Sekvenování DNA

Sekvenování (čtení) DNA je určení pořadí nukleotidů (A,T,C,G) v molekule DNA. Molekula DNA byla poprvé izolována Friedrichem Miescherem již v roce 1869 (a již tehdy tušil, že by mohla hrát roli v dědičnosti), ale její struktura je známa teprve od roku 1953. První snahy o její čtení začaly krátce poté, ale teprve v 70. letech došlo k prvním průlomův v sekvenování DNA. V první polovině 70. let bylo postupně přečteno několik fragmentů DNA o délce maximálně desítek bp; teprve v druhé polovině 70. let dva týmy nezávisle na sobě vyvinuly skutečné sekvenační metody (Sangerova a Maxam-Gilbertova), obe metody byly publikované v roce 1977. Prvním kompletně přečteným DNA virem byl bakterofág (virus napadající bakterie) Phi X174 o délce 5386 bp. Sangerova metoda byla využívána až do konce 20. století, kdy nastoupila celá řada metod tzv. nové generace (Next Generation Sequencing, NGS) – vývoji techto metod velmi pomohl Human Genom Project (HGP), který běžel od roku 1990 do cca 2003 (některé jeho součásti běží doteď). Rozvoj NGS je skutečně enormní – nasekvencovat celý lidský genom trvalo v rámci HGP cca 10 let a stálo 3 miliardy dolarů, dnes je přečtení celého lidského genomu trvá a stojí řádově méně. V rámci HGP byla sekvencovaná DNA získána od řady anonymních dárců – prvním kompletně osekvencovaným konkrétním jedincem byl James Watson (jeden z objevitelů DNA), a to v roce 2007.

Maxam-Gilbertova metoda je založena na chemické modifikaci DNA a následném štěpení na specifických nukleotidech. Vzorek krátké sekvence DNA je označen radioaktivním fosforem na 5’ nebo 3‘ konci a dále je rozdělen na fragmenty, které jsou dále vystaveny chemickému působení. Každý z použitých enzymů specificky štěpí DNA v místě rozpoznání jednoho z čtyř nukleotidů. Fragmenty ze čtyř reakcí jsou následně uspořádány vedle sebe na polyakrylamidovém gelu a podrobeny elektroforéze. Pro vizualizaci je gel po ukončení elektroforézy vystaven γ-záření, což způsobí zesvětlení těch částí sekvence DNA, která byla na začátku označena radioaktivním fosforem. Výsledkem reakce je tedy soubor fragmentů, jejichž délka by se měla vzájemně lišit o jediný nukleotid.

Sangerova metoda se stala klasickou metodou první generace, především pro svou praktičnost, jednoduchost, škálovatelnost, možnosti automatizace, ... Tato metoda je často označována také jako dideoxy sekvenace, enzymová sekvenace, plus a mínus sekvenace a další. Jedná se o nejčastěji používanou sekvenační technologii založenou na elektroforéze a stala se základem pro Human Genome Project. Je založena na sekvenaci pomocí detekce ukončení prodlužujícího se vlákna DNA. K tomuto ukončení vlákna DNA se využívá dideoxynukleotid (ddNTP). Na začátku experimentu je potřeba příprava tzv. DNA knihovny. Řetězec DNA je sestřižen na kratší úseky, naklonován do DNA vektorů a amplifikován in vivo. Amplifikace probíhá v bakteriálních buňkách, ze kterých se následně extrahují plazmidy nesoucí klonované fragmenty. ddNTP obsahují na 3‘ uhlíku vodík hydroxylové skupiny; jsou-li tyto modifikované nukleotidy začleněny do sekvence, znemožní vytvoření fosfodiesterové vazby s dalším nukleotidem, čímž ukončuje prodlužující se řetězec DNA. V prvních verzích této metody reakce probíhá ve čtyřech oddělených nádobách, do kterých je přidána reakční směs obsahující primery, templát, DNA fluorescenčně značené dNTP a přidané ddNTP. V každé ze čtyř reakčních směsí je přítomen jeden charakteristický ddNTP – ddATP, ddCTP, ddGTP nebo ddTTP. Aby mohla zároveň probíhat i klasická syntéza DNA, jsou ddNTP ve směsi v nízké koncentraci (poměr ddNTP : dNTP zpravidla odpovídá 1:100). Výsledkem reakce jsou fragmenty DNA o různé délce, zakončené fluorescenčně značenými ddNTP.

Fragmenty jsou denaturovány a roztříděny podle velikosti elektroforézou na polyakrylamidovém gelu (každá reakční směs má svou dráhu). Po ozáření vzniká obraz s typickými proužky, ze kterých lze odvodit sekvenci DNA. Novější modifikace Sangerovy metody, kapilární elektroforéza, umožňuje sekvenaci v jedné reakci. Každý ddNTP (popř. primer) je značen jiným fluorescenčním barvivem a elektroforéza probíhá v jedné skleněné kapiláře. I současné automatizované analyzéry jsou založeny na Sangerově metodě, např. Applied Biosystems 3730xl DNA Analyzer dokáží přečíst Sangerovou metodou 2 mil. bp/den. Sangerova metoda je nejpoužívanější a nejpřesnější sekvenační metodou. Nevýhodou oproti NGS je časová a finanční náročnost této metody. Nové metody navíc nevyžadují klonování DNA do bakteriálních vektorů.

První z NGS metod je metoda *MPSS - Massively Parallel Signature Sequencing*. Je to základní sekvenační metoda nové generace, která na základě počtu jednotlivých mRNA molekul, produkovaných každým genem, analyzuje úroveň exprese těchto genů ve vzorku. Technologie MPSS řeší dvě zásadní otázky: fyzické oddělení jednotlivých fragmentů nukleových kyselin a vlastní čtení sekvence. První otázka je řešena pomocí emulzní PCR (emPCR) nebo amplifikace ve shlucích. Při emPCR dochází k vytvoření tzv. mikroreaktorů, v nichž jsou uzavřeny všechny potřebné komponenty pro PCR (nukleotidy, enzym polymeráza, primery). Na mikrokuličky je navázaný fragment nukleové kyseliny. Při emPCR dochází ke klonální mplifikaci – všechny molekuly DNA vytvořené v mikroreaktoru pochází z jediné molekuly templátu. Druhá otázka – čtení sekvence – je u dostupných zařízení řešena dvěma metodami: sekvencováním syntézou („sequencing by synthesis“) a sekvencováním založeném na ligaci („ligation based sequencing“). Označené PCR produkty (amplikony) získané z cDNA jsou amplifikovány tak, aby každá příslušná mRNA molekula dávala asi 100000 amplikonů s jedinečnou značkou. Značky se používají k upevnění amplikonů mikrokuličkami. Po několika kolech ligace sekvence se pomocí restrikční endonukleázy stanovuje posloupnost značek. Na každé kuličce je identifikováno asi 16 - 20 bp (kvalitní sekvence běžně obsahuje 17 bp). Během jednoho experimentu je získáno až jeden milion sekvencí. Každé označení sekvence (MPSS značka) v datovém souboru MPSS se srovnává se všemi ostatními značkami a identické značky jsou počítány. Úroveň exprese jakéhokoli genu se vypočítá vydělením počtu označení určitého genu celkovým počtem označení pro všechny mRNA, přítomné v datovém souboru.

Pyrosekvenace je jednou z prvních sekvenačních metod nové generace, která se začala rozvíjet jako alternativa ke klasické Sangerově metodě. Již v roce 1987 P. Nyrén navrhl, jak lze monitorovat aktivitu DNA polymerázy pomocí bioluminiscence. Trvalo téměř jedenáct let, než byla pyrosekvenace sama o sobě uvedena do praxe. V roce 2000 byla založena společnost 454 Life Siences (http://454.com), která v květnu roku 2005 se spojila s Roche Diagnostics. Tato firma měla zajistit sekvenační přístroje a činidla. Technologie firmy Roche FLX Genome Sequencer (FLX GS) byla první komerčně dostupné zařízení pro masivní paralelní sekvencování. V roce 2011 přichází na trh další FLX GS série Titanium, který umožnil zlepšení výkonu prodloužením čtecího rámce, který je srovnatelný s kapilární Sangerovou metodou.