

**Česká zemědělská univerzita v Praze**

**Fakulta agrobiologie, potravinových a přírodních  
zdrojů**

**Katedra genetiky a šlechtění**



**Metody sekvenování nové generace (NGS)**

**Bakalářská práce**

**Gabriela Guthová**

**Speciální chovy**

**Ing. Jakub Vašek, Ph.D.**

© 2018-2019 ČZU v Praze

## **Čestné prohlášení**

Prohlašuji, že svou bakalářskou práci "Metody sekvenování nové generace (NGS)" jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autorka uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 18.4.2019

---

## **Poděkování**

Ráda bych touto cestou poděkovala vedoucímu bakalářské práce panu Ing. Jakubu Vaškovi Ph.D., za cenné rady a připomínky při vypracování této práce.

# Metody sekvenování nové generace (NGS)

## Souhrn

Cílem této bakalářské práce je shrnout základní principy fungování genetických sekvenačních metod a možnosti jejich využití ve výzkumu. Dále jsou v práci zohledněny rozdílné parametry jednotlivých metod, jako například cena sekvenování, způsoby využití a chybovost čtení.

V úvodní kapitole jsou popisovány tradiční metody sekvenování, které stály na počátku rozvoje tohoto odvětví. V rámci této kapitoly je také popsána metoda PCR (Polymerázová řetězová reakce), jejíž znalost je stěžejní pro pochopení principů sekvenačních metod.

Následující kapitoly jsou věnovány metodám NGS (Sekvenování nové generace). U každé metody jsou popsány způsoby přípravy vzorků a následně samotného sekvenování. V práci jsou také zohledněny možnosti využití jednotlivých technologií.

Jedna kapitola je věnována studiím v oblasti zoologie, pro které byly využity sekvenační metody. V rámci této kapitoly je demonstrována důležitost těchto metod pro další obory. Jedná se například o medicínu, evoluční biologii nebo šlechtění užitkových zvířat.

Poslední kapitola je věnována srovnání jednotlivých generací sekvenačních metod.

**Klíčová slova:** Sangerova metoda, Illumina, 454 sekvenace, Ion Torrent, nanopórové sekvenování

# Next generation sequencing methods (NGS)

## Summary

The main goal of this thesis is to summarize basic principles of genetic sequencing methods and their applications in research. Furthermore, different parameters of individual methods are taken into account in this thesis, such as the costs of sequencing, area of application and reading error rate.

The introduction applies to traditional methods of sequencing, which stood at the foundation of this industry. Principle of PCR (Polymerase Chain Reaction), which is crucial for understanding of genetic sequencing, is also described in this chapter.

Further chapters are dedicated to the NGS (New Generation Sequencing) methods. For each of them, methods of samples preparation and sequencing are described. Possibilities of utilization of different technologies are also described in this part.

One chapter is dedicated to applications of sequencing methods in zoological researches. The importance of genetic sequencing for medicine, evolutionary biology and breeding programmes and other biological fields is mentioned.

The last chapter is devoted to a comparison of sequencing methods.

**Keywords:** Sanger sequencing, Illumina, 454 sequencing, Ion Torrent, nanopore sequencing

# Obsah

<b>1 Úvod.....</b>	<b>1</b>
<b>2 Cíl práce.....</b>	<b>2</b>
<b>3 Literární rešerše.....</b>	<b>3</b>
<b>3.1 Sekvenování první generace.....</b>	<b>3</b>
3.1.1 Sangerova metoda sekvenování.....	3
3.1.2 Postup sekvenování.....	4
3.1.2.1 Příprava knihovny.....	4
3.1.2.2 Sekvenování.....	5
<b>3.2 Sekvenování druhé generace.....</b>	<b>7</b>
3.2.1 Pyrosekvenace.....	8
3.2.1.1 Postup sekvenování.....	8
3.2.1.2 Chybovost čtení.....	10
3.2.2 Illumina Solexa.....	10
3.2.2.1 Postup sekvenování.....	11
3.2.3 SOLiD.....	14
3.2.3.1 Postup sekvenování.....	15
3.2.4 Ion torrent.....	17
3.2.4.1 Postup sekvenování.....	18
<b>3.3 Sekvenování třetí generace.....</b>	<b>19</b>
3.3.1 Pacific Bioscience.....	20
3.3.1.1 Postup sekvenování.....	21
3.3.2 Oxfordský Nanopór.....	23
3.3.2.1 Postup sekvenování.....	23
3.3.2.2 Sekvenování s využitím exonukleázy.....	25
<b>3.4 Sekvenace čtvrté generace.....</b>	<b>26</b>
<b>3.5 Příklady použití nových sekvenačních metod.....</b>	<b>26</b>
3.5.1 Studie kolapsu kolonií u včel v USA.....	26
3.5.2 Výzkum přenosné rakoviny d'áblů.....	27
3.5.2.1 Genetická analýza přenosné rakoviny u d'áblů.....	28
3.5.3 Výzkum rezistence vůči viru SIV u primátů.....	29
3.5.3.1 Genetická analýza hostitelů viru SIV.....	29
3.5.4 Fylogenetická studie čeledi Hominidae.....	30
3.5.4.1 NGS analýza podčeledi Ponginae.....	31

3.5.5	Genetické studie hospodářských zvířat.....	32
3.5.5.1	Genetická analýza slintavky a kulhavky .....	32
<b>3.6</b>	<b>Srovnání generací sekvenování.....</b>	<b>33</b>
<b>4</b>	<b>Závěr.....</b>	<b>37</b>
<b>5</b>	<b>Literatura.....</b>	<b>38</b>
5.1	Knižní zdroje .....	38
5.2	Internetové zdroje .....	41

# 1 Úvod

V roce 1953 byla vědci popsána struktura DNA. Za tento objev získali v roce 1962 Nobelovu cenu Maurice Wilkis, James Watson a Francis Crick (Gann & Witkowski 2013). Již v 70. letech 20. století byly následně představeny první metody čtení genetické sekvence. Po dalších třicet let se vývoj genetického sekvenování zaměřoval na modernizaci jedné z nich, a to Sangerovy metody (McGinn & Gut 2012). V posledních dvaceti letech bylo naproti tomu, díky rozvoji technologií, představeno velké množství sekvenačních metod s různými způsoby mapování genomu. Také zásluhou stálého zlepšování sekvenačních technik patří dnes molekulární genetika k nejrychleji se rozvíjejícím oborům.

Na základě společných principů jsou metody sekvenování nové generace (z anglického **New Generation Sequencing - NGS**) řazeny do několika generací. Sangerova metoda představuje takzvanou první generaci sekvenování. Dnes jsou již testovány metody čtvrté generace (Feng et al. 2015). Kvůli vysoké finanční náročnosti a poměrně vysoké chybovosti nejnovějších metod však většina laboratoří využívá přístroje, které zakoupila již dříve. V praxi se tedy stále můžeme setkat s využitím většiny metod, které jsou předmětem této bakalářské práce.

Využití mapování genomu pomocí NGS pomohlo k rozvoji řady biologických vědních oborů. Nejvíce výzkumů je samozřejmě vedeno ve zdravotnictví. V zoologii se využívá například v oblastech šlechtění, veterinární medicíny nebo evoluční biologie.



## **2 Cíl práce**

Cílem této bakalářské práce zpracování přehledu nových technologií sekvenování a také metod, které dnes již nemají praktické využití, nicméně byly z historického hlediska velmi důležité pro rozvoj v oblasti genomového sekvenování. V bakalářské práci se autor zabývá představením jednotlivých metod a následně popsáním principu jejich fungování. Část práce je věnována praktickému využití těchto metod ve vědních oborech. Na popisovaných výzkumech v oblasti zoologie jsou zmíněny možnosti využití sekvenačních metod v oblasti genetiky, evoluční biologie či medicíny.

## **3 Literární rešerše**

### **3.1 Sekvenování první generace**

V 70. letech minulého století došlo k revoluci na poli molekulární biologie objevem metod sekvenování DNA. Dva týmy vědců v tu dobu zveřejnily své objevy, a představily odlišné postupy pro sekvenaci genomu. Dle příjmení jejich vynálezců dnes jednu z metod nazýváme Maxam-Gilbertova a druhou metodu Sangerova. (Ahmadian et al. 2006). Sangerova metoda zůstala po dalších 30 let jedinou možností pro sekvenování (McGinn & Gut 2012).

#### **3.1.1 Sangerova metoda sekvenování**

Sangerova metoda byla popsána roku 1977. Převládla oproti Maxam-Gilbertově metodě, a to díky menší technologické náročnosti a potenciálu pro zvětšování celkové délky sekvenovaných úseků a objemu sekvenčních dat (Ahmadian et al. 2006). Díky těmto vlastnostem se počítalo s tím, že bude do budoucna možné plně automatizovat proces sekvenování.

Touto technikou lze osekvenovat vlákno DNA dlouhé 800 bp až 1000 bp (párů bází) (Schadt et al. 2010). Po třiceti letech zdokonalování této techniky je přesnost této metody až 99,99 % (Shendure & Ji 2008).

V pozdějších letech byla Sangerova metoda využita k sekvenaci celého lidského genomu. Kvůli možnosti zpracování relativně malého množství sekvenčních dat za jednotku času, stál tento projekt tři miliardy dolarů a trval deset let (Schadt et al. 2010).

Dnes je Sangerova metoda využívána hlavně při projektech, které nevyžadují osekvenování velkého množství genetické informace. Konkrétně se aplikuje při kontrolním čtení, aby se ověřily výsledky dosažené jinými sekvenačními technikami, nebo například na školách, jako praktická ukázka z tohoto oboru, kterou je možné si finančně dovolit (Ahmadian et al. 2006).

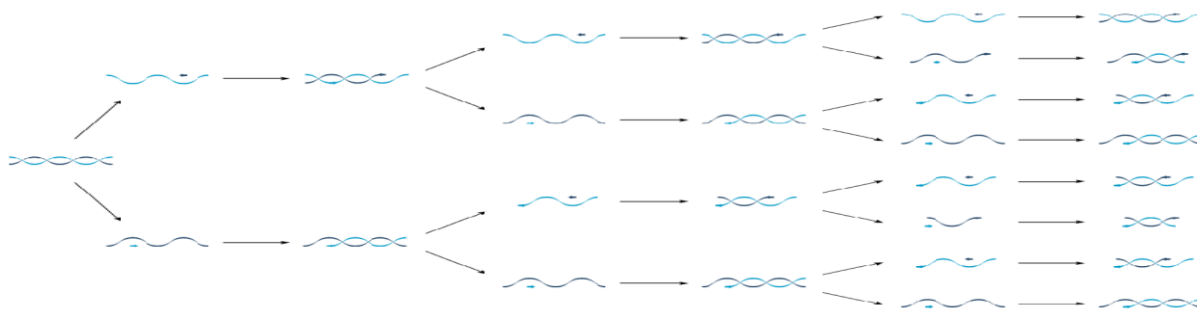
### 3.1.2 Postup sekvenování

#### 3.1.2.1 Příprava knihovny

Před samotným sekvenováním bylo nutné namnožit studovanou DNA. O tomto procesu mluvíme jako o přípravě knihovny. Dostatečný počet kopií je důležitý pro zesílení signálu a schopnost jeho detekce v průběhu sekvenace (Sanger et al. 1977).

K namnožení fragmentů molekuly DNA se využívaly například plazmidy bakterií, bakteriální chromozomy nebo kvasinkové chromozomy, do kterých byla pomocí enzymu DNA ligázy zanesena DNA zkoumaného organismu (Shendure & Ji 2008). Vložením úseku DNA do bakterie, která se přirozeně replikuje, se dosáhlo výsledného namnožení požadovaných úseků DNA. Jedná se o takzvané molekulární klonování. Požadovaná DNA byla následně z bakterie či kvasinky izolována a sekvenována. Tento způsob namnožení genetické informace byl dříve jedinou možností pro vytvoření genetické knihovny (Lodish et al. 2000). V pozdějších letech se takto postupovalo při sekvenování *de novo*, tedy při sekvenování, kde dosud nebyl k dispozici referenční genom.

Díky technologickému vývoji se v roce 1983 začalo využívat k namnožení fragmentů genetické informace metody polymerázové řetězové reakce (z anglického Polymerase Chain Reaction - PCR). Princip této reakce je založen na tepelné denaturaci dvouřetězové DNA a opětovné syntéze druhého řetězce díky enzymu polymeráze (Garibyan & Avashia 2013). Tímto způsobem je, v opakujících se cyklech, dosaženo exponenciálního zmnožení vzorku, jak můžeme vidět na Obrázku 1.



Obrázek 1: Princip PCR

Zdroj: Brown & Brown 2011

Celá reakce probíhá v termocykleru. Cyklické změny teploty v přístroji zajišťují střídání tří fází. První fází je denaturace dvouřetězové DNA za teploty 95 °C, při které se rozrušují

vodíkové můstky. Následně je teplota snížena na 50-60 °C, kdy k jednořetězové molekule dosedají primery. Opětovné zvýšení teploty na 65-75°C je potřeba pro práci DNA polymerázy, která při této teplotě zahajuje syntézu nového řetězce (Garibyan & Avashia 2013). Díky PCR bylo například možné odhalit odchylky v genetické informaci studovaného jedince oproti modelovému organismu (Shendure & Ji 2008).

### **3.1.2.2 Sekvenování**

Sangerova metoda sekvenování je původně založena na odečítání genetické sekvence z polyakrylamidového gelu (Karger & Guttman 2009).

Namnožené kopie původního řetězce DNA jsou pipetovány do jedné ze čtyř zkumavek. V nich je přítomná reakční směs s DNA polymerázou, nukleotidy, a také čtyřmi terminačními dideoxynukleotidy. Do každé zkumavky je přidán jiný dideoxynukleotid. Tedy ddATP, ddTTP, ddGTP a ddCTP. Dideoxynukleotidy jsou analogy klasických deoxynukleotidtrifosfátů zabudovávaných do řetězce při jeho polymeraci s tím rozdílem, že na 3' konci řetězce je místo hydroxylové skupiny pouze vodík, a tím pádem v tomto místě dochází k terminaci polymerázové reakce, jelikož se nemůže připojit další nukleotid (Sanger et al. 1977).

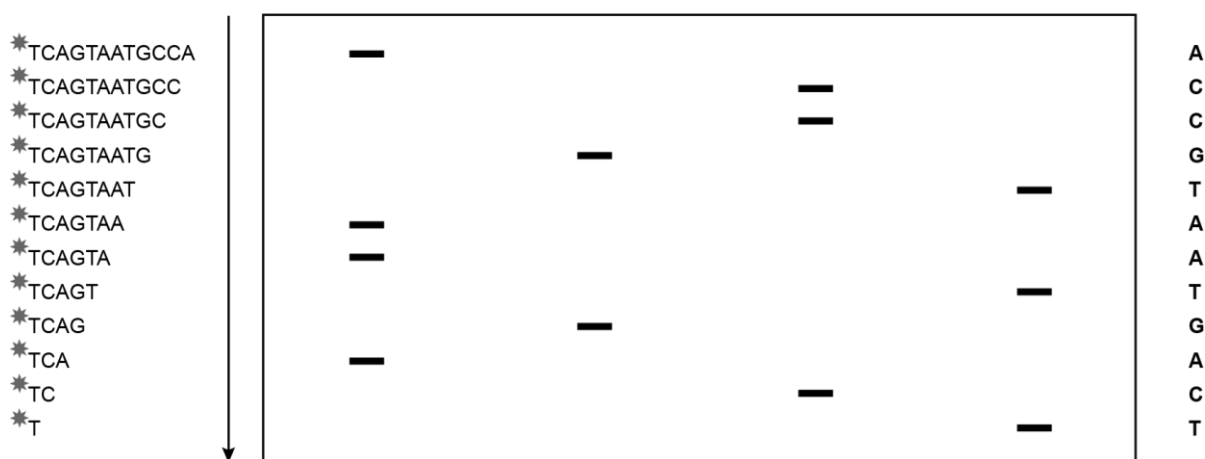
Při vzniku komplementárních řetězců se do nově vznikajícího vlákna mohl zabudovat dNTP, tedy některý z deoxynukleotidtrifosfátů nebo ddNTP, tedy terminační dideoxynukleotidtrifosfát, ukončující syntézu vlákna (Ahmadian et al. 2006). Syntéza DNA začínala vždy od jednoho místa v řetězci, ovšem terminace syntézy byla závislá na navázání konečného ddNTP, který byl ve zkumavce přítomen. Proto se jednotlivé fragmenty DNA získané z jedné zkumavky lišily svojí délkou (Schadt et al. 2010).

Kdybychom například měli zkumavku, kde bude přidána molekula ddATP, tedy dideoxyadenosintrifosfátu, mohla by se tato částice navázat v řetězci nukleotidů na jakékoliv místo, kde by na komplementárním vlákně byla přítomna báze thyminu. Na prvním syntetizovaném vlákně se ddATP mohla navázat hned na první takovou pozici. Na dalším vlákně se však na dané místo naváže dATP a syntéza řetězce tedy pokračuje dál do doby, kdy nebude navázán výše zmíněný ddATP.

Produkty reakcí ze čtyř zkumavek, obsahující různě dlouhé fragmenty DNA, jsou následně naneseny na denaturační polyakrylamidový gel a elektroforeticky separovány. Záporný náboj molekul DNA v zásaditém prostředí gelu zapříčiní pohyb fragmentů jedním směrem od katody k anodě (Ahmadian et al. 2006).

Na Obrázku 2 je vidět výsledek Sangerovy metody sekvenování a způsob čtení sekvence DNA. Díky odlišné mobilitě, která je ovlivněna velikostí fragmentů, doputuje každý fragment do jiné vzdálenosti (Schadt et al. 2010). Kratší fragmenty se pohybují gelem rychleji a doputují do delší vzdálenosti. Vidíme, že do největší vzdálenosti doputoval nejkratší fragment, který je zařazen ve sloupci, kam byl na gel nanesen vzorek ze zkumavky s ddTTP.

Posledním krokem je přečtení nukleotidové sekvence z gelu (Bräutigam & Gowik 2010). Po odečtení tedy víme, že v případě zobrazeném na Obrázku 2 bude první nukleotid sekvence thymin. Dalším nukleotidem v řetězci bude vzorek, který doputoval do druhé největší vzdálenosti. V tomto případě tedy cytosin. Tímto způsobem se postupuje při čtení celého vzorku.



Obrázek 2: Čtení sekvence DNA z polyakrylamidového gelu

Zdroj: Brown & Brown 2011

Dnes se již pro detekci výsledků Sangerovy metody využívá kapilární elektroforéza. Tato technika vede hlavně díky menším nárokům na práci laboranta při přípravě a vyhodnocování vzorků. Všechny typy ddNTP jsou vloženy do jedné zkumavky a fluorescenčně označeny, každý dideoxynukleotid jinou barvou (McGinn & Gut 2012). Podle své velikosti jsou poté fragmenty DNA přes kapiláru dopraveny k detektoru zařízení, který je vyhodnotí jako barevné signály. Díky technickému vylepšení je dnes CCD kamera schopna snímat signál i z většího počtu kapilár najednou. Tímto způsobem lze osekvenovat vzorky o délce 1000 bp (Karger & Guttman 2009).

Principiálně se od sebe oba způsoby sekvenace příliš neliší. Dříve byly produkty detekovány díky radioaktivnímu značení. Dnes je však Sangerova metoda prováděna hlavně pomocí kapilární elektroforézy, kdy jsou vzorky značeny fluorescenčně a jejich detekce a následná sumarizace dat probíhá díky počítačovým technologiím (McGinn & Gut 2012).

## 3.2 Sekvenování druhé generace

S výrazným rozvojem v oblastech robotiky, informatiky, chemie, biotechnologií, optiky a také počítačových databází, postupně klesala cena za jednu sekvenační reakci (Mardis 2008 b). Nové poznatky umožnily rozvoj dalších technik, které jsou schopné v porovnání s dideoxyterminační metodou vytvořit velké množství sekvenačních dat za jednotku času. Sangerova metoda tedy pomalu ustupovala do pozadí (Schadt et al. 2010).

Ke snížení ceny sekvenace přispěl v roce 2004 americký Národní institut zdraví (National Institut of Health). Vypsál soutěž s vysokou peněžní odměnou o to, kdo jako první osekvenuje haploidní genom člověka za 1000 USD. V tu dobu by osekvenování takového množství dat stálo 100 000 USD (Spencer 2004). Tato soutěž rapidně přispěla k vylepšování technologií sekvenace. Cena za reakci však stále zůstala příliš vysoká na to, aby se tyto technologie druhé generace mohly používat rutinně (Clarke et al. 2009).

Techniky druhé generace sekvenování mají společný základní princip postupu přípravy vzorků a jejich následného sekvenování. V první řadě je nutné fragmentovat templátovou DNA na úseky, které je daná metoda schopna osekvenovat. Dalším krokem je příprava knihovny ze získané DNA (Shendure & Ji 2008).

Technologie využívané při detekci potřebují větší intenzitu světelných signálů pro jejich úspěšné zachycení detektorem. Proto musí dojít k amplifikaci templátových vzorků, podobně jako v případě první generace sekvenování. Po dokončení přípravy knihovny jsou všechna získaná vlákna paralelně osekvenována. Díky kapacitám přístrojů tak mohou být osekvenovány i miliony sekvencí najednou (Shendure & Ji 2008).

Se zlepšováním rychlosti a výkonnosti celogenomických sekvenačních technologií se také rychle posouvá výzkum v oblasti genetické diagnostiky nebo evoluční biologie. Z oblasti zemědělství můžeme zmínit například využití sekvenace při mapování genomu plodin a následném vytváření genetických modifikací, které kupříkladu zajišťují vyšší rezistenci vůči škůdcům (Esposito et al. 2016).

Druhá generace otevřela také možnost pro rozvoj komparativní analýzy. Ta se využívá například při porovnání genové exprese u rostlin stejného druhu v různých podmínkách. Při klinických studiích díky ní můžeme porovnávat genovou expresi nemocných jedinců v porovnání s jedinci zdravými. Porovnáním exonů zdravých a nemocných zjišťujeme, jaké mutace mohly podmínit vznik nemoci (Pop & Salzberg 2008). Díky možnosti sekvenace transkriptomů, na základě přepisu RNA do copy DNA (cDNA), je možné sledovat, jaké

bílkoviny buňka vyrábí ve chvíli odběru vzorku (Bräutigam & Gowik 2010). Tím odpadá nutnost sekvenování celého genomu pro klinickou diagnostiku.

### **3.2.1 Pyrosekvenace**

Ačkoliv byla tato metoda zveřejněna v devadesátých letech minulého století, nebyla často využívána kvůli své malé výkonnosti z hlediska sekvenovaného počtu bází. V rámci technologického zdokonalování této metody začala být pyrosekvenace užívána v roce 2004, jako první metoda nové generace (Ahmadian et al. 2006). Dnes se již tyto sekvenátory od firmy Roche neprodávají. Příčinami jsou technologická náročnost procesu, vysoká cena enzymů potřebných k reakci i možnost zakoupení technologicky vyspělejších přístrojů.

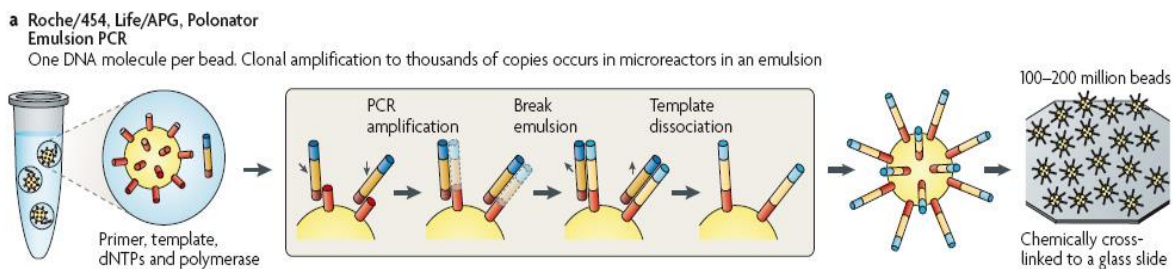
Touto technikou lze osekvenovat vlákno DNA dlouhé 300 bp až 800 bp. Díky možnosti osekvenovat najednou až 100 000 molekul DNA dosahuje objem dat za čtení 400-600 Megabází (Mb) (Liu et al. 2012). Za jeden den tedy může být přečtena až 1/8-1/5 lidského genomu.

Pyrosekvenace byla využívána při mutační analýze k detekci polymorfismů jediného nukleotidu (z anglického Single Nucleotide Polymorphisms - SNPs), nebo například k diagnostice chorob a stanovení klinických prognóz pacienta pomocí genotypizace bakterií (Rothberg & Leamon 2008). Také byla první metodou, mimo Sangerovy, která byla využita při sekvenaci lidského genomu (Metzker 2010).

Zásadním nedostatkem této metody byla zvyšující se chybovost detekce při čtení homopolymerních sekvencí (Margulies et al. 2005).

#### **3.2.1.1 Postup sekvenování**

Prvním krokem je fragmentace templátové DNA nebulizací na délku 300-800 bp (Liu et al. 2012). Následná příprava knihovny probíhá v laboratorních zkumavkách za přítomnosti potřebných reaktantů. Jednotlivá vlákna s tupými konci jsou z každé strany napojena na adaptéry, u nichž známe jejich sekvenci. Takto upravené fragmenty jsou imobilizovány na kuličky se zabudovanými známými sekvencemi oligonukleotidů, které jsou komplementární k sekvencím adaptérů. Kuličky jsou enkapsulovány do emulze vody a oleje. V této chvíli probíhá emulzní amplifikace fragmentů, přichycených na povrch kuličky (Margulies et al. 2005).



Obrázek 3: Emulzní PCR

Zdroj: Metzker 2010

Na Obrázku 3 je zobrazena zkumavka s reagensii potřebnými k polymerázové řetězové reakci. Fragment s adaptérem se naváže k sekvenci oligonukleotidu pevně přichyceného na kuličce. Dochází k syntéze paralelního řetězce DNA za pomoci sekvence primeru a enzymu DNA polymerázy. Původní řetězec je poté odmyt a na kuličce zůstávají kovalentně navázané komplementární řetězce DNA.

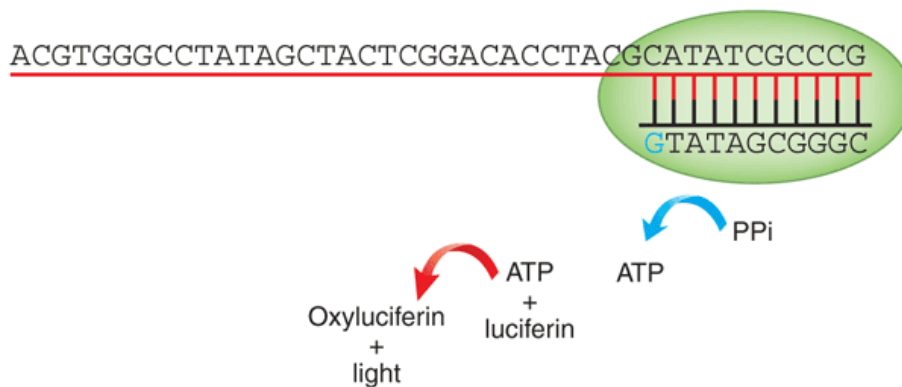
Po ukončení emulzní PCR jsou kuličky naneseny na pikotitrační destičku, na které probíhá pyrosekvenace. Destička je vybavena jamkami. Do každé z těchto jamek je umístěna jedna kulička a následně i enzymy, které se využívají pro přípravu sekvenace. Konkrétně se jedná o tekutinu s obsahem enzymů potřebných k sekvenaci, tedy o DNA polymerázu a deoxyribonukleotidtrifosáty (Shendure & Ji 2008).

Prvním krokem sekvenace je nasazení primeru na 3' konec původního vlákna. Na sekvenci primeru následně dosedá DNA polymeráza, která připojí další nukleotid k řetězci v případě, že na dané místo patří (Metzker 2010).

Postupnou aplikací jednotlivých roztoků, které vždy obsahují právě jeden typ nukleotidu, je dosaženo syntézy nového řetězce. Z Obrázku 4 je patrné, že k detekci signálu je využíváno bioluminiscence, která vznikne díky inkorporaci jednoho ze čtyř nukleotidů do řetězce (Ahmadian et al. 2006).

Vyzáření světla předchází kaskáda chemických reakcí enzymů. Ve chvíli, kdy je jeden z nukleotidů pomocí DNA polymerázy zařazen do nově vznikajícího vlákna, dojde k uvolnění pyrofosfátu. Ten reaguje s adenosinfosfosulfátem, který je přítomen v roztoku a za pomoci enzymu sulfurylázy vytváří ATP. Vzniklé ATP následně využije enzym luciferáza, který mění luciferin na oxyluciferin (Metzker 2010). Oxyluciferin poté vysílá světelné záření, které je zachyceno CCD kamerou (Shendure & Ji 2008). Jedna kamera detekuje v reálném čase záblesky emitované ve všech místech pikotitrační destičky.





Obrázek 4: Princip detekce zakomponovaného nukleotidu

Zdroj: Rothberg & Leamon 2008

Nukleotidy jsou na destičku aplikovány ve formě nukleotidtrifosfátů. Jelikož jsou ATP, GTP, CTP a TTP na destičku rozlévány postupně a po detekci jednoho cyklu jsou opět odmyty, víme, který z nukleotidů způsobil světelné záření při zabudování do řetězce v daném cyklu (Margulies et al. 2005).

### 3.2.1.2 Chybovost čtení

Problémem tohoto způsobu detekce je zpracování homopolymerních sekvencí. Ve chvíli, kdy se určitá sekvence vícekrát opakuje, měla by se úměrně počtu stejných opakujících se bází zvedat i světelná intenzita signálu. Tento princip však funguje pouze při opakování omezeného počtu nukleotidů. Obecně platí, že je-li v řetězci více než osm stejných opakujících se nukleotidů, je chybovost čtení velká a výsledky nemohou být důvěryhodné (Margulies et al. 2005).

Často se opakující sekvence nalezneme například v chloroplastech nebo v telomerách chromozomů (Esposito et al. 2016). V případě sekvenování těchto struktur je tedy doporučeno využít jiných technik nebo kontrolního čtení, které nám odhalí chyby v sekvencích.

### 3.2.2 Illumina Solexa

V roce 2007 byla vynalezena metoda, která jako první využívá reverzibilně vázaných terminačních značek na fluorescenčně značených nukleotidech. Vynálezce této metody, firma Solexa, byla později koupěna firmou Illumina (Mardis 2013).

Tato technologie umí přečíst pouze krátké sekvence. V začátcích se jednalo pouze o vzorky délky 35 bází. Dnes se v některých sekvenačních přístrojích již mohou využívat

i vzorky o délce 200-300 bází. Při párovém čtení, které je také jednou z novinek některých sekvenačních metod druhé generace, může být finální objem dat z jednoho vzorku až dvojnásobný.

Díky masovému sekvenování velkého počtu fragmentů najednou však získáváme mnohonásobně větší objem dat za jeden sekvenační běh. Metzker (2010) tvrdí, že právě tento fakt dělá z Illuminy současně nepoužívanější sekvenační metodu.

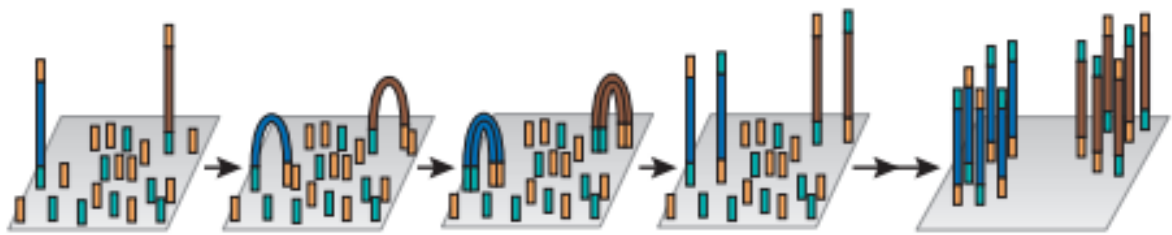
Ačkoliv se chybovost výrazně zvedá s počtem přidaných bází, je tato metoda statisticky přesná na 98 % (Liu et al. 2012).

### **3.2.2.1 Postup sekvenování**

Prvním krokem je úprava DNA do podoby, kterou přístroj dokáže osekvenovat. Až 10 kilobází dlouhé dvouvláknové fragmenty DNA jsou nejprve fragmentovány na segmenty o délce 200 bp. Tak se děje díky označení (tagmentování) pomocí transpozónů a následnému nastříhání vlákna DNA enzymem transponázou (Bräutigam & Gowik 2010).

K takto upraveným segmentům jsou následně připojeny adaptéry, tedy uměle připravené oligonukleotidy se známou genetickou sekvencí. Tato sekvence je komplementární vůči sekvenci oligonukleotidu, který je přichycen na pevném povrchu takzvaných reakčních komůrek („flow cells“). Tyto komůrky jsou umístěny v osmi mikrofluidních kanálcích skleněné destičky, kterou vkládáme k vyhodnocení do sekvenačního přístroje. Po navázání segmentů DNA k daným oligonukleotidům dochází ke klonální amplifikaci vláken, tedy sestavování nového řetězce DNA na základě komplementarity s templátovým řetězcem (Metzker 2010).

Amplifikace segmentů je poté dosaženo díky takzvané můstkové polymerázové řetězové reakci (Bridge PCR), jejíž princip je vysvětlen na Obrázku 5 (McGinn & Gut 2012).



Obrázek 5: Příprava knihovny Illumina (Bridge PCR)

Zdroj: Shendure & Ji 2008

Obrázek 5 již zobrazuje nově nasyntetizované vlákno DNA po odmytí vlákna původního. Nově vzniklé vlákno je pevně uchyceno k povrchu reakční komůrky díky oligonukleotidovému nosiči. Ten v původním kroku sloužil jako primer pro syntézu tohoto vlákna.

V dalším kroku se volný konec vlákna naváže ke komplementárnímu oligonukleotidu, opět pevně přichycenému k destičce. Z něj je poté opět vytvářeno komplementární vlákno k vláknu původnímu.

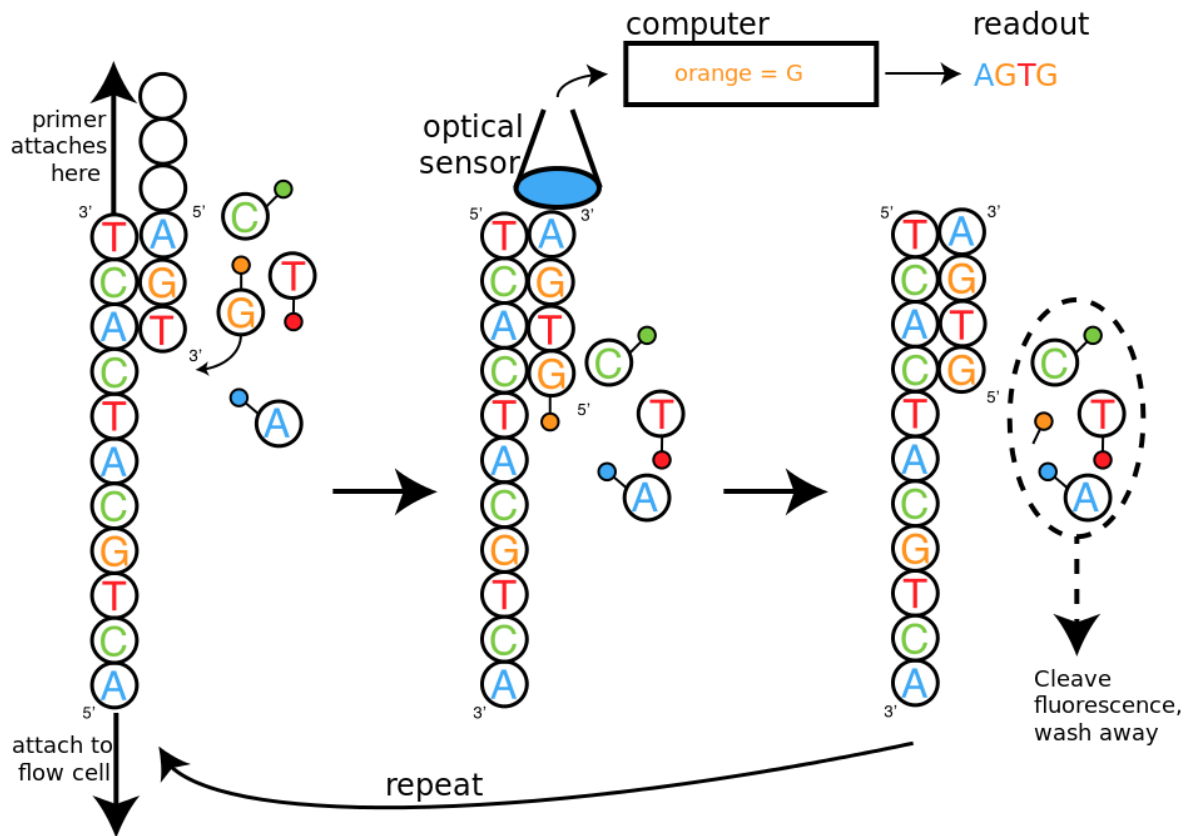
Po vytvoření komplementárního vlákna, které je stejně jako vlákno původní jedním koncem pevně připojeno k destičce, je původní vlákno odmyto. Tento děj se cyklicky opakuje, dokud není dosaženo dostatečného počtu kopií vzorku. Na konci jsou odmyta všechna vlákna, která jsou navázána na jeden ze dvou typů adaptérů. V oblasti na destičce tedy zůstávají pouze vlákna stejného typu.

Tímto způsobem se díky DNA polymeráze tvoří na určitých místech destičky až 1000 amplikonů jediného vlákna DNA. Tato místa nazýváme klastry (Shendure & Ji, 2008). Jednotlivé amplikony při sekvenaci emitují stejný signál. To slouží k efektivnějšímu zachycení signálů z jednotlivých klastrů při prodlužování nově vznikajícího vlákna DNA (Kchouk et al. 2017).

Po vytvoření jednotlivých klastrů následuje samotné sekvenování. Do vzorku jsou přidány primery pro umožnění začátku polymerace a speciálně upravené nukleotidy. Vlákna, která byla vytvořena v průběhu přípravy knihovny, slouží jako předloha pro vytvoření vlákna komplementárního (Shendure & Ji 2008).

Princip syntézy vlákna a detekce nukleotidů je zobrazen na Obrázku 6. V každém cyklu polymerace je dosaženo přidání pouze 1 nukleotidu. Tento krok je umožněn díky modifikaci jednotlivých nukleotidů, kdy na 3' hydroxylový konec je navázán reverzibilně vázaný

terminátor, označený jednou ze čtyř barev (Kchouk et al. 2017). V případě příkladu na Obrázku 6 se jedná o guanin. Po jeho připojení do řetězce je CCD kamerou detekován fluorescenční signál oranžové barvy. Poté je terminační i fluorescenční částice odmyta společně se zbývajících nukleotidy a celý cyklus se může opakovat (Mardis 2008 a).



Obrázek 6: Princip syntézy vlákna a detekce nukleotidů

Zdroj: binf.snipcademy.com

Po nasyntetizování komplementárního vlákna je původní vlákno odmyto. Následně se celý proces opakuje. Jako templát pro syntézu nového řetězce slouží vlákno, které bylo právě nasyntetizováno. Opětovná syntéza obou komplementárních vláken nám může odhalit chyby vznikající zařazením jiného nukleotidu do řetězce (Zhou et al. 2010).

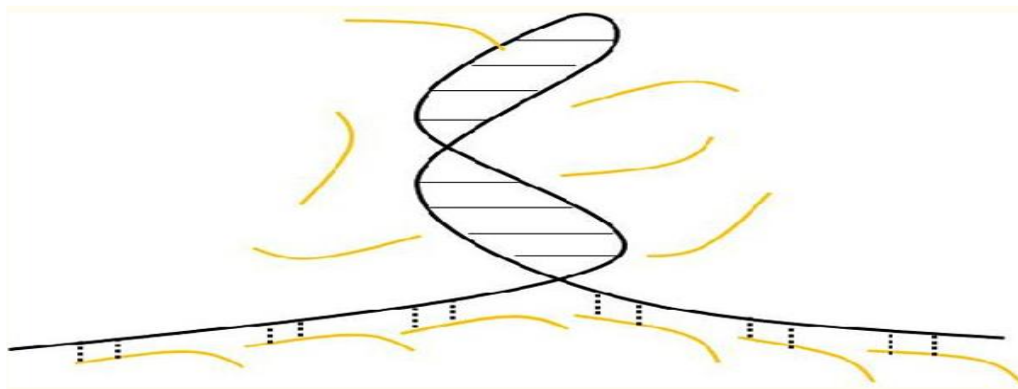
CCD kamera na závěr detekuje záblesky příslušné barvy u jednotlivých klastrů. Laborant je díky získaným datům schopen určit, jaký nukleotid byl do řetězce vložen. Kvůli schopnosti přečíst pouze krátké sekvence, je potřeba využít způsobu, který nám dovolí z těchto krátkých úseků sestavit sekvenci původního templátového DNA. Při sestavování výsledného genomu se využívá takzvaných kontigů, což jsou série bází, které jsou totožné na konci jednoho vzorku a na začátku druhého (Zhou et al. 2010). Ve chvíli, kdy je takto nalezena shoda,

je velice pravděpodobné, že se jedná o pokračování daného vlákna. Dle tohoto klíče je v počítači možné poskládat celé vlákno, které je komplementární k původnímu.

### 3.2.3 SOLiD

Tato metoda druhé generace byla vynalezena firmou Applied Biosystems (dnes již ThermoFisher) v roce 2008. Dnes se již prakticky nevyužívá. Příčinou je poměrně náročná technologie sekvenace a nižší objemy dat za sekvenační běh. Vzorky musí být před samotným sekvenováním upraveny na délku 50-75 bp (Mardis 2008 a). Jelikož je v průběhu sekvenování řetězec čten několikrát, počet sekvenovaných bází za běh je 320 000 000. Objem výchozích dat je téměř třikrát nižší, než při sekvenaci Illuminou (Goodwin et al. 2016).

Výhodou této metody je, díky opakovanému čtení řetězce, vysoká přesnost. Ta dosahuje až 99,9 %. SOLiD systém je rovněž schopen detekovat homopolymerní sekvence s menší chybovostí, než ostatní metody (Goodwin et al. 2016). Má naopak problém detekovat palindromické sekvence, tedy takové, které se na přímém i zpětném vlákně čtou stejně. Nukleotidy vzorku DNA si v tomto případě vytvoří vodíkové můstky mezi sebou a vytvoří tak kličku, jak je zobrazeno na Obrázku 7. Sondy, které jsou na obrázku představeny žlutými čarami, se tedy nemají možnost napojit a v místě palindromu nemůže probíhat sekvenace (Huang et al. 2012).



Obrázek 7: Schéma palindromu na vlákně DNA

Zdroj: Huang et al. 2012

Název této metody je přímo odvozen od techniky, kterou sekvenování funguje. SOLiD (Sequencing by Oligonucleotide Ligation and Detection) nevyužívá ke zjištění DNA sekvence enzymu DNA polymerázy, jako předchozí metody, ale DNA ligázy (Goodwin et al. 2016). Ta připojí již vzniklou sekvenci jednořetězové DNA ke stávajícímu řetězci.

Hlavně kvůli potížím při čtení palindromatických sekvencí není tato metoda vhodná k sekvenaci dosud neznámých genomů. Používá se hlavně k sekvenování transkriptomů a malých molekul DNA (Huang et al. 2012).

### 3.2.3.1 Postup sekvenování

Prvním krokem je úprava řetězce DNA na úseky o délce 50-75 bp. K nastříhání vzorků je možné využít restričních enzymů, již zmiňované nebulizace, nebo ultrazvukových vln, které vytvářejí ve vzorku bubliny a následně DNA rozstříhají díky zvukovým vibracím. V rámci přípravy vzorku se na vlákno DNA opět připojí z každé strany adaptéry o známých sekvencích (Mardis 2008 a).

Dalším krokem je PCR. SOLiD využívá metody emulzní PCR, která je blíže popsána u technologie pyrosekvenace (Shendure & Ji 2008). Po namnožení získaných vzorků vzniká 700–1400 milionů sekvencí (Goodwin et al. 2016). Tyto DNA fragmenty jsou nanášeny na sklíčko, na kterém následně probíhá čtení sekvence pomocí nukleotidových sond.

K samotnému procesu sekvenace potřebujeme primery, enzym ligázu, namnožené vzorky a osminukleotidové sondy, označené jednou ze čtyř fluorescenčních barev. Počátečním krokem je navázání primeru k templátovému vláknu (Zhou et al. 2010). Sekvence primeru je komplementární k sekvenci adaptéru, který byl připojen k 5' konci původního vlákna. Na primer se pomocí DNA ligázy na základě komplementarity vláken připojí první sonda (Mardis 2008 b).



Obrázek 8: Nukleotidová sonda

Zdroj: Thermo Fisher 2006

Sonda, zobrazená na Obrázku 8, je navržena tak, že zjišťujeme kombinaci dvou nukleotidů na čtvrté a páté pozici vlákna (Mardis 2008 a). Složení nukleotidů na prvních a posledních třech místech neznáme. Obrázek 8 dále naznačuje, že poslední tři nukleotidy sondy jsou nositeli barevného označení. Pro označení sond se využívají čtyři barvy, a to zelená, oranžová, modrá a červená. Vzorek je zespu osvětlen laserem. Ve chvíli zaligování

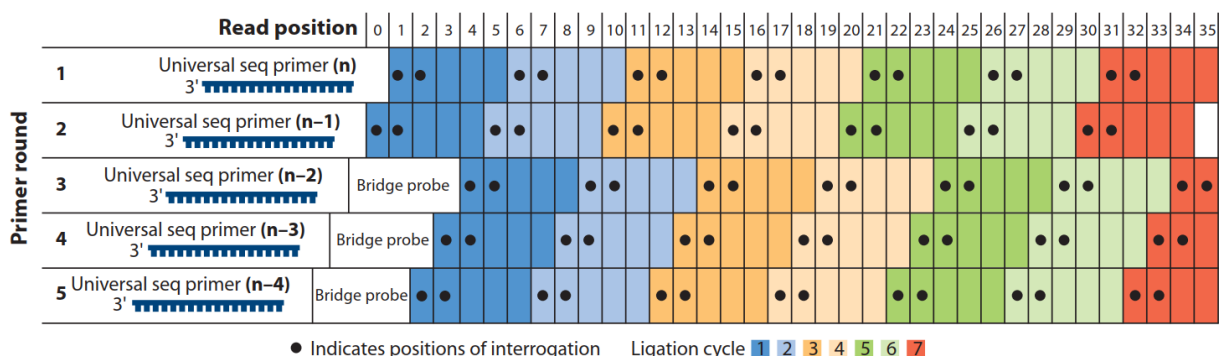
sondy do řetězce je díky vyzáření fotonu fluorescenční barvy CCD kamerou detekován signál. Poté jsou poslední tři nukleotidy odmyty a následuje připojení další sondy (Glenn 2011).

Jak již bylo zmíněno, využívají se vzorky o délce 35 bází. Po odmytí posledních tří bází určujících barvu sondy, zůstává 5 bází napojených k původnímu vláknu DNA. Sedminásobným opakováním procesu navázání sondy tedy zjistíme postupně celou sekvenci vzorku.

Obrázek 9 demonstruje způsob, kterým je dosaženo kompletního přečtení sekvence vzorku. Z každé sondy jsme při prvním čtení identifikovali pouze 2 nukleotidy (Mardis 2008 a).

Nově vzniklé vlákno, z něhož bylo určeno pouze 14 nukleotidů, se tedy odmyje a k templátovému vláknu se připojí nový primer, který je nasyntetizovaný tak, aby se připojil do pozice (n-1). Skládání nového vlákna tedy začalo o jeden nukleotid dříve než v předchozím sekvenačním běhu. Opakovaným posunutím primeru po odmytí dalšího vzniklého vlákna se dostaneme do pozice, kdy se první sonda navazuje na místo (n-4). Tento posun počátku syntézy nového řetězce nám umožní připojovat stále stejný typ sond, ovšem na různá místa řetězce (Kchouk et al 2017). Postupně jsou tedy nasekvenovány všechny nukleotidy.

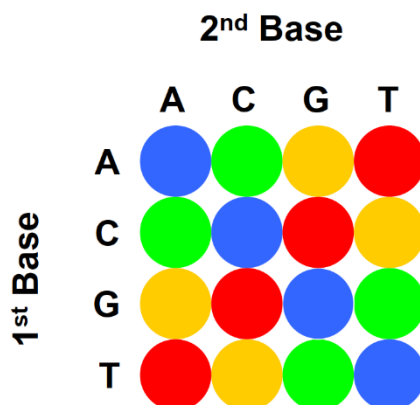
Na Obrázku 9 jsou kvůli vyšší přehlednosti jednotlivé cykly ligace sond zobrazeny různými barvami. Černé tečky označují nukleotidy, které byly identifikovány v daném cyklu čtení.



Obrázek 9: Princip sekvenace řetězce DNA při napojení první sondy v pozici n až v pozici (n-4)

Zdroj: Mardis 2008 a

Barva sondy závisí na složení dvou nukleotidů, máme tedy šestnáct možných variací, které jsou zastoupeny čtyřmi barvami (Mardis 2008 a). Klíč pro barevné rozlišení sond je zobrazen na Obrázku 10.



Obrázek 10: Kódování dvounukleotidových sond pomocí čtyř barev

Zdroj: Thermo Fisher 2006

Z Obrázku 10 je zřejmé, že jedna barva může představovat jednu ze čtyř dinukleotidových kombinací. Proto vždy musíme znát aspoň jeden nukleotid, abychom mohli s jistotou určit ten druhý. První nukleotid známe. Je totiž součástí napojeného primeru (Mardis 2008 a). Víme-li, že na první pozici je například nukleotid A a detekujeme oranžovou barvu, na druhé pozici bude nukleotid G. Ve chvíli, kdy známe tento klíč pro značení sond a zároveň jsme při sekvenaci postupovali tak, jak bylo popsáno, jsme schopni určit velmi přesně sekvenci DNA.

### 3.2.4 Ion torrent

Tato technologie sekvenování byla firmou Life Technologies představena v roce 2010 (Nayariseri et al. 2013). Způsob přípravy knihovny, stejně jako postup při sekvenování, je podobný pyrosekvenaci. Obě metody se ovšem liší principem, na němž je založena detekce signálu. Ion torrent využívá detekce vodíkového iontu, který je uvolňován při zařazení komplementárního nukleotidu do nově vznikajícího vlákna (Nayariseri et al. 2013). Chemický signál je tedy přímo elektronicky detekován a odpadá potřeba využití optického detektoru. Tento způsob detekce snižuje cenu sekvenace a čas sekvenačního sekvenačního běhu (Shendure & Ji 2008).

Řetězce DNA je před sekvenováním nutné fragmentovat na vzorky o délce 200-400 bp. Za jednu sekvenaci lze přečíst až 10 Gigabází (Gb) (Kchouk et al. 2017).



Při využití této metody k celogenomovému sekvenování může působit potíže, podobně jako u Illuminy, skládání jednotlivých kontigů. Obzvláště u částí genomu kde se nachází větší množství homopolymerních sekvencí, je ideální mít pro porovnání výsledků již dříve získaný referenční genom (Nayariseri et al. 2013). Chyby způsobené hlavně inzercemi nebo delecemi v rámci čtení homopolymerů tvoří z celkového čtení až 1 % dat (Kchouk et al. 2017).

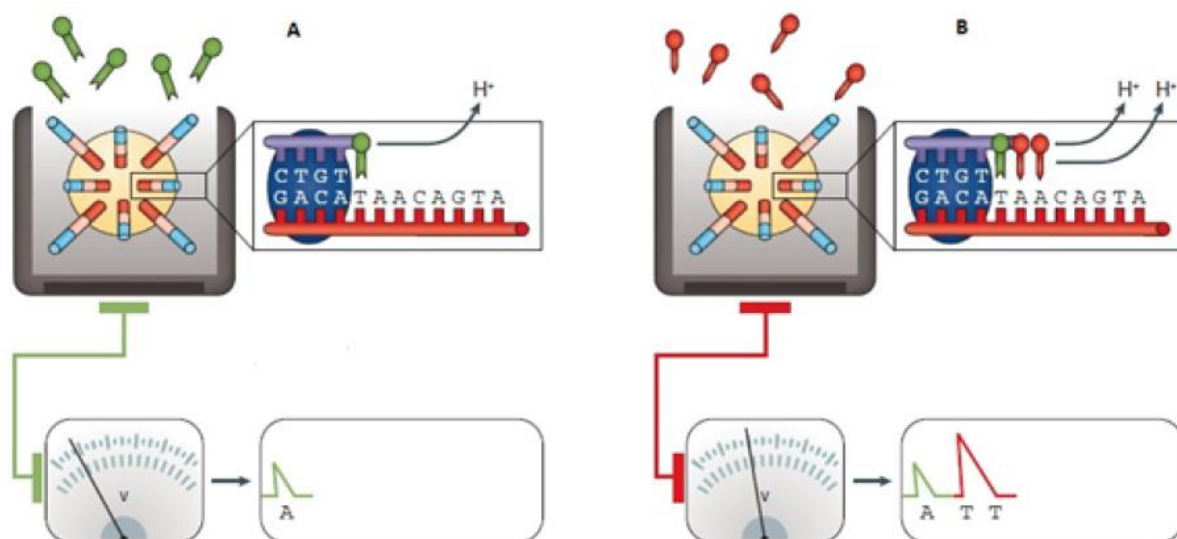
Tato metoda je také využívána při metagenomických studiích, kdy jsou na základě identifikace nukleových kyselin (NK) určeny organismy přítomné ve zkoumaném biologickém vzorku. V rámci metagenomiky jsou studovány například půdy, ovzduší nebo gastrointestinální soustavy živočichů (Zhou et al. 2010).

#### **3.2.4.1 Postup sekvenování**

Podobně jako u ostatních metod druhé generace je ještě před sekvenací potřeba vytvořit genetickou knihovnu. Namnožení vzorků se provádí emulzní PCR. Na fragmenty o délce 200-400 bp jsou z obou stran ligovány adaptéry o známé genetické sekvenci. Samotná PCR probíhá na kuličkách plovoucích v emulzi s volnými nukleotidy (Nayariseri et al. 2013). Detailně je princip emulzní PCR vysvětlen u metody pyrosekvenace. Po dokončení PCR má kulička na svém povrchu na základě komplementarity namnožené fragmenty jediného vzorku DNA (Margulies et al. 2005).

Kuličky jsou následně pipetovány na polovodičový čip, který je opatřen jamkami. Do každé z jamek zapadá jediná kulička.

Na čip jsou při následné sekvenaci postupně aplikovány roztoky, vždy obsahující pouze jeden ze čtyř nukleotidů. Obrázek 11 zobrazuje případ, kdy jsou v roztoku obsaženy nukleotidy adeninu, což vede k elongaci vznikajícího řetězce. Z dinukleotidtrifosátu je při tomto ději uvolněn vodíkový proton. Ten způsobí změnu pH v roztoku o 0,02, která vychýlí ukazatel na voltmetru z klidové polohy. Změna je patrná po dobu 4 sekund. Následně se ručička voltmetru vrátí do klidové polohy a může dojít k navázání dalšího nukleotidu. V případě homopolymerních sekvencí je změna pH vyšší kvůli většímu množství volných vodíkových protonů, jak je níže demonstrováno zařazením dvou thyminů do řetězce (Kchouk et al. 2017).



Obrázek 11: Princip detekce chemického signálu na voltmetru

Zdroj: Kchouk et al. 2017

Na dně každé jamky je tento chemický signál detekován senzory a následně převeden do digitální podoby (Rothberg et al. 2011). Následně je zbylý obsah jamky odmyt a celý proces se opakuje.

### 3.3 Sekvenování třetí generace

Vývoj metod třetí generace probíhal téměř paralelně s vývojem metod druhé generace. U některých z nich se zatím zdokonalují postupy, a proto se v laboratořích často nevyužívají (Glenn 2011).

Metody třetí generace jsou schopné osekvenovat mnohem delší úseky DNA než tomu bylo u předchozích metod. Lze s nimi analyzovat za jeden čtecí cyklus i několikanásobek počtu bází, které by připadly na genom eukaryotního organismu. Například můžeme zmínit genom člověka, u něhož je v jádře uloženo  $3,2 \times 10^9$  párů bází organizovaných do chromozomů (Spencer 2004). Genetické analýzy lidského genomu dnes, díky výkonným technologiím sekvenace, probíhají v řádu dnů až týdnů.

Velkou výhodou sekvenování třetí generace je schopnost sekvenace jediné templátové molekuly. Tento postup umožnilo zdokonalení technologií využitých při detekci světelného signálu. Detektory jsou již schopné spolehlivě zaznamenat záření vyzařované jedním fotonem, a tedy nepotřebují zesilovat signály namnožením vláken DNA (Glenn 2011).

Obrovskou nevýhodou těchto metod však zůstává velká chybovost. U Oxfordského Nanopóru zatím může dosáhnout až 90 % (Schadt et al. 2010). V tomto případě se však jedná o extrémní případ. Kupříkladu u metody PacBio dosahuje chybovost po prvním čtení 10-15 % (Kchouk et al. 2017).

Kvůli vysoké chybovosti tyto metody vyžadují kontrolní čtení pomocí metod předchozích generací, nebo opakovaného čtení jedním přístrojem, které zmenší chybovost až na polovinu (Schadt et al. 2010). Do budoucna se s očekávaným zlepšením technologií však počítá se zkvalitněním výsledků těchto metod.

### **3.3.1 Pacific Bioscience**

Společnost Pacific Bioscience představila v roce 2009 metodu sekvenování SMRT (Single Molecule Real Time Sequencing) (Mardis 2013). SMRT byla jako první metoda schopna v jednom sekvenačním cyklu překonat v délce sekvenovaných bází původní Sangerovu metodu. Oproti Sangerovým tisíci párům bází byla tato metoda v začátcích schopna přečíst sekvence o délce 1500 bp (Zhou et al. 2010). Dnes dokáže přečíst patnáct až dvacet tisíc párů bází za sekvenační cyklus. Dohromady se objem dat získaných jedinou sekvenací blíží hodnotě 1 Gigabáze. Čtení však může mít až patnáctiprocentní chybovost, jak již bylo zmíněno (Kchouk et al. 2017). Musí se tedy využívat kontrolního čtení kupříkladu pomocí Illuminy nebo opakovaných čtení molekuly DNA.

Oproti jiným technikám má však tato metoda také spoustu výhod. Je schopna sekvenovat delší úseky DNA, které se díky tomu lépe skládají do větších celků a navíc dokáže rozlišit jednotlivé isoformy DNA (Rhoads & Au 2015). Také rozpoznává, díky neměnné síle signálu při sekvenaci homopolymerů, repetitivní sekvence s nižší chybovostí než u metod druhé generace.

Tato technika nevyžaduje, aby byla před samotnou sekvenací vytvořena knihovna v podobě mnoha kopií jedné molekuly DNA. Zjednodušení přípravy vzorků představuje i možnost vkládat do sekvenátoru jak jednovláknové, tak dvouvláknové molekuly DNA, získané přímo ze studovaného organismu. Samotná sekvenace následně probíhá díky přirozenému procesu replikace s využitím enzymu DNA polymerázy. Na základě komplementarity jsou volně plovoucí nukleotidy napojovány k nově vznikajícímu vlákně. Díky detekci pořadí těchto nukleotidů je následně určena sekvence vzorku (Coupland et al. 2012). Čtení této sekvence a zpracování dat přitom probíhá v reálném čase (Mardis 2008 a).

### 3.3.1.1 Postup sekvenování

Podobně jako u předchozích metod jsou na dvouvláknovou molekulu DNA z obou stran přichyceny dva adaptéry. Z Obrázku 12 je patrné, že sekvence konců těchto adaptérů je navržena tak, aby byly vůči sobě komplementární. Díky enzymu ligáze dochází následně k napojení adaptérů z obou stran dvouvláknové DNA. Tím vzniká kruhová molekula (Rhoads & Au 2015). Tento tvar je ideální pro opakované čtení jednoho vzorku. DNA polymerázu stačí k molekule DNA napojit pouze jednou, ale sekvenace lze provádět opakovaně.

Na obrázku 12 je dále oranžovou barvou zobrazen primer, který přisedl k jednomu z adaptérů. Na něj se poté naváže DNA polymeráza. Takto upravený vzorek je vložen do jamky o rozměrech sto krát deset nanometrů. Na dně této jamky se enzym DNA polymerázy fixuje k místu, kudy je do jamky emitováno laserové záření a zároveň prováděna detekce pomocí CCD kamery (Eid et al. 2009).



Obrázek 12: Vlákno DNA s napojenými adaptéry, primerem a DNA polymerázou

Zdroj: Rhoads & Au 2015

K detekci se využívají fluorescenčně značené nukleotidy, přičemž každý z nich je značen jinou barvou. Na základě komplementarity bází dochází k syntéze nového řetězce. Obrázek 13 představuje princip tvorby nového řetězce. Zeleně značený deoxyadenosintrifosfát (dATP) se dostává k místu na dně jamky, kde je zafixována DNA polymeráza. Jelikož je komplementární k thyminové bázi na původním vlákně, je díky DNA polymeráze zařazen do nově vznikajícího řetězce (Kchouk et al. 2017). Zároveň je dATP v tuto chvíli ozářen laserem. Díky monochromatickému záření laseru je CCD kamera schopna detekovat zelenou barvu odraženého světla (Eid et al. 2009). Ve chvíli, kdy je nově zařazený nukleotid spojen pomocí vodíkových můstků s nukleotidem na původním vlákně, dochází k posunu řetězce. Zároveň nukleotid ztrácí své fluorescenční označení. Světlo laseru již osvětluje dTTP, který se řadí do řetězce a je detekován oranžovým zářením (Rhoads & Au 2015).

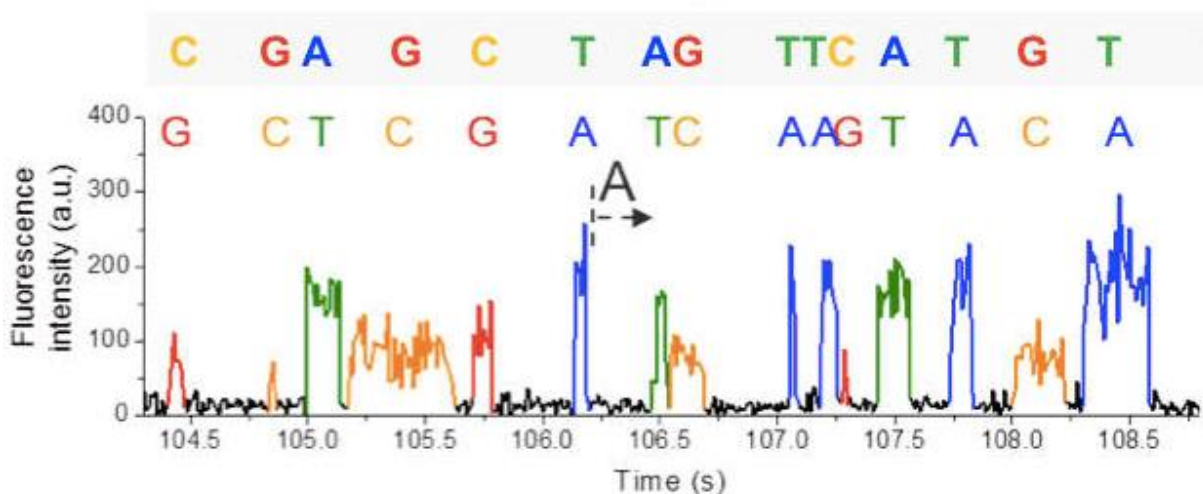


Obrázek 13: Fluorescenčně značené nukleotidy tvoří řetězec DNA

Zdroj: Pacific Bioscience 2009

Světlo laseru proniká do jamek o velikosti pouhých desítek nanometrů, zvaných jako zero mode waveguides (ZMWs). Díky malým rozměrům ZMWs jsou dovnitř propouštěny pouze krátké světelné vlny. Ty osvětlují DNA polymerázu, ale ne zbytek jamky a tím pádem snižují fluorescenční zátěž emitovaného světelného signálu pro CCD kameru (Schadt et al. 2010). Jsou-li osvětleny volně plovoucí nukleotidy, emitují světelný signál. Vlnová délka světla je však příliš velká a signál je tedy pohlcen díky ZMWs. V CCD kameře nevzniká žádný šum díky velmi malým rozměrům jamky (Rhoads & Au 2015).

Na Obrázku 14 můžeme sledovat, jak jsou signály zpracovány graficky. První nukleotid byl vyhodnocen jako guanin. Do záznamu v počítači se rovnou zapisuje sekvence nukleotidů na templátovém vlákně DNA, tedy sekvence na obrázku zobrazena v prvním řádku. Ve chvíli, kdy se do vlákna zařadí guanin, v počítači je zobrazen cytosin.



Obrázek 14: Zápis sekvence nukleotidů do grafu v počítači

Zdroj: Pacific Bioscience 2012

### 3.3.2 Oxfordský Nanopór

Oxfordský nanopór je metoda třetí generace, u které nemusí být před počátkem sekvenace nijak modifikovány nukleotidy používané k sekvenaci. Sekvenujeme přímo vlákno DNA, které bylo izolováno. Příprava vzorku pro sekvenaci je v některých přístrojích i automatizovaná. Sekvenace probíhá v reálném čase (Schadt et al. 2010).

První údaje o této technice sekvenování byly zveřejněny již v roce 2008, nicméně se tato metoda začíná více využívat až v posledních letech. Za pomalý nástup pravděpodobně může poměrně vysoká chybovost, která u této techniky dosahuje až 90 %. Opakovaným čtením vzorku může být chybovost snížena až na 38 % z celkového čtení, což je stále poměrně vysoké číslo (Clarke et al. 2009). Přesto je tato technika nyní poměrně populární. Firma Oxford Nanopore technologies totiž uvedla jako novinku přenosný přístroj pro sekvenaci genomu. Má velikost většího flash-disku, sekvenuje v reálném čase poměrně dlouhé úseky DNA a pomocí konektoru se dá připojit k počítači. Vědci tedy nyní mohou dostat výsledky sekvenování přímo v terénu (Laver et al. 2015).

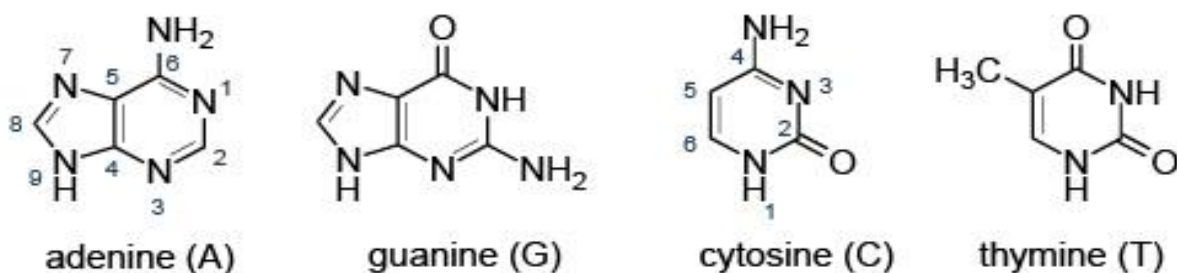
Díky technice nanopóru je možno přečíst i sekvence o délce 2 Mb (Clarke et al. 2009). Proto je tato technika nejvíce využívána ke čtení velkých genomů (Schadt et al. 2010). A to při *de novo* sekvenování, analýzách chromozomálních variací nebo například při mapování haplotypů (Clarke et al 2009).

#### 3.3.2.1 Postup sekvenování

Na vlákno DNA je napojen adaptér, který navede a ukotví molekulu přímo v oblasti póru. Enzym DNA helikáza v tomto místě odděluje obě vlákna dvoušroubovice tím, že rozrušuje vodíkové můstky mezi nimi. Na sekvenovaném vlákně vždy pórem prochází až 6 nukleotidů najednou (McGinn & Gut 2012).

Princip funkce Oxfordského nanopóru je založen na protažení jednovláknové molekuly DNA membránovým proteinovým pórem  $\alpha$ -hemolysinu, který je vyplněn elektrolytem. Elektrický potenciál, aplikovaný na tento pór, umožňuje sledovat průtok iontů pórem, a tedy změnu elektrického proudu v čase (Schadt et al. 2010).

Na Obrázku 15 je zobrazena chemická struktura jednotlivých bází DNA. Tyto báze tvoří společně s pětiuhlíkatými cukry a zbytky kyseliny fosforečné nukleotidy, které tvoří řetězec DNA (Sanger 1977). Každý nukleotid má jinou velikost a způsobí jinou výchylku na amplitudě toku iontů pórem. Právě na základě výchylek toku iontů je nakonec detekován příslušný nukleotid (Clarke et al 2009).

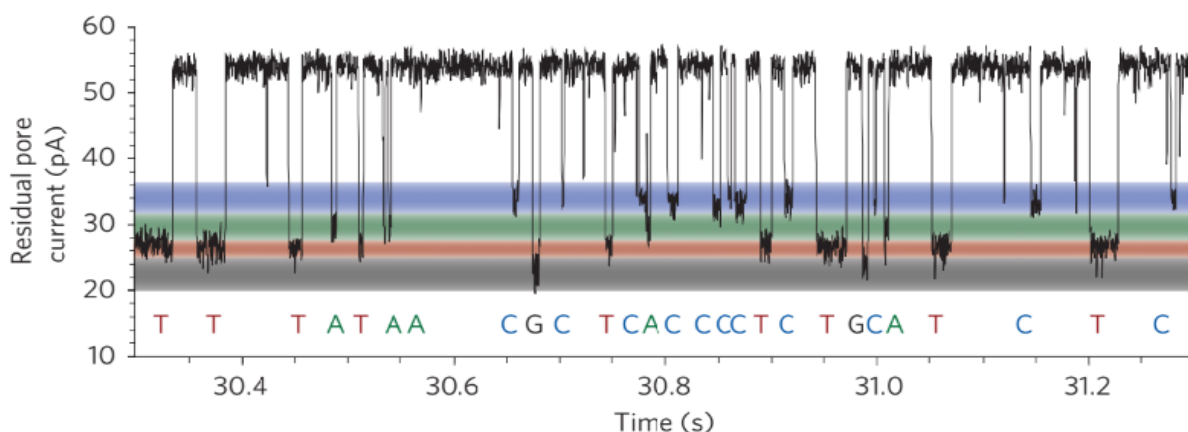


Obrázek 15: Chemické struktury bází DNA

Zdroj: Brown & Brown 2011

Výchylka elektrického proudu při průchodu pórem závisí jak na velikosti procházejících bází, tak na prostorové konformaci nukleotidů v póru (Clarke et al 2009). Porovnáme-li velikosti jednotlivých bází s výchyly elektrického proudu na grafu na Obrázku 16, je dále zřejmé, že prostorově nejmenší cytosin při svém průchodu ucpává pór nejméně a zbytkový průtok částic pórem je nejvyšší. Naopak guanin blokuje pór nejvíce.

Graf na Obrázku 16 má pro ulehčení čtení signálu na svém pozadí čtyři barevné pruhy, které symbolizují čtyři nukleotidy. Při průchodu prvního nukleotidu se signál vychýlí do oblasti červeného pruhu, který signalizuje thymin. Následně se vrátí na hodnotu kolem 55 pikoampérů (pA). To značí, že pór žádný nukleotid v tuto chvíli neprochází.



Obrázek 16: Graf změny elektrického proudu v čase v závislosti na průchodu nukleotidů pórem

Zdroj: Clarke et al. 2009

Na pomalu se pohybujícím vlákně by tato technologie fungovala dobře (Clarke et al. 2009). Rychlost pohybu řetězce skrz pór je však tak vysoká, že často dochází

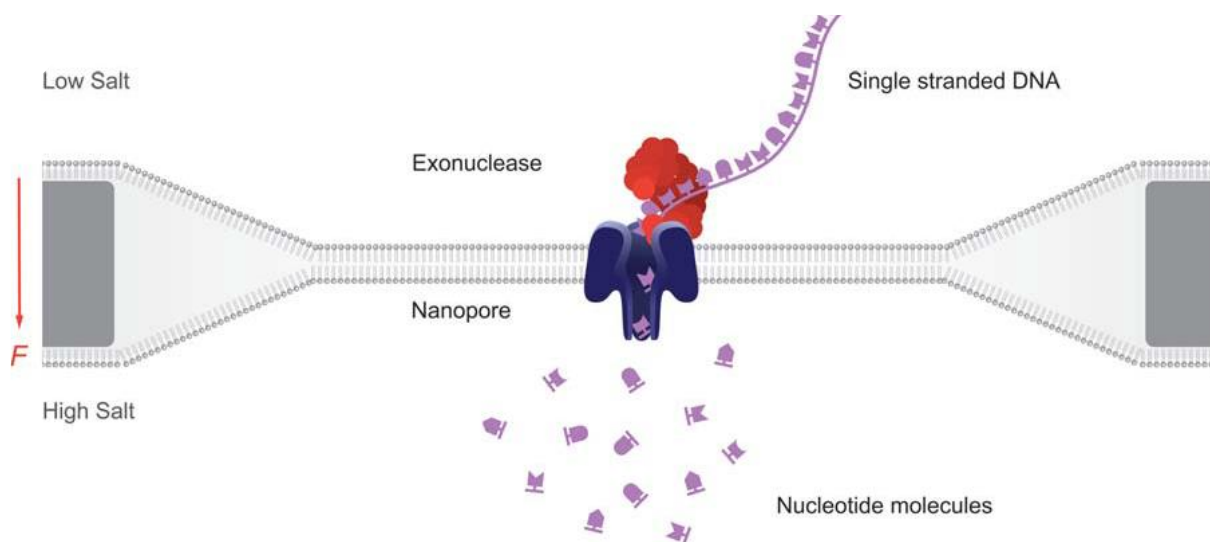
k záměně jednoho nukleotidu za jiný při čtení dat (McGinn & Gut 2012). Proto se dále studují enzymy, které by svou přítomností v nanopóru omezily počet translokací.

### 3.3.2.2 Sekvenování s využitím exonukleázy

Firma Oxford Nanopore technologies má patentovanou ještě jednu obdobnou metodu sekvenování s využitím póru zobrazenou na Obrázku 17. Technologie se oproti té předchozí liší v použití exonukleázy, která by před průchodem řetězce zkrz pór nastříhala vlákno DNA na jednotlivé nukleotidy. Tyto nukleotidy by poté jednotlivě procházely pórem a byla by tedy ulehčena jejich detekce (Clarke et al. 2009). Koncept využití exonukleázy však zatím nebyl uveden do praxe.

Problémy při detekci spočívaly v čase, kdy pórem neprochází žádný nukleotid, a tedy není vytvářen záznam, a dále ve vysoké chybovosti metody a v principu funkce enzymu, který zajišťuje oddělení nukleotidů před průchodem pórem (Glenn 2011).

Enzym exonukleáza, který byl získán z bakterie *Escherichia coli*, při teplotě 37 °C štěpí 275 nukleotidů za sekundu. Taková rychlost může vést ke značným chybám v detekci signálu v nanopóru. Ovšem činnost exonukleázy může být regulována teplotou (Clarke et al. 2009).



Obrázek 17: Princip funkce nanoporu za použití exonukleázy

Zdroj: Schadt et al. 2010



### 3.4 Sekvenace čtvrté generace

Ačkoliv se technika Oxfordského nanopóru začala využívat až v posledních letech, stojí už na jejích principech vývoj další generace sekvenování. Cílem čtvrté generace je opět snížit cenu sekvenování a také práci s přípravou vzorků (Feng et al. 2015).

Vývoj se tedy zaměřuje na testování různých druhů nanopórů. A to jak biologických, jako byl nanopór  $\alpha$ -hemolysinu, tak i uměle vytvořených. Ty jsou zastoupeny chemickými látkami, které tvoří membránu, do níž je uměle vytvořen nanopór (McGinn & Gut 2012). Využívá se například  $\text{SiO}_2$ , který je chemicky stabilní a zároveň pevný, na rozdíl od biologických nanopórů. Vědci se snaží modifikovat tyto membrány tvořené pevnými látkami pomocí bílkovin a tuků. Vznikají tak nanopóry, které kombinují biologické a uměle vytvořené složky. Ty jsou při kombinaci s tranzistory a dalšími složkami zařízení méně náročné na prostor. Poskytují tedy možnost zmenšení sekvenačního přístroje (Feng et al. 2015).

Momentálně se také pracuje na metodě, kdy se nukleotid protáhne mezi dvěma elektrodami. U této techniky však není dokonale vyladěn způsob detekce jednotlivých nukleotidů, a pracuje se tedy na optimalizaci velikosti elektrod a vzdálenosti mezi nimi (Bai et al. 2014).

Až bude technika nanopóru dostatečně zdokonalena, má potenciál stát se levnou a rychlou metodou sekvenování, která je schopna dosáhnout délky čtení 10 000-50 000 bp. Mohla by tedy dosáhnout cíle v podobě osekvenování lidského genomu za 100 amerických dolarů (Feng et al. 2015).

### 3.5 Příklady použití nových sekvenačních metod

#### 3.5.1 Studie kolapsu kolonií u včel v USA

Technika pyrosekvenace byla využita při studii Colony Collapse Disorder (CCD), tedy kolapsu v koloniích včel, kdy včelstva přicházejí o své dělnice (Rothberg & Leamon, 2008). Tento jev měl na svědomí ztrátu 50-90 % kolonií včel medonosných (*Apis mellifera* Linnaeus 1758) napříč Spojenými státy. Masivní úhyny dělnic začaly od roku 2004 u včelstev, která byla do USA importována z Austrálie nebo u těch, která byla ve styku s těmito včelstvy (Cox-Foster et al. 2007).

U nakažených včelstev přežívaly pouze dělnice s nejvyšší imunitou. Ty se však nebyly schopné starat o potravu a zároveň o nově nakladené larvy, a tak včelstvo vyhynulo.

Studie porovnávala mikroflóru zástupců čtyř včelstev z USA nakažených CCD, dvou nenakažených včelstev z USA, zdravého včelstva importovaného z Austrálie a mateří kašičky importované z Číny. Vědci se zaměřili na získání RNA, aby bylo zohledněno patogenní působení RNA virů. Jednalo se o metagenomickou studii, kdy se analyzovala veškerá genetická informace získaná ze střev včel (van Engelsdorp et al. 2009).

U včel nakažených CCD bylo nalezeno větší množství patogenů v porovnání s kontrolními včelstvy (van Engelsdorp et al. 2009). V rámci výzkumu bakteriální kolonizace včelstev byly analyzovány také rody bakterií jako například *Lactobacillus* nebo *Neisseria*, které se nacházejí ve střevech včel přirozeně. Ovšem nebylo tím vyloučeno, že by jejich namnožení nemohlo mít vliv na propuknutí nemoci. Také byla analyzována rRNA u 81 zástupců říše Fungi, ovšem jejich podíl na propuknutí CCD nebyl prokázán (Cox-Foster et al. 2007).

Genetickou analýzou bylo analyzováno sedm ssRNA kandidátních virů, které by mohly mít vliv na propuknutí CCD. Výsledky výzkumu identifikovaly jako nejpravděpodobnější, že virus způsobující CCD je tzv. Izraelský virus akutní paralýzy (IAPV), který se ve vysoké míře nacházel právě u včelstev nakažených CCD. Tento virus byl poprvé objeven v roce 2004 v Izraeli, kde jednotlivým zástupcům způsoboval třes křídel a včely se tedy nemohly vrátit do úlu (Hou et al. 2014). Ačkoliv vir IAPV nebyl nakonec identifikován jako patogen, který přímo způsobuje CCD, z výsledků je pravděpodobné, že jeho působení má vliv na obranyschopnost organismu. Společně s dalšími faktory, jako je nedostatečná výživa nebo působení pesticidů, způsobuje CCD (van Engelsdorp et al. 2009).

V rámci tohoto výzkumu se nepovedlo určit patogen, který může za propuknutí CCD u amerických včelstev. Byl jím však demonstrován způsob, jak provádět epidemiologickou studii infekčních nemocí s nejasným patogenním agens.

### **3.5.2 Výzkum přenosné rakoviny d'áblů**

U druhu d'ábla medvědovitého (*Sarcophilus harrisii* Boitard 1841) byla nalezena infekční forma rakoviny zvaná Devil facial tumour disease (DFTD). Poprvé byla pozorována faciální rakovina d'áblů v roce 1996. Dnes se již rozšířila na 80 % území Tasmánského ostrova. Díky vysoké nakažlivosti se d'áblové řadí mezi druhy ohrožené vyhynutím (Nowak 2017).

Přenosný typ rakoviny byl objeven pouze u psů, mlžů a tasmánských čertů. Karcinomy představují buněčné linie, které jsou při kontaktu dvou jedinců přenosné. Jelikož jsou tasmánští d'áblové převážně mrchožraví, dochází k přenosu hlavně při boji o mršinu, případně při pohlavním styku. DFTD vytváří zpočátku malé nádory na obličejí nebo ústní sliznici.

Tyto útvary rostou a metastázuji do mízních uzlin po celém těle. Nakažený jedinec často umírá po několika měsících na podvýživu, jelikož je pro něj příliš bolestivé přijímat potravu (Pye et al. 2015).

V populaci d'áblů je patrný efekt zakladatele, tedy nízká genetická variabilita jedinců. Problém vysoké nakažlivosti spočívá v nízké variabilitě genů pro hlavní histokompatibilní komplex (Major Histocompatibility Complex, MHC) a podobném složení alel pro MHC v rakovinotvorných buňkách (Pye et al. 2015). MHC by za normálních okolností měl aktivovat NK buňky a další obranné mechanismy, ovšem při nákaze je tento proces oslaben. Náhylnějšími k získání rakoviny jsou starší jedinci, jejichž imunita je více oslabena (Peel et al. 2018).

Dnes již díky rozsáhlým studiím, které se problematikou přenosu rakoviny zabývají, známe dva typy této rakoviny, a to DFT1 a DFT2. Na pohled nejsou tyto dvě formy od sebe odlišitelné, ovšem geneticky ani histologicky si nejsou podobné. Rakovinotvorné buňky typu DFT1 mají na rozdíl od DFT2 aneuploidní karyotyp se čtyřmi nadpočetnými chromozomy. U DFT1 buněk nejsou na rozdíl od DFT2 přítomné gonozomy. Histologicky se oba typy liší uspořádáním buněk. DFT1 vytváří v tkáních shluky buněk, kdežto DFT2 buňky se rozptylují tkáněmi a nevytváří ucelené struktury (Pye et al. 2015).

### **3.5.2.1 Genetická analýza přenosné rakoviny u d'áblů**

V posledních letech byla v několika případech u nákazy DFT1 sledována přirozená regrese tumoru a případně i následné vyléčení jedince (Wright et al. 2017). K výzkumu toho, které geny podmiňují regresi tumoru, bylo využito sekvenace přístrojem Illumina HiSeq. Analyzované vzorky byly získány biopsií z ušních boltců 10 jedinců, kteří byli nakaženi DFT1. U sedmi z těchto jedinců došlo k regresi tumoru, u zbývajících tří ne. Komparativní analýzou těchto dvou skupin byly zjištěny úseky genomu, konkrétně na třetím a pátém chromozomu jedinců, které podmiňují regresi tumoru. Geny velkého účinku, nacházející se v těchto oblastech, pravděpodobně polygenně podmiňují regresi tumoru. Zároveň byly u těchto genů zjištěny SNPs v oblasti intronů výše zmíněných chromozomů, které pravděpodobně ovlivňují expresi těchto genů (Margres et al. 2018).

Tímto výzkumem bylo demonstrováno, že exprese genů souvisejících s imunitní odpovědí jedince je zásadní pro spuštění regrese tumoru. Identifikace kandidátních genů a metody využívané při komparativní onkologii přispívají ke studiím souvislostí mezi genotypem a ději probíhajícími při rakovinném bujení (Wright et al. 2017). Studie, jako je

tato, mají velký vliv na humánní onkologii. Z etického hlediska je totiž vyloučeno studovat vývoj rakoviny bez léčení u člověka. Komparativní onkologie je tedy prováděna pouze u zvířat, ovšem její poznatky slouží ke stanovení principů biologické léčby v humánní medicíně (Peel et al. 2018).

### 3.5.3 Výzkum rezistence vůči viru SIV u primátů

Mangabej kouřový (*Cercocebus atys* Audebert 1797) z čeledi kočkodanovitých (Cercopithecidae) žijící při pobřeží západní Afriky je nyní jedním z druhů, který vědce extrémně zajímá. Důvodem je jeho rezistence vůči viru SIV (Simian immunodeficiency virus). Mezidruhovým přenosem tohoto viru na člověka vznikl virus HIV (Human immunodeficiency virus). Nakažení mangabejové, jako přirození hostitelé viru, však nevykazují klinické známky nemoci (Milush et al. 2011). Proto by mohli být klíčovými organismy pro stanovení léčebných metod lidí nakažených HIV (Dunham et al. 2006).

Imunologické studie se dříve zaměřovaly na funkci T-lymfocytů v průběhu nákazy SIV. Bylo zjištěno, že T-buňky se v organismu mangabeje nemnoží, jak by měly v rámci imunitní reakce vůči množícímu se viru (Milush et al. 2011). U makaka rhesuse (*Macaca mulatta* Zimmerman 1780) se naopak bez léčby antivirotiky při SIV nákaze vyvinul syndrom imunodeficiency AIDS (Palesch et al. 2018). Další studie se proto zabývají mechanismy, které způsobují imunologickou rezistenci mangabejů vůči AIDS (Dunham et al. 2006).

#### 3.5.3.1 Genetická analýza hostitelů viru SIV

Pro výzkum genetického materiálu mangabejů kouřových bylo v první řadě využito technologie Illuminy HiSeq 2000. Následně byly k získaným vzorkům přirovnávány vzorky sekvenace Pacific Bioscience RS platformy. Jelikož technologie PacBio je schopná přečíst delší fragmenty DNA, byla využita ke komparativnímu čtení původních vzorků a doplnění chybějících sekvencí původního čtení Illuminy (Palesch et al. 2018).

DNA byla získána z krve samice mangabeje kouřového žijící v Yerkesově výzkumném centru pro primáty. Ze zde žijících mangabejů měla tato samice ve skupině nejčastěji zastoupený haplotyp MHC (Major Histocompatibility Complex), a byla tedy vybrána jako ideální zástupce pro výzkum. Zároveň byla kvůli zjišťování polymorfismů v rámci druhu izolována DNA z krve i jater 10 dalších mangabejů. V rámci skupiny byli tito jedinci vybíráni tak, aby pokryli 90% diverzitu v rámci dané kolonie (Palesch et al. 2018).

Sekvenací bylo odhaleno 20 829 kódujících genů a 4 464 genů nekódujících. Množství získaných dat umožnilo provádět komparativní genomové analýzy s 25 druhy primátů, jejichž genom již byl osekvenován (Palesch et al. 2018).

Díky komparativní analýze mangabeje a primátů náchylných k Acquired Immune Deficiency Syndrome (AIDS), bylo možné identifikovat geny, které mohou mít roli při náchylnosti k propuknutí AIDS u jedinců nakažených SIV (Dunham et al. 2006). Vědci takto odhalili 34 kandidátních genů ovlivňujících imunitní systém, které se lišily u mangabeje kouřového a makaka rhesuse (Palesch et al. 2018). Pomocí genomové analýzy byly odhaleny rozdíly v proteinech vyskytujících se při nákaze HIV virem, receptorech buněk i dalších imunitních mediátorech (Milush et al. 2011).

Na těchto výzkumech můžeme vidět efektivitu využití genetického sekvenování. Z několika desítek tisíc nasekvenovaných genů bylo určeno 34 genů imunitního systému, které mohou mít spojitost se vznikem imunodeficience u nepřirozených hostitelů viru SIV. Následné fyziologické studie poté vychází z těchto dat a zkoumají imunologické funkce bílkovin vzniklých expresí těchto genů (Palesch et al. 2018). Výsledky takových výzkumů v budoucnosti mohou zásadně přispět k vývoji vakcíny proti působení viru HIV na organismus, a tedy i k potlačení propuknutí nemoci AIDS.

#### **3.5.4 Fylogenetická studie čeledi Hominidae**

Ke stanovení nového druhu, nebo více poddruhů živočicha dříve byly zásadní hlavně morfologické analýzy jedinců.

Genetická analýza se před využitím NGS metod prováděla pomocí proteinových markerů nebo zjišťováním délkových polymorfismů restričních fragmentů (Locke et al. 2011). Při rozlišení druhů orangutana sumaterského (*Pongo abelii* Lesson 1827) a orangutana bornejského (*Pongo pygmaeus* Linnaeus 1760) se konkrétně využívalo mitochondriálních a nukleárních markerů. Tímto způsobem byla například identifikována neocentromera, která se u obou druhů orangutana na různých místech chromozomu 12 (Tolomeo et al. 2017).

Také již bylo zjištěno, že rod *Pongo* je z fylogenetického hlediska nejvzdálenější člověku (Nater et al. 2017). Genetické studie se z toho důvodu zaměřovaly na genomické rozdíly v rámci čeledi hominidů (Hominidae) (Locke et al. 2011).

Z této čeledi má podčeď Ponginae nejméně změněný karyotyp. Velké změny v genetické informaci orangutanů byly často způsobeny zdvojením určitého segmentu DNA. V porovnání se šimpanzy (*Pan* sp.) nebo člověkem u orangutanů docházelo s menší intenzitou k přeskupování těchto segmentů (Locke et al. 2011).

Zajímavé je například srovnání Alu sekvencí těchto tří rodů. Jedná se o transpozonální sekvence DNA o délce kolem 300 bp, které obsahují časté repetitivní sekvence. Slouží například k regulaci exprese genů. U orangutanů tvoří tyto sekvence pouze malou část genomu. U šimpanzů sledujeme nárůst Alu retropozic. Lidský genom je z hlediska Alu sekvencí ještě mnohem plastičtější oproti předchozím dvěma rodům. Právě na zjišťování retropozic u Alu sekvencí je možné dokázat fylogenetický vývoj primátů (Walker et al. 2012).

#### **3.5.4.1 NGS analýza podčeledi Ponginae**

V roce 2011 byla technologie od firmy Illumina využita k sekvenování genomu orangutana sumaterského. V rámci výzkumu byl tento genom srovnáván s genomem orangutana bornejského. Jejich předek žil v Jihovýchodní Asii, ovšem oba druhy žily izolovaně na různých ostrovech a byly ovlivňovány vzrůstající lidskou populací (Walker et al. 2012).

Zkoumaný materiál byl odebrán od samice orangutana sumaterského. Pro vytvoření genetické knihovny byly dále odebrány vzorky deseti zástupců orangutanů sumaterských a orangutanů bornejských, odchycených z volné přírody. Dále se k sestavení výsledného genomu orangutana využilo již známého genomu člověka. Díky příbuznosti obou druhů byl lidský genom využíván k ověřování správnosti sestavení výsledného pořadí jednotlivých krátkých segmentů DNA u orangutana (Locke et al. 2011).

Ve studii se nejprve díky čtení krátkých úseků DNA zhotovily téměř kompletní sekvence genomu obou druhů. Oba genomy byly následně porovnávány hlavně z hlediska SNPs, tedy single nucleotide polymorphisms (polymorfismů jediného nukleotidu). U sumaterského orangutana bylo porovnáváním genomů nalezeno o 36 % SNPs více než u bornejského orangutana oproti jejich společnému předkovi (Locke et al. 2011). Z hlediska samostatného vývoje obou druhů bylo více mutací v rámci genomu nalezeno u orangutana bornejského. Ovšem v rámci hodnocení jedinců jedné podčeledi byla u orangutana sumaterského zjištěna vyšší variabilita ve zkoumaných genomech (Locke et al. 2011).

Díky analýze nukleotidových polymorfismů, která však může mít poměrně vysokou chybovost, byl vývoj obou samostatných druhů datován od doby před 400 000 lety.

Dnes již díky dalším genetickým a morfologickým studiím prováděným na rodu *Pongo* víme, že existuje další druh orangutana, a sice orangutan tapanulijský (*Pongo tapanuliensis* Nowak et al. 2017) (Nater et al. 2017).

### 3.5.5 Genetické studie hospodářských zvířat

Poměrně často se metody NGS využívají k sekvenování genomů dobytka. Začaly tak přispívat k selekci zvířat pro zvyšování mléčné nebo masné užitkovosti, která se dříve dělala hlavně na základě fenotypu. Umělý výběr dal vzniknout velkému počtu plemen, která se liší užitkovostí, adaptací na životní prostředí a zdravotními charakteristikami (Rosse et al. 2017).

Velký vliv na tyto aspekty mají u mnoha druhů zvířat takzvané Copy Number Variations (CNVs), tedy úseky DNA, kdy se v různém počtu kopií opakuje jeden typ sekvence. Přímou tedy ovlivňují kvalitativní i kvantitativní hodnoty u zvířete a mohou být důvodem vzniku onemocnění (da Silva et al. 2016). Díky NGS se CNVs efektivněji určují v rámci celého genomu. Díky tomu bylo zjištěno, že například u člověka je druhová variabilita způsobena ze 77,97 % právě variabilitou počtu kopií úseků DNA (da Silva et al. 2016).

Největší variabilita v CNVs byla u skotu nalezena u skupiny zebu (*Bos taurus indicus* Linnaeus 1758), zahrnující asijská a africká plemena. Konkrétně byla studie prováděna na plemeni Nelore (*Bos indicus* Linnaeus 1758). Vzorky, získané od 1717 zvířat z chovů v Brazílii, byly porovnávány s osmi vzorky nepříbuzných býků, jejichž plemenitbou toto plemeno vzniklo. Ke studii byl použit sekvenátor HiSeq2000 od Illuminy (da Silva et al. 2016).

Z celkového počtu bylo 24 nalezených CNVs sledováno u více než 10 % zkoumaných zvířat. K variabilitě ostatních CNVs přispívaly hlavně SNPs, tedy mutace v jediném nukleotidu řetězce (da Silva et al. 2016).

Velmi důležité z hlediska plemenářské práce jsou CNVs identifikované v lokusech autozomů, které ovlivňují užitkovost, velikost těla a resistenci vůči parazitům. V rámci plemena Nelore jsou tedy důkladněji zkoumány sekvence genů podmiňujících masnou užitkovost (da Silva et al. 2016). Výsledky tohoto výzkumu tedy bude možné využít k další plemenářské práci v rámci tohoto plemene i druhu (Rosse et al. 2017).

#### 3.5.5.1 Genetická analýza slintavky a kulhavky

U skotu je velká pozornost věnována RNA virovému onemocnění slintavky a kulhavky (SLAK). Nákaza způsobuje vysoké ztráty v užitkovosti a i po odeznění příznaků je jedinec nositelem viru. Díky vysoké odolnosti a schopnosti přežít ve vnějším prostředí je SLAK nejobávanější chorobou sudokopytníků (Paton et al. 2018).

Ke studii bylo využito sekvenátoru Illumina Miseq a Sangerovy metody sekvenování pro kontrolní čtení. Sekvenace nové generace umožnila studovat evoluci viru v průběhu infekce,

množení v buňkách, selekce a migrace do dalších tkání. I z malého množství vzorku je možné detailně studovat různé genetické varianty viru. U viru SLAK, jehož genom je 8 300 nukleotidů dlouhý, tedy můžeme sledovat až 25 000 bodových mutací. Při pravidelném odběru vzorků takto můžeme stanovit mutační rychlost (Wright et al. 2010).

Zájem vědců vzbudila evoluce viru, přežívajícího v hostiteli mezi nakažovými situacemi. Vzorky ke studii SLAK byly získávány stěry slin zvířatům na třiceti pákistánských farmách, chovajících asijské buvoly (*Bubalus bubalis* Linnaeus 1758), kteří nevykazovali klinické projevy nemoci. Jelikož se zkoumala mutační rychlost virů, probíhal sběr vzorků u 12 nakažených jedinců každých 28 dnů po dobu jednoho roku.

U jednoho zvířete byly prokázány dva sérologické typy viru. Tento vzorek byl použit k další studii. Díky analýze vzorků bylo zmapováno 21 téměř kompletních genomů viru. Studie ukázala, že při dlouhodobé nákaze s vymizením příznaků nemoci, je variabilita genotypů viru nízká. Po 98 dnech od inokulace viru genetická informace virů nezměněna, a to až z 99 % (Wright et al. 2010).

Studie evoluce viru v organismu v rámci chronických nález bez klinických příznaků je důležitá pro vývoj efektivnějších vakcín, které zvýší obranyschopnost organismu vůči nakažení (Ramirez-Carvajal et al. 2018).

### **3.6 Srovnání generací sekvenování**

Generace sekvenování jsou rozděleny na základě podobností v postupech pro přípravu vzorků, v délce osekvenovatelných úseků, v ceně sekvenování i z hlediska období počátku jejich využívání.

Metody první generace jsou nejnáročnější z hlediska poměru času vloženého do příprav vzorků k objemu výchozích dat. Kupříkladu kompletní genom člověka se Sangerovou metodou zpracovával 10 let (Schadt et al. 2010). První dvě generace sekvenování vyžadují nejprve mnohonásobné namnožení vzorků. Před samotným sekvenováním je nutné připravit roztok s reagensy, přístroj nebo destičku, ve kterých bude samotná sekvenace probíhat.

Pro první generaci, tedy Sangerovu metodu, se historicky pro přípravu knihovny používalo množení vzorků za pomoci ligování do kruhových molekul DNA, které jsou v bakterii přirozeně množeny (Clarke et al. 2009). Později začala být využívána PCR, čímž se velmi urychlila příprava vzorků. Dnes se k Sangerově sekvenaci využívá automatizované kapilární elektroforézy, kdy přístroj zasílá data do počítače a na základě vyhotoveného elektroforeogramu je zřejmá sekvence řetězce (McGinn & Gut 2012). Maximálně může tímto



způsobem být přečteno 96 fragmentů za jedno čtení. Výstupní množství dat je tedy mnohem nižší než při sekvenaci NGS metodami (Kchouk et al. 2017).

Výhodou této metody je vysoká spolehlivost výsledků, nevýhodou vysoká cena za sekvenaci. Využívá se proto hlavně ke kontrolnímu čtení po sekvenacích jinými, méně přesnými technikami (Liu et al. 2012).

Pyrosekvenaci můžeme chápat jako přechod mezi Sangerovou metodou sekvenování a sekvenováním nové generace. Cena sekvenace je vyšší než u ostatních metod druhé generace, a to kvůli pořizovacím nákladům enzymů potřebných k sekvenaci. Délka DNA fragmentů dosahuje 250 bp, tedy se na počátku využívání metod NGS nejvíce blížila Sangerovým 1000 bp (Pettersson et al. 2009).

Příprava knihovny je u metod druhé generace opět prací laboranta. Sekvenaci a vyhodnocení vzorků dělají přístroje, které jsou paralelně schopny osekvenovat tisíce vzorků najednou. Ovšem, jak již bylo zmíněno, metody druhé generace mají statisticky o něco větší chybovost, než Sangerova metoda (Glenn 2011). Za to mohou chyby lidského faktoru, enzymu DNA polymerázy nebo například chyby v detekci signálu přístroji.

Metody druhé generace jsou schopné zpracovávat kratší řetězce genetické informace. Oproti Sangerově metodě, která dokázala zvládnout při jednom čtení i 900 bp, například metoda SOLiD v jednom čtení přečte pouze 75 bp (Mardis 2008 a). Díky masivnímu paralelnímu sekvenování je však výstupní hodnota dat sekvenací SOLiD 160 Gb, zatímco u Sangerovy metody pouze 2,1 Mb (Kchouk et al. 2017).

V Tabulce 1 můžeme vidět srovnání metod první, druhé i třetí generace. Vidíme, že se liší technikou sekvenace, velikostí fragmentů i cenou, která je potřeba na jeden sekvenační cyklus. Na rozdíl od sekvenačních technik třetí generace jsou ovšem všechny velmi spolehlivé, a proto se využívají ke kontrolnímu čtení.

Tabulka 1: Srovnání charakteristik přístrojů

Přístroj	Počet čtení/běh	Délka vzorků (bp)	Přesnost %	Množství dat/čtení (Gb)	Rok výroby
Sanger (3730 xl)	96	400-900	99,97	0,00069 - 0,0021	2002
454 (FLX)	400	250	99,9	0,1	2007
Illumina (HiSeq)	5 Bilionů	150	99,99	1,5 Tb	2012
SOLiD (5500 W)	3 Biliony	75	99,99	160	2011
Ion Torrent (Ion Proton)	60-80 Milionů	200	99	10	2012
PacBio (RS II P5 C3)	528	8500	87	0,5-1	2014
ON (MinION Mk)	100	9545	88	1,5	2015

Zdroj: Kchouk et al. 2017

Hlavním důvodem pro nahrazení metod druhé generace těmi ze třetí generace je délka čtených úseků. Délka fragmentů pro sekvenování pomocí technologie od Illuminy může mít délku kolem 150 až 300 bp. Nicméně metoda PacBio pracuje s fragmenty o velikosti až 15 000 bp. Největší potíž u technik druhé generace představuje finální skládání sekvenovaných úseků, které je technologicky náročné a přispívá k větší chybovosti (Mardis 2013).

Sekvenace třetí generace je ze všech doposud zmíněných metod nejefektivnější z hlediska ušetření lidského faktoru práce a objemu výchozích dat. Sekvenujeme přímo původní vlákno DNA a není tedy potřeba PCR. Příprava vzorků trvá pouze 4-6 hodin. Náklady na sekvenaci jedné nukleotidové báze se zmenšují díky jednodušší přípravě vzorků, kratšímu času sekvenace i díky jednoduššímu procesu skládání nasekvenovaných úseků do sekvence kompletního genomu (Kchouk et al. 2017). Vývoj ceny sekvenování při využití různých technik je zanesen do Tabulky 2.

Tabulka 2: Vývoj ceny sekvenování v závislosti na charakteristikách přístrojů

	Sanger 3730 xl	454 GS FLX	Illumina HiSeq 2000	PacBio RS II:P6-C4	ON MinION
Délka vzorků	400-900 bp	700 bp	150 bp	15 000 bp	5 000 bp
Objem dat/čtení	1,9-8,4 Kb	0,7 Gb	600 Gb	0,5-1 Gb	1,5 Gb
Čas/čtení	20 min - 3 hod	24 hod	3-10 dní	0,5 – 4 hod	50 hod
Cena/milion bází	24 000 USD	10 USD	0,07 USD	0,4-0,8 USD	6,44-17,90 USD

Zdroj: Liu et al. 2012; Rhoads & Au 2015

Samotná sekvenace zabere řádově hodiny, což je oproti Illumině, u které může čtení trvat až deset dní, velká výhoda. Třetí generace sekvenování zatím ovšem není ve větší míře používána, jelikož se vyznačuje poměrně velkou chybovostí ve čtení dat. U PacBio může být až 13% chybovost, jak lze vidět v Tabulce 2 (Kchouk et al. 2017).

Kvůli vysoké kapacitě přístrojů druhé i třetí generace se tyto metody využívají hlavně v laboratořích, které studují velký objem dat. A to k sekvenování celých genomů, detekci strukturálních mutací chromozomů, sekvenování populací virů nebo nádorových buněk i epigenetickým studiím (Schadt et al. 2010).

Dnes již mluvíme také o čtvrté generaci sekvenování. Tyto metody se zatím pohybují hlavně v experimentální a teoretické rovině. Jak již bylo zmíněno, základ pro vývoj čtvrté generace tvoří technologie nanopóru. Hlavní zaměření je na zlepšování zobrazovací techniky a zvyšování objemu sekvenovaných dat (McGinn & Gut 2012).

V laboratořích jsou dnes nejčastěji užívány přístroje Illuminy, jak je zřejmé i z výzkumů popisovaných výše. Pyrosekvenace i metoda SOLiD a Ion Torrent se příliš nevyužívají. Metody PacBio i Oxford Nanopore jsou v rozmachu. Nedá se určit, která z metod je nejlepší a záleží tedy na laboroři, co od pořízení sekvenátoru očekává (Glenn 2011). Jednotlivé laboratoře také mohou využít možnosti zaslat vzorky do firem, které přístroje vlastní a zakázkově osekvenují zaslané vzorky za přijatelnou cenu (Brown & Brown 2011).

## 4 Závěr

Na základě informací z dostupné odborné literatury byly v bakalářské práci popsány principy sekvenačních technologií. Charakteristiky jednotlivých metod zahrnovaly přípravu genetických knihoven, způsoby sekvenace genetické informace i její detekce a následného zpracování dat. U každé technologie byla popsána specifika použití, se kterými souvisí chyby při čtení sekvencí DNA.

První dostupná metoda sekvenování byla popsána v 70. letech minulého století. Po nástupu nového tisíciletí začaly být využívány první metody masivního paralelního sekvenování. Společně se stálým zlepšováním technologií také klesala cena sekvenování. Naopak stoupalo množství vzorků a zjednodušila se jejich příprava. Byla také zmíněna možnost využití kapesního sekvenátoru, který umožňuje přečtení genetické informace přímo v terénu.

Byly popsány možnosti resekvenace i kontrolního čtení, které zvyšují kvalitu dat. Ačkoliv nebyla žádná z metod zhodnocena jako nejlepší, v praxi se zatím nejvíce používají sekvenátory firmy Illumina.

V rámci zoologických studií byly popsány příklady použití sekvenačních metod v praxi. Metagenomické studie přispěly k identifikaci virů snižujících užitkovost hospodářských zvířat. Ve studii regrese rakovinného bujení u d'áblů medvědovitých byl nastíněn způsob identifikace kauzálních genů. V rámci studie mangabejů kouřových byl popsán přínos genetických studií zvířat pro humánní medicínu. Na výzkumu orangutanů bylo přiblíženo, jak přispívá čtení genomu k fylogenetickým studiím.

V bakalářské práci bylo popsáno, jak rychle se posouvá vývoj sekvenačních metod, a jejich význam pro biologické obory.

## 5 Literatura

### 5.1 Knižní zdroje

- Ahmadian A, Ehn M, Hober S. 2006. Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta* **363**:83–94.
- Bai JW, Wang DQ, Nam SW, Peng HB, Bruce R, Gignac L. 2014. Fabrication of sub-20 nm nanopore arrays in membranes with embedded metal electrodes at wafer scales. *Nanoscale* **6**:8900–8906.
- Bräutigam A, Gowik U. 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology* **12**:831–841.
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* **4**:265–270.
- Coupland P, Chandra T, Quail M, Reik W, Swerdlow H. 2012. Direct sequencing of small genomes on the pacific biosciences RS without library preparation. *BioTechniques* **53**:365–372.
- Cox-Foster DL. 2007. A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. *Science* **318**:283-287.
- da Silva JM, Giachetto PF, da Silva LO, Cintra LC, Paiva SR, Yamagishi MEB, Caetano AR. 2016. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics* **17**:454.
- Dunham R et al. 2006. The AIDS resistance of naturally SIV-infected sooty mangabeys is independent of cellular immunity to the virus. *Blood* **108**:209–217.
- Eid J et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**:133–138.
- Esposito A, Colantuono Ch, Ruggieri V, Chiusano ML. 2016. Bioinformatics for agriculture in the Next-Generation sequencing era. *Chemical and Biological Technologies in Agriculture* **3**:1-12.
- Feng Y, Zhang Y, Ying C, Wang D, Du CH. 2015. Nanopore-based Fourth-generation DNA Sequencing Technology. *Genomics Proteomics Bioinformatics* **13**:4-16.
- Gann A, Witkowski JA. 2013. DNA: Archives reveal Nobel nominations. *Nature* **496**:434.
- Garibian L, Avashia N. 2013. Research Techniques Made Simple: Polymerase Chain

- Reaction (PCR), *The Journal of investigative dermatology* **133**:1-4.
- Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Molecular ecology resources* **11**:759-769.
- Goodwin S, Mcpherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies: *Nature Reviews Genetics* **6**:333-351.
- Hou Ch, Rivkin H, Slabezki Y, Chejanovsky N. 2014. Dynamics of the Presence of Israeli Acute Paralysis Virus in Honey Bee Colonies with Colony Collapse Disorder. *Viruses* **6**:2012–2027.
- Huang YF, Chen SC, Chiang YS, Chen TH, Chiu KP. 2012. Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC Systems Biology* DOI: 10.1186/1752-0509-6-S2-S10.
- Karger BL, Guttman A. 2009. DNA Sequencing by Capillary Electrophoresis. *Electrophoresis* **75**:495-549.
- Kchouk M, Gibrat JF, Elloumi M. 2017. Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine* DOI: 10.4172/0974-8369.1000395.
- Laver T, Harrison J, O’Neill PA, Moore K, Farbos A, Pasziewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification* **3**:1-8.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of Next-Generation Sequencing System. *Journal of Biomedicine and Biotechnology* **9**:395-403.
- Locke DP et al. 2011. Comparative and demographic analysis of orangutan genomes. *Nature* **469**:529-533.
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. 2000. *Molecular Cell Biology* - 4th edition. W. H. Freeman and Company. New York.
- Mardis ER. 2008 a. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics* **9**:387-402.
- Mardis ER. 2008 b. The impact of next-generation sequencing technology on genetics. *The Impact of Next-Generation Sequencing Technology on Genetics* **13**:35–37.
- Mardis ER. 2013. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry* **6**:287–303.
- Margres MJ et al. 2018. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisi*) infected with a transmissible cancer. *Molecular Ecology* **27**:4189-4199.
- Margulies M et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.

- McGinn S, Gut IG. 2012. DNA sequencing - spanning the generations. *New Biotechnology* **30**:366–372.
- Metzker ML. 2010. Sequencing technologies the next generation. *Nature Reviews Genetics* **11**:31–46.
- Milush JM et al. 2011. Lack of clinical AIDS in SIV-infected sooty mangabeys with significant CD4+ T cell loss is associated with double-negative T cells. *JCI* **121**:1102-1110.
- Nater A et al. 2017. Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current Biology* **27**:3487-3498.
- Nayarisseri A, Yadav M, Bhatia M, Pandey A, Elkunchwar A, Paul N, Sharma D, Kumar G. 2013. Impact of Next-Generation Whole-Exome sequencing in molecular diagnostics. *Drug Invention Today* **5**:327-334.
- Palesch D et al. 2018. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature* **553**:77-81.
- Paton DJ, Gubbins S, King DP. 2018. Understanding the transmission of foot and mouth disease virus at different scales. *Current Opinion in Virology* **28**:85-91.
- Peel E, Belov K. 2018. Lessons learnt from the Tasmanian devil facial tumour regarding immune function in cancer. *Mammalian Genome* **29**:731-738.
- Pettersson E, Lundeberg J, Ahmadian A. 2009. Generations of sequencing technologies. *Genomics* **93**:105-111.
- Pop M, Salzberg SL. 2008. Bioinformatics Challenges of New Sequencing Technology. *Trends in Genetics* **24**:142-149.
- Pye RJ et al. 2015. A second transmissible cancer in Tasmanian devils. *PNAS* **113**:374-379.
- Ramirez-Carvajal L, Pauszek SJ, Ahmed Z, Farooq U, Naeem K, Shabman RS, Stockwell TB, Rodriguez LL. 2018. Genetic stability of foot-and-mouth disease virus during long-term infections in natural hosts. *PLoS One* (e 0190977) DOI: 10.1371/journal.pone.0190977.
- Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics, proteomics & bioinformatics* **13**:278–289.
- Rosse IC et al. 2017. Whole genome sequencing of Guzerá cattle reveals genetic variants in candidate genes for production, disease resistance, and heat tolerance. *Mammalian Genome* **28**:66-80.
- Rothberg JM, Leamon JH. 2008. The development and impact of 454 sequencing. *Nature Biotechnology* **10**:1117–1124.

- Rothberg JM et al. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**:348-352.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors, *PNAS* **12**:5463-5467.
- Shendure J & Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* **36**:99–103.
- Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Human Molecular Genetics* **19**:227-240.
- Tolomeo D et al. 2017. Epigenetic origin of evolutionary novel centromeres. *Scientific Reports* DOI: 10.1038/srep41980.
- van Engelsdorp D et al. 2009. Colony Collapse Disorder: A Descriptive Study. *PLoS ONE* (e 6481) DOI: 10.1371/journal.pone.0006481.
- Walker JA, Konkel MK, Ullmer Brygg, Monceaux ChP, Ryder OA, Hubley R, Smit AFA, Batzer MA. 2012. Orangutan Alu quiescence reveals possible source element: support for ancient backseat drivers. *Mobile DNA* DOI: 10.1186/1759-8753-3-8.
- Wright CF, Morelli MJ, Thébaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP. 2010. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *Journal of Virology* **85**:2266-2275.
- Wright B, Willet CE, Hamede R, Jones M, Belov K, Wade CM. 2017. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. *Scientific Reports* DOI:10.1038/s41598-017-00439-7.
- Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. 2010. The next-generation sequencing technology and application. *Protein Cell* **6**:520-536.

## 5.2 Internetové zdroje

- Brown T & Brown T. 2011. Sequencing, forensic analysis and genetic analysis. ATDBio. Southampton. Available from: <https://www.atdbio.com/nucleic-acids-book> (accessed February 2019).
- Hawkins CE, McCallum H, Mooney N, Jones M, Holdsworth M. 2008. *Sarcophilus harrisii*. The IUCN Red List of Threatened Species 2008. IUCN. Available from [www.iucnredlist.org](http://www.iucnredlist.org) (accessed January 2019).



- ITIS. 2017. *Apis mellifera* Linnaeus, 1758. Available from [www.its.gov](http://www.its.gov) (accessed September 2018).
- ITIS. 2017. *Bos indicus* Linnaeus, 1758. Available from [www.its.gov](http://www.its.gov) (accessed April 2019).
- ITIS. 2017. *Bos taurus indicus* Linnaeus, 1758. Available from [www.its.gov](http://www.its.gov) (accessed April 2019).
- ITIS. 2017. *Bubalus bubalis* Linnaeus, 1758. Available from [www.its.gov](http://www.its.gov) (accessed April 2019).
- ITIS. 2017. *Cercocebus atys* Audebert, 1797. Available from [www.its.gov](http://www.its.gov) (accessed January 2019).
- ITIS. 2017. *Macaca mulatta* Zimmermann, 1780. Available from [www.its.gov](http://www.its.gov) (accessed March 2019).
- ITIS. 2017. *Pongo abelii* Lesson, 1827. Available from [www.its.gov](http://www.its.gov) (accessed March 2019).
- ITIS. 2017. *Pongo pygmaeus* Linnaeus, 1760. Available from [www.its.gov](http://www.its.gov) (accessed March 2019).
- Nowak MG, Rianti P, Wich SA, Meijaard E, Fredriksson G. 2017. *Pongo tapanuliensis*. The IUCN Red List of Threatened Species 2017. IUCN. Available from [www.iucnredlist.org](http://www.iucnredlist.org) (accessed January 2019).
- Pacific Bioscience. 2012. Detecting DNA Base Modifications Using Single Molecule, Real-Time Sequencing. Pacific Bioscience. California. Available from [https://www.pacb.com/wp-content/uploads/2015/09/WP\\_Detecting\\_DNA\\_Base\\_Modifications\\_Using\\_SMRT\\_Sequencing.pdf](https://www.pacb.com/wp-content/uploads/2015/09/WP_Detecting_DNA_Base_Modifications_Using_SMRT_Sequencing.pdf) (accessed January 2019).
- Pacific Bioscience. 2009. Single Molecule Real Time (SMRT) DNA Sequencing. Pacific Bioscience. California. Available from <https://www.pacb.com/smrt-science/smrt-sequencing/> (accessed February 2018).
- Snipcademy. 2019. Illumina Sequencing-By-Synthesis (SBS) Technology. Snipcademy. Available from <https://binf.snipcademy.com/lessons/ngs-techniques/illumina-solexa/> (accessed March 2019).
- Spencer G. 2004. NHGRI Seeks Next Generation of Sequencing Technologies, NIH, Available from: <https://www.genome.gov/12513210/> (accessed October 2018).
- Thermo Fisher. 2006. SOLiD Data – 2 Base Encoding. Applied Biosystems. USA. Available from <https://www.thermofisher.com/cz/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing/solid-next-generation-sequencing-publications-literature.html> (accessed July 2018).