představují medicínsky významné bakterie, tzv. bakterie patogenní, které ovšem tvoří pouze nepatrnou část bakteriálních druhů žijících na Zemi. Důvody výběru jsou zřejmé – sekvenace bakteriálního genomu představuje finančně náročný projekt a je zde zřejmá snaha sekvenovat takové bakterie, u kterých se předpokládá, že by se výdaje mohly vrátit ve formě nových diagnostických postupů, vakcín anebo terapeutik. Ačkoliv většina sekvenovaných genomů jsou genomy patogenů, u celé řady genomů je motivace přece jen odlišná. U archaebakterií (mezi nimiž není zatím znám žádný bakteriální patogen) je hlavní motivací výzkumu schopnost těchto bakterií žít za extrémních podmínek (např. za vysokých teplot – i nad 100 °C, v prostředí s vysokou koncentrací solí nebo v prostředí radioaktivním). Tyto bakterie přitahují pozornost průmyslu zejména kvůli slibnému použití jejich proteinů, např. jejich enzymů – třeba pro výrobu ekologicky šetrných pracích prášků. Zajímavou bakterií v tomto ohledu je *Deinococcus radiodurans*, bakterie mimořádně odolná k ionizujícímu záření, vysušení, UV záření, oxidačním látkám atp. Oproti běžné bakterii *Escherichia coli*, je *D. radiodurans* v růstové fázi rezistentnější vůči ionizujícímu záření nejméně padesátkrát a během stacionární fáze je odolnější dokonce stokrát. Ukazuje se, že klíčem k těmto zajímavým vlastnostem je přítomnost několika genomových kopií v buňce spolu s mimořádně efektivním systémem reparace (opravy) DNA. Genom původce skvrnitého tyfu, *Rickettsia prowazekii*, byl sekvenován zejména proto, že má potenciálně nejbliže ke genomu mitochondrií, organel eukaryontních buněk. Rickettsiální genom tak nepřímo podporuje endosymbiotickou teorii, která předpokládá, že některé organely eukaryontních organismů se vyvinuly

z buněk prokaryontních. Bakteriální genomika zde úzce prorůstá s evoluční biologii a mapuje evoluční události ve vývoji živých organismů.

Velikost bakteriálních genomů kolísá v rozmezí více než jednoho řádu. Nejmenším známým genomem je genom *Mycoplasma genitalium* (tato bakterie způsobuje urogenitální infekce) o velikosti pouhých 0,6 Mb a největším sekvenovaným genomem je genom půdní bakterie fixující plynný dusík *Mesorhizobium loti* (8 Mb). Je přitom zajímavé, že velikost genomu přímo úměrně koresponduje s počtem genů dané bakterie. Je to dáno tím, že u bakterií se nesetkáváme (na rozdíl od genomů eukaryontních) s rozsáhlejšími intra- a intergenovými úseky DNA. Zdá se, že efektivita uložení genů v bakteriální DNA je společným znakem všech bakterií. Současně ještě platí, že patogenní bakterie adaptované na úzký okruh hostitelů mají nejmenší genomy. Příkladem může být genom bakterie *Treponema pallidum*, který je jen 1,1 Mb velký. Tato spirocheta způsobuje syfilis, onemocnění, které se plně manifestuje jen u člověka. Výjimky z pravidla o úzké korelaci mezi velikostí bakteriálního genomu a počtem funkčních genů znamenají, že se něco zajímavého děje s genomem dané bakterie. Relativně malý počet genů najdeme např. u *Mycobacterium leprae*, původce malomocenství, kde genom patrně prochází redukční evolucí. Tato bakterie se pravděpodobně evolučně adaptuje na jednoho hosta a ztrácí nejdříve funkční geny (hromadí mutace v nepotřebných genech) a později ztratí i samotné úseky DNA, ve kterých tyto nefunkční geny leží. Opačným příkladem je genom bakterie *Aeropyrum pernix*, jejíž genom obsahuje relativně vyšší počet genů, než by odpovídalo velikosti chromozomu. Genom této bakterie obsahuje vysoké množství zdvojených (duplikovaných) genů. To snad může souviset s tím, že tato bakterie je schopna žít za extrémních teplot (90-100 °C) v aerobních podmínkách.

Zatím nejmenší známý genom *M. genitalium* obsahuje jen 517 genů. Je zajímavé, že tento genom lze ještě experimentálně zmenšovat, aniž by tato bakterie ztratila životaschopnost za příznivých kultivačních podmínek. Odhaduje se, že minimální počet genů, které dostačují za daných laboratorních podmínek (byť velmi příznivých) k životu, je pouhých 265-350 genů. Oproti relativně malým genomům obsahují větší genomy více paralogních (příbuzných) genů a umožňují bakteriím, které je nesou, růst za širšího spektra různých podmínek okolního prostředí (umožňují více „životních stylů“).

Odhaduje se, že každý genom obsahuje 20-30 % nových genů, které se nevyskytují v žádného dalšího bakteriálního druhu; toto číslo je překvapivě konstantní u téměř každého nově sekvenovaného bakteriálního genomu a tak se někdy neformálně označuje jako „první genomický zákon“. Na základě těchto předpokladů lze odhadnout, že už teď je v databázích okolo 100.000 jedinečných bakteriálních genů a v nich je ukryto obrovské množství informace o prostředí a životě na této planetě. Přitom počet se-

kvenovaných bakterií je zcela zanedbatelný vzhledem k celkovému počtu bakterií, které se vyskytují na Zemi (např. jeden gram půdy obsahuje 5.000 – 38.000 různých druhů bakterií!). Současně se ukazuje, že genomy bakterií jsou velmi dynamické a že jsou tvořeny genetickým materiálem různého původu. Jsou doslova poskládány z kousků DNA pocházejících z nejrůznějších jiných bakterií, virů a snad i eukaryontních buněk. Tuto skutečnost vystihuje „druhý genomický zákon“, který říká, že genomy jsou mozaiky různorodého genetického materiálu.

Bakteriální genomika už začíná přinášet první výsledky. Např. pro bakterii *Helicobacter pylori* (jíž je mimochodem infikována asi polovina celosvětové lidské populace), která způsobuje řadu různých závažných onemocnění od asymptomatické gastritidy (bezpriznakového zánětu žaludku) po žaludeční vředy a karcinom žaludku, bylo prokázáno, že stupeň virulence je geneticky podmíněn, a to kmenově specifickou genovou výbavou. Podobně pro streptokoky skupiny A, které způsobují různé infekce od zánětů hrtanu až po velmi závažné invazivní onemocnění – nekrotizující fasciitidu (tyto streptokoky jsou někdy označovány jako tzv. masožravé bakterie), byl prokázán fágový (virový) horizontální přenos jako určující faktor v evoluci streptokokálních virulentních kmenů. Je přitom zajímavé, že u virulentních kmenů je detegována nejen přítomnost genů (kódujících tzv. faktory virulence) navíc oproti kmenům nevirulentním, ale také někdy zjišťujeme absenci určitých genů. Příkladem může být genom patogenní bakterie *Escherichia coli* O157:H7, způsobující epidemie průjmových onemocnění. Ve srovnání s laboratorními (nepatogenními) kmeny *E. coli*, chybí této bakterii asi 500 genů (má však asi 900 dalších genů navíc). Ukazuje se tedy, že jak geny přítomné navíc, tak i geny, které chybí ve srovnání s laboratorní *E. coli*, se mohou podílet na nárůstu virulence *Escherichia coli* O157:H7. Komparativní genomika u původce tuberkulózy, *Mycobacterium tuberculosis*, odhalila u různých kmenů způsobujících závažné plicní projevy tuberkulózy korelaci mezi sníženou schopností tvořit plicní kavítace a velikostí deletované (chybějící) chromozomální DNA. Všechny tyto případy ukazují, že schopnost bakterie vyvolávat onemocnění (virulence) je podmíněna geneticky a je metodami funkční a komparativní genomiky poměrně rychle detegovatelná. Odpověď na zásadní otázku, které genetické rozdíly jsou rozhodující pro virulenci bakterií, však ještě většinou neznáme. Navíc je tu možnost, že rozdíly (alespoň některé) mezi virulentními a nevirulentními bakteriemi nejsou primárně způsobeny nedostatkem či prezencí celých genů, ale spíše jsou způsobeny změnami mnohem jemnějšími, t.j. nukleotidovými změnami, které jsou obtížněji detegovatelné. Např. genomy bakterií *T. pallidum* ssp. *pallidum* a *T. pallidum* ssp. *pertenue*, které způsobují odlišná onemocnění (syfilis a yaws), liší se jen ve čtyřech relativně krátkých úsecích DNA (v řádu stovek nukleotidů) a asi v 500-750 nukleotidových změnách. Určení, který genetický rozdíl je rozhodující pro odlišnou

patogenezi uvedených onemocnění, je však úkol klíčový a také podstatně komplikovanější než je prostá identifikace genetických rozdílů. Je to úkol pro řadu specializovanějších genomických metod a pro klasická molekulárně-biologická zkoumání.

Jaké jsou možnosti aplikací výsledků zjištěných pomocí metod bakteriální genomiky?

První významnou aplikací je vývoj nových rychlých diagnostických testů, které jsou na základě odlišností v genetické výbavě schopny selektivně detegovat jednotlivé bakteriální kmeny, poddruhy a druhy a jejich vlastnosti (např. rezistenci a citlivost na antibiotika), a zlepšit tak odhad vývoje onemocnění a umožnit nasazení odpovídající léčby. Předpokládá se, že půjde zejména o diagnostiku založenou na technice řetězové polymerázové reakce (PCR), popř. o detekci na speciálních DNA čipech. Je tak například možné diagnostikovat odlišná onemocnění, jejichž původce nebylo možno identifikovat na základě mikroskopických a sérologických technik.

Druhou významnou aplikací genomického přístupu je vývoj vakcín. Znalost sekvence všech genů daného organismu umožňuje použít systematické klonování všech genů do rekombinantních plasmidů a jejich systematické testování na schopnost vyvolávat imunitní odpověď u infikovaných jedinců. Výsledkem takového skríningu je několik kandidátních genů, které jsou konzervované v širokém spektru patogenních kmenů a jejichž produkty vyvolávají u hostitele protektivní imunitu. Systematické testování genů původce meningitid a septikémií, *Neisseria meningitidis*, identifikovalo 5 genů kódujících proteiny s kompletní evoluční konzervací napříč spektrem virulentních kmenů a indukujících protektivní imunitu. Lze předpokládat, že v dohledné době bude vyvinuta vakcína, která bude obsahovat kombinace těchto několika proteinů (antigenů), což pomůže zabránit vzniku bakteriálních mutant rezistentních na vakcinaci. Podobné projekty jsou v běhu pro další patogeny, např. pro bakterie *T. pallidum*, *R. typhi*, *Francisella tularensis* (původce tularémie) a *Bacillus anthracis* (původce antraxu).

Třetí významnou aplikací genomického přístupu je vývoj nových terapeutik se selektivním zaměřením na nemoc vyvolávající bakterii. Tato aplikace je však vázána na identifikaci klíčových faktorů virulence, které dnes ještě nejsou povětšinou známy. Proti nim je třeba vyvinout selektivní inhibitory. Spolu s přesnou diagnostikou by tato terapie měla minimální vedlejší účinky a nenarušovala by křehkou rovnováhu komensálních bakterií v lidském těle.

Závěrem je možno shrnout, že možnosti bakteriální genomiky jsou nebyvale široké, a to i přes jistá omezení, která z nynějších metodik vyplývají. Je však zřejmé, že možné aplikace bakteriální genomiky budou výrazně stimulovat její další rozvoj. A to tím spíše, že se bakteriální genomika teprve rodí a že je na svém samém začátku.