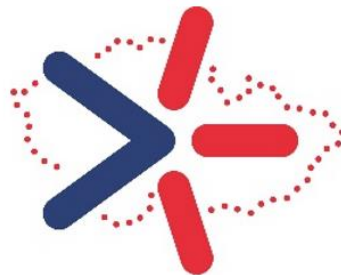




Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Metody sekvenování DNA a RNA

Jiří Fajkus

Molekulární komplexy chromatinu, Mendelovo centrum genomiky a proteomiky rostlin, CEITEC MU

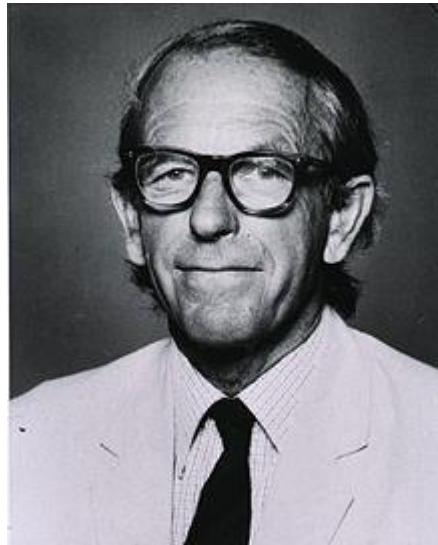
Národní centrum pro výzkum biomolekul, Lab. FGP, PřF MU

fajkus@sci.muni.cz



MUNI
SCI

Nobelova cena za chemii (1958) – amino acid sequence of insulin
Nobelova cena za chemii(1980) – dideoxy method of DNA sequencing



Frederick Sanger (13. 8. 1918 – 19. 11. 2013)

Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors". Proc. Natl. Acad. Sci. U.S.A. 74 (12): 5463–7.

Historicky první metoda: 1975 - Sanger and Coulson – **Plus-minus sekvenování**

(4 **plus reakce** (1 dNTP) a 4 **mínus reakce** (3 další dNTPs), restriční fragmenty použity jako primery pro elongaci DNA polymerázou I. Elfo detekce, ve kterých plus a mínus reakcích došlo k zabudování.

– metoda použita pro sekvenování 5,386 nt genomu - jednovláknové **DNA bakteriofága ϕ X174**

Chemické sekvenování / Maxam-Gilbertovo sekvenování (Nobelova cena 1980)

1976–1977- Allan Maxam and Walter Gilbert (Maxam AM, Gilbert W (Feb 1977). "A new method for sequencing DNA". *Proc. Natl. Acad. Sci. U.S.A.* **74** (2): 560–4).

1) radioaktivní **značení** na 5' konci DNA (typicky pomocí kinázy s použitím gama- ^{32}P ATP) a purifikace fragmentu DNA, který se má sekvenovat .

2) Chemické působení vytváří zlomy v malé frakci u jedné nebo dvou bází ze 4, v každé se 4 reakcí: (**G**, **A+G**, **C**, **C+T**).

puriny (**A+G**) se depurinují s použitím kyseliny mravenčí,
guaniny (**G**) (a částečně i adeniny) se methylují dimethylsulfátem
pyrimidiny (**C+T**) – báze se odštěpují hydrazinem.

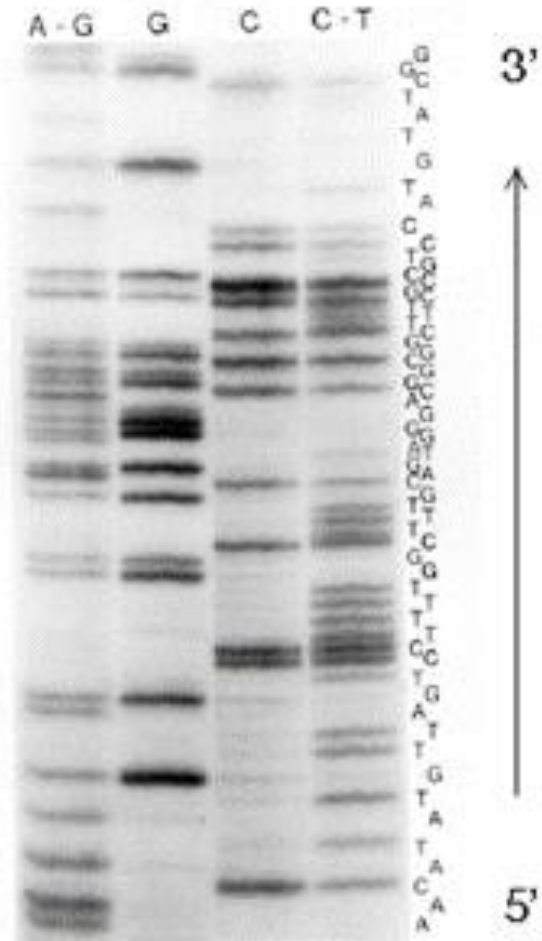
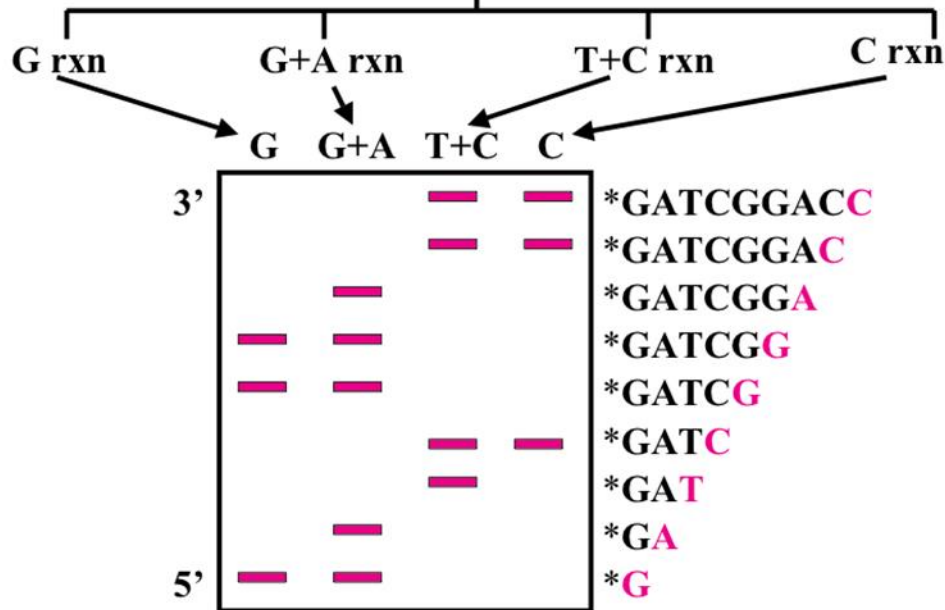
Přídavek soli (NaCl) k hydrazinu inhibuje methylaci thyminu, reakce je pak specifická pro **C**.

3) Modifikované DNA jsou pak **štěpeny** v **horkém piperidinu** v pozici modifikované báze. Tím se vytváří soubor značených fragmentů od značeného konce až po první štěpení v každé molekule.

4) Fragmenty z jednotlivých reakcí jsou nanášeny na elfo vedle sebe v denaturujícím **PAGE pro separaci podle velikosti**. Vizualizace pomocí autoradiografie (rtg film)

Maxam-Gilbert sequencing

5' *pGpApTpCpGpGpApCpC 3'

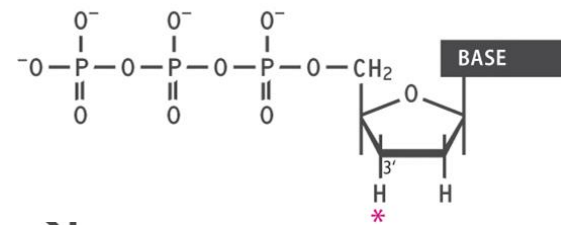


Sangerovo sekvenování (Chain-termination seq., **dideoxy-seq.**)

Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors". Proc. Natl. Acad. Sci. U.S.A. 74 (12): 5463–7. (Nobel Prize 1980)

- používá **dideoxynukleosidtrifosfáty (ddNTPs)**

jako **terminátory** syntézy řetězce DNA.

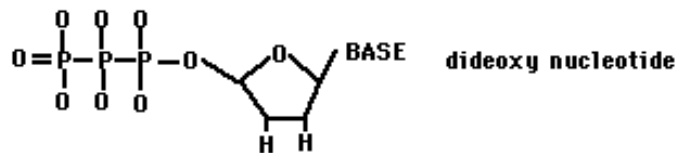


- Požadavky: ss- (tepelně denaturovaná) templátová DNA, DNA primer, DNA polymeráza, normální deoxynukleosidtrifosfáty (dNTPs) a modifikované nukleotidy (dideoxynTPs, ddNTPs) které terminují prodlužování řetězce DNA.

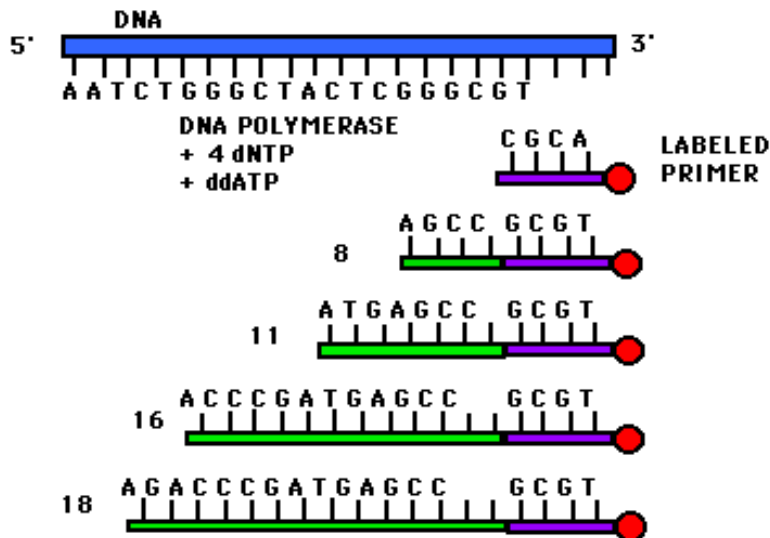
- Vzorek DNA se rozdělí na **4 oddělené sekvenační reakce**, obsahující všechny 4 standardní deoxynukleotidy (dATP, dGTP, dCTP and dTTP) a DNA polymerázu. Do každé reakce se přidá jen 1 ze 4 **ddNTPs** (ddATP, ddGTP, ddCTP, or ddTTP) jako terminační, postrádá **3'-OH skupinu** nezbytnou pro vznik fosfodiesterové vazby mezi dvěma nukleotidy, čímž ukončí prodlužování řetězce DNA, vznikají fragmenty různé délky (končící ale v dané reakci vždy stejným dNTP).

- Nově syntetizované a značené řetězce jsou tepelně denaturovány a separovány podle velikosti (s rozlišením 1 nt) **gelovou elektroforézou na denaturujícím polyakrylamidovém gelu s močovinou**. Každá ze 4 reakcí běží ve své dráze (dráhy A, T, G, C);

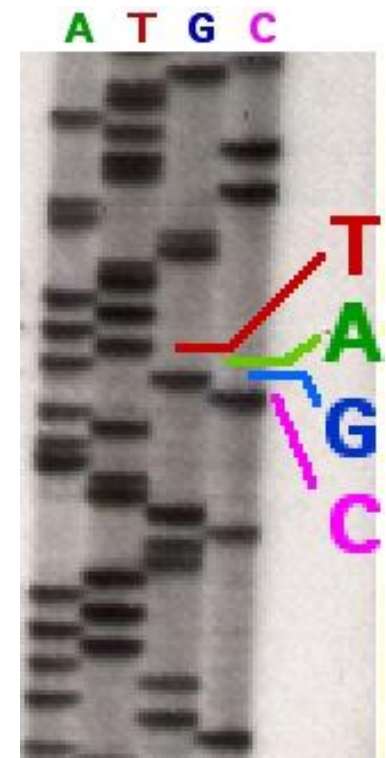
- proužky DNA se vizualizují autoradiograficky. DNA sekvenci lze číst přímo z rtg filmu (v klasické verzi)



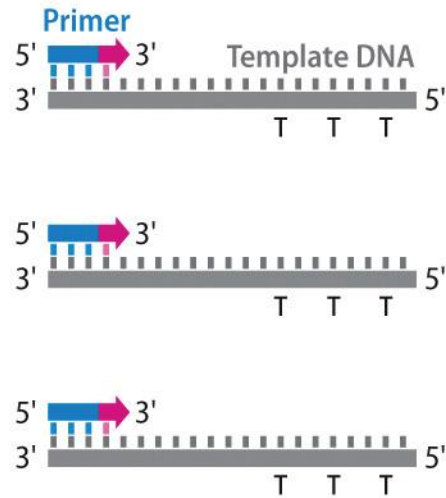
Jedna ze 4 sekvenačních reakcí



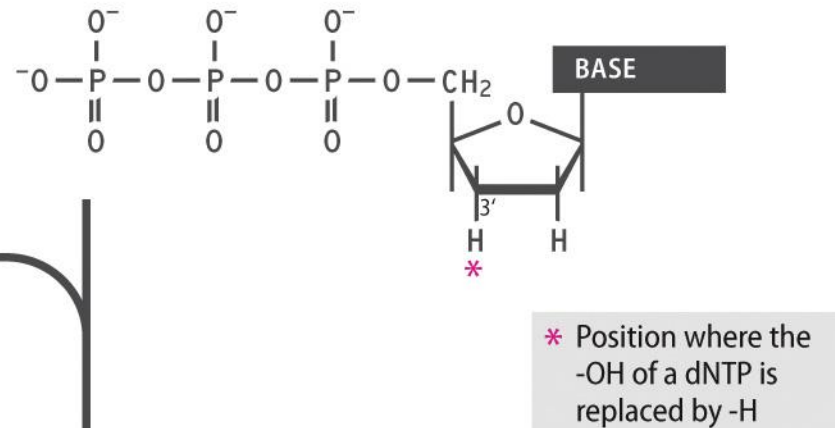
- Na obrázku vpravo byl rtg film exponován na gelu. Tmavé proužky odpovídají fragmentům DNA různých délek. Tmavý proužek v dráze odpovídá fragmentu DNA, který vznikl **terminací řetězce po zabudování ddNTP** (ddATP, ddGTP, ddCTP, or ddTTP). Vzájemná poloha různých proužků mezi 4 dráhami se pak použije ke čtení sekvence DNA (zdola nahoru).



(A) Initiation of strand synthesis



(B) A dideoxynucleotide



(C) Strand synthesis terminates when a ddNTP is added

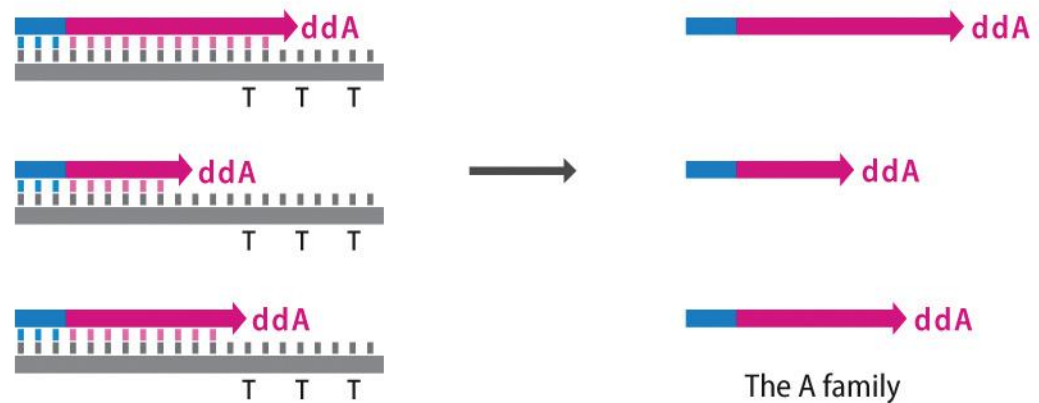
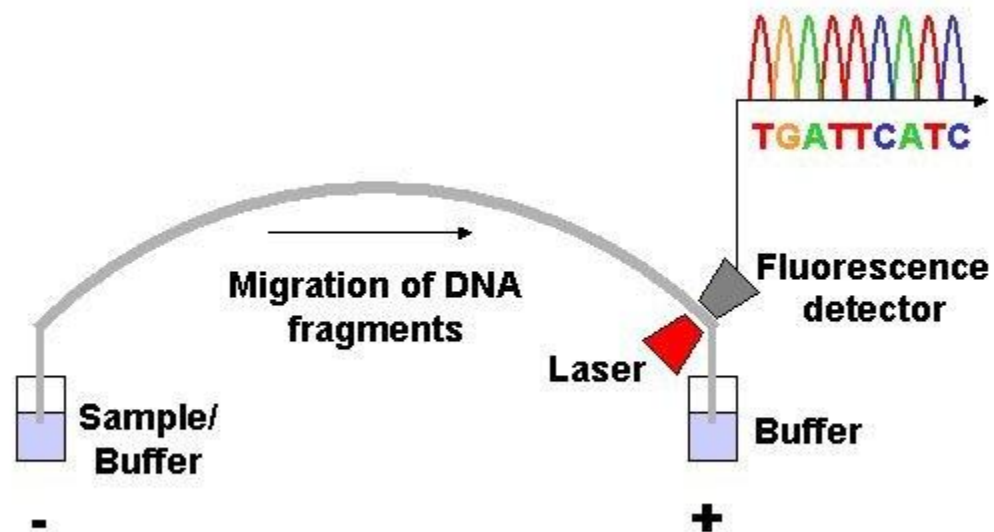
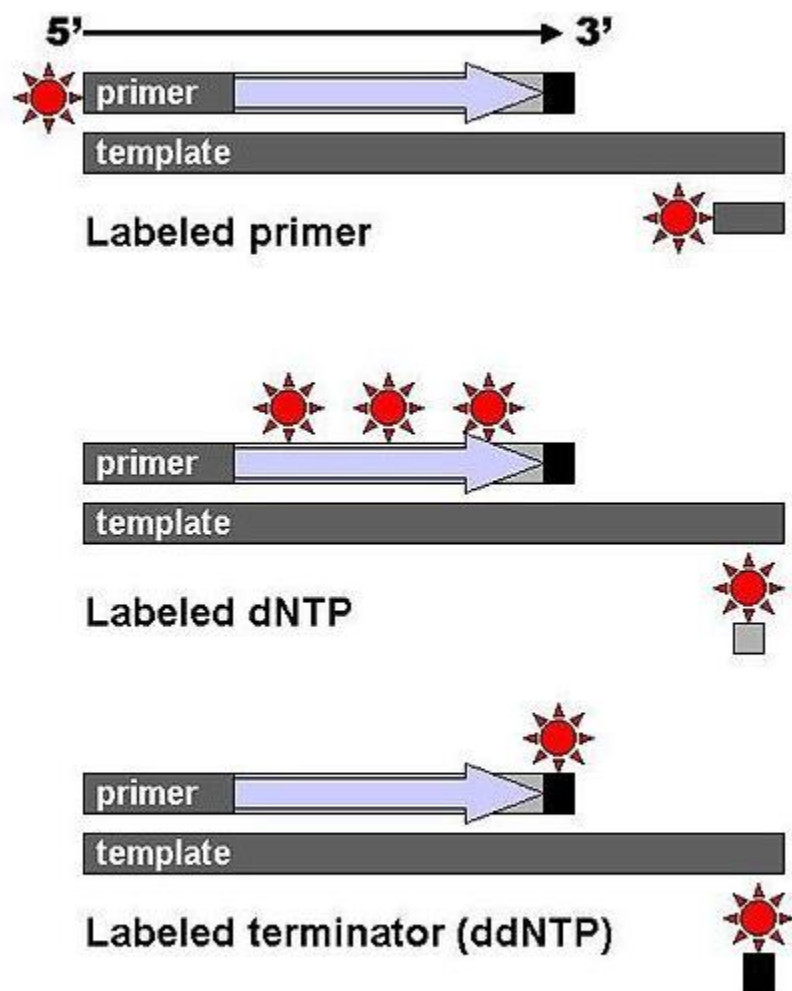
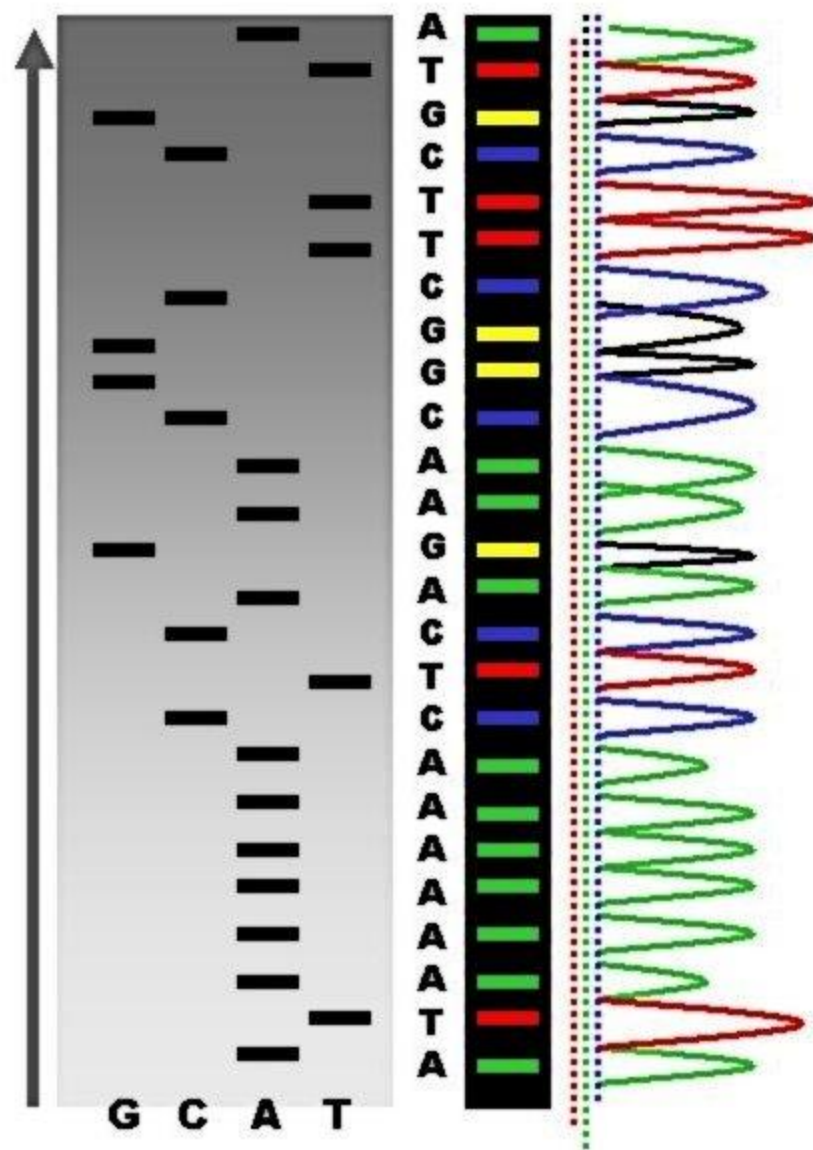
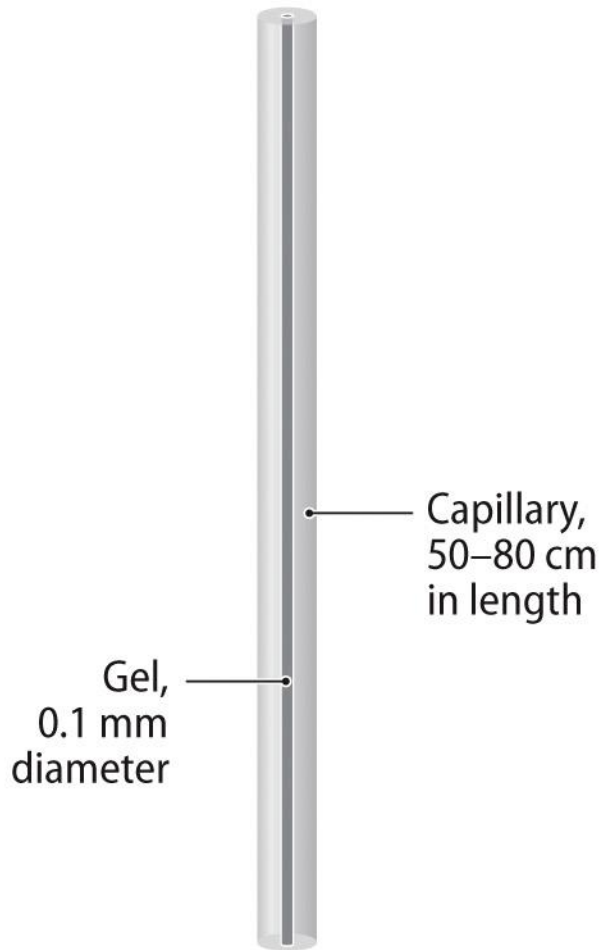


Figure 4.2 Genomes 4 (© Garland Science 2018)





(A) Capillary gel



(B) Electrophoresis of chain-terminated DNA molecules

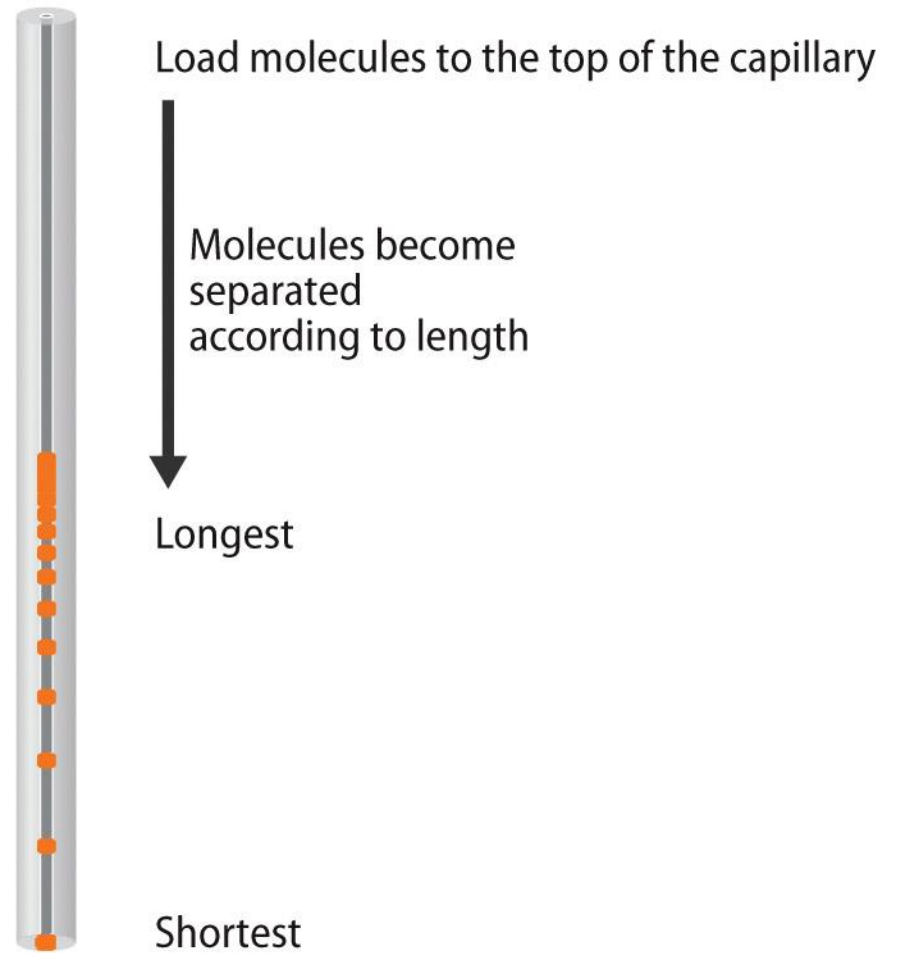
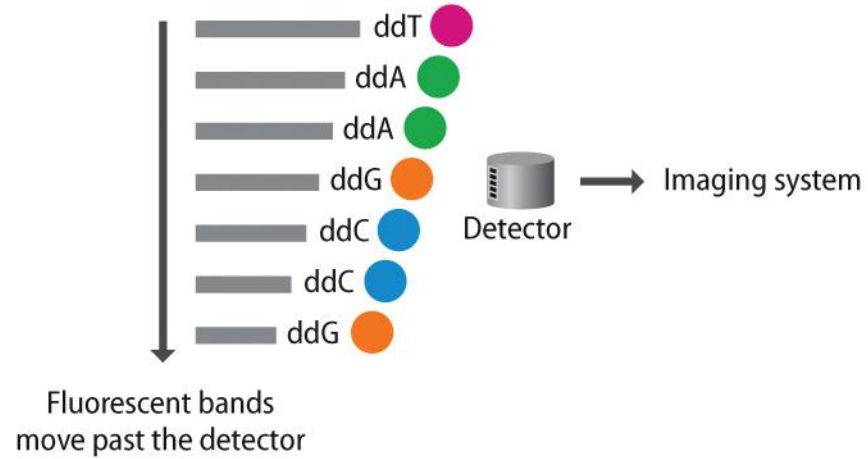


Figure 4.1 Genomes 4 (© Garland Science 2018)

(A)

ddA ● ddC ● ddNTPs – each with a
different fluorescent label
ddT ● ddG ●

↓ Sequencing reactions,
fractionation of products



(B)

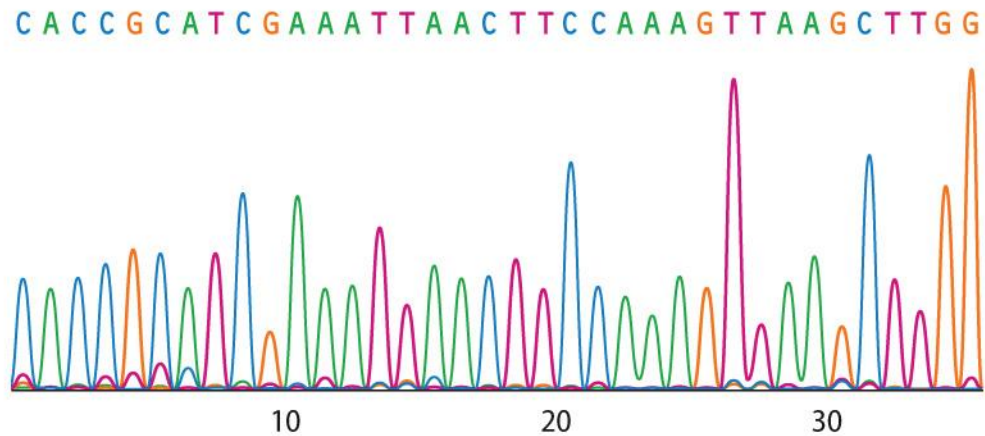
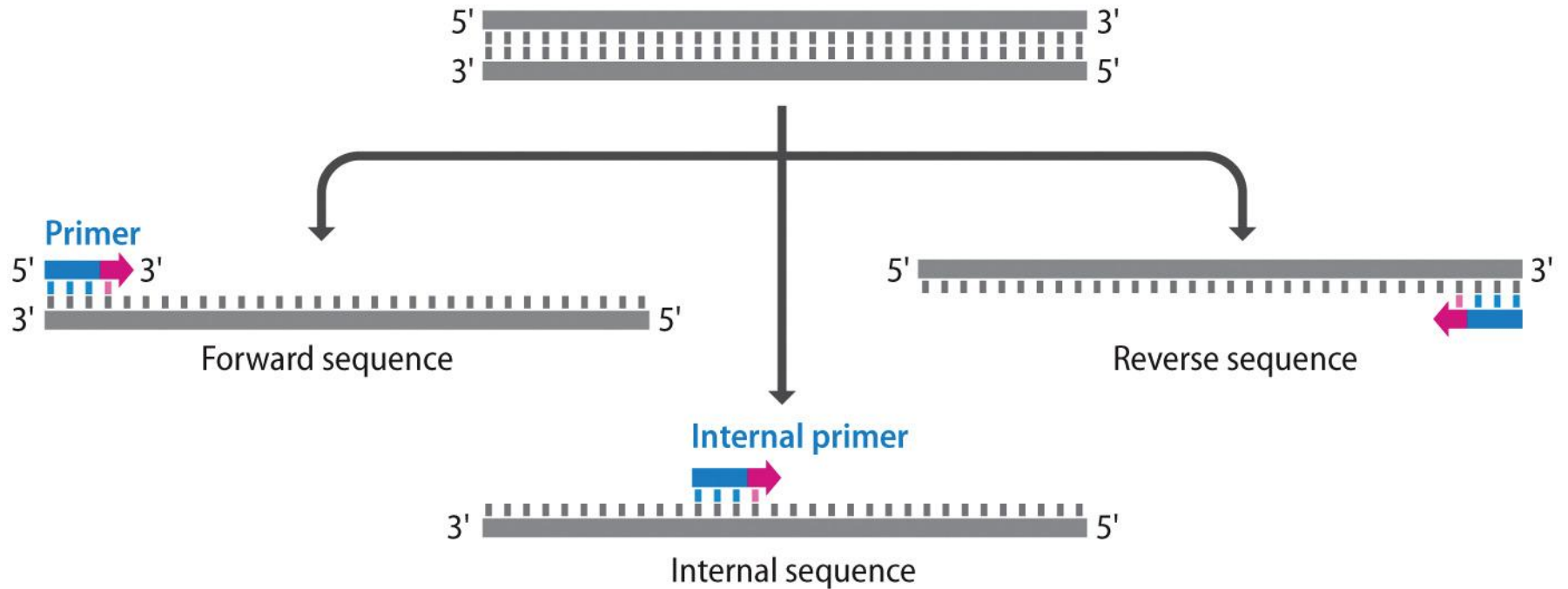


Figure 4.3 Genomes 4 (© Garland Science 2018)

(A) Forward, reverse, and internal primers



(B) Universal primer

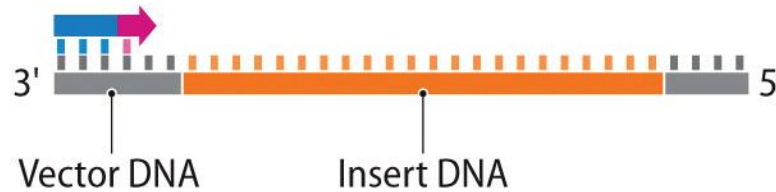
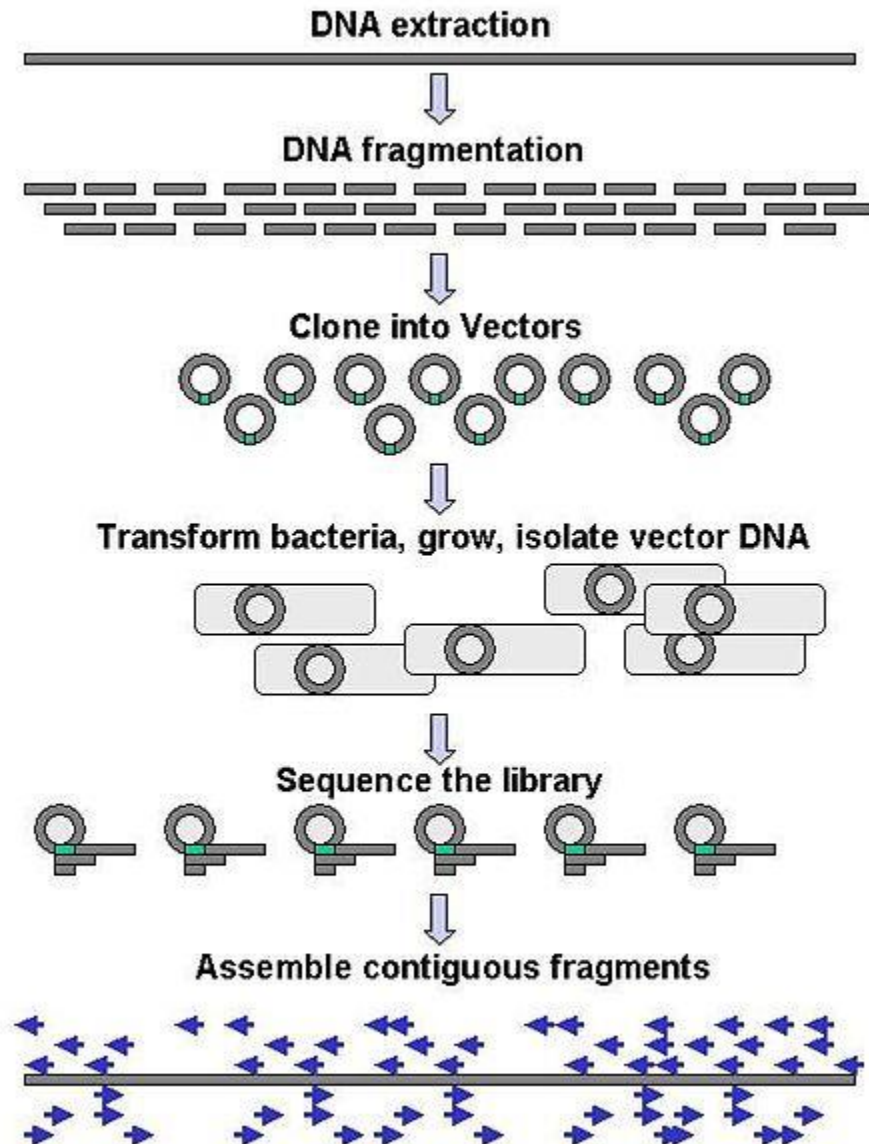


Figure 4.5 Genomes 4 (© Garland Science 2018)

ACCATCGTAGCTTCAGTATGTATGTACTAG	Read 1
ATGTTTGTAGCTAGGATCGTAGCTACC	Read 2
TTTGTAGCTAGGATCGTAGCTACCATCGTAGCTT	Read 3
TTGTAGCTTCAGTATGTATGTACTAG	Read 4
GGATCGTAGCTACCATCGTAGCTTCAGT	Read 5
ATGTTTGTAGCTAGGATCGTA	Read 6
ATGTTTGTAGCTAGGATCGTAGCTACCATCGTAGCTTCAGTATGTATGTACTAG	Deduced sequence

Figure 4.6 Genomes 4 (© Garland Science 2018)

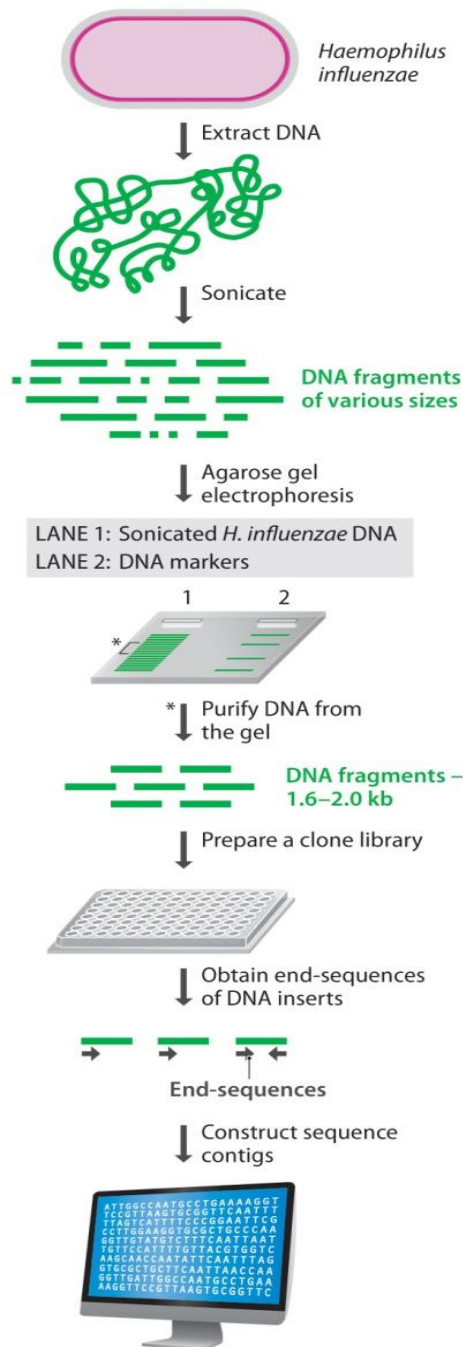
Obecná sekvenační strategie v 90. letech (před nástupem NGS a TGS)



Pro větší genomy je třeba (kvůli usnadnění sestavování výsledných kontigů) nejprve klonovat větší fragmenty do vektorů s vysokou kapacitou (**YAC, BAC**), a teprve jejich inzerty rozklonovat **do běžných plazmidových vektorů** s kapacitou 2-3 kb vhodných pro čtení Sangerovým sekvenováním.

U velmi velkých, zejména polyploidních rostlinných genomů (např. pšenice) se ukázalo jako výhodné pro tvorbu BAC knihoven využít specifické (průtokovým tříděným izolované) chromozomy. Sekvenování **chromozom-specifických knihoven** snížilo komplexitu problému a odstranilo problém s příslušností sekvenční na konkrétní chromozom. (Jaroslav Doležel, UEB AVČR Olomouc).

I potom je ale extrémně obtížné správně sestavit oblasti vysoce repetitivní DNA (např. centromery, rDNA). To umožnily až techniky TGS (viz dále).



První sekvence celého genomu – bakterie *Haemophilus influenzae* – 1.83 Mb

(tým J. C. Ventera - The Institute for Genome Research (TIGR), *Science*, 1995)

Důležité rysy:

1. Začíná se z **náhodných fragmentů** (není zapotřebí předběžná genetická mapa)
2. Následuje čtení koncových sekvencí fragmentů
3. S využitím softwarových nástrojů (***TIGR assembler***), jsou tisíce fragmentů porovnávány, klastrovány a skládány na základě shod do *kontigů* pro sestavení celého genomu (contigs). Nejvíce informativní a neopakující se sekvenční fragmenty jsou identifikovány nejdříve, repetitivní fragmenty jsou porovnávány následně. Delší fragmenty pomohly seřadit některé repetitivní a téměř identické sekvenční fragmenty. Menší mezery, které zůstaly po zpracování pomocí *TIGR assembleru* byly pak doplněny pomocí jiných – pomocných- strategií.

Pro sekvenování **eukaryotických genomů** bylo nutné před klonování do plazmidových vektorů (s kapacitou cca 2 kb, tedy vhodných pro samotné sekvenování pomocí Sangerovy metody) klonování větších fragmentů do vektorů typu **YAC** nebo **BAC** (kapacita 2Mb, resp. 100 kb). Teprve jejich inzerty se subklonují do plazmidových vektorů – snadnější sestavování genomu z větších kusů.

Pro velmi velké genomy (pšenice – **15 Gb**) strategie **chromozom-specifických knihoven** (J. Doležel).

Figure 4.17 Genomes 4 (© Garland Science 2018)

Projekt sekvenování lidského genomu - HUGO

1983 – začátek mapování lidského genomu (Laboratoř Los Alamos a Laboratoř Lawrence Livermora)

1987 – vyvinuty kvasinkové arteficiální chromozomy (**YAC** – yeast artificial chromosomes), klonovací kapacita až 2 Mb

HUGO (HUMAN Genome Organisation) – založena 1988

1990 – zahájení projektu mapování lidského genomu - U.S. Department of Energy (měli zájem sledovat vliv záření na lidskou populaci) a National Institute of Health U.S.). Byly připraveny **umělé bakteriální chromozómy, tzv. BAC**, které umožňují klonovat dlouhé úseky DNA (cca 100 kb) – sice nižší kapacita než YAC, ale pohodlnější práce

1996 – zveřejněna kompletní genetická mapa myši domácí (*Mus musculus*), byl sekvenován kompletní genom *Saccharomyces cerevisiae* a první archaebakterie *Metanococcus jannaschi*

1998 - soukromá společnost **Celera Genomics** se rozhodla také sekvenovat lidský genom

2001 - zveřejněna 1. verze sekvence lidského genomu oběma paralelními projekty, HUGO - cca. 3 miliardy dolarů, **Celera vystačila s částkou 10x nižší**

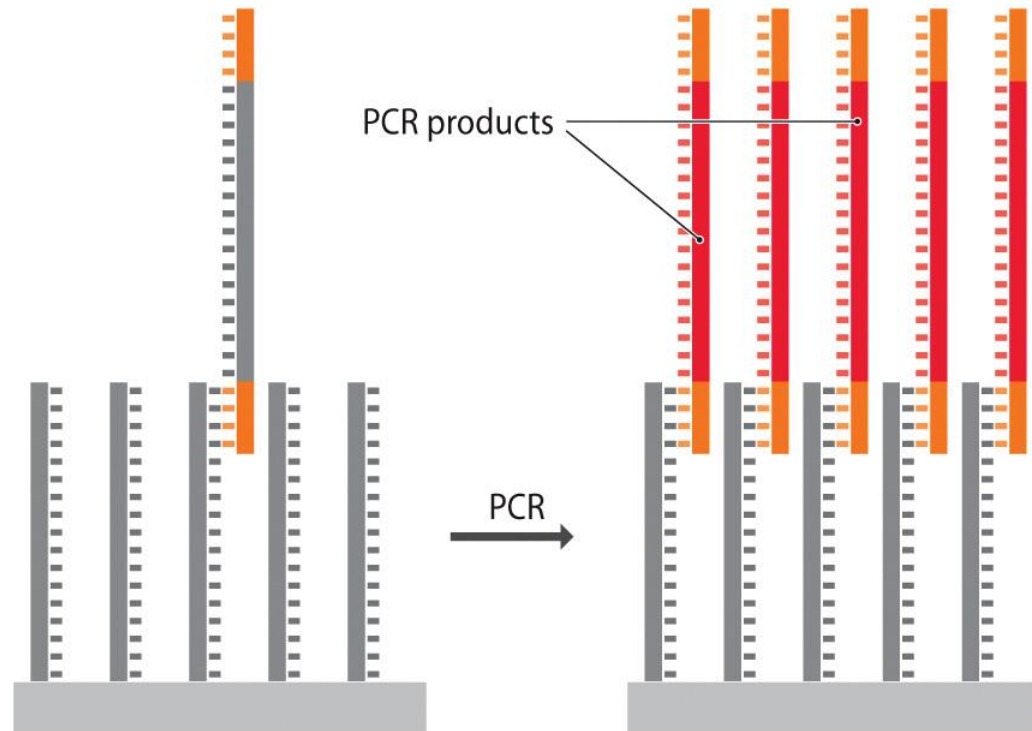


Next generation

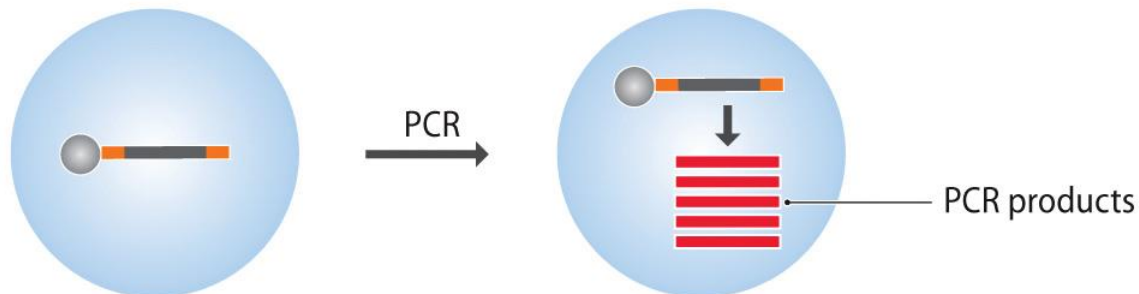
sequencing

Obečné přístupy k imobilizaci a amplifikaci fragmentů genomové DNA pro NGS

(A) Glass slide method



(B) Metallic bead method



Obečné přístupy k amplifikaci fragmentů genomové DNA pro NGS

(A) Attachment to beads by streptavidin–biotin linkages

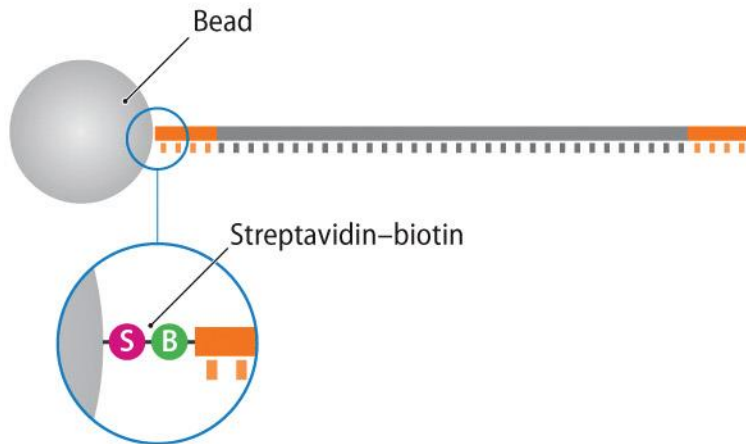
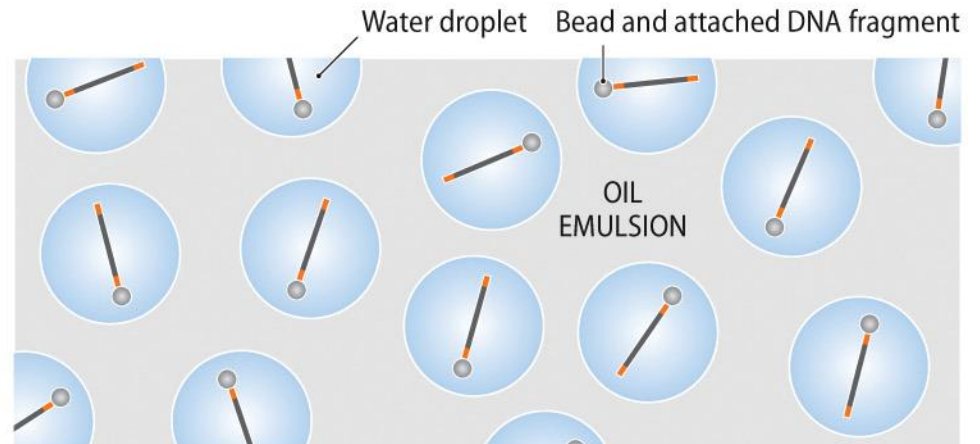
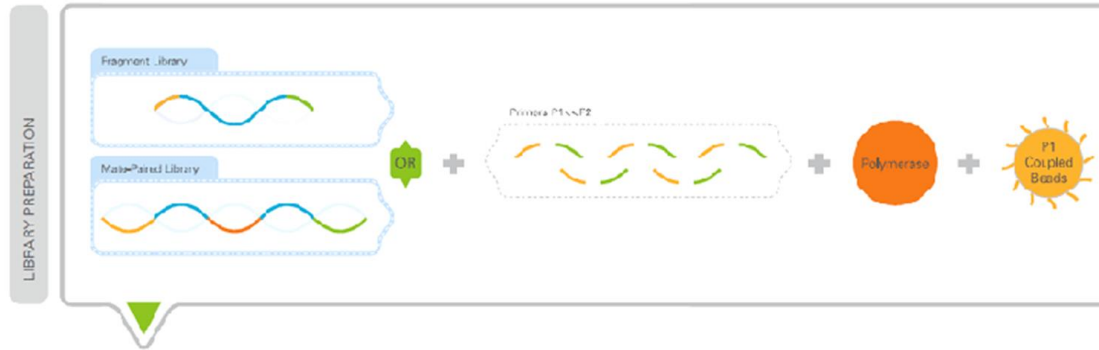


Figure 4.9 Genomes 4 (© Garland Science 2018)

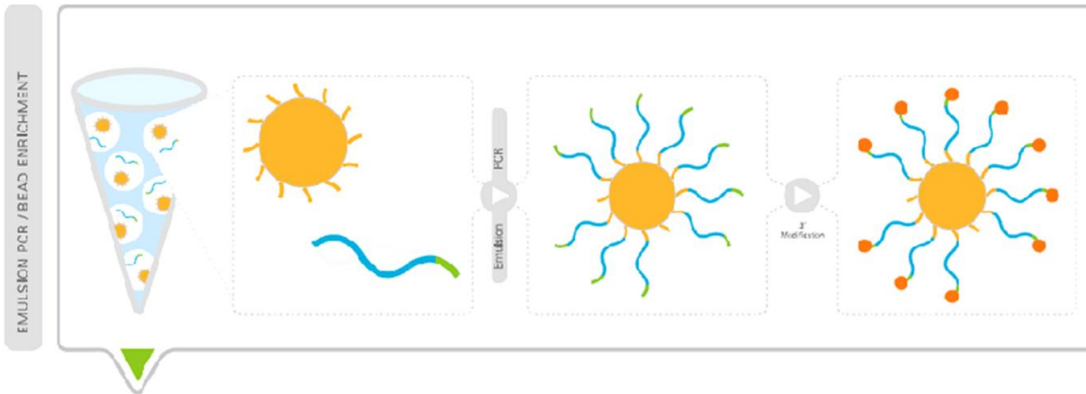
(B) Beads in water droplets in an oil emulsion



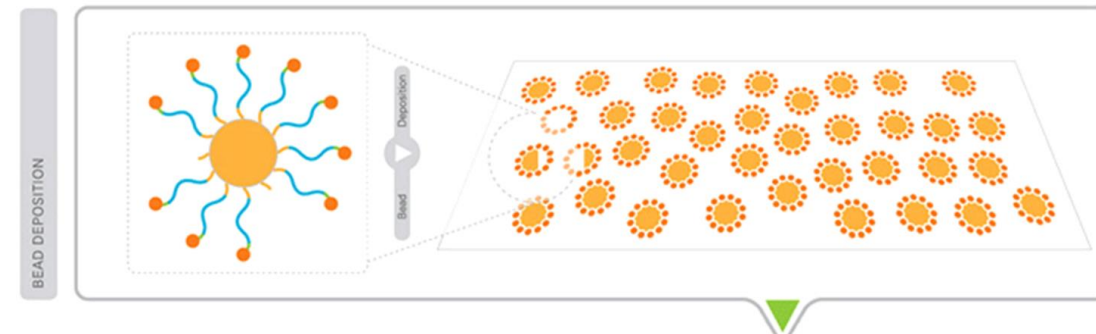
SOLiD sekvenování (ukončeno) – technologie vypracovaná jako první NGS technika (**první MPSS na trhu**). Vzhledem k velmi krátkým čtením však nenašla takové rozšíření, jako po ní pyrosekvenování a později Illumina. Vypracovala však koncept NGS přístupů používaný i v jiných platformách (příprava knihovny – ligace adaptorů, emulzní PCR na kuličkách, imobilizace...)
SOLiD technologie (Applied Biosystems) využívá jako jediná k sekvenování **ligaci**.



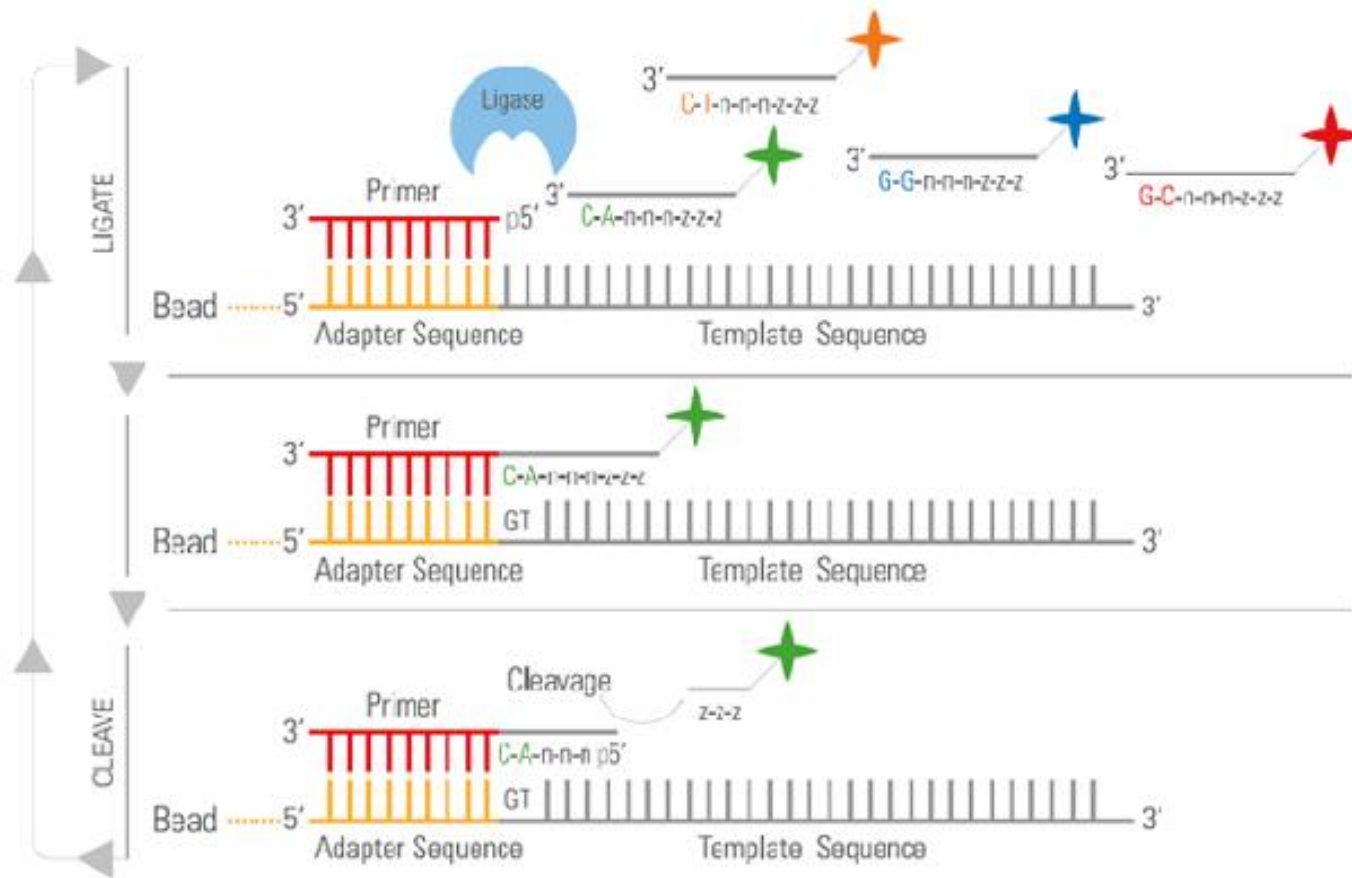
Příprava NGS knihovny z fragmentů DNA



Emulzní PCR na kuličkách, modifikace templátové DNA na 3'konci pro připojení na sklíčka



Nanesení kuliček s modifikovanou DNA na 3'konci na sklíčko



Sekvenování ligací

1. Primery hybridizují k P1 adaptorové sekvenci na kuličkách s templátovou DNA.
2. Sada čtyř fluorescenčně značených dvoj-bázových sond soutěží o ligaci k sekvenačnímu primeru. Specifita dvoj-bázové sondy je docílena zkoušením každé 1. a 2. báze v každé ligační reakci.
3. Opakované cykly ligace, detekce a štěpení se dělají v počtu, který určuje potřebnou délku čtení.
4. Po dokončení ligačních cyklů je extendovaný produkt odstraněn a na templát se hybridizuje primer komplementární k n-1 pozici na druhé kolo ligačních cyklů.

Princip:

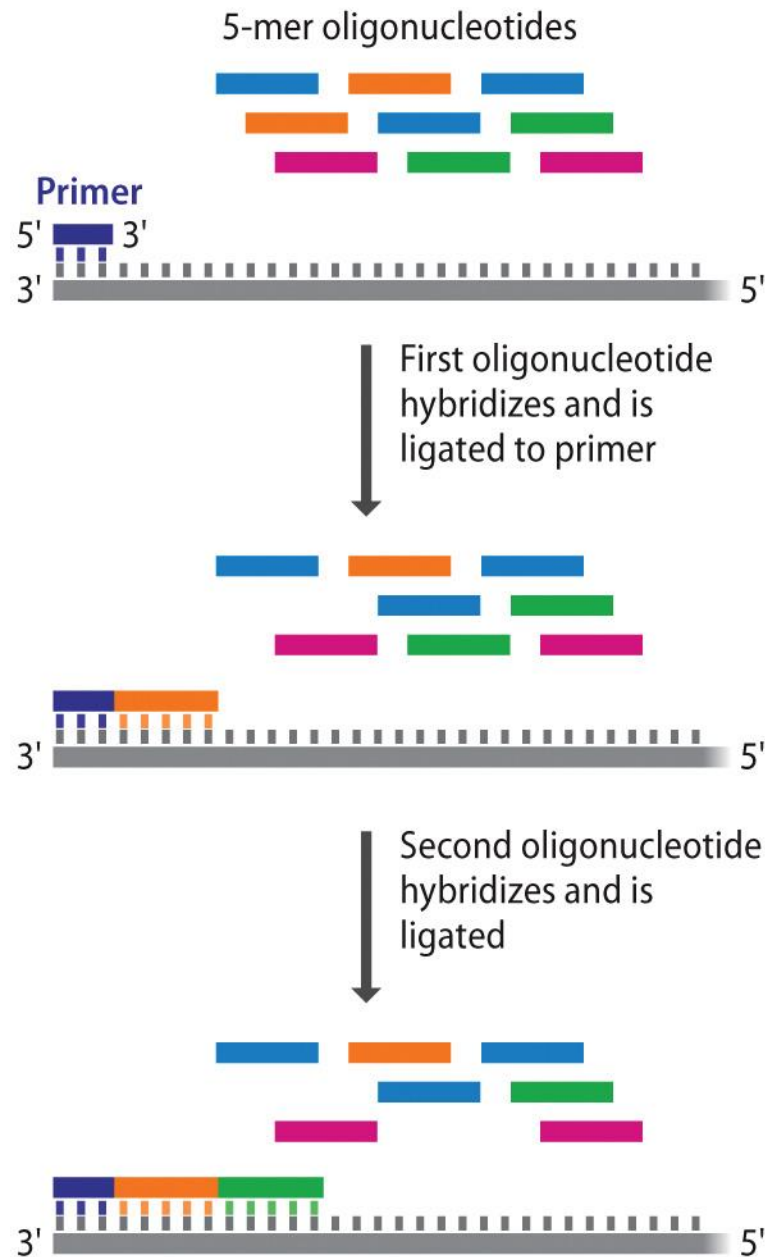
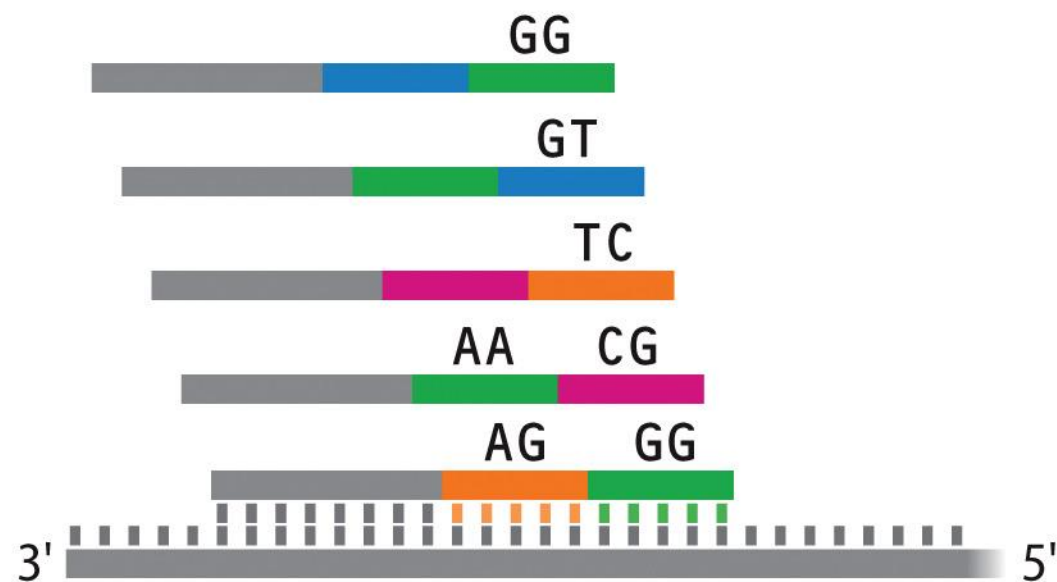
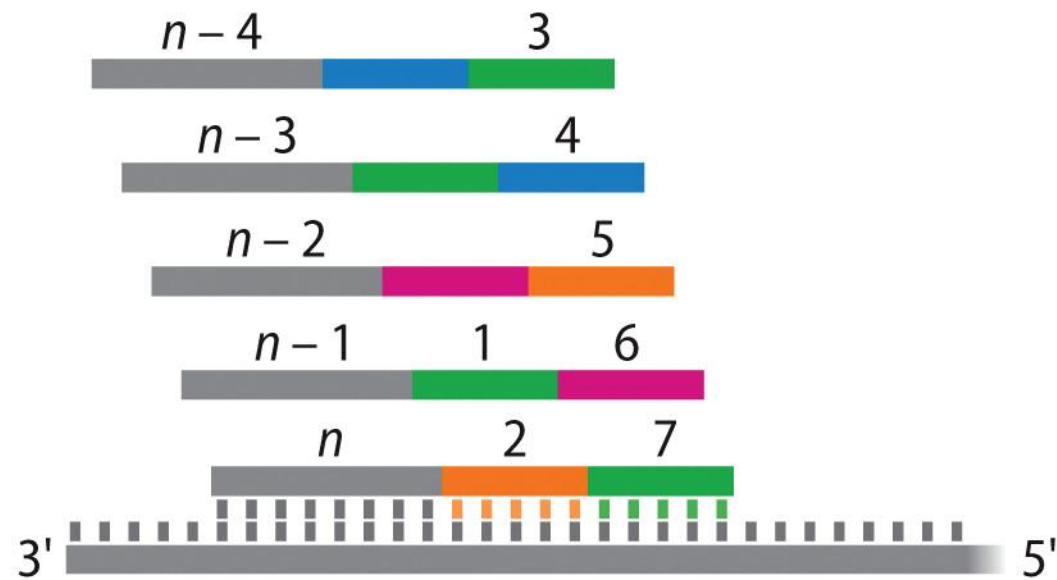


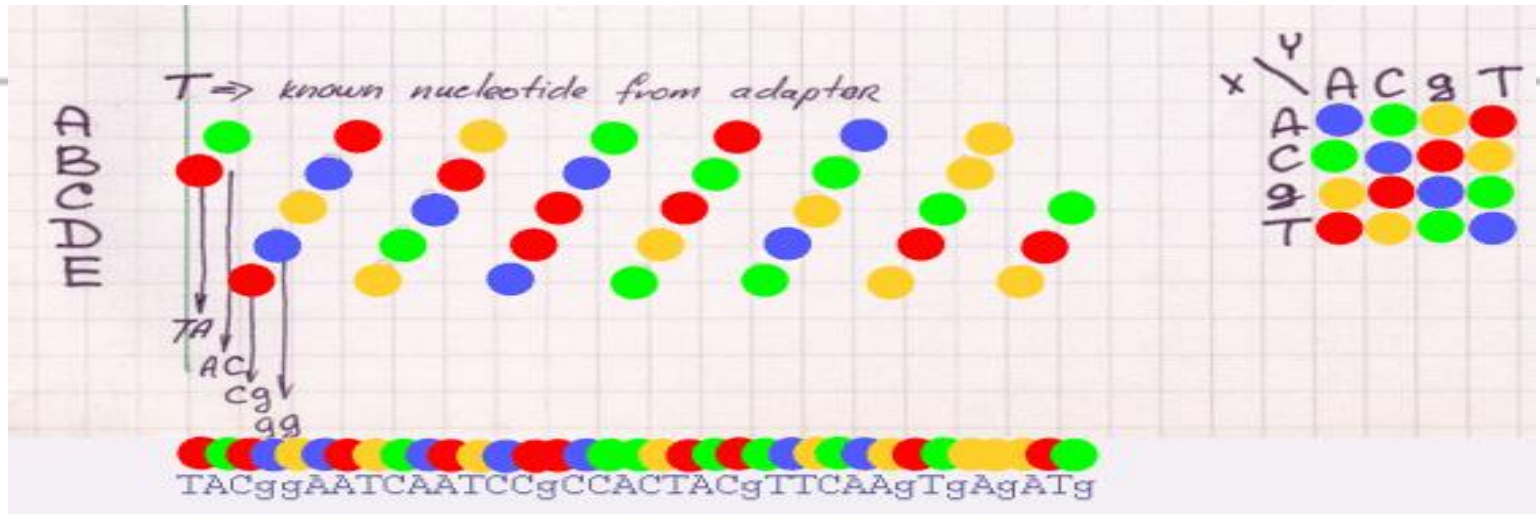
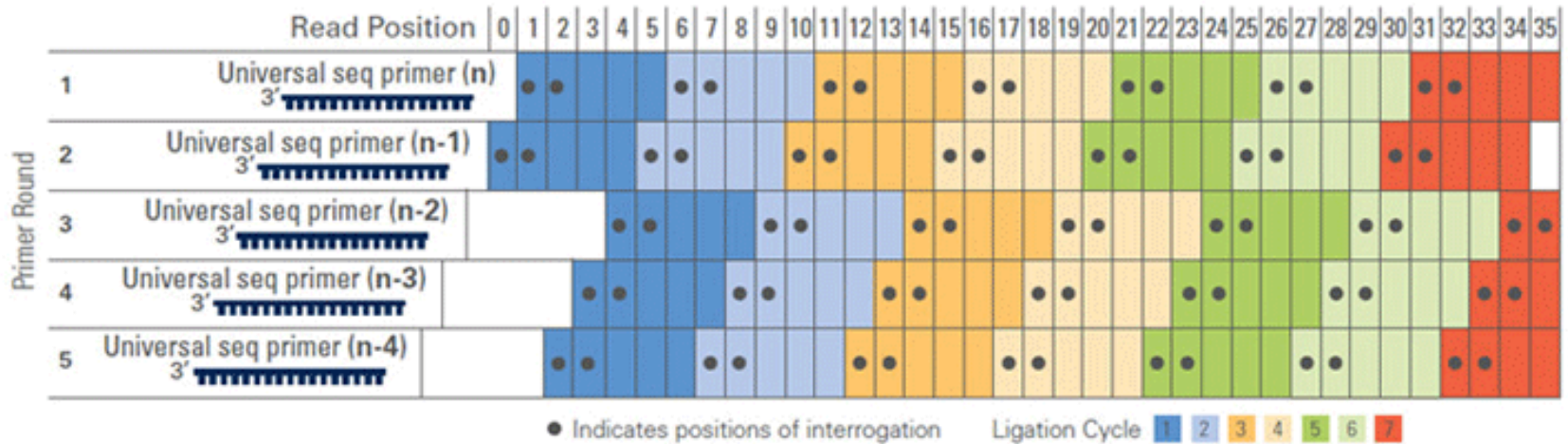
Figure 4.13 Genomes 4 (© Garland Science 2018)



KEY					
		2nd base			
		A	C	G	T
1st base	A				
	C				
	G				
	T				

Figure 4.14 Genomes 4 (© Garland Science 2018)

SOLiD sequencing



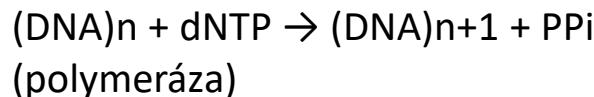
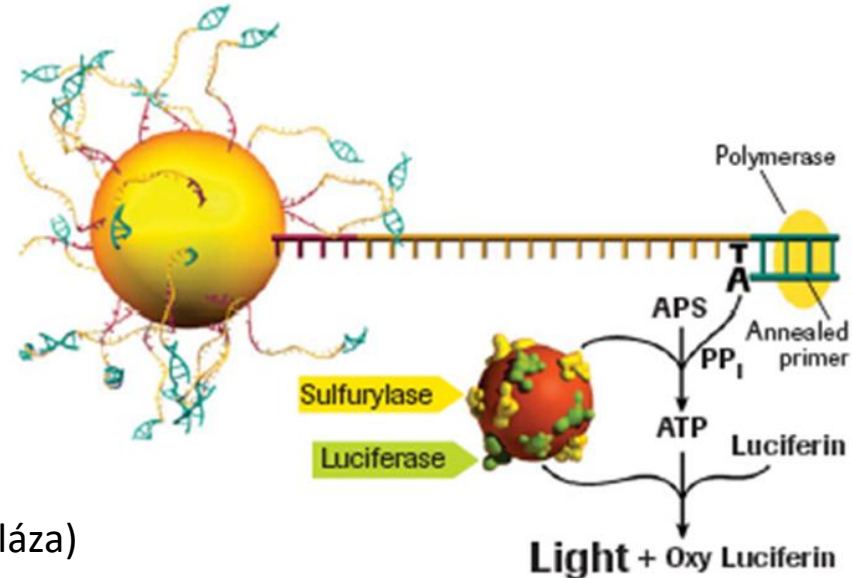
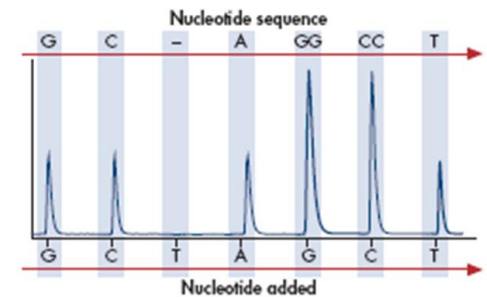
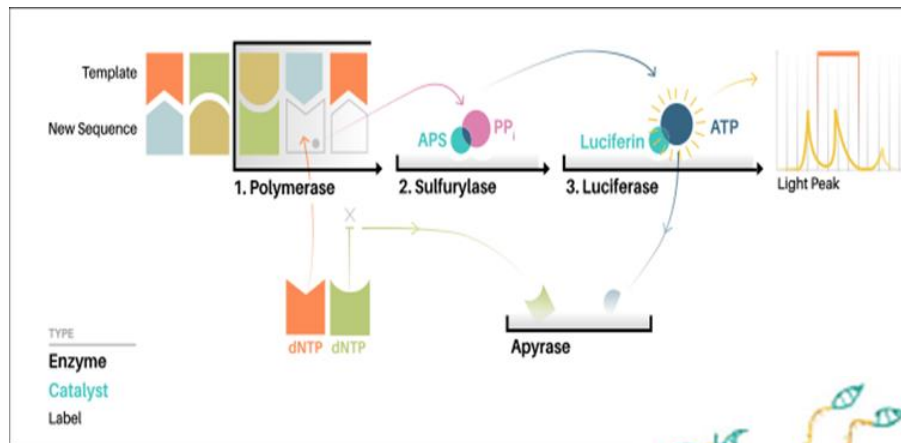
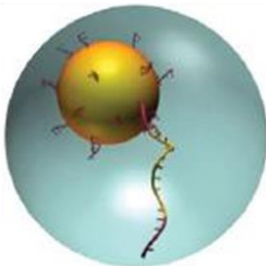
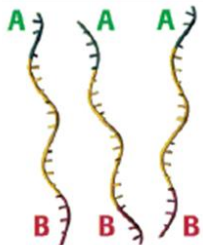
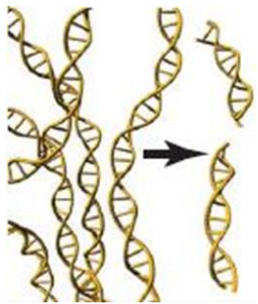
Primer Reset

Five rounds of primer reset are completed for each sequence tag (Figure 5). Through the primer reset process, virtually every base is interrogated in two independent ligation reactions by two different primers. For example, the base at read position 5 is assayed by primer number 2 in ligation cycle 2 and by primer number 3 in ligation cycle 1. This dual interrogation is fundamental to the unmatched accuracy characterized by the SOLiD™ System.

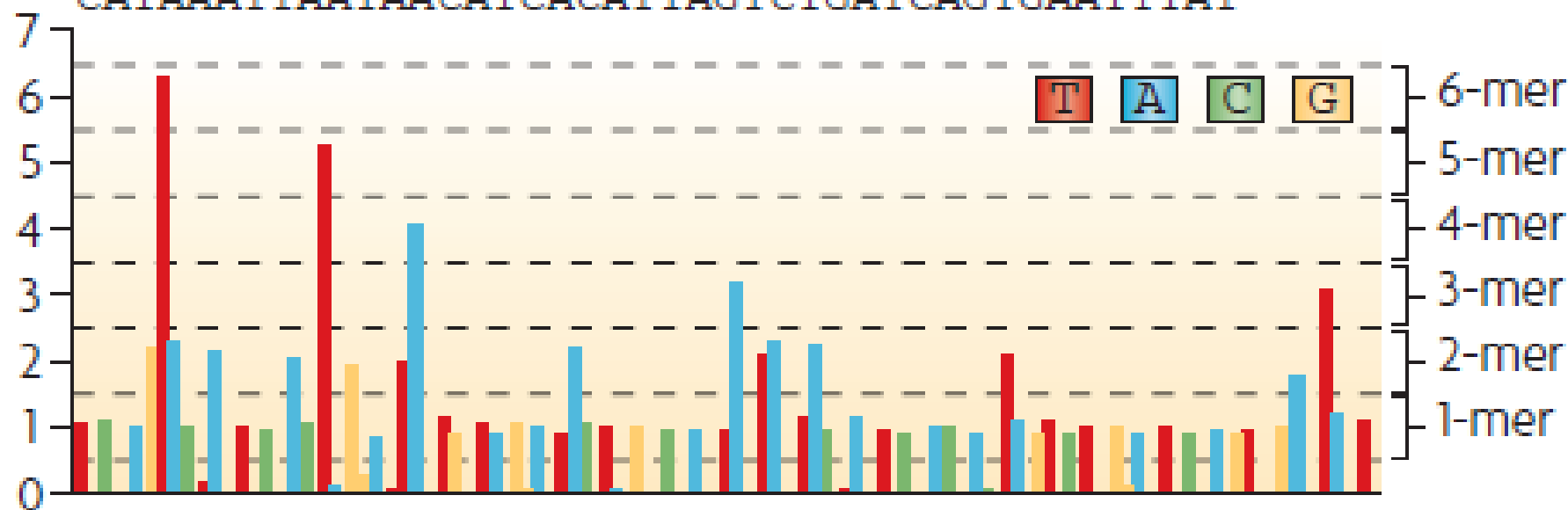
Pyrosekvenování – celkově až druhá popsaná NGS, ale zaznamenala rychlý rozmach

Pyrosequencing (*Ronaghi and Nyérén, 1996*), *Ukončena podpora*

Paralelizovaná verze pyrosekvenování vyvinuta firmou 454 Life Sciences, (nyní součást *Roche Diagnostics*) – odtud název **454 Pyrosequencing**. Detekuje se difosfát (pyrofosfát) při zabudování nukleotidu do DNA. Jednotlivé dNTP se přidávaly postupně, vždy jeden v jednom reakčním cyklu. V homopolymerních traktech je světelný signál úměrný počtu zařazených nukleotidů, tedy i uvolněných PPi.



TCAGGTTTTTAAACAATCAACTTTTTGGATTAAAATGTAGATAACTG
 CATAAATTAATAACATCACATTAGTCTGATCAGTGAATTTAT



Illumina (Solexa) sekvenování

Solexa, nyní součást Illumina, vyvinula sekvenační technologii založenou na reverzibilních barevně značených terminátorech.

DNA molekuly se nejprve připojí k primerům zakotveným k destičce a amplifikují se tzv. můstkovou izotermální amplifikací, čímž se tvoří klonální kolonie (**isothermal bridge amplification**).

Viz též videa, např.:

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

<https://www.youtube.com/watch?v=woMkfikWlxM>

<https://www.youtube.com/watch?v=HMyCqWhwB8E>

Glass slide method

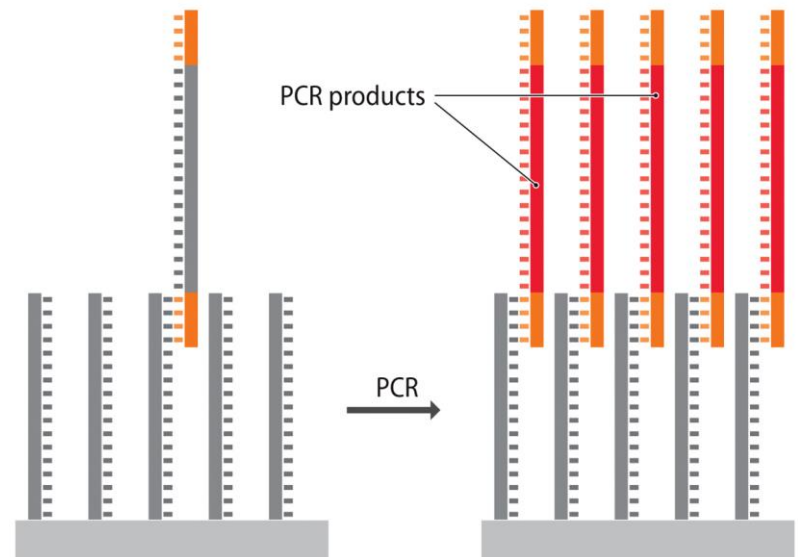
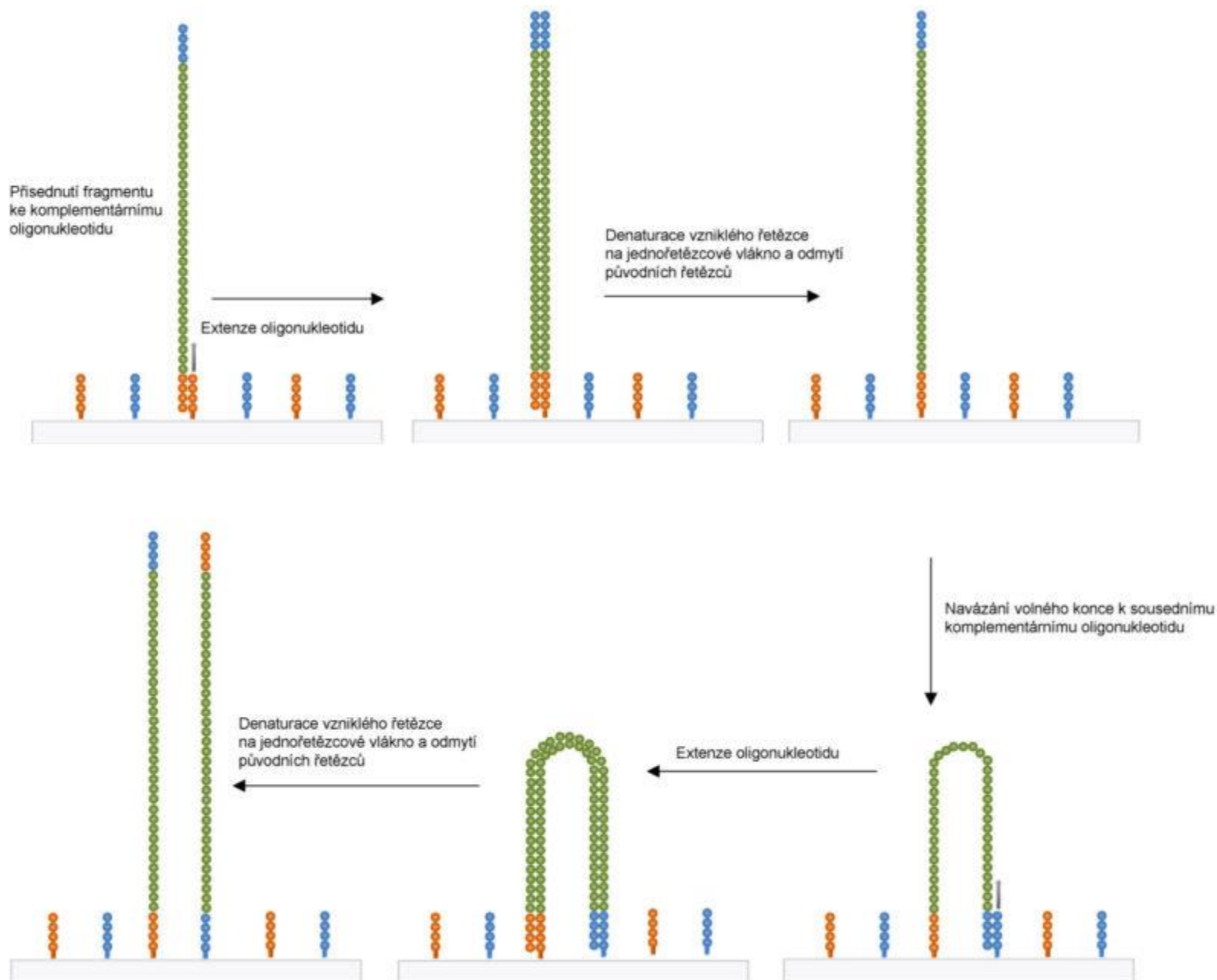
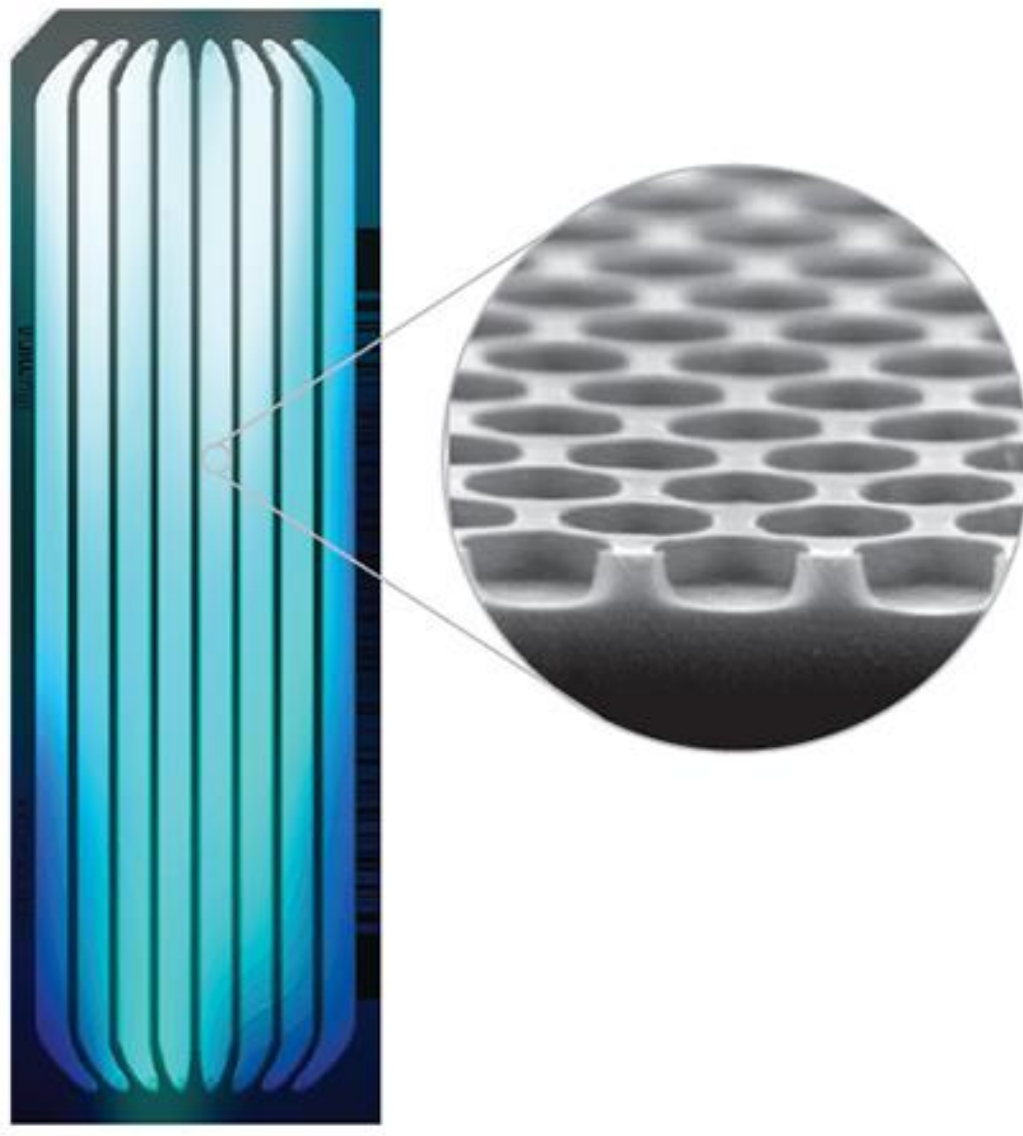


Figure 4.10a Genomes 4 (© Garland Science 2018)

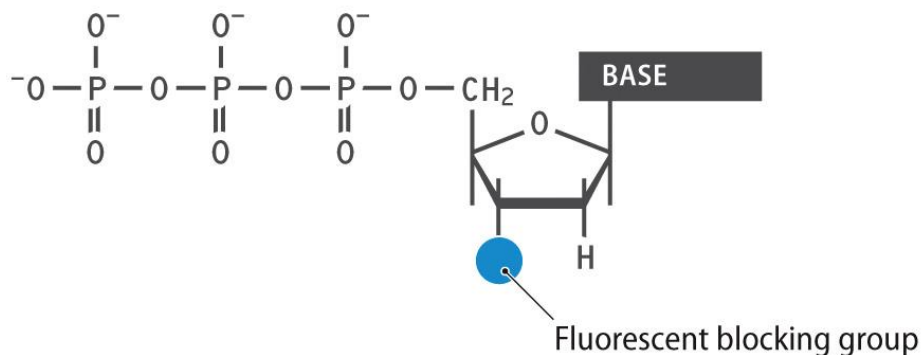
KLASTROVÁ (MŮSTKOVÁ) PCR



Průtočná cela pro Illumina sekvenování



(A) A reversible terminator nucleotide



(B) Reversible terminator sequencing

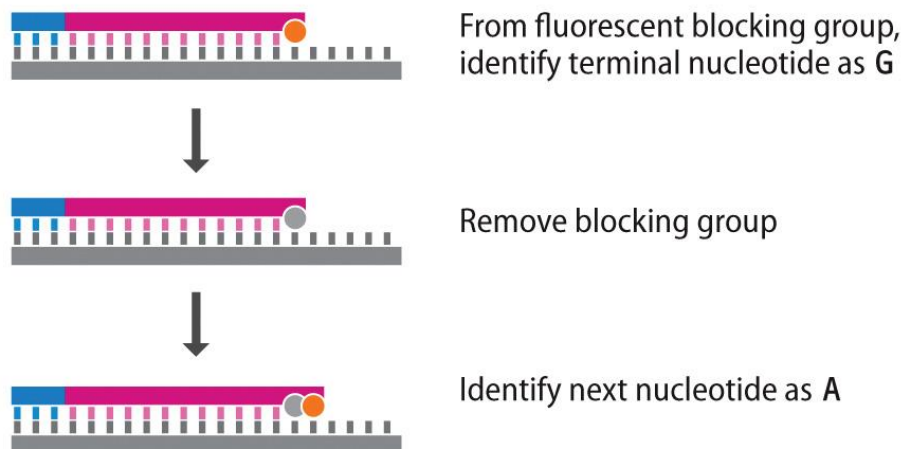
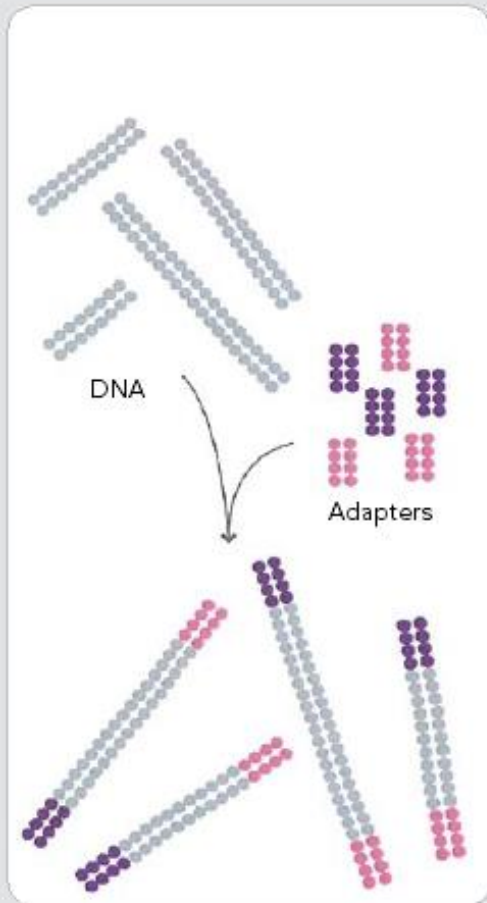


Figure 4.11 Genomes 4 (© Garland Science 2018)

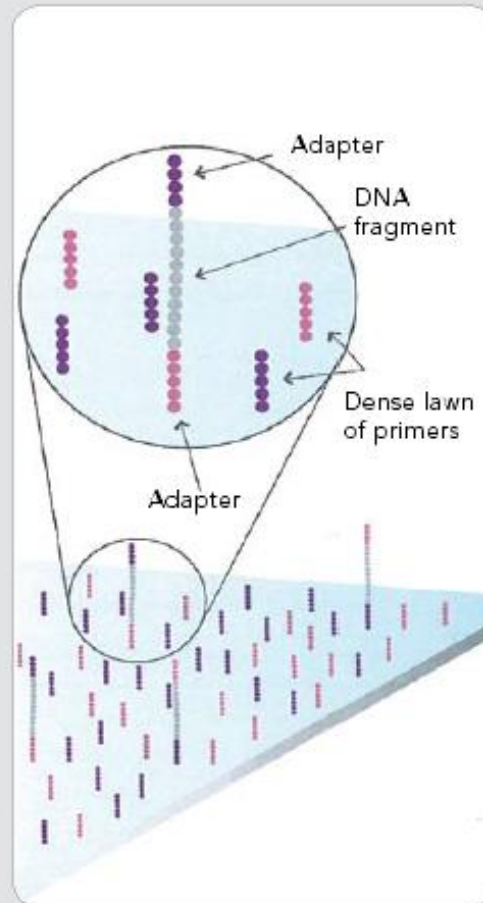
Přidají se všechny 4 typy reverzibilních nukleotidových terminátorů (RT-bases) (podobně jako ddNTP u Sangerovy metody), a nezabudované nukleotidy jsou odmyty. Na rozdíl od pyrosekvenování může být **DNA prodloužena v každém reakčním cyklu jen o jeden nukleotid**. Kamera sejme obraz fluoreskujících nukleotidů, pak je barvička společně s koncovou blokující skupinou v poloze 3' chemicky odstraněna z DNA, čímž se umožní další reakční cyklus.

1. PREPARE GENOMIC DNA SAMPLE



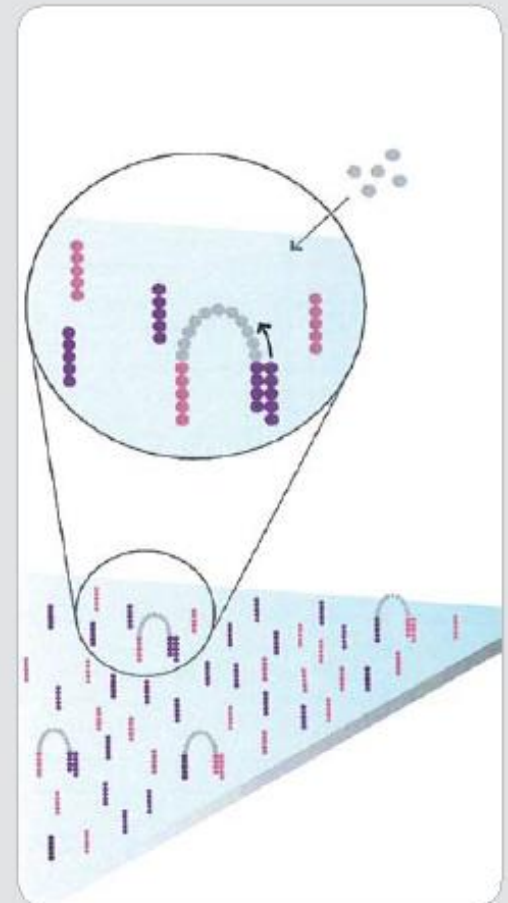
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



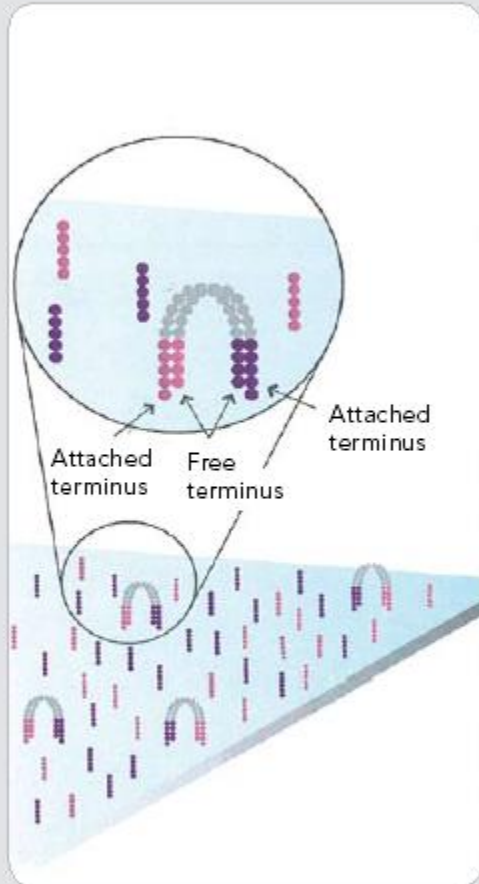
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



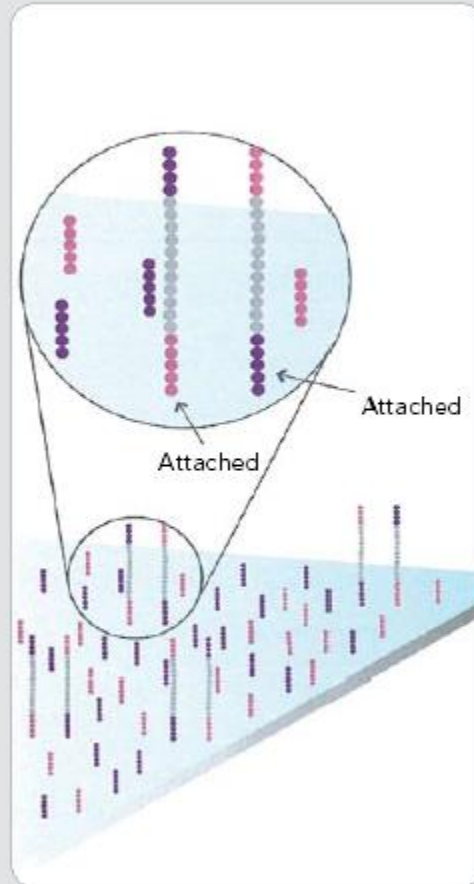
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME
DOUBLE-STRANDED



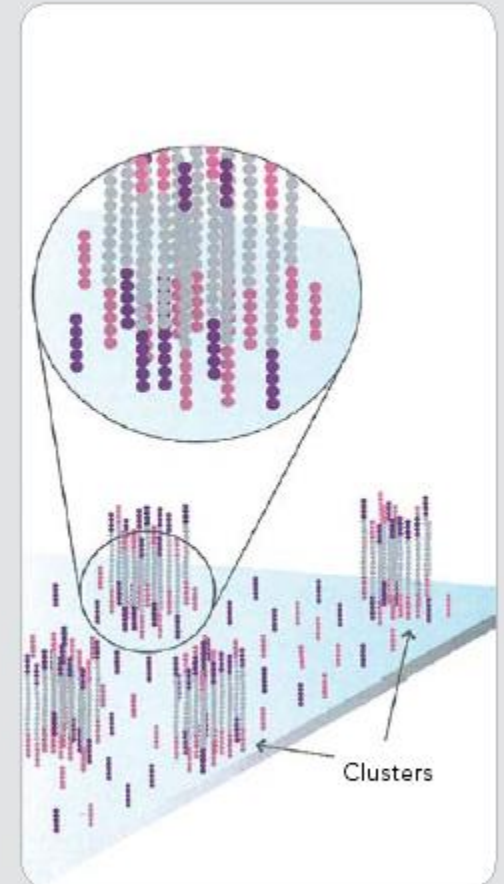
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-
STRANDED MOLECULES



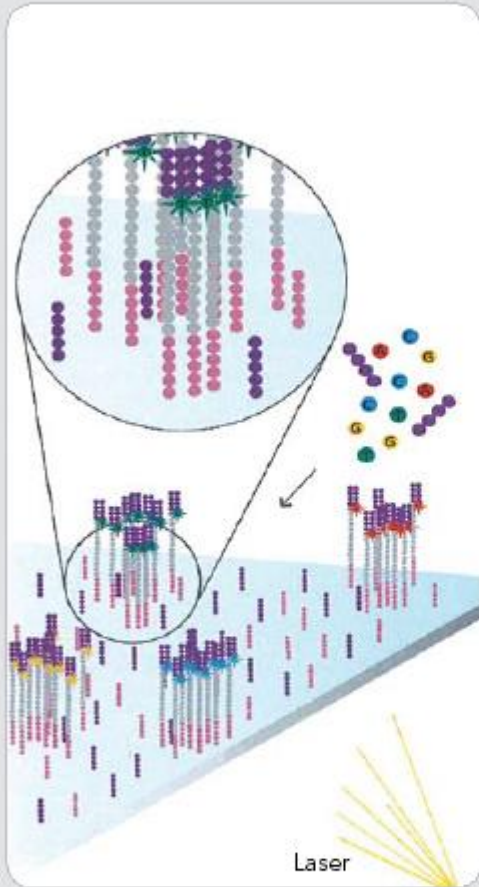
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



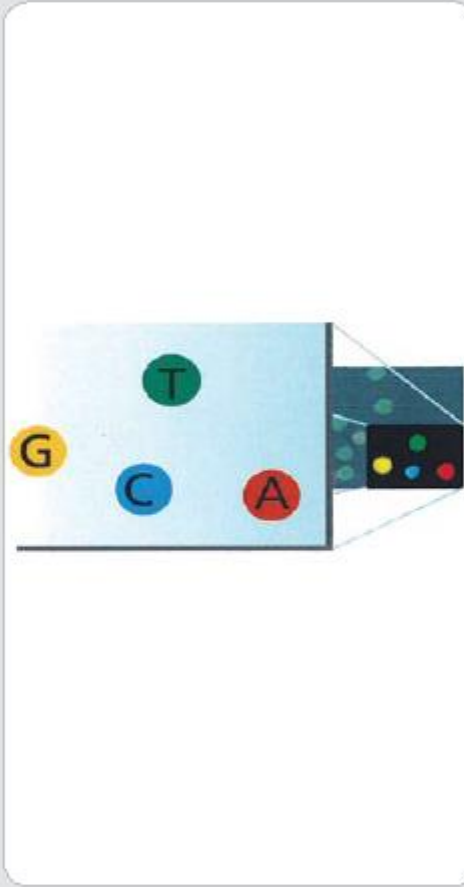
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

7. DETERMINE FIRST BASE



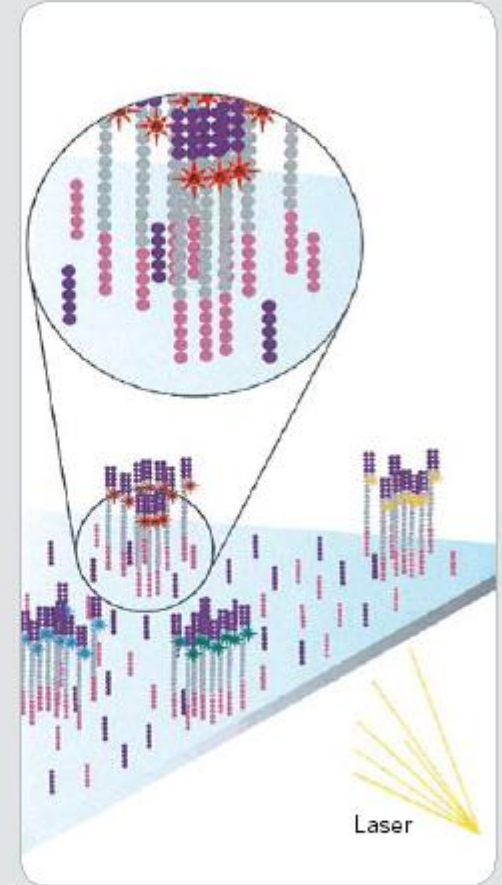
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE



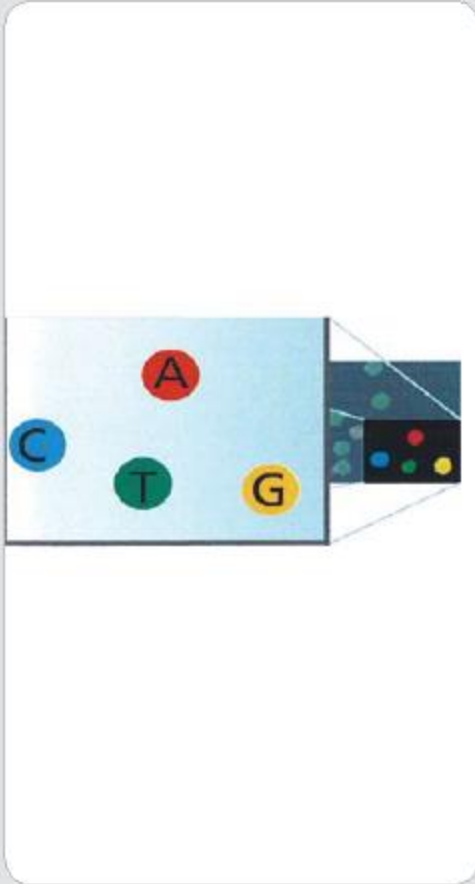
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE



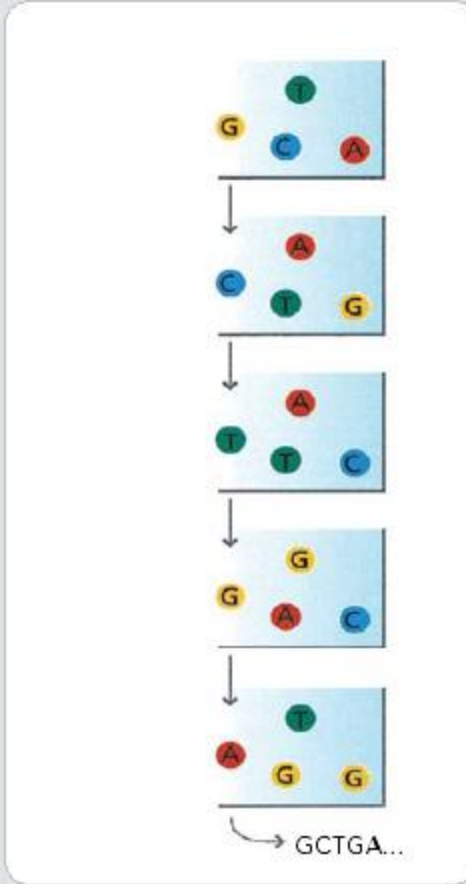
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

10. IMAGE SECOND CHEMISTRY CYCLE



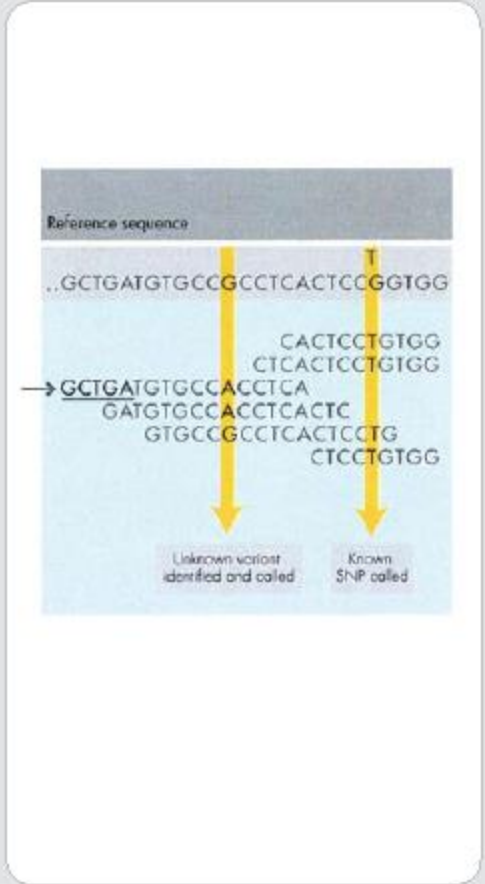
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

11. SEQUENCING OVER MULTIPLE CHEMISTRY CYCLES



The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

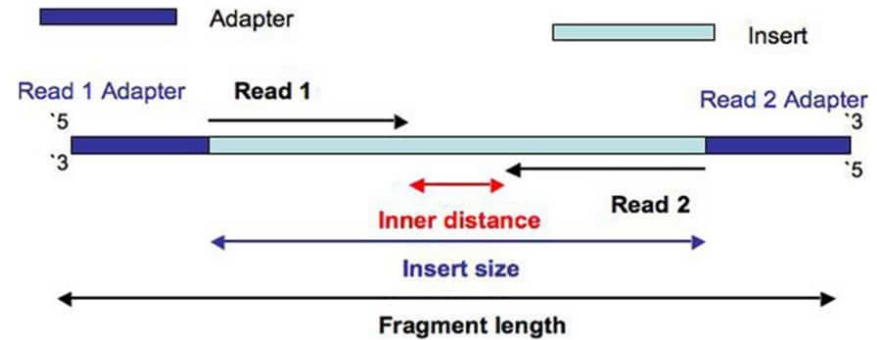
12. ALIGN DATA



The data are aligned and compared to a reference, and sequencing differences are identified.

Paired - End sekvenování

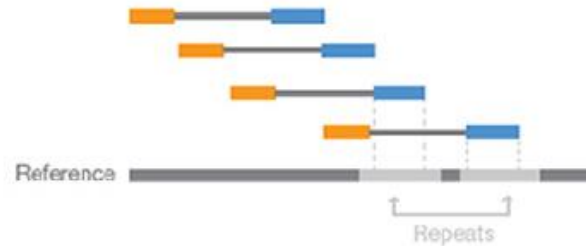
Čtení z Illumina platformy jsou poměrně krátká (100-150 nt, výjimečně 400 nt). Pokud chceme vylepšit informaci o vzájemné poloze čtených sekvencí, používá se tzv. paired-end sekvenování. Při něm se čtou oba konce daného fragmentu (po můstkové amplifikaci klastru). Tím získáme kromě čtení samotných i **informaci o jejich vzájemné poloze na genomu** (jejich vzdálenost lze vyhodnotit z velikosti fragmentů použitých v dané NGS knihovně).



Paired-End Reads



Alignment to the Reference Sequence



Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

Sekvenování na polovodičovém čipu

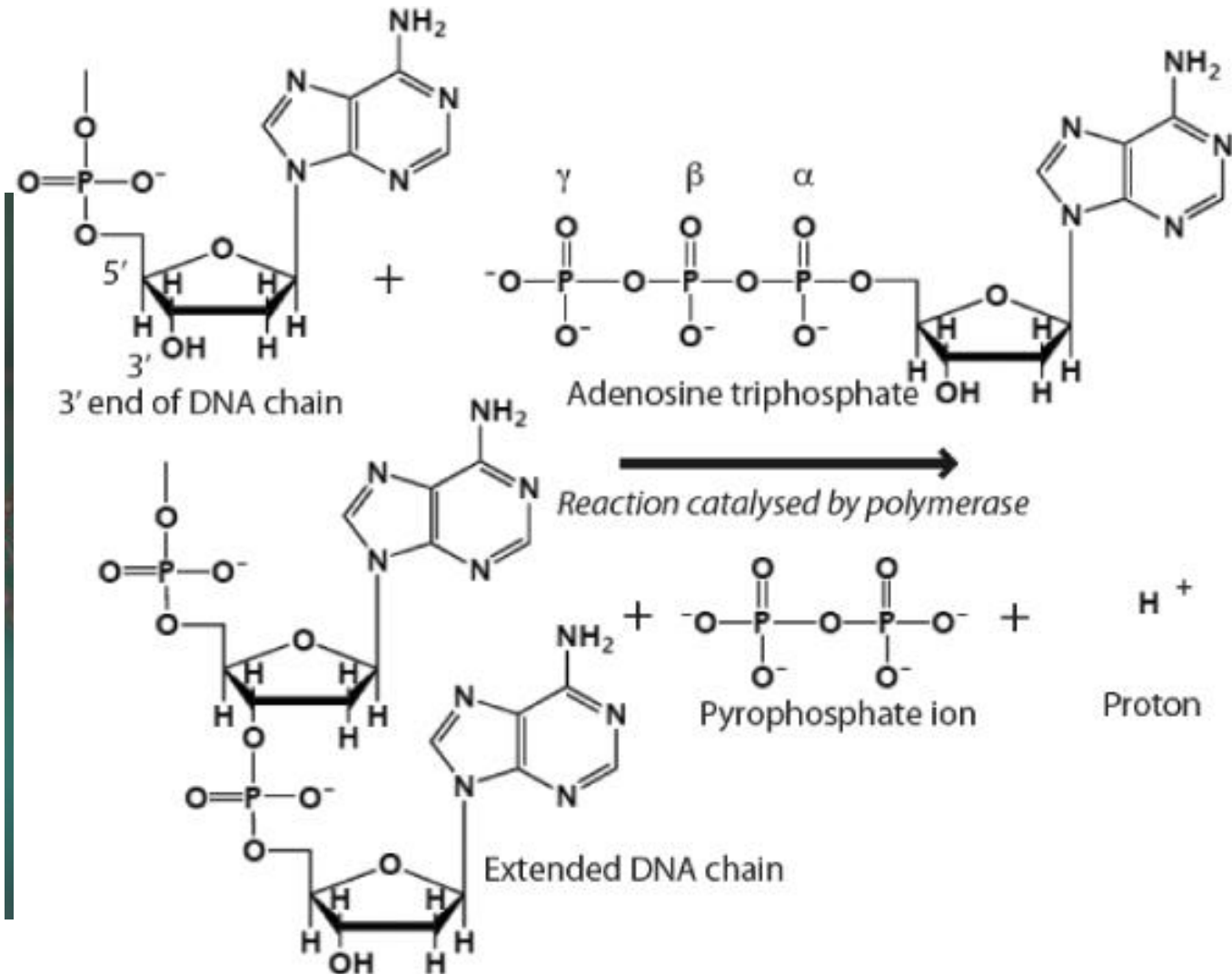
Ion Torrent Systems Inc. (nyní vlastněna Life Technologies) vyvinula systém využívající standardní sekvenační chemie, avšak s novým **systémem detekce na polovodičovém čipu**. Tato metoda je založena na **detekci vodíkových iontů**, které se uvolňují při polymeraci DNA, namísto optické detekce používané ostatními sekvenačními systémy.

Mikrojamka obsahující řetězec templátové DNA k sekvenování je zaplavena jedním typem nukleotidu. Pokud je tento komplementární k templátovému nukleotidu, je zabudován do rostoucího komplementárního řetězce. To způsobí uvolnění vodíkového iontu, který je detekován vysoce citlivým iontovým senzorem (v podstatě mikroverzí pH detektoru), což ukazuje, že reakce proběhla. V případě homopolymerních úseků v templátu se zabuduje příslušný počet nukleotidů v jednom cyklu, což vede k uvolnění odpovídajícího počtu vodíkových iontů a úměrně vyššímu elektrickému signálu.

<http://www.iontorrent.com/>

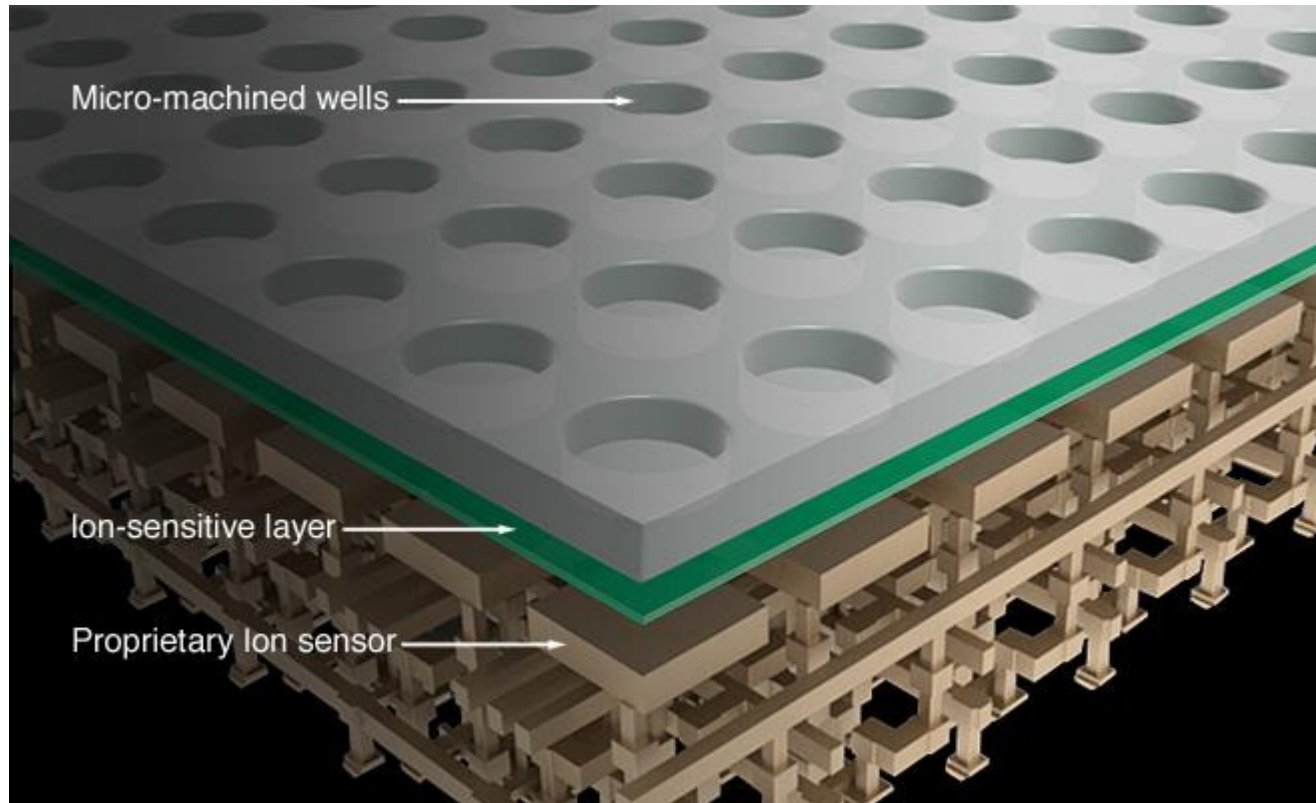
Hlavní aplikace: genotypování, cílené resekvenování (diagnostika, forensní DNA analýza)....**ukončeno**

Ion Torrent



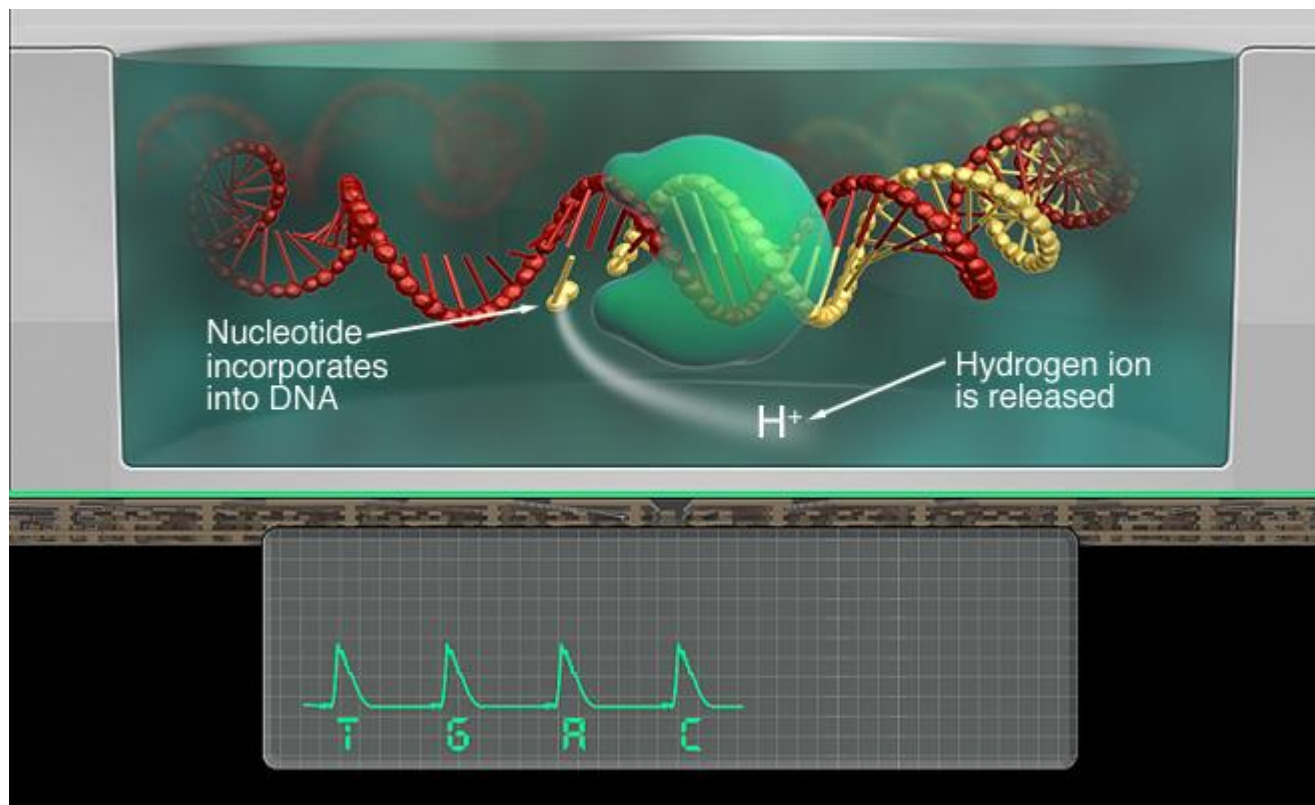
Když je nukleotid zabudován do řetězce DNA polymerázou, uvolní se **vodíkový ion** jako vedlejší produkt a může být detekován citlivým pH detektorem

Ion Torrent



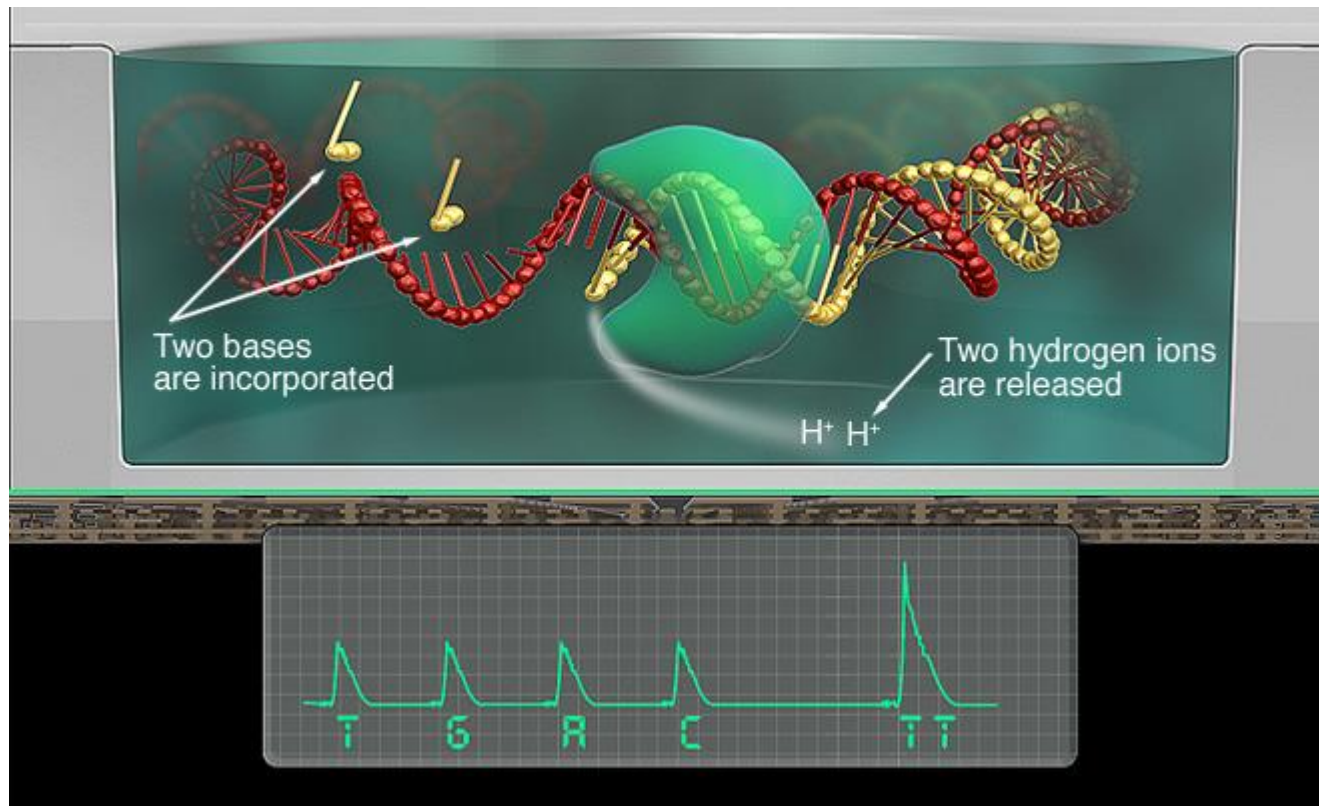
Ion Torrent™ využívá vysokohustotní síť mikrojamek, aby byl možný vysoce paralelní způsob provádění reakcí. **Každá jamka obsahuje jiný templát DNA.** Pod jamkami je iontově-citlivá vrstva a pod ní speciální iontový senzor.

Ion Torrent



Jestliže je nukleotid, např. dCTP, přidán k templátové DNA a je do něj zabudován, uvolní se H^+ iont. Jeho náboj změní pH roztoku, což je detekováno iontovým senzorem. Tento sekvenátor je v podstatě miniaturním vysoce paralelním *solid-state* pH metrem. Pak se zaplavuje destička postupně dalšími dNTP a zaznamenávají případné signály, když dojde k jejich zabudování.

Ion Torrent



V případě výskytu dvou stejných bází po sobě bude zaznamenána změna potenciálu dvojnásobná. Díky přímosti detekce – bez skenování, videokamer a snímání světelných záblesků – je zabudování každého nukleotidu snímáno poměrně rychle - během sekund.

Sekvenování třetí generace (TGS): dlouhá čtení a sekvenování jednotlivých molekul bez amplifikace (*single molecule sequencing*).

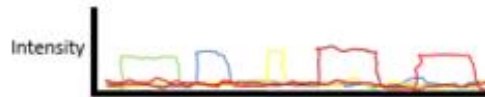
V současnosti se jako nejperspektivnější jeví:

PacBio SMRT seq

DNA passes thru
polymerase in an
illuminated volume



Raw output is fluorescent signal
of the nucleotide incorporation,
specific to each nucleotide

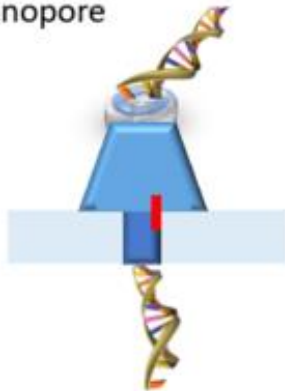


A,C,T,G have known pulse
durations, which are used to
infer methylated nucleotides



Oxford Nanopore

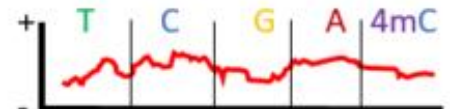
DNA passes thru
nanopore



Raw output is electrical signal
caused by nucleotide blocking
ion flow in nanopore

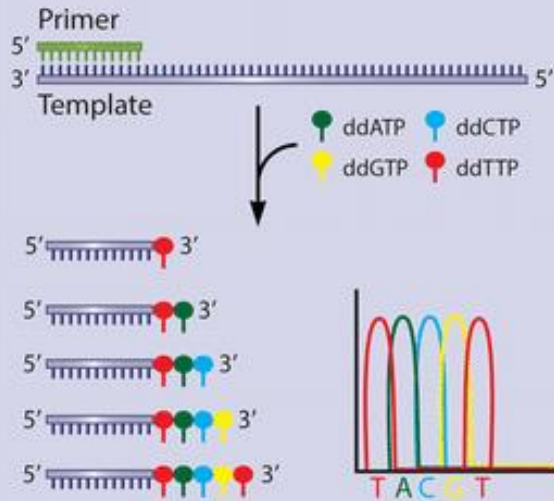


Each nucleotide has a specific
electric "signature"



First Generation

Shotgun Sequencing

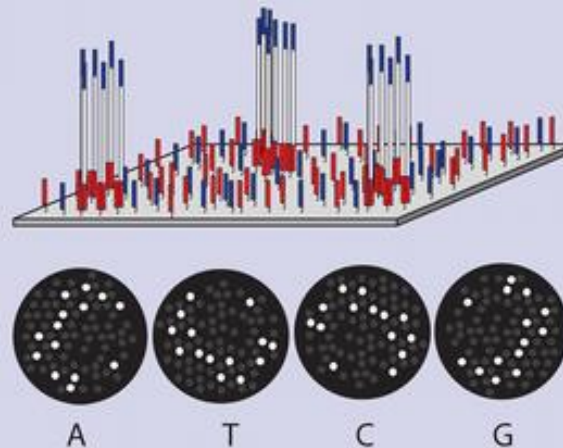


- Sequencing by synthesis
- High accuracy
- Long read lengths
- Relatively small amount of data generated

e.g., ABI capillary sequencer (ABI)

Second Generation

Massively Parallel Sequencing

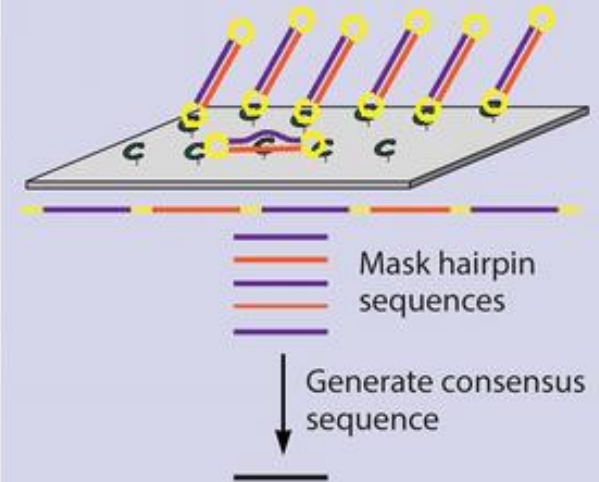


- Sequencing by synthesis
- Amplified templates are generated during sequencing, reducing the requirements for starting material
- High accuracy
- Short read lengths

e.g., MiSeq (Illumina), Ion Torrent (Thermo Fisher Scientific)

Third Generation

Single-molecule Sequencing

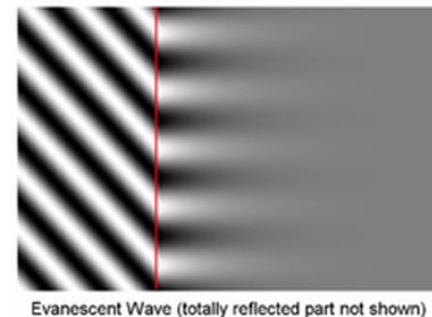
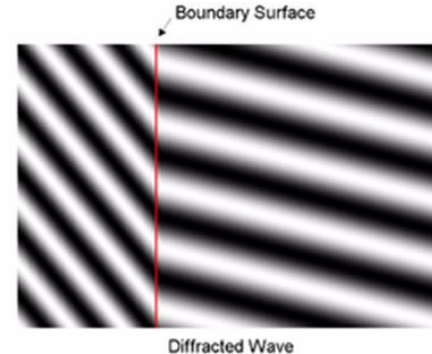
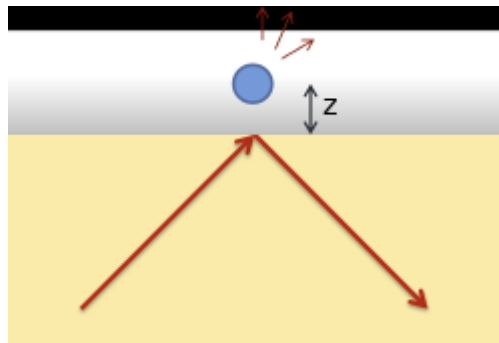


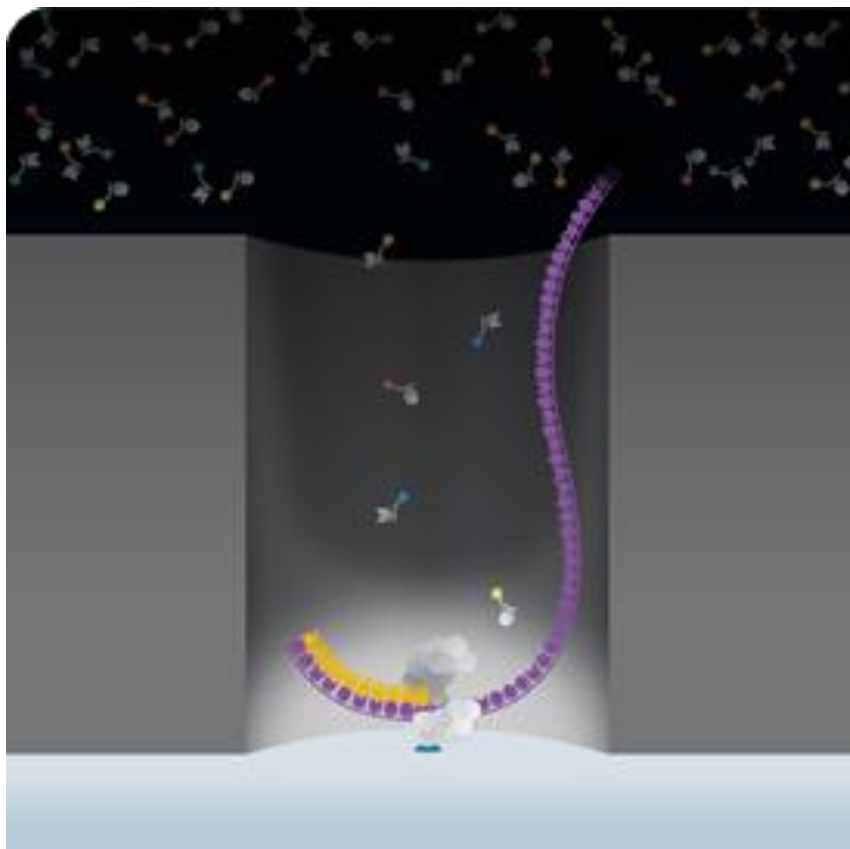
- Single-molecule templates
- Low accuracy
- Long read lengths

e.g., Single-Molecule Real-Time (SMRT) — Sequencing (Pacific Biosciences), MinION (Oxford Nanopore Technologies)

Single molecule real time (SMRT™) sequencing

SMRT sekvenování je založeno na sekvenování syntézou. DNA je syntetizována v mikrojamkách **zero-mode wave-guides** (ZMWs) s polymerázou zachycenou na dně ZMW jamky. Fluorescenčně značené nukleotidy plavou volně v roztoku. Jamky jsou konstruovány tak, že **jen fluorescence, která se objeví přímo na dně jamky je detekována**. Fluorescenční značka je odštěpena při zabudování nukleotidu, takže rostoucí řetězec DNA zůstává nemodifikovaný. Tato metodika umožňuje kromě čtení sekvence také **detekci modifikací** (např. methylace cytosinu – z pomalejší kinetiky polymerace). Tento přístup umožňuje čtení i přes **100,000 nt**, průměrná délka čtení kolem **45 kb**.

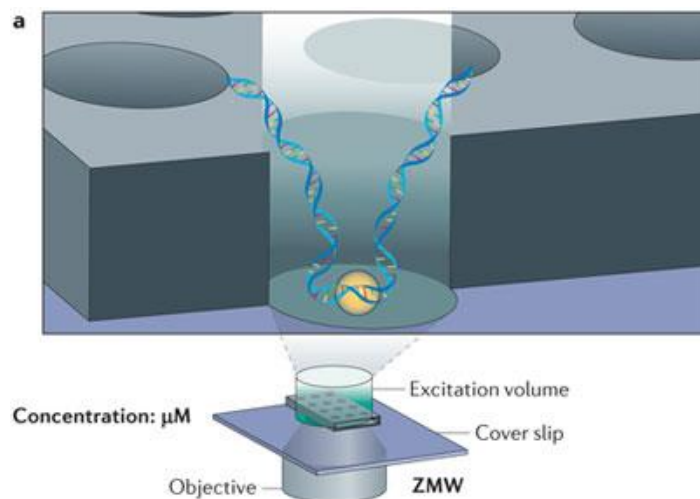




Aktivní **polymeráza** je **imobilizovaná na dně každé ZMW**, nukleotidy difundují dovnitř ZMW komůrky. Pro detekci zabudování a identifikaci báze je každý ze čtyř nukleotidů A, C, G a T značen jinou fluorescenční barvičkou s **odlišným emisním spektrem**. Protože **excitační záření směřuje na dno ZMW**, nukleotid držený polymerázou před zabudováním emituje signál, který identifikuje zabudovávanou bázi.

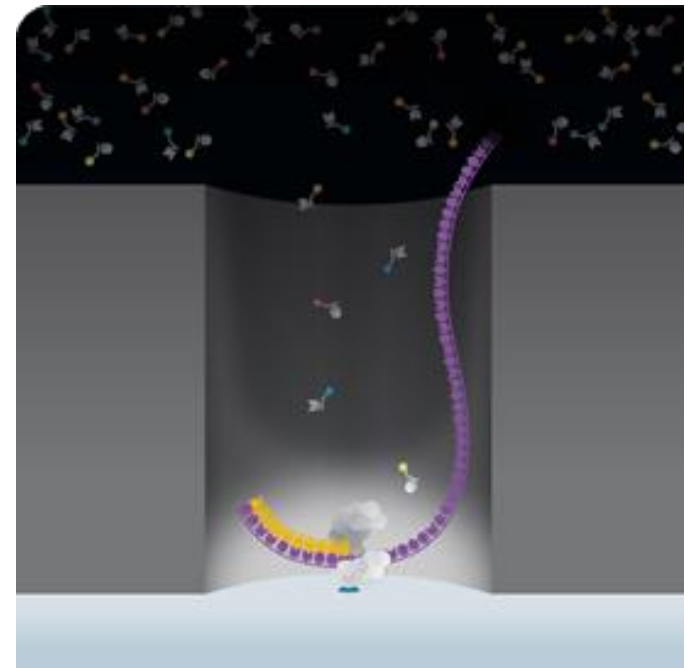
Reakční komůrka pro SMRT

Klíčovým problémem pozorování činnosti **DNA polymerázy v reálném čase** je možnost **zabudovat jediný nukleotid** vzatý z velké zásoby potenciálních nukleotidů během syntézy, aniž by ostatní značené rušily. Stejný princip se využívá u mikrovlnné trouby, kde relativní velikost otvorů v perforované desce na skleněném okénku průchod mikrovln, ale mnohem kratší vlnové délky viditelného světla propustí a je možné pozorovat potraviny při ohřívání. Tento princip, miniaturizovaný do nanorozměrů, využívá **zero-mode waveguide**, or ZMW.



ZMW je nano-jamka, desítky nm v průměru, vytvořená ve 100nm kovovém filmu, deponovaném na skleněném podkladu. Malá velikost ZMW brání viditelnému paprsku laseru, jehož vlnová délka je asi 600nm, v průchodu celou ZMW. Místo aby paprsek prošel celou ZMW, se světlo při vstupu do ZMW exponenciálně tlumí. Proto při posvícení laserem přes sklo dovnitř ZMW, **dojde pouze k osvětlení dna ZMW do vzdálenosti asi 30 nm**. V každém ZMW je připojena ke dnu jedna molekula DNA polymerázy. Nukleotidy (každý značený jiným fluoroforem) pak tečou nad mikrodestičkou se ZMW a rychle difundují do ZMW, takže během mikrosekund se dostanou až ke dnu ZMW a k polymeráze. Obklopí ji a pak opět difundují z jamky. Protože laserové světlo neprojde nad dno dovnitř jamky dál než 30 nm, nedochází k excitaci volných dNTP a prostředí uvnitř ZMW je tmavé. Když je polymerázou detekován správný nukleotid, **je během milisekund zabudován** do rostoucího řetězce DNA, na rozdíl od **difúze**, která trvá **mikrosekundy**. Tento rozdíl v čase způsobuje vyšší intenzitu signálu zabudovávaných versus nezabudovávaných nukleotidů, což zvyšuje poměr signálu vůči šumu. Tím pádem má **ZMW schopnost detekovat jednotlivou inkorporaci nukleotidu na pozadí roztoku všech fluorescenčně značených nukleotidů**.

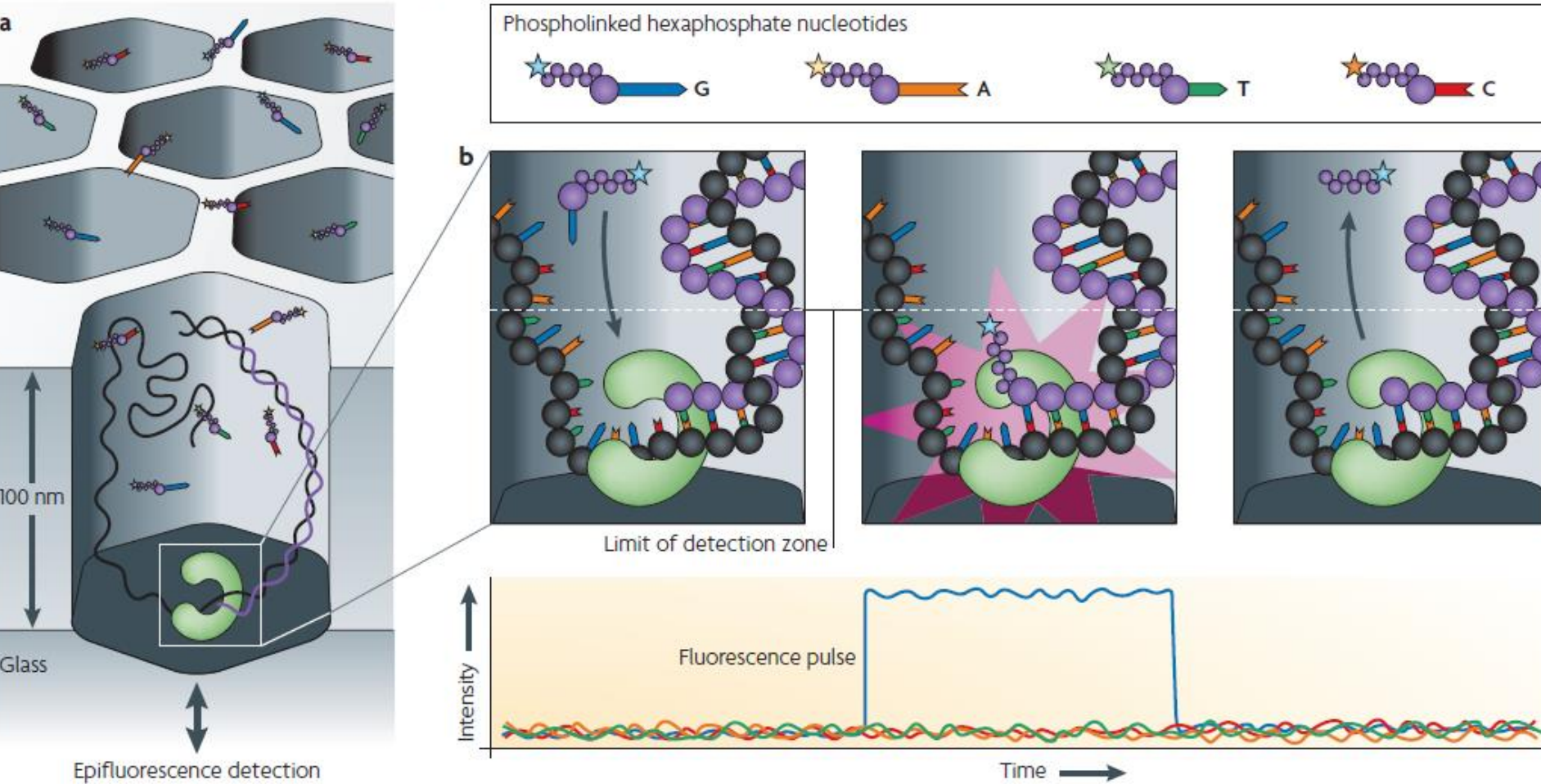
Sekvenování DNA se provádí ve speciálních SMRT komůrkách, z nichž každá obsahuje pole asi 75000 jednomolekulových sekvenačních reakcí vedle sebe.

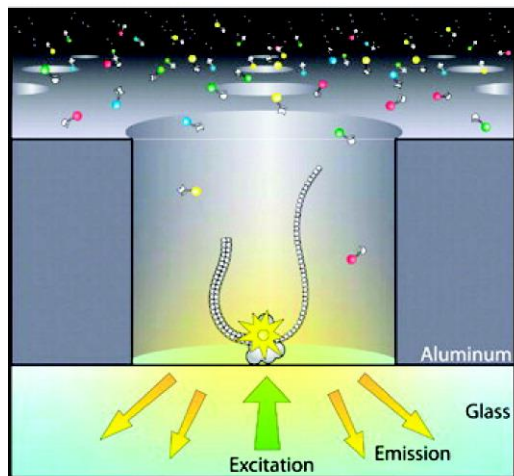
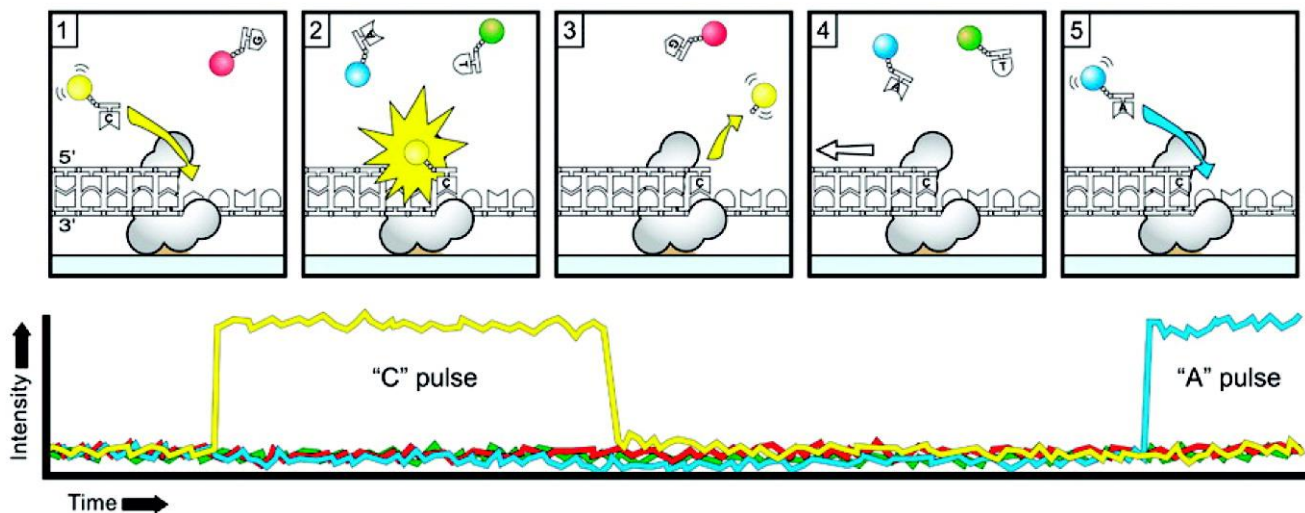


Phospholinked nukleotidy – fluorescenční značka připojena k fosfátovému řetězci, nikoli k bázi

- fosfátový řetězec je při inkorporaci báze odštěpen spolu s barvičkou, difunduje pryč a řetězec zůstává neznačený – díky tomu značka neblokuje další práci polymerázy (při SMRT sekvenování se provádí čtení opakovaně – viz další obrázek)

Pacific Biosciences — Real-time sequencing

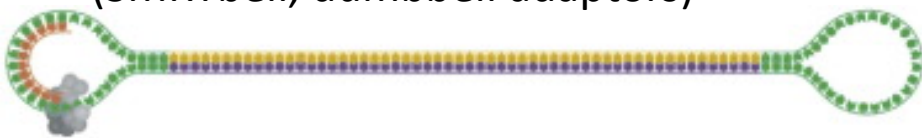


A**B**

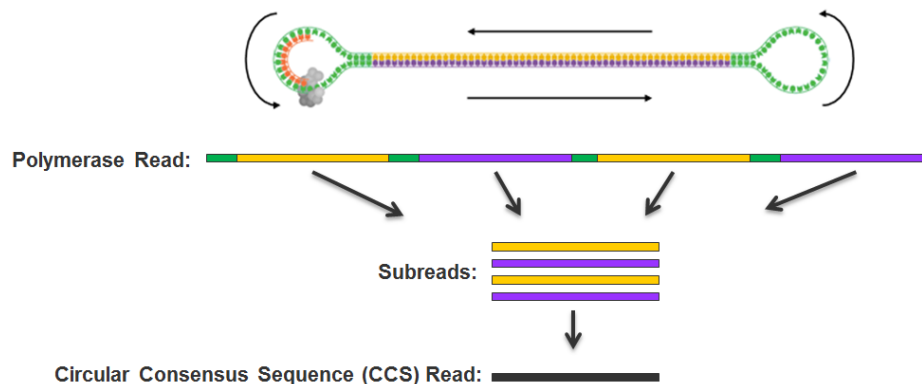
Schema PacBio real-time single molecule sequencing. (A) Boční pohled na jednu ZMW nanostrukturu obsahující jednu molekulu DNA polymerázy ($\Phi 29$) vázané ke dnu skleněného povrchu. ZMW a konfokální zobrazovací systém umožňuje detekci fluorescence jen na povrchu dna každé ZMW. (B) Znázornění zabudovávání fluorescenčně značeného nukleotidového substrátu do sekvenovaného templátu. Odpovídající časový průběh detekce fluorescence vzhledem ke každému z 5 inkorporačních kroků je ukázán dole.

Thomas P. Niedringhaus; Denitsa Milanova; Matthew B. Kerby; Michael P. Snyder; Annelise E. Barron; *Anal. Chem.* **2011**, 83, 4327-4341.
DOI: 10.1021/ac2010857

Adaptory mají vlásenkovou strukturu (SMRTbell, dumbbell adaptors)



To umožňuje cyklické čtení polymerázou při vlastní sekvenaci



Nanopore sekvenování (Oxford Nanopore Technology)

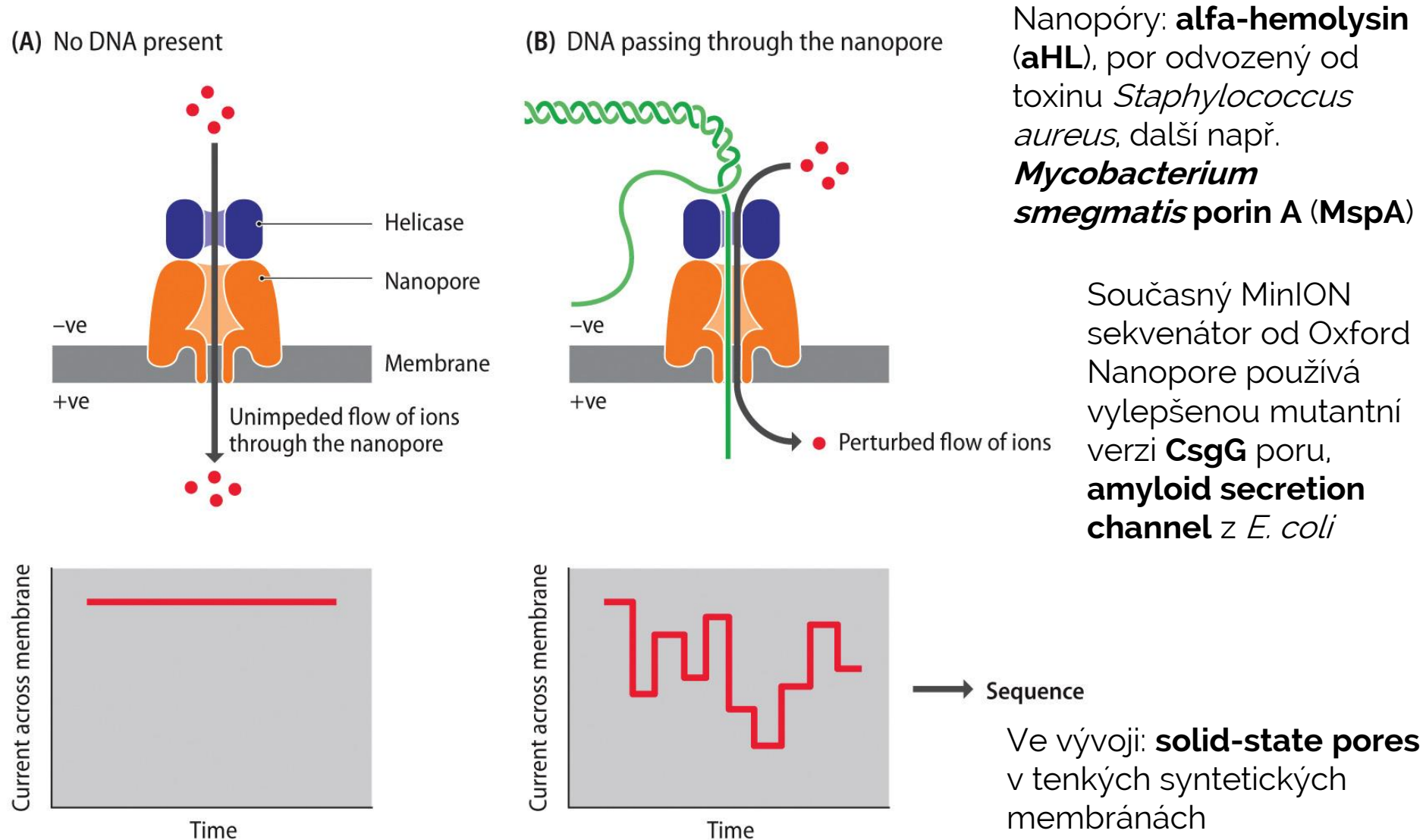
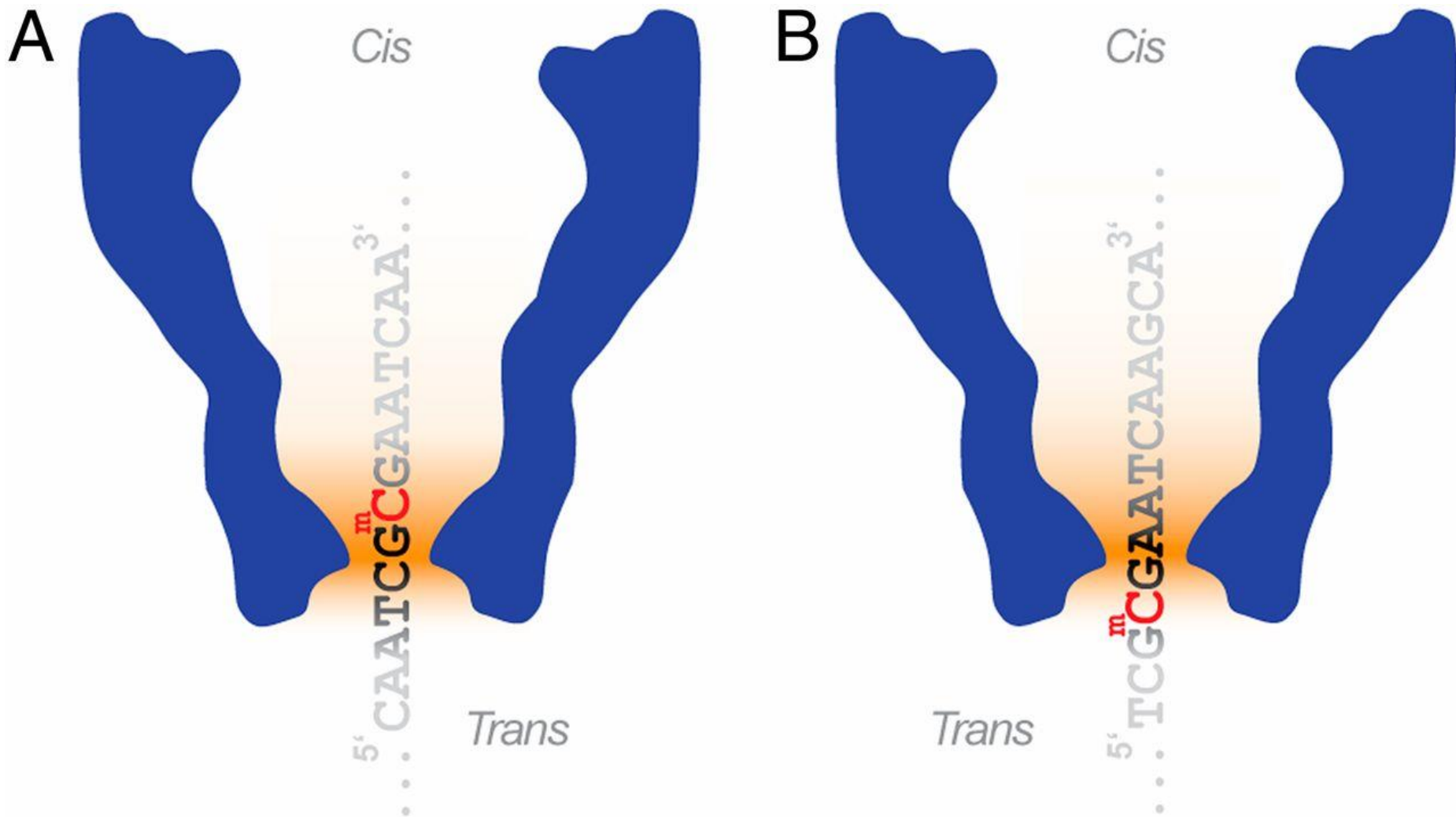


Figure 4.16 Genomes 4 (© Garland Science 2018)

Detaily viz např. <https://nanoporetech.com/resource-centre>

Nečtou se jednotlivé báze, ale celá skupina bází limitující proud. Oranžově vyznačená oblast v největším zúžení uvnitř nanoporu má největší proudovou hustotu, v tomto úseku DNA je tedy vliv bází a jejich modifikací na velikost el. proudu největší.



Vyvíjejí se ONT aplikace na **přímou detekci modifikací RNA** – těch je asi 170 !!! – klíčový nástroj pro rozvoj **EPITRANSKRIPТОMIKY**.

Hlavní výhody nanopore sekvenování:

- V současnosti umožňuje **nejdelší čtení** (omezeno jen schopností připravit dostatečně dlouhé vysokomolekulární DNA – většinou stovky kb) – snazší sestavování kontigů, chromosomálních scaffoldů a genomů, možnost pročtení rozsáhlých repetitivních oblastí (např. centromer, rDNA klastrů)
- **Metoda bez PCR (nebo podobné amplifikace)** – nedochází ke zkreslení zastoupení různých oblastí genomu ve výsledcích
- Možnost **přímé detekce modifikací nukleotidů** (např. methylace cytosinu)
- Možnost **přímého sekvenování nativní RNA** (bez reverzní transkripce) – tím pádem i možnost přímé detekce modifikací RNA, sestřihových forem, vyloučení artefaktů spojených s RT-PCR.



DNA was extracted from giant sequoia (*Sequoiadendron giganteum*) needle tissue, prepared for sequencing using a ligation-based method, and sequenced for 48 hours on 13 MinION Flow Cells. Hybrid assembly was then performed, with the incorporation of chromatin conformation capture data.



RESEARCH SPOTLIGHT

*Salzberg et al.,
2019*

The use of long-read sequencing data generated on the MinION™ device significantly improved the assembly of the giant sequoia genome, and facilitated the generation of the largest chromosome scaffold to date.

22x

genome depth of coverage with nanopore reads

8.2 Gb

assembled genome size

348 kb

increase in N50 from addition of long nanopore reads

985 Mb

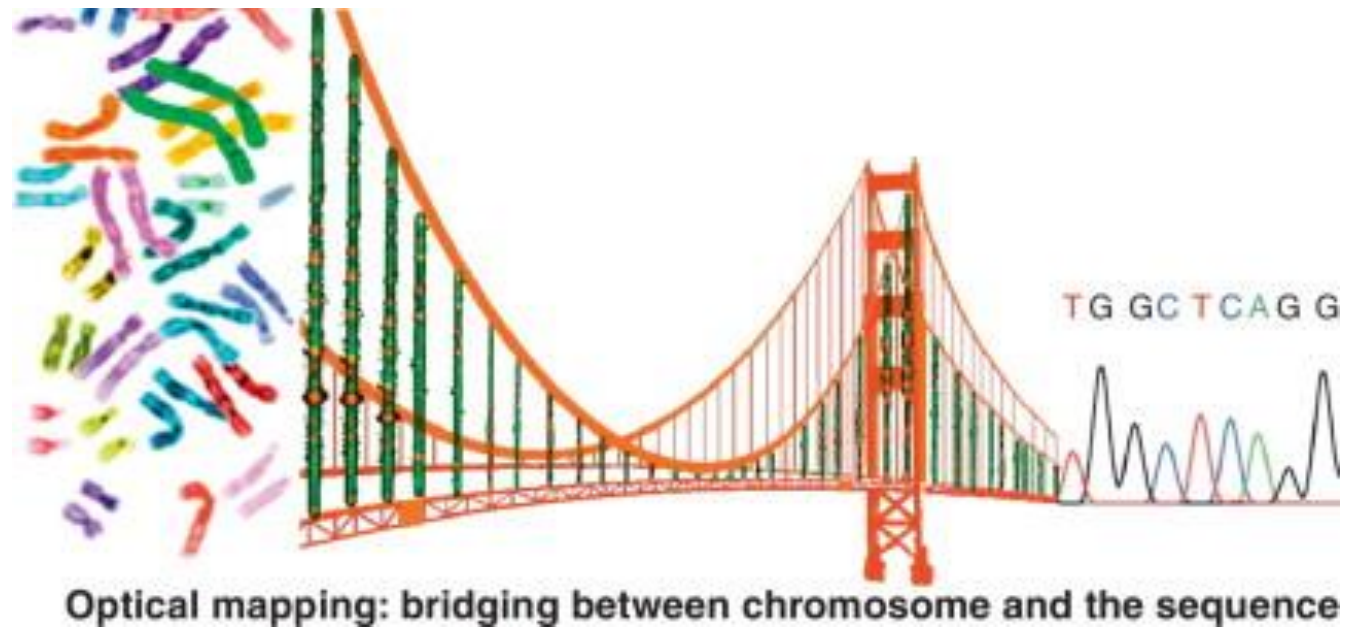
largest scaffolded chromosome

30-fold

reduction in contig number by including nanopore data

Optické mapování

Technika překlenující mezeru mezi sekvenčními a zobrazovacími technikami (FISH, fibre-FISH)

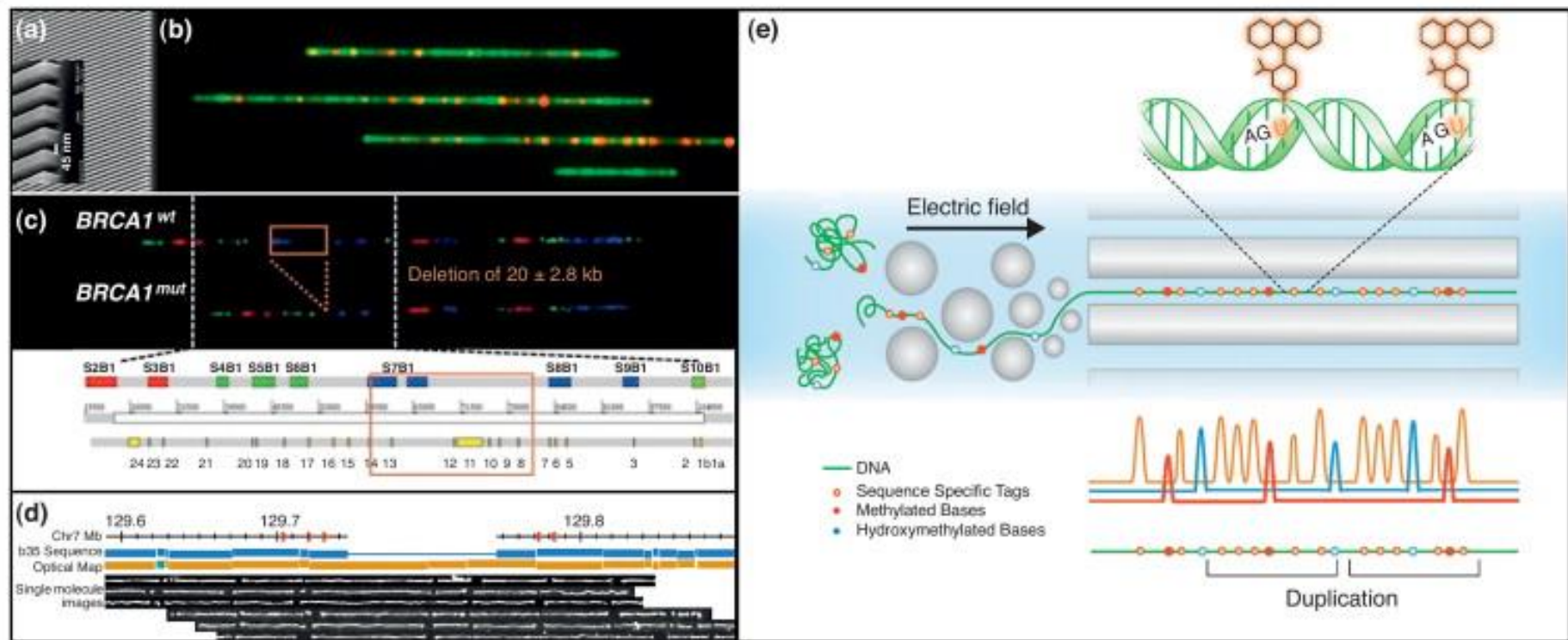


- ▶ Optický čarový kód se vytváří fluorescenčním značením krátkých sekvenčních motivů podél natažené DNA
- ▶ Optické mapy doplňují sekvenční výsledky pro účely uzavírání mezer, dokončování a validaci *de novo* sestavených genomů.
- ▶ Strukturní varianty genomů nepřístupné sekvenčním technikám mohou být přímo vizualizovány.
- ▶ Lze mapovat i epigenetické značky jako metylaci DNA, vazbu DNA-vazebných proteinů přímo na jednotlivých fragmentech genomu.

Michal Levy-Sakin , Yuval Ebenstein

Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy

Current Opinion in Biotechnology, Volume 24, Issue 4, 2013, 690 - 698



Current Opinion in Biotechnology

Schema optického mapování. (a) Snímek vysoce paralelního nanokanálkového čipu (BioNano Genomics) získaný pomocí skenovací el. mikroskopie (SEM) (b) DNA molekuly BACů (zeleně) jsou štěpeny v jednom vlákně DNA specifickým enzymem v určitých sekvenčních motivech a opraveny polymerázou, která zabudovává fluorescenčně značené nukleotidy (oranžové tečky) jsou snímány v nanokanálku. (c) Molekulární „česání“ DNA (combing). Rozsáhlá přeuspořádání v lokusu BRCA1 jsou detekována u pacientek s rakovinou prsu. Del ex 8–13 je viditelná jako delecce modrého signálu S7B1. (d) Restrikčně-optické mapování. 90 kbp inzerce (červené šipky). Pod mapou je fotomontáž několika jednomolekulových snímků které dokazují tuto inzerci. (e) optické mapování v nanokanálkách: molekuly DNA (zelené) jsou specificky značeny různobarevnými sondami ukazujícími různé genomické/epigenomické informace jako sekvenční motivy a místa methylace (barevné tečky). Vložené elektrické pole táhne molekuly přes stále jemnější nano-překážky (šedé kuličky), které usměrňují molekuly do nanokanálků. Jakmile je DNA natažena a uvnitř kanálku, lze přesně měřit vzdálenosti mezi značkami pomocí fluorescenční mikroskopie. Software generuje konsenzuální mapu sekvenčních motivů a epigenetických modifikací jednotlivých molekul. Tyto mapy usnadňují analýzu strukturních a apigenetických variant, např. rozsáhlých duplikací a variant v methylačním obrazu.

Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy

Current Opinion in Biotechnology, Volume 24, Issue 4, 2013, 690 - 698 Michal Levy-Sakin, Yuval Ebenstein

Hlavní milníky v sekvenování DNA

- 1953** Objev struktury DNA jako dvojité šroubovice.
- 1972** Vývoj technologií rekombinantní DNA, umožňující izolaci definovaných fragmentů DNA
- 1977** První kompletní sekvenovaný DNA genom - bakteriofág ϕ X174.
- 1977** Allan Maxam a Walter Gilbert publikují "DNA sequencing by chemical degradation".
Frederick Sanger, nezávisle, publikuje "DNA sequencing with chain-terminating inhibitors".
- 1984** vědci z Medical Research Council rozluštili kompletní sekvenci DNA Epstein-Barr viru, 170 kb.
- 1986** Laboratoř Leroye E. Hooda, CalTech oznamuje první poloautomatický sekvenátor DNA.
- 1987** Applied Biosystems uvádí na trh první automatický sekvenátor, model ABI 370.
- 1990** U.S. National Institutes of Health (NIH) začíná rozsáhlé sekvenační projekty na *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, a *Saccharomyces cerevisiae* (cena US\$0.75/base).
- 1991** laboratoř Craiga Ventera začíná sekvenovat lidské EST (expressed sequence tags) ve snaze charakterizovat kódující frakci lidského genomu.
- 1995** Craig Venter, Hamilton Smith a další z Institute for Genomic Research (TIGR) publikují první úplný genom volně žijícího organismu, bakterie *Haemophilus influenzae*. Kružnicový chromozom má 1,830,137 bp. První použití celogenomového shotgun sekvenování, bez potřeby předchozího mapování.
- 1996** Pål Nyrén a jeho student Mostafa Ronaghi v Royal Institute of Technology ve Stockholmu publikují svoji metodu pyrosekvenování.
- 1998** Phil Green a Brent Ewing, University of Washington, publikují „pyrex“ - nástroj pro analýzu dat ze sekvenátoru.
- 2000** Lynx Therapeutics publikuje a uvádí na trh "MPSS" – paralelizované sekvenování založené na ligaci (první next-generation" sequencing).
- 2001** První verze sekvence lidského genomu publikována.
- 2004** 454 Life Science uvádí na trh paralelizovanou verzi pyrosekvenování. První verze jejich sekvenátoru snížila náklady na sekvenování 6x v porovnání s automatickými Sangerovými sekvenátory, stala se druhou (ale rozšířenější) NGS technologií, po MPSS.

Porovnání v současnosti nejužívanějších NGS a TGS technologií/platform

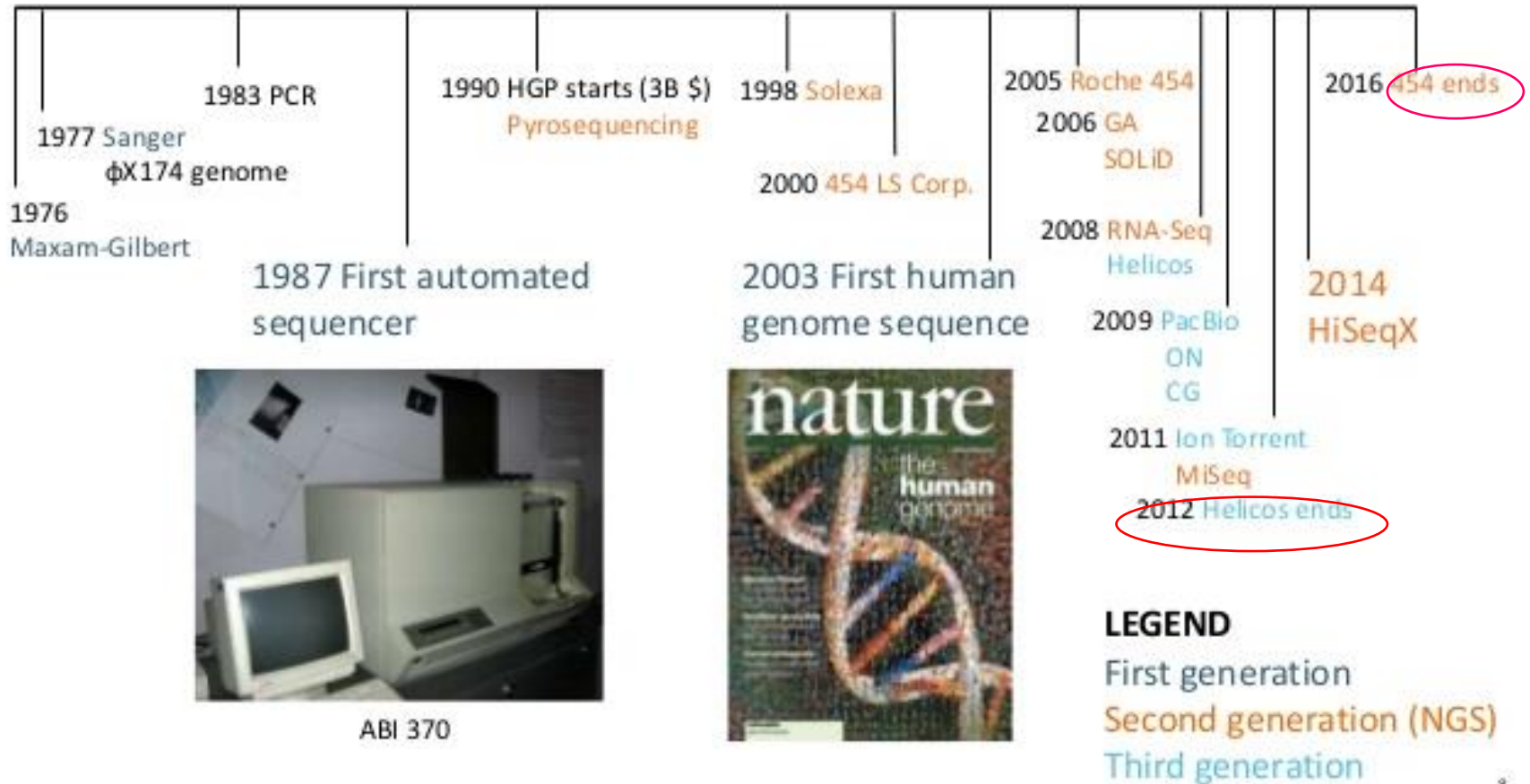


	Illumina (HiSeq 4000)	PacBio (Sequel)	Oxford Nanopore (MinION)
Read length	Up to 150 bp	10-15kb	Up to 900kb
Number of reads	2.5-5 Million	500 K	Up to 1 M
Processing time	<1-3.5 days	Up to 10 hours	~ 6 hours
Error rate	<1%	10-15%	5-15%
Cost per run	~\$3000	~\$850	\$500-\$900
Instrument price	\$900 K	\$350K	\$1K
Advantages	Highly accurate	Sequence long reads	Sequence long reads Portable device

Four decades, three generations

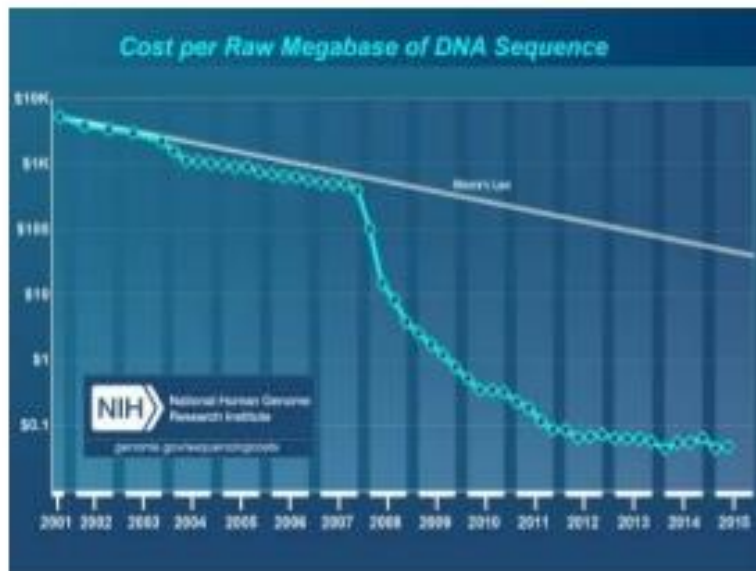
1976

2016



Next-Generation Sequencing for Clinical Bioinformatics

- NGS revolutionised our access to genomic information
- 2nd generation technology allows WGS for less than 1000 GBP
- However, a number of challenges exist
 - Data creation
 - Data analysis/processing
 - Data (clinical) interpretation



Některé základní pojmy a strategie sekvenování

Referenční genom – důležitý, když provádíme **re-sekvenování** genomu nebo jeho vybrané oblasti (např. v molekulární diagnostice) nebo když vyhodnocujeme **TRANSKRIPTOM**, přepisované oblasti, TSS, úrovně transkriptů.

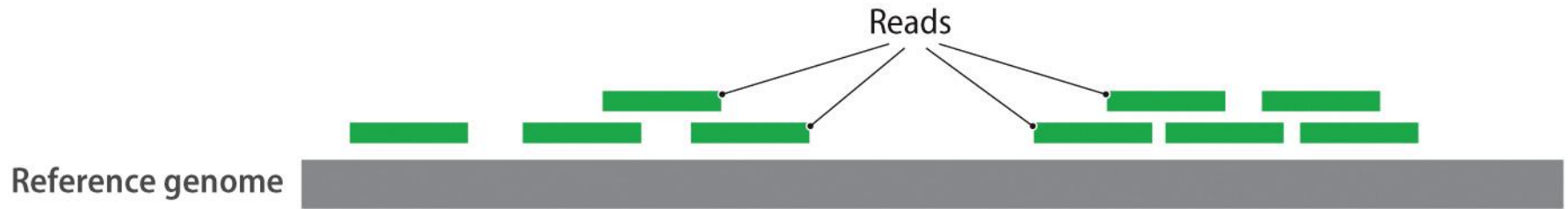


Figure 4.19 Genomes 4 (© Garland Science 2018)

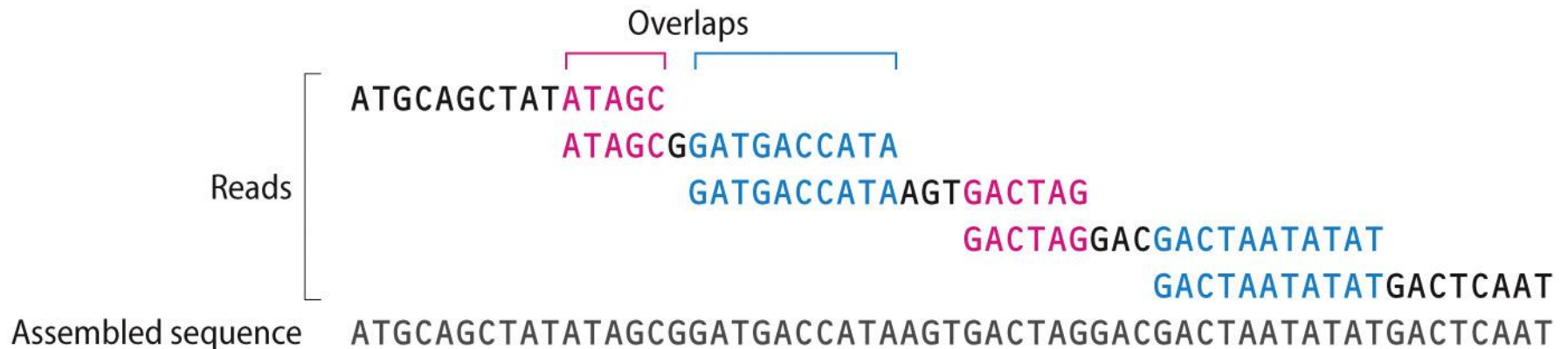
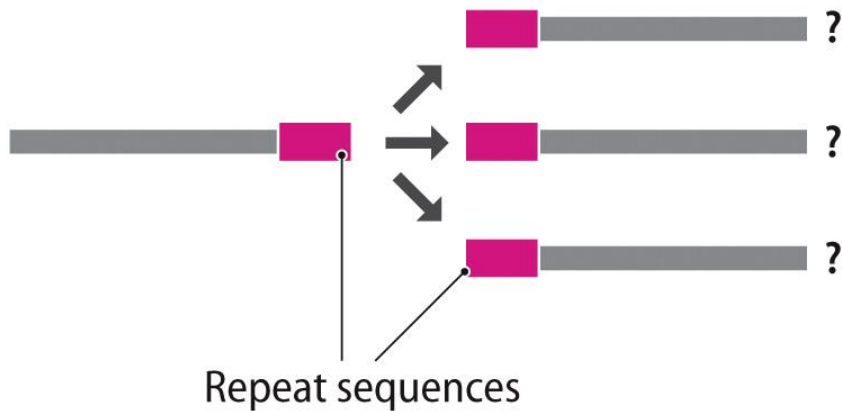


Figure 4.21 Genomes 4 (© Garland Science 2018)

(A) Ambiguous sequence assembly



(B) Paired-end reads resolve the assembly

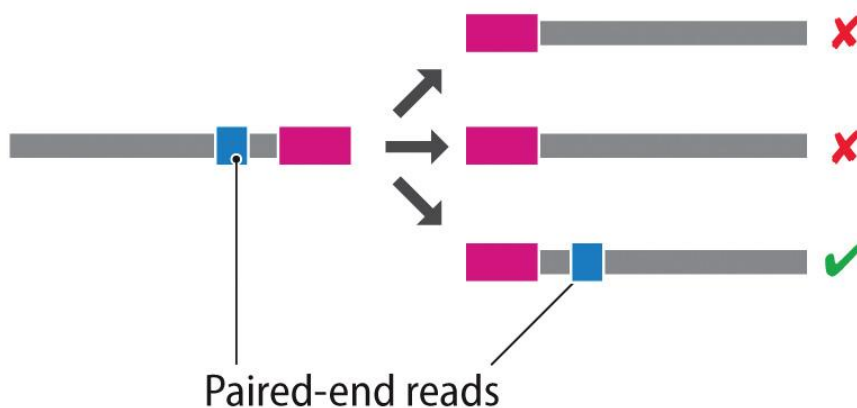


Figure 4.23 Genomes 4 (© Garland Science 2018)

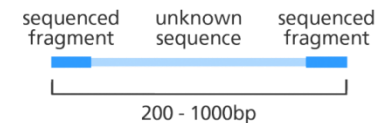
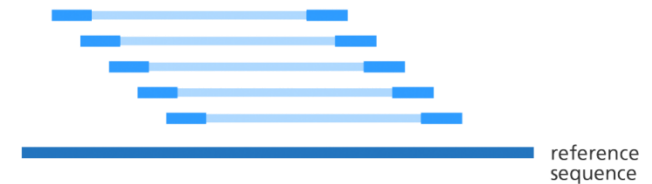
Repetice všeho druhu jsou problematické pro sestavování genomu – nejednoznačné možnosti sestavení.

Sekvenování s párovými konci čtení je nejběžnějším řešením.

Single-end reads



Paired-end reads



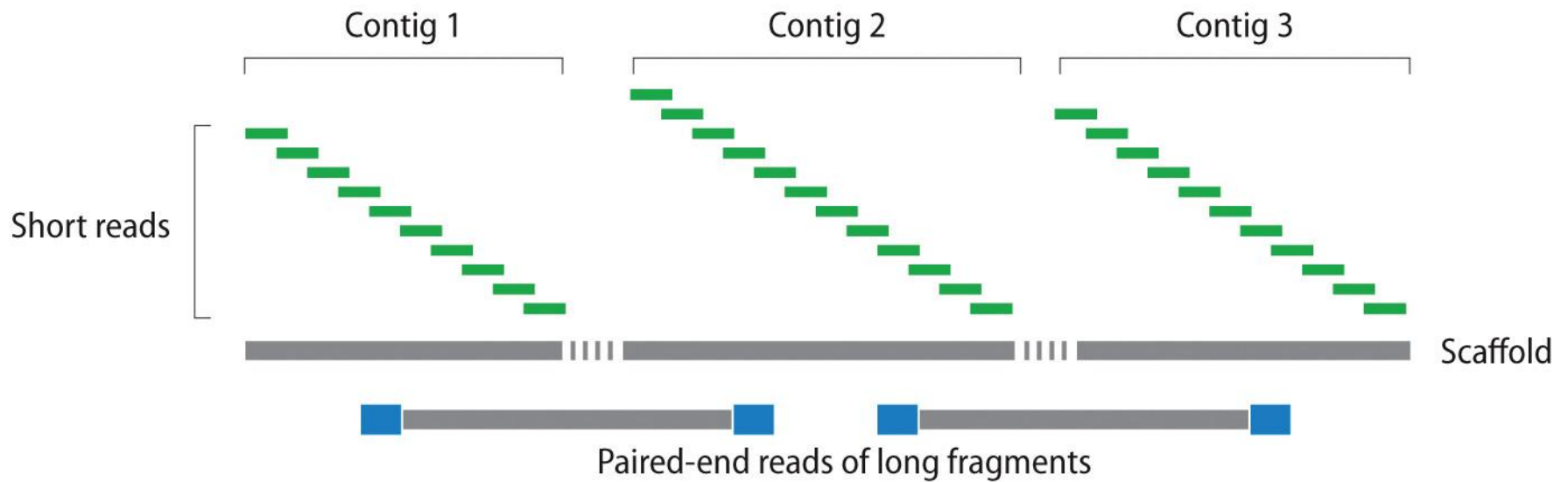
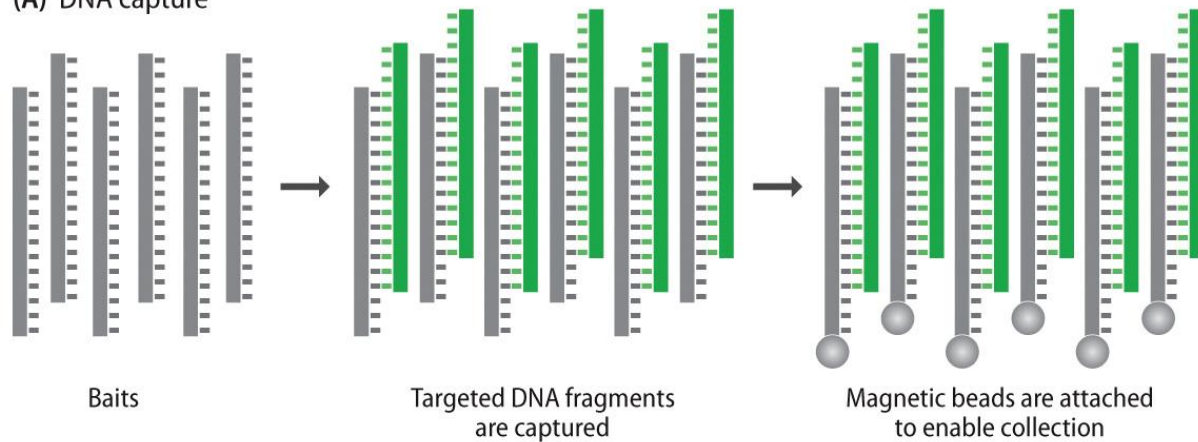


Figure 4.24 Genomes 4 (© Garland Science 2018)

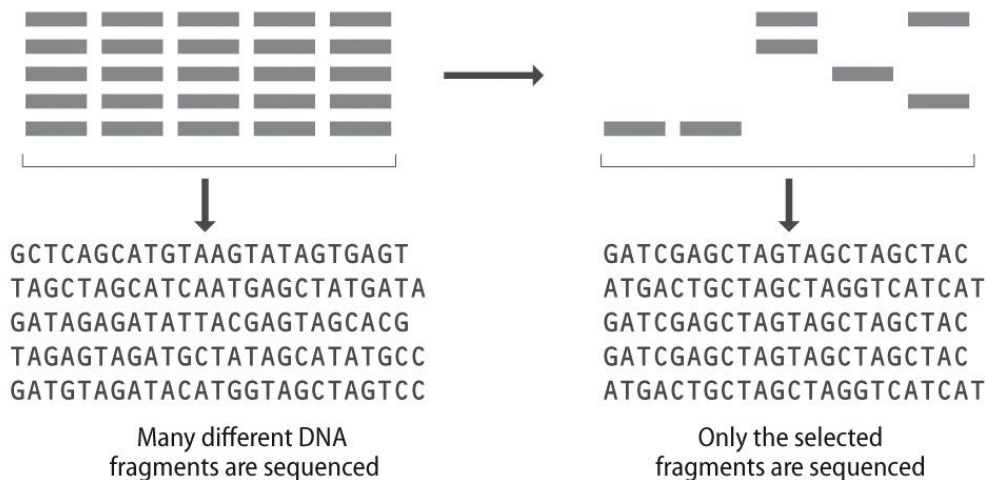
PE čtení ale nemohou vyřešit problém dlouhých oblastí tandemových repetit. Současným řešením je optické mapování, nanopore sekvenování...

Sequence capture – vychytání žádoucí frakce genomu: možnost zaměřit sekvenační kapacitu na žádoucí oblasti genomu (např. geny spojené s určitým onemocněním, exony – exome sequencing) nebo omezit plýtvání kapacitou na čtení vysoce repetitivní oblasti. Při sekvenování celého transkriptomu se naopak provádí **deplete rRNA** (která jinak tvoří přes 90% veškeré RNA v buňce).

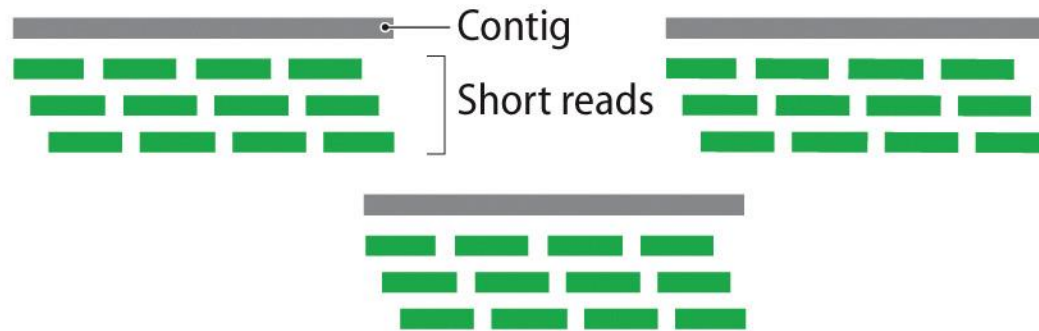
(A) DNA capture



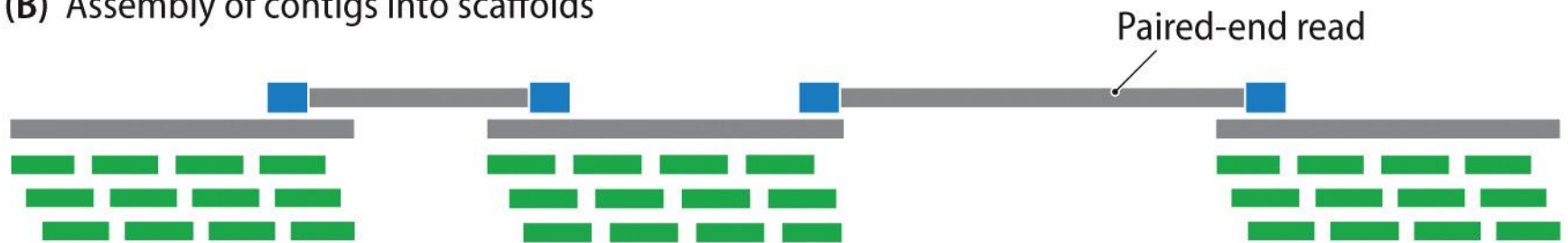
(B) Comparison between sequencing before and after capture



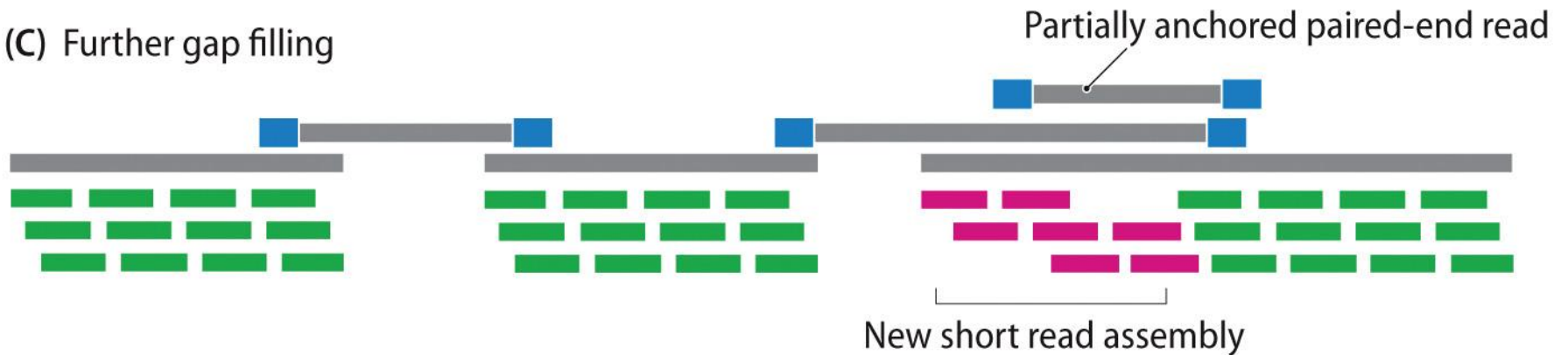
(A) Assembly of short reads into contigs



(B) Assembly of contigs into scaffolds



(C) Further gap filling



Aplikace

De novo sekvenování: v současnosti pomocí kombinace Nanopore, PacBio, optického mapování a Illumina.

Resekvenování: Solexa/**Illumina**, u jednotlivých sekvencí Sanger

Sekvenování transkriptomu – nahradilo dříve používané cDNA/RNA microarrays, většinou **Illumina**

- representace čtení transkriptů (normalizovaná k referenčnímu transkriptu) odráží hladinu transkriptu

- Přesnější než microarrays, možnost detekce neznámých transkriptů – výhoda oproti RT-qPCR. Cena za sekvenování transkriptomu odpovídá ceně analýz několika desítek různých transkriptů pomocí RT-qPCR

Celogenomová sekvenování, 1000 genome project (variace lidského genomu), obdobné projekty u jiných modelových organismů

Metagenomika – analýza kompletního genetického materiálu ze vzorků prostředí **bez kultivace** – ukazuje skutečnou rozmanitost (např. složení mikroorganismů střevního traktu a jiných orgánů – jen asi 1 z 10 buněk v lidském těle je skutečně lidská buňka: většina buněk našich těl jsou mikroby!)

Epigenetika / epigenomika – Analýzy DNA methylation, ChIP-Seq, Hi-C – viz další přednáška.

Děkuji Vám za pozornost...
.....a těším se na Vaše dotazy