

Optické mapování aneb Jak vidět sekvenci DNA na vlastní oči

Poznání struktury a funkce dědičné informace je klíčem k porozumění podstaty veškerého života. Díky moderním technologiím sekvenování DNA roste počet organismů, u nichž byla dědičná informace přečtena. Sestavené genomové sekvence mají však obvykle daleko k dokonalosti. Optické mapování představuje nástroj, který může přispět k jejich vylepšení, ale své uplatnění nalézá také na poli studia strukturní variability mezi genomy.

Genetická informace – základ života

Nositelkou genetické informace živých organismů je dvouvláknová molekula DNA, jejímiž základními stavebními jednotkami jsou čtyři typy nukleotidů lišící se dusíkatou bází. Její přečtení představuje klíč k odhalení struktury genomu, identifikaci genů zodpovídajících za jednotlivé vlastnosti organismů a k osvětlení podstaty biologických procesů, které v živých organismech probíhají. Vzájemné porovnání genomů jedinců stejných i různých druhů a hledání genetické variability mezi nimi má nezastupitelný význam při studiu evoluce a vzniku druhů.

Genomy různých druhů se mezi sebou více či méně liší, a to jednak velikostí a počtem chromozomů, mezi které se dědičná informace rozděluje, ale také množstvím nekódujících repetitivních sekvencí (tedy sekvencí opakujících se ve velkém počtu kopií). Vůbec největší genom o velikosti téměř 700 miliard párů bází (700 Gb) nese ve své jediné buňce sladkovodní měňavka *Polychaos dubium* (dříve nazývaná *Amoeba dubia*). Naopak nejmenší známý genom má endosymbiotická bakterie *Carsonella ruddii* s pouhými 160 tisíci páry bází (160 kb). I genom člověka tvořený „pouhými“ 3,2 miliardami párů bází (3,2 Gb) je z pohledu sekvenování považován za složitý, i když zdaleka nedosahuje např. velikosti geno-

mu některých rostlin – jako pšenice seté (*Triticum aestivum*, 17 Gb; viz Živa 2012, 4: 155–157) nebo rostliny s největším genomem – vraního oka japonského (*Paris japonica*, 150 Gb; Živa 2015, 1: 4). Velikost genomu, a zejména pak množství repetitivních sekvencí značně ovlivňuje kvalitu výsledku sekvenování. I přesto, že sekvenční technologie nové generace (next generation sequencing) dnes umožňují v krátkém čase získat obrovské množství dat a algoritmy používané pro sestavování sekvencí se neustále zdokonalují, není sestavení úplné genomové sekvence snadným úkolem (např. Živa 2016, 2: 61–63). Nabízí se otázka, jak výsledky sekvenování dovést k co největší dokonalosti. A právě zde přichází na scénu optické mapování.

Optické mapování – nic nového pod sluncem

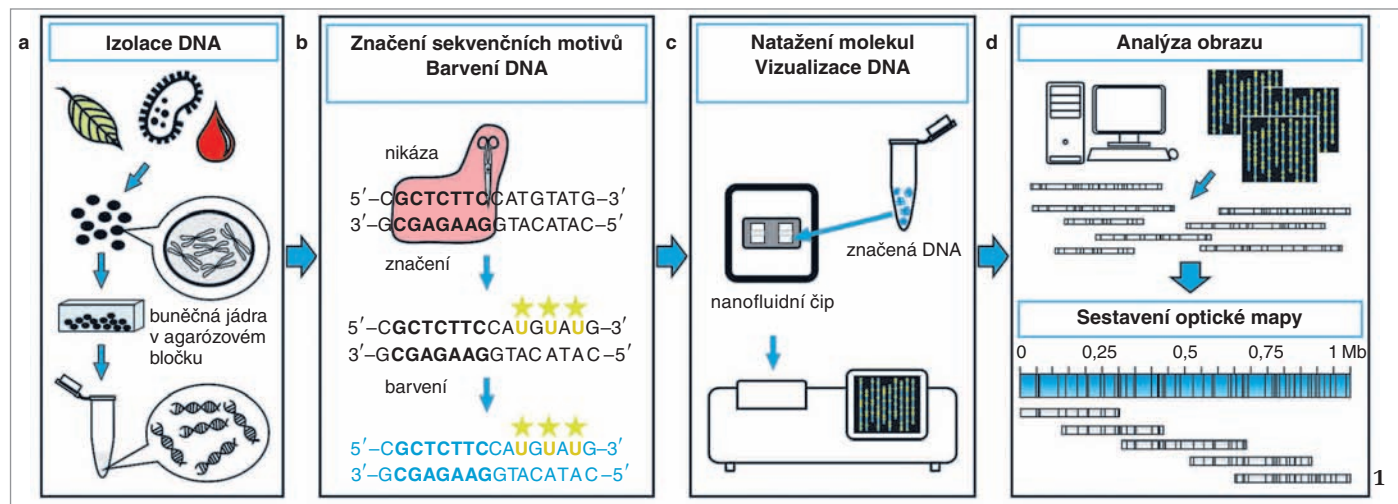
Optickou mapu lze popsat jako mapu výskytu krátkého sekvenčního motivu roztroušeného v desítkách až stovkách tisíců kopií podél dlouhých fragmentů molekuly DNA. Podstata její konstrukce tkví ve zviditelnění zmíněných motivů a přesném určení jejich vzájemných vzdáleností, udávaných v jednotkách párů bází (bp, base pairs). Výsledná mapa připomíná zlomky dlouhého čárového kódu, který charakterizuje celý genom daného jedince nebo

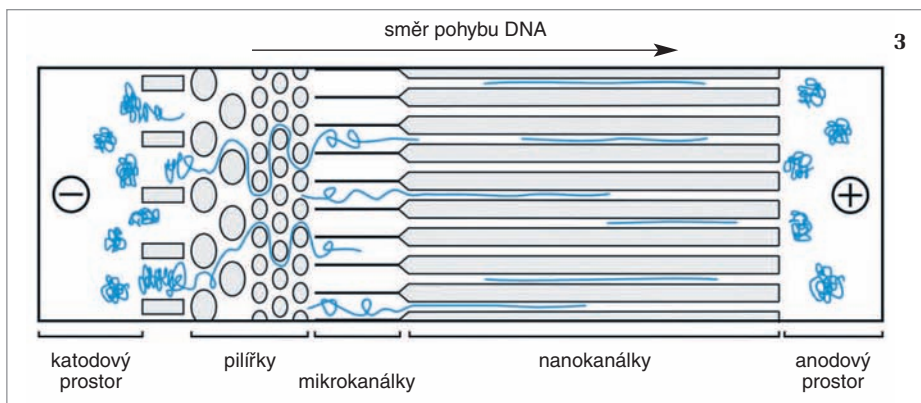
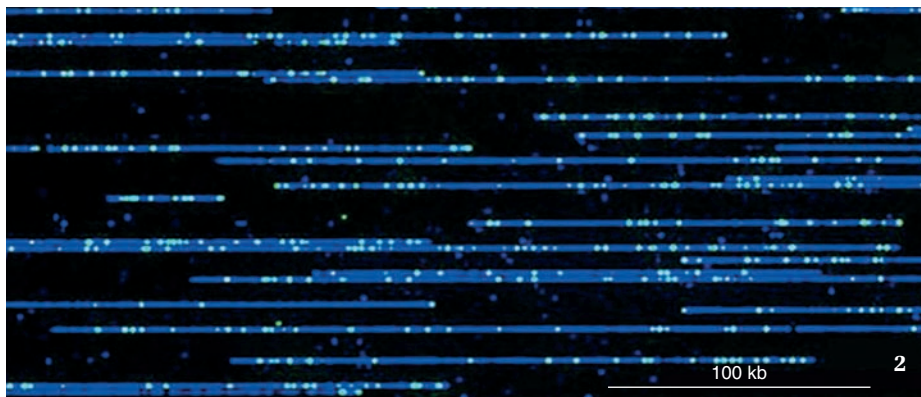
jeho úsek. Tvorba optické mapy zahrnuje několik kroků – izolaci DNA, označení sekvenčních motivů a obarvení celé DNA, natažení molekul DNA, vizualizaci DNA a sekvenčních motivů, analýzu získaného obrazu a sestavení optické mapy na základě překryvů spekter sekvenčních motivů (čárových kódů) charakteristických pro každý z fragmentů DNA.

Princip optického mapování není na poli technologií zaměřených na mapování genomů žádnou novinkou. Jeho historie sahá do poloviny 90. let 20. stol., kdy byla publikována vůbec první optická mapa, a to mapa jednoho z chromozomů kvasinky pивní (*Saccharomyces cerevisiae*). Jejím autorem a současně vynálezcem technologie byl prof. David C. Schwartz, v té době působící na New York University v USA (Schwartz a kol. 1993). Původní technika využívala k natažení molekul DNA tzv. mikrofluidní zařízení (později čip) obsahující kanálky o šířce v řádu mikrometrů. Působením elektrostatických sil byly záporně nabitě molekuly vázány na kladně nabitý povrch zařízení. Proudění roztoku skrze zařízení současně zapříčinilo částečné natažení molekul DNA. Imobilizované molekuly pak byly vystaveny restriktivnímu enzymu, který způsobil zlomy dvoušroubovice jen v místech výskytu specifického sekvenčního motivu. Molekuly DNA byly následně fluorescenčně barveny, vizualizovány pomocí fluorescenčního mikroskopu a obraz obarvených molekul nasnímán kamerou s vysokým rozlišením. Po vyhodnocení snímků byla stanovena celková délka každé molekuly a zároveň vzdálenosti mezi sekvenčními motivy, tedy mezerami vzniklými štěpením. Na základě restriktivních

1 Princip konstrukce optické mapy s použitím nanofluidního čipu.

- Buněčná jádra jsou uvolněna z buněk a zachycena v agarózovém gelu, z kterého je extrahována DNA (a).
- Po rozpuštění gelu jsou fluorescenčně označeny specifické sekvenční motivy a zároveň obarvena celá molekula DNA (b).
- Značená DNA je nanesena na čip, poté za pomoci přístroje natažena v nanokanálcích a vizualizována (c).
- Získaný obrazový materiál se převede do souboru nesoucího informaci o délce molekul a pozicích sekvenčních motivů (d). Na základě této informace je sestavena optická mapa. Blíže v textu





2 Fotografie molekul DNA v nanokanálkách. Lze je pozorovat v podobě modrých vláken se zeleně zářícími body vyznačujícími sledovaný sekvenční motiv.

3 Schéma stavby nanofluidního čipu. Modře je znázorněna DNA pohybující se strukturami čipu od katody směrem k anodě. V oblasti pilířků a mikrokanálek dochází k postupnému rozplétání a natahování molekul. V nanokanálkách se již pohybují zcela natažené molekuly DNA.

spekter jednotlivých fragmentů DNA byla na závěr sestavena optická mapa. Tato starší technika trpí dvěma zásadními nedostatky – kapacitou mikrofluidního čipu, který lze použít pouze jednou, a nereprodukovatelným stupněm natažení molekul DNA. Ten lze částečně vyvážit mnohonásobným opakováním experimentu, ale v kombinaci s malou kapacitou mikrofluidního čipu metoda nutně naráží na značnou finanční i časovou náročnost. Zmíněné nedostatky po dlouhou dobu omezovaly aplikaci optické mapování převážně na malé a střední genomy. Svě místo si proto metoda našla zejména při identifikaci a charakterizaci bakteriálních druhů a kmenů. I přesto však byly sestaveny mapy také pro několik organismů s velkými genomy, mezi něž patří např. pšstros (~1,25 Gb) nebo kukuřice (~2,3 Gb), ale i člověk (~3,2 Gb, viz výše).

Optické mapování v nanokanálkách

Průlomem v metodice se stal vynález nanofluidního čipu, který obsahuje soubor tisíců paralelně uspořádaných nanokanálků, v nichž dochází k přesně regulovanému natažení molekul DNA. Použití nanofluidního čipu bylo poprvé publikováno teprve před 6 lety (Das a kol. 2010) a dnes již existuje na trhu kompletní platforma zahrnující chemikálie, čip, přístroj pro analýzu vzorku i software pro zpracování dat. Jednotlivé

kroky zpracování vzorku DNA a konstrukce optické mapy s využitím nanofluidního čipu (obr. 1) se velmi podobají výše popsané původní metodě. Mapování v nanokanálkách však přináší značný pokrok, zejména díky rovnoměrnému natažení molekul DNA, které má dopad na přesné určení jejich délky a také vzdáleností mezi sekvenčními motivy. Nesporný přínos představuje také výrazně vyšší kapacita čipu daná možností cyklické analýzy vzorku – z jednoho čipu nové generace lze získat až 15× více dat než u předchozí technologie.

Superdlouhá DNA – klíč k úspěchu

Na počátku celého procesu přípravy optické mapy stojí vysoce kvalitní DNA, jejíž fragmenty dosahují délky stovek tisíc až milionů párů bází. Taková DNA je předpokladem pro sestavení kvalitní optické mapy, jejíž dílčí části dosahují velikosti několika milionů párů bází a snadno tak délkou předčí většinu genomových sekvencí sestavených pomocí počítačových programů. Jak ale kvalitní vysokomolekulární DNA získat a zároveň ji nepoškodit? Úspěch spočívá v postupu, kdy jsou izolovaná jádra smíchána s tekutou agarózou, jejímž ztuhnutím dochází k ukotvení jader v agarózovém gelu. Ve vzniklých agarózových bločcích pak probíhá extrakce a přečištění DNA (obr. 1a). Pevná gelová matrice DNA chrání před působením mechanických sil. Promýváním bločků v příslušných roztocích dochází k lyzi jaderné membrány, odstranění proteinů navázaných na DNA a k degradaci molekul RNA, které byly spolu s DNA přítomny v jádře. Očištěnou DNA lze v bločcích uchovávat neporušenou po dobu několika měsíců.

Stojí za to zmínit, že kromě běžně používaných metod purifikace mohou být jádra, ale také izolované mitotické chromozomy přečištěny průtokovou cytometrií (blíže viz Živa 2012, 4: 155–157). Tím se

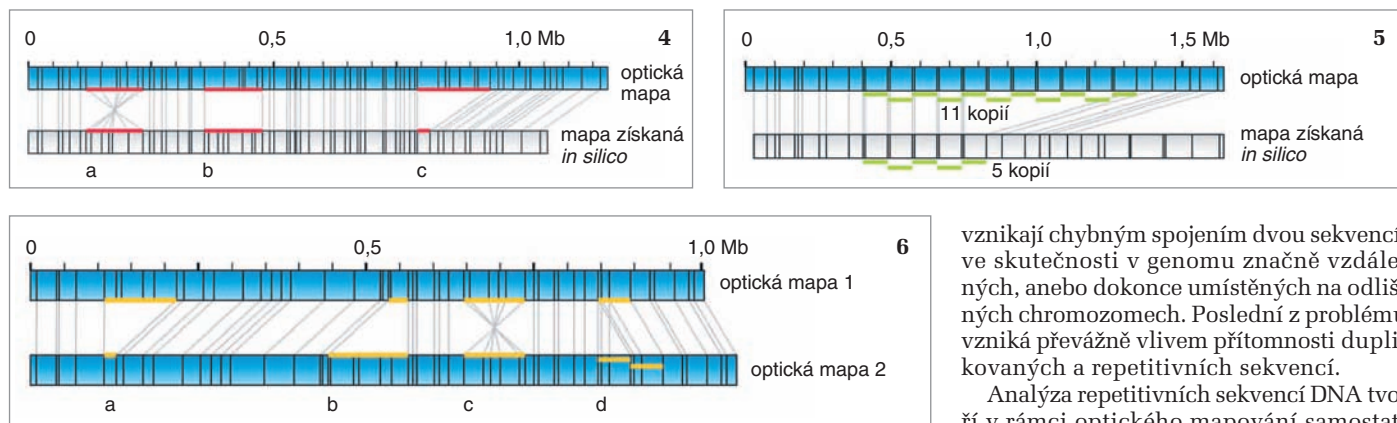
otvírá cesta k vytvoření optické mapy přímo z určitého chromozomu. DNA získaná z takto připraveného materiálu vyniká vysokou čistotou, jelikož je účinně zbavena veškerých buněčných komponent, především organel obsahujících vlastní genetickou informaci (mitochondrii, plastidů) a látek, které by mohly inhibovat reakce v následných krocích přípravy vzorku.

Jak rozsvítit DNA

Dalším krokem na cestě k optické mapě je označení specifického sekvenčního motivu a současně obarvení molekuly DNA po celé délce (obr. 1b). Na rozdíl od původní metody dochází při mapování v nanokanálkách ke značení a barvení ještě před samotnou analýzou DNA na čipu. Jelikož zmíněné reakce probíhají v roztoku, předchází jim enzymatické uvolnění DNA z agarózového gelu. Značení sekvenčních motivů začíná naštípnutím DNA v místech jejich výskytu enzymem zvaným nickáza. Ta, obdobně jako restriční endonukleáza, rozpoznává určitou krátkou sekvenci DNA (obvykle o velikosti 6–7 bp), avšak na rozdíl od ní štěpí pouze jedno ze dvou vláken a vytváří v molekule DNA zářezy (nicks). Druhé vlákno zůstává nedotčeno, čímž je zaručena integrita fragmentů DNA. V místech zářezů se následně vymění původní nukleotidy za nukleotidy nesoucí fluorescenční značku. Na závěr je celá molekula DNA obarvena fluorescenčním barvivem. Značení a barvení umožní pozorovat DNA v podobě modrých vláken nesoucích zeleně zářící body, označující sledovaný sekvenční motiv (obr. 2). Ačkoli je proces značení poměrně náchylný k mechanickému poškození DNA, díky vysoké kvalitě vstupního materiálu zůstávají ve vzorku stále přítomny molekuly, jejichž délka se pohybuje v průměru ve stovkách kilobází, což představuje minimálně stonásobek délky čtení (reads) dnešních sekvenačních přístrojů.

Nanofluidní čip – srdce celého systému

Fluorescenčně označená DNA se v roztoku ve svém přirozeném stavu vyskytuje ve formě stočených klubíček. V této podobě ji sice můžeme pozorovat pomocí fluorescenční mikroskopie, rozhodně však není možné měřit velikost jednotlivých fragmentů, natož určit vzájemné vzdálenosti značených sekvenčních motivů. Zde se do procesu tvorby optické mapy zapojuje nepostradatelná součást systému – nanofluidní čip. Toto drobné zařízení nese dvě průtokové komůrky (flowcells) vyrobené z kombinace křemíku a skla, v nichž dochází k přesně regulovanému a reprodukovatelnému natažení molekul DNA. Každou komůrku o rozměru 9 × 1,5 mm tvoří soubor 13 tisíc paralelně uspořádaných nanokanálků, jimž předchází soustava speciálních struktur – pilířků a mikrokanálek, pro rozpletení klubíček DNA (obr. 3). Důležitou vlastností nanokanálků je jejich průměr, který činí pouhých 42 nm. Projde jimi pouze jediná molekula DNA a zároveň je při jejím pohybu zajištěno natažení na předem přesně stanovené procento skutečné délky (zpravidla okolo 85 %). Díky této skutečnosti lze s velkou přesností určit skutečnou délku molekul i vzdálenosti mezi fluorescenčními značkami.



Neměně důležitou součástí systému je samotný přístroj, který obstarává pohyb molekul ve strukturách čipu a jejich následné zviditelnění (obr. 1c). Záporně nabitě molekuly DNA se v čipu pohybují na principu elektroforetické separace – v elektrickém poli putují od záporně nabitých elektrod (katody) ke kladné (anodě). Ve chvíli, kdy molekuly vstoupí do nanokanálek, se elektroforéza zastaví a lasery o specifických vlnových délkách ozáří DNA, čímž dochází k excitaci fluorescenčních barviv a následné emisi (vyzáření) světla. Citlivá kamera s vysokým rozlišením postupně snímá celou plochu souboru nanokanálek a spojením více než tisíce snímků vznikne obraz průtokové komůrky. Jakmile se dokončí snímání, molekuly jsou z nanokanálek vypláchnuty a proces může začít znovu s další dávkou DNA. Na jednom čipu lze provést až 30 cyklů analýzy molekul DNA a procedura pak trvá přibližně 24 hodin. V každém cyklu jsou detekovány a snímány na každé ze dvou průtokových komůrek desetitisíce fragmentů DNA. Množství čipů nezbytné pro nasbírání dostatečného množství dat je přímo úměrné velikosti zkoumaného genomu. Např. pro bakterii *Escherichia coli*, jejíž genom čítá pouhých 4,6 Mb, lze dostatek dat nasbírat z jednoho až dvou cyklů. V případě analýzy jednoho ramene chromozomu pšenice o velikosti 400 milionů párů bází potřebujeme jeden až dva čipy; a pokud zkoumáme např. genom hrachu, jehož velikost činí přibližně 4,5 Gb, je nutných až 15 čipů a analýza potrvá 15 dní nepřetržitého běhu přístroje.

Posledních několik kroků k cíli

Skládání optické mapy představuje výpočetně poměrně náročný proces, jehož délka exponenciálně roste s velikostí genomu. U menších genomů postačí pro sestavení mapy jen několik hodin, u větších však může trvat i měsíce. Před samotným skládáním se nasbíraný obrazový materiál převede do podoby textového souboru, jenž obsahuje informaci o délce každé molekuly a přesné pozici všech značek, které na ní leží. Spektra fluorescenčních značek (čárové kódy) odpovídající každé molekule jsou vzájemně porovnávána a v případě překryvu se molekuly spojí (obr. 1d). Jelikož jsou vstupní data generována ve velkém nadbytku (okolo 200 ekvivalentů velikosti genomu), vzniká překryvem spekter mnoha molekul velmi spolehlivá konsenzuální mapa. I přesto, že přítomnost tzv. fragilních míst, v nichž dochází téměř vždy ke zlomům DNA, nedovoluje vytvořit jed-

4 Identifikace chyb v sekvencích.

Porovnání optické mapy s mapou získanou *in silico* (vytvořenou pomocí počítačového programu na základě sestavené sekvence). Svislé černé linie uvnitř každé mapy znázorňují pozice sekvenčních motivů, šedé spojovací linie značí shodu mezi motivy v obou mapách. Porovnáním byla v *in silico* získané mapě představující sestavenou sekvenci nalezena problematická místa – nesprávně orientovaný úsek sekvence (a), nesprávně sestavený (b) a chybějící úsek sekvence (c).

5 Analýza oblasti tandemově uspořádané repetice DNA – porovnání optické mapy s mapou získanou *in silico* štěpením sestavené sekvence (blíže viz obr. 4). Porovnání obou map umožnilo určit skutečný počet jednotek tandemově uspořádané repetice a délku celého úseku.

6 Příklad strukturní variability mezi genomy dvou jedinců. Srovnáním optických map dvou blízkých příbuzných jedinců můžeme identifikovat několik typů přestaveb DNA: delece (a), inserce (b), inverze (c) nebo duplikace (d). Svislé černé linie uvnitř obou map znázorňují pozice sekvenčních motivů, šedé propojující linie představují shodu mezi motivy optických map obou jedinců. Všechny orig. H. Toegelová

nu spojitou optickou mapou, poskytují dílčí mapy hodnotnou informaci o struktuře genomu v kontextu stovek kilobází až několika megabází.

Optická mapa – užitečný pomocník

Optické mapy nacházejí široké uplatnění zejména díky jednoduchosti přípravy a následného použití. Nejčastěji se využívají ke kontrole kvality sestavených genomových sekvencí, ale slouží také jako předloha pro jejich prodloužení (propojování kratších sekvencí). Vše lze realizovat pouhým porovnáváním sekvenčních dat s optickou mapou. Aby bylo možné srovnání provést, musejí být sekvence nejprve převedeny do podoby optické mapy. Sekvence jsou *in silico* naštěpeny, což znamená, že jsou v nich pomocí počítačového programu nalezeny cílové motivy enzymu nikázy použité při konstrukci optické mapy. Jejich rozložení je porovnáváno s optickou mapou, čímž odhalíme chybně sestavená místa. Nejčastěji jde o úseky, které byly do kontextu delší sekvence vloženy ve špatné orientaci nebo zcela chybně, či o chybějící úseky sekvence (obr. 4). Významné potíže představují tzv. chimérické sekvence, jež

vznikají chybným spojením dvou sekvencí, ve skutečnosti v genomu značně vzdálených, anebo dokonce umístěných na odlišných chromozomech. Poslední z problémů vzniká převážně vlivem přítomnosti duplikovaných a repetitivních sekvencí.

Analýza repetitivních sekvencí DNA tvoří v rámci optického mapování samostatnou kapitolu. Je obecně známo, že tyto opakující se motivy představují pro jakýkoli algoritmus používaný při skládání sekvencí velmi tvrdý oříšek. Příčinou je zejména maximální délka sekvenačního čtení, jež bývá kratší než tandemově uspořádané úseky repetice. Tyto oblasti DNA jsou tvořeny specifickou sekvencí uspořádanou v několika či mnoha kopiích těsně za sebou. Většinou je možné poskládat pouze několik kopií tandemově uspořádané sekvence, a poté se skládání zastaví. Software jednoduše neví jak dále. Optická mapa díky délce analyzovaných molekul DNA umožňuje překlenout repetitivní úsek, určit jeho délku a počet tandemově uspořádaných repetitivních jednotek DNA (obr. 5). Často tak odhalí poměrně velké mezery v sestavené genomové sekvenci.

Kromě hledání a opravy chyb nabízí optická mapa i možnost prodloužení stávajících sekvencí pomocí tzv. hybridního skládání (hybrid scaffolding). Tento přístup kombinuje informace uložené v optické mapě s informací sekvenční. Optická mapa slouží jako most k propojení sekvencí, které spolu sousedí, ale nepřekrývají se. Výsledkem jsou dlouhé úseky (scaffolds) tvořené dílčími kratšími sekvencemi, mezi nimiž byly přesně určeny délky mezer.

Významnou aplikací optického mapování je studium strukturní variability mezi genomy jedinců stejného druhu, případně velmi blízkých příbuzných druhů. Cenově dostupná optická mapa může zastoupit pracné a nákladné sekvenování. Stačí pouze vytvořit optické mapy pro studované jedince a provést vzájemné porovnání, které ukáže na strukturní odlišnosti mezi genomy v podobě delecí, insercí, duplikací, inverzí, stromokací nebo variability v počtu kopií určitého sekvenčního motivu (viz obr. 6). Ve spojitosti s referenční sekvencí daného druhu pak můžeme interpretovat nalezenou změnu na sekvenční úrovni. Toho se využívá např. v medicínském výzkumu při hledání spojitosti mezi strukturními přestavami genomu a některými typy rakovinných onemocnění u člověka. Další oblastí uplatnění je evoluční biologie, kde nalezená strukturní variabilita poskytuje informaci o míře příbuznosti studovaných jedinců.

Závěrem lze říci, že optické mapování představuje důmyslnou, spolehlivou a perspektivní metodu. Do budoucna jistě přispěje k přesnější informaci o sekvenci mnoha genomů a praktické uplatnění najde i v návazných genetických studiích.

Použitá literatura uvedena na webu Živý.