

**UNIVERZITA KARLOVA**

**1. lékařská fakulta**



## **BAKALÁŘSKÁ PRÁCE**

**Nové techniky sekvenace lidského genomu a jejich uplatnění ve studiu molekulární podstaty a diagnostiky dědičně podmíněných onemocnění**

*Next generation sequencing technologies: applications in human genome research and potential impact on a field of molecular medicine.*

**Autor:** Anna Přistoupilová  
**Vedoucí bakalářské práce:** Ing. Stanislav Kmoch, CSc.

**Praha 2008**

### **Poděkování**

Ráda bych poděkovala Ing. Stanislavu Kmochovi, CSc. za pomoc, cenné rady a připomínky při vypracování méjí bakalářské práce.

**Čestné prohlášení**

Prohlašuji, že jsem svou bakalářskou práci vypracovala samostatně, s použitím pramenů uvedených v seznamu literatury.

.....  
Anna Přistoupilová

## Souhrn

Předmětem této bakalářské práce je historie analýzy genomu, nové metody sekvenování a jejich srovnání.

Úvod se zabývá základními pojmy analýzy genomu a její historií. Jsou zde zmíněny metody izolace a manipulace DNA a popsány klasické techniky sekvenování, Maxam-Gilbertova a Sangerova metoda.

Další kapitola popisuje nové, komerčně dostupné, metody sekvenace genomu firm 454 Life Science, Applied Biosystems, Illumina a Helicos. Zabývá se jejich vzájemným porovnáním a srovnáním s klasickou Sangerovou metodou. Následuje popis dalších perspektivních metod sekvenace, které jsou teprve ve stádiu vývoje.

Následující část je věnována dědičným metabolickým poruchám a možnému využití nových metod sekvenování na tomto poli.

V závěru jsou zmíněny etické otázky týkající se masivního rozšíření metod celogenomového sekvenování.

**Klíčová slova:** analýza genomu, sekvenování, nové metody sekvenování

## Summary

The goal of this thesis is to describe the history of the genome analysis, next-generation sequencing technologies and its comparison. The thesis is divided into four parts.

In the first part, there are explained fundamental terms regarding the genome analysis and its history. This part contains a description of DNA isolation and manipulation methods, traditional sequencing methods and the Maxam-Gilbert's and Sanger's method.

In the second part, there are described next-generation technologies, that are commercially affordable. These are produced by companies 454 Life Science, Applied Biosystems, Illumina and Helicos. These methods are compared with other ones and beside that confronted with the traditional Sanger's method. This comparison is supplemented by a description of the potential perspective sequential methods, which are still in a stage of development.

The third part is aimed on a problematic of the metabolic disorders and the potential of using the sequences methods in this area.

The last part is focused on the development and spreading of the next-generation sequencing technologies from the ethic point of view.

**Keywords:** genome analysis, sequencing, next-generation sequencing

# OBSAH

<b>1 ÚVOD .....</b>	<b>8</b>
<b>2 ÚVOD DO GENETIKY .....</b>	<b>9</b>
2.1 HISTORIE .....	9
2.2 CENTRÁLNÍ DOGMA MOLEKULÁRNÍ BIOLOGIE .....	10
2.2.1 DNA .....	10
2.2.2 Replikace (DNA → DNA) .....	12
2.2.3 Transkripce (DNA → RNA) .....	13
2.2.4 RNA .....	13
2.2.5 Translace (RNA → protein) .....	14
2.3 JAKÉ INFORMACE DNA OBSAHUJE A JEJICH VÝZNAM .....	15
2.4 METODY IZOLACE A MANIPULACE DNA .....	18
2.4.1 Izolace DNA .....	18
2.4.2 Kontrola koncentrace a kvality DNA .....	19
2.4.3 Štěpení DNA .....	20
2.4.5 Spojování DNA .....	20
2.4.6 Kopírování DNA .....	21
<b>3 KLASICKÉ METODY SEKVENOVÁNÍ.....</b>	<b>23</b>
3.1 MAXAM- GILBERTOVA METODA .....	23
3.2 SANGEROVA METODA .....	23
<b>4 NOVÉ METODY SEKVENOVÁNÍ.....</b>	<b>25</b>
4.1 ROCHE (454) GS FLX SEQUENCER .....	27
4.1.1 Příprava knihovny .....	27
4.1.2 Amplifikace DNA .....	27
4.1.3 Sekvenace .....	28
4.1.4 Zpracování dat .....	29
4.2 ILLUMINA GENOME ANALYZER .....	31
4.2.1 Příprava knihovny .....	31
4.2.2 Amplifikace .....	31
4.2.3 Sekvenace .....	32
4.2.4 Zpracování dat .....	33
4.3 APPLIED BIOSYSTEMS SOLiD™ SYSTEM .....	34
4.3.1 Příprava knihovny .....	34
4.3.2 Amplifikace .....	34
4.3.3 Sekvenace .....	36
4.3.4 Zpracování dat .....	37
4.4 HÉLICOS HELISCOPE .....	39
4.4.1 Příprava knihovny .....	39
4.4.2 Amplifikace .....	39
4.4.3 Sekvenace .....	39
4.4.4 Zpracování dat .....	40
<b>5 POROVNÁNÍ.....</b>	<b>41</b>
5.1 CENA .....	41
5.2 VÝKONNOST .....	41
5.3 DÉLKA ČTENÝCH ÚSEKŮ A ÚPLNOST .....	41
5.4 PŘESNOST .....	42
5.5 DOBA CYKLU .....	42
5.6 POČET A MNOŽSTVÍ VZORKŮ .....	43
5.7 MOŽNOSTI ROZŠÍŘENÍ SEKVENÁTORU .....	43
5.8 ZPRACOVÁNÍ DAT .....	44

<b>6 CO ZVÁŽIT PŘED NÁKUPEM SEKVENÁTORU .....</b>	<b>45</b>
<b>7 DALŠÍ PERSPEKTIVNÍ METODY .....</b>	<b>46</b>
7.1 NANOPÓROVÉ SEKVENOVÁNÍ.....	46
7.2 FRET (FLUORESCENT RESONANCE ENERGY TRANSFER).....	47
<b>8 DĚDIČNÉ METABOLICKÉ PORUCHY .....</b>	<b>48</b>
<b>9 ETICKÉ PROBLÉMY .....</b>	<b>49</b>
<b>10 ZÁVĚR .....</b>	<b>51</b>
<b>11 SEZNAM POUŽITÉ LITERATURY .....</b>	<b>52</b>
<b>12 PŘÍLOHY .....</b>	<b>54</b>
PŘÍLOHA 1- POROVNÁNÍ SEKVENÁTORŮ .....	54
PŘÍLOHA 2- POROVNÁNÍ TYPŮ DAT U JEDNOTLIVÝCH FIREM .....	55

# 1 ÚVOD

Znalost genotypu má význam pro diagnostiku a výzkum geneticky podmíněných onemocnění a predikci u komplexních (polygenních) onemocnění. Sekvence DNA je základní metodou v identifikaci těchto funkčně významných variant.

Klasickou metodou sekvenování je Sangerova enzymatická metoda, která je využívána již několik desetiletí. Dodnes je laboratořemi používána k určování sekvence organismů. V poslední době se objevila řada perspektivních metod celogenomového sekvenování, které slibují vyhnout se úskalím klasické metody. Především zvýšením rychlosti sekvenování a radikálním snížením ceny.

Je jasné, že tyto metody se v budoucnu masově rozšíří a nastane období personalizované medicíny, kdy bude známá sekvence genomu každého lidského jedince. Tato znalost přinese mnoho zlepšení, ale zároveň spolu nese mnoho etických problémů.

Popis a porovnání nových metod celogenomového sekvenování a zamyšlení nad důsledky jejich rozšíření je tématem této bakalářské práce.



## 2 ÚVOD DO GENETIKY

### 2.1 Historie

Zakladatelem genetiky je brněnský opat Johann Gregor Mendel, který na základě svých pokusů s křížením zahradního hrachu popsal a vysvětlil základní zákonitosti, jimiž se řídí přenos dědičných znaků z jedné generace do generace další. Práci se získanými poznatky publikoval již v roce 1866, ale její správnost byla potvrzena až roku 1900, kdy se genetika stala samostatným vědním oborem.

Roku 1869 izoloval švýcarský chemik Friedrich Miesher molekulu, kterou pojmenoval nuclein. Jednalo se o kyselinu deoxyribonukleovou (DNA, *deoxyribonucleic acid*). Termín genetika použil poprvé britský vědec William Bateson roku 1906. O tři roky později byl použit dánským botanikem Wilhelmem Johansenem termín pro jednotku mendelovské dědičnosti- gen.[7] V roce 1919 dokázal americký genetik Thomas Hunt Morgan na pokusech s mouchou ocotníčkou (*Drosophila melanogaster*), že chromosomy jsou dlouho hledanými nositeli genetické informace. Za tento objev obdržel Nobelovu cenu.

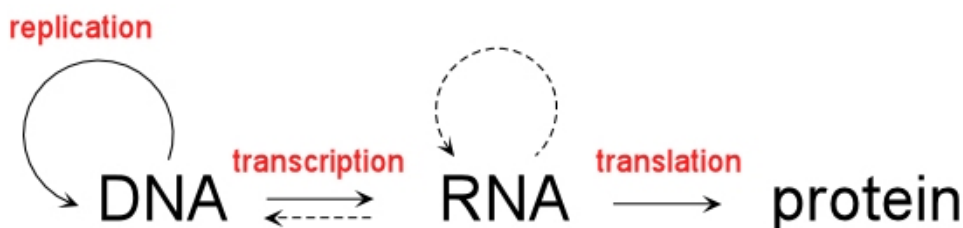
Důležitým mezníkem molekulární genetiky byl rok 1953, kdy byla objevena struktura DNA. Tohoto roku předložili vědci James Watson a Francis Crick strukturní model dvoušroubovice DNA. Velkou měrou k tomu přispěla rentgenová difrakční fotografie DNA č.51, kterou zhotovila Rosalind M. Franklinová. Za tento objev byla oběma vědcům udělena Nobelova cena. [4][20]

V roce 1955 byl určen správný počet lidských chromozomů, tedy 46. Zasloužil se o to Joe Hin Tjio. Ve stejném roce byl izolován Arthurem Kornbergem enzym DNA polymeráza, který je dodnes používán pro manipulaci DNA a její sekvenování. V roce 1966 objevil Marshall Nierenberg genetický kód, tedy systém, jakým jednotlivé báze určují sled dvaceti aminokyselin v syntéze proteinů. [7]

Klíčovým okamžikem v oblasti genetiky byl rok 2003, kdy byl ukončen Projekt lidského genomu, tedy byla přečtena kompletní sekvence lidského genomu.

## 2.2 Centrální dogma molekulární biologie

Centrální dogma molekulární biologie bylo formulované roku 1958 Francísem Crickem. Popisuje cesty přenosu dědičné informace. Umožňuje transkripci (přepis) mezi nukleovými kyselinami a translaci (překlad) kyseliny ribonukleové (RNA, *deoxyribonucleic acid*) do proteinů. (Obr.1)



Obr.1: Centrální dogma molekulární biologie

Zdroj: [http://guweb2.gonzaga.edu/faculty/cronk/CHEM360/images/central\\_dogma\\_only.png](http://guweb2.gonzaga.edu/faculty/cronk/CHEM360/images/central_dogma_only.png)

### 2.2.1 DNA

DNA je nositelkou genetické informace. Ve své struktuře kóduje a buňkám zadává jejich program a tím předurčuje vývoj a vlastnosti celého organismu.

Je uložena v každé buňce živých organismů (s výjimkou zralých erytrocytů). U eukaryotických organismů je uvnitř jádra, zatímco u prokaryotických organismů se nachází volně v cytoplazmě. U eukaryot existují další druhy DNA, uložené v semiautonomních organelách, což jsou mitochondrie a plastidy.

Jaderná DNA je uložena v jádře v podobě chromosomů. Chromozomy jsou tvořeny chromatinem, což je komplex DNA a proteinů, především histonů. Tento komplex se nazývá nukleohistonové vlákno. V průběhu jaderného dělení dochází ke kondenzaci (spiralizaci) chromosomů v důsledku mnohonásobného ohýbání, skládání a stáčení nukleohistonového vlákna. Molekula DNA se nejprve obtáčí kolem histonů, čímž vytvoří nukleozom. Nukleozom se dále spiralizuje až do vzniku útvaru nazývaného solenoid. Solenoid vytváří chromatidu, která tvoří chromosom. Každá chromatida obsahuje pouze jedinou molekulu DNA. Před započítím jaderného dělení dochází k replikaci molekuly DNA, při které vznikají dvě vlákna nesoucí zcela shodnou genetickou informaci. Následkem replikace dochází ke zdvojení chromatidy a vytvoření dvouchromatidového chromozomu. (Obr.2)

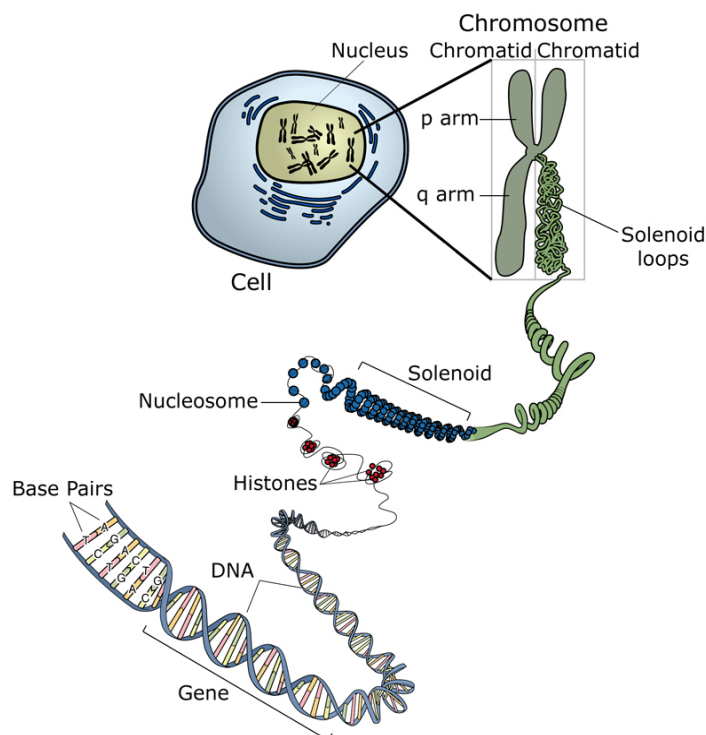


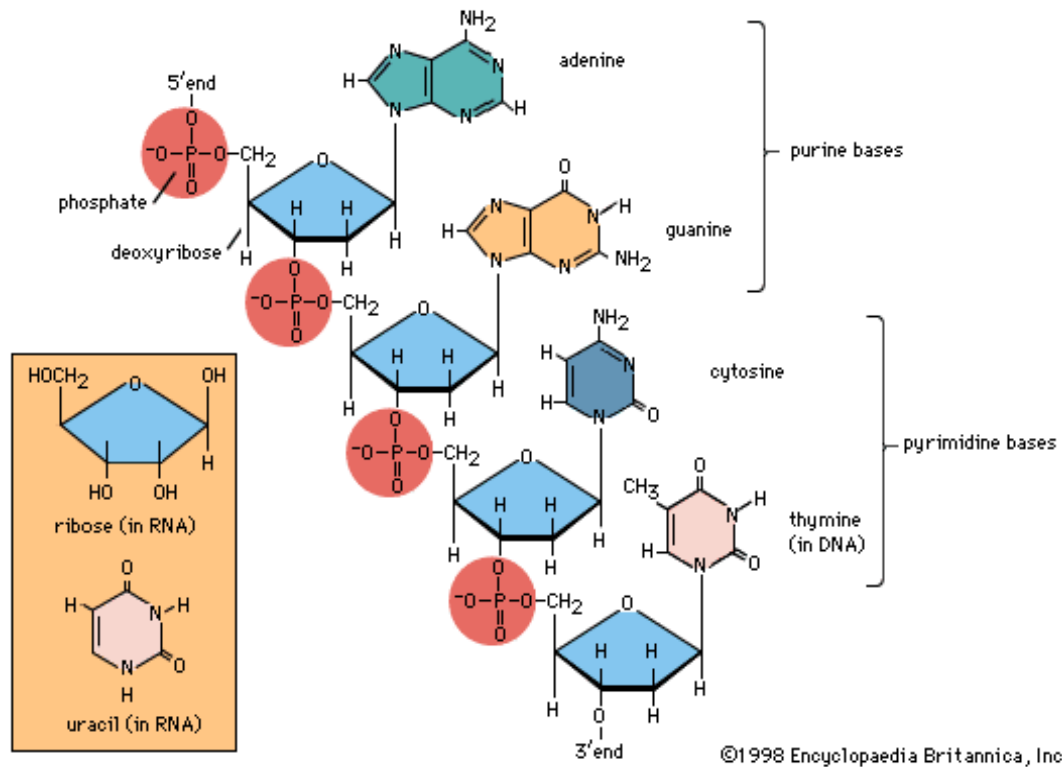
Image adapted from: National Human Genome Research Institute.

Obr.2: Struktura nukleohistonového vlákna

Zdroj: [http://images2.clinicaltools.com/images/gene/chromosomebreakdown\\_large.jpg](http://images2.clinicaltools.com/images/gene/chromosomebreakdown_large.jpg)

### Struktura DNA

Molekula DNA je tvořena dvěma opačně orientovanými polynukleotidovými řetězci stočenými do dvoušroubovice. Jeden řetězec má směr fosfodiesterových vazeb mezi nukleotidy  $5' \rightarrow 3'$  a druhý  $3' \rightarrow 5'$ . Nukleotid obsahuje tři základní stavební kameny: cukernou složku 2-deoxy-D-ribosou, fosfát a dusíkatou bázi. Báze jsou heterocyklické sloučeniny odvozené od purinu (adenin, guanin) nebo pyrimidinu (cytosin, thymin). (Obr.3) Obě vlákna DNA jsou vzájemně spojena na základě komplementarity bází. Adenin se páruje s thyminem dvěma vodíkovými můstky a cytosin s guaninem třemi vodíkovými můstky. V pořadí těchto bází je uložena genetická informace.



Obr.3: Polynukleotidový řetězec  
Zdroj: <http://www.britannica.com/eb/art-2658>

## 2.2.2 Replikace (DNA → DNA)

Replikace zajišťuje přenos genetické informace při rozmnožování buněk. Při replikaci vznikají z původní molekuly DNA dvě strukturně naprosto stejné dceřiné molekuly DNA. Jedná se o semikonzervativní proces, kdy nově vzniklá dvoušroubovice DNA má vždy jedno vlákno původní a druhé vlákno nově syntetizované. Replikace probíhá u eukaryotických buněk najedenou na více místech, která se nazývají replikony. Nejprve jsou enzymem helikázou rozvolněny vodíkové můstky, které spojují obě vlákna dvoušroubovice a poté je zahájen proces replikace enzymem zvaným DNA-dependentní-RNA-polymeráza. Poté jiný enzym, DNA polymeráza, připojuje nové nukleotidy podle komplementarity bází ve směru 5' → 3' (z hlediska nově vznikajícího vlákna).

Obdobným způsobem dochází i k replikaci RNA.

### 2.2.3 Transkripce (DNA → RNA)

Jedná se o přenos informace z DNA do mediátorové RNA (mRNA, *messenger RNA*). Transkripce je enzymatický proces, při kterém má hlavní úlohu enzym zvaný DNA-dependentní-RNA-polymeráza, který prozkoumává řetězec DNA ve směru 5' → 3' a hledá startovní sekvenci nukleotidů zvanou promotor. Nejčastěji se jedná o takzvaný TATA box nebo CCAAT box, obsahující vyšší množství daných nukleotidových sekvencí. Ve chvíli, kdy polymeráza vyhledá promotor, jsou vodíkové můstky mezi vlákny DNA rozpojeny. Dochází k syntéze vlákna RNA podle komplementarity bází v DNA, přičemž thymin je vyměněn za uracil. Syntéza pokračuje do doby, než polymeráza narazí na ukončující sekvenci bází zvanou terminátor, čímž zastaví přepis. Vzniklé vlákno mRNA prochází post-transkripčními úpravami, během nichž jsou vystříženy nekódující oblasti genu, tzv. introny. Exony, úseky, které nesou informaci o pořadí aminokyselin v polypeptidu, jsou spojeny a takto upravená mRNA putuje z jádra do cytoplazmy.

### 2.2.4 RNA

Molekula RNA se od DNA liší tím, že je až na výjimky, tvořena pouze jedním polynukleotidovým řetězcem. Nukleotid je tvořen stejnými stavebními kameny jako DNA s tím rozdílem, že cukernou složkou je v tomto případě D-ribosa a thymin je nahrazen uracilem. Přítomnost hydroxylové skupiny na druhém uhlíku D-ribózy (na rozdíl od vodíku v 2-deoxy-D-ribose) činí RNA méně stabilní než DNA. Existují dva základní typy RNA:

#### a) protein kódující RNA

- **mRNA**- přenáší informaci o pořadí aminokyselin z jaderné DNA do cytoplazmy, kde dochází k proteosyntéze.

#### b) nekódující RNA (ncRNA, *non-coding RNA*)

Existuje mnoho druhů ncRNA a jejich funkce není zcela prozkoumána. Je ale jasné, že jejich funkcí je regulace genové exprese.

- **trasferová RNA** (tRNA *transfer RNA*)- zařazuje aminokyseliny na správné místo vznikajícího proteinu
- **ribosomální RNA** (rRNA, *ribosomal RNA*)- tvoří stavební složku ribosomálních podjednotek. Vyskytuje se několik velikostně odlišných typů.

### **2.2.5 Translace (RNA → protein)**

Při translaci je genetická informace přenášena z mRNA do pořadí aminokyselin. Translace probíhá v cytoplasmě na ribosomech. Aminokyseliny jsou přiřazovány na základě genetického kódu. Jedná se o systém, ve kterém vždy trojici bází (triplet, kodon) přísluší jedna určitá aminokyselina. Genetický kód je univerzální pro všechny organismy na Zemi (s výjimkou lidských mitochondrií) a také degenerovaný, což znamená, že některé aminokyseliny jsou kódovány více tripletami. Existují triplety, které mají speciální funkce. Jedná se o iniciační triplet AUG od kterého začíná proteosyntéza (zároveň kóduje methionin) a triplety UAG, UAA, UGA, takzvané stop-kodony, které proteosyntézu ukončují.

Proteosyntéza začíná navázáním tRNA nesoucí methionin na molekulu mRNA. Na následující kodony nasedají další tRNA podle komplementarity bází (systém kodon na mRNA - antikodon na tRNA). Mezi přiřazovanými aminokyselinami vznikají peptidové vazby a vzniklé polypeptidové vlákno je v Golgiho aparátu nebo endoplazmatickém retikulu dále upravováno na požadovanou bílkovinu.

## **2.3 Jaké informace DNA obsahuje a jejich význam**

V roce 1990 započal Projekt lidského genomu<sup>1</sup>, který měl za úkol přečtení kompletní sekvence lidského genomu do roku 2005 za cenu 3 miliard dolarů. V čele tohoto veřejného konsorcia stál Francis Collins. V roce 1998 oznámil J. Craig Venter založení nové společnosti Celera Genomics s cílem osekvenovat lidský genom o několik let dříve. Následující roky byly soubojem těchto dvou společností, ukončené dohodou a společným publikováním výsledků v roce 2001. Celera publikovala výsledky své práce v časopise Science a veřejné konsorcium v časopise Nature. V roce 2003 byla dokončena fyzikální mapa lidského genomu a projekt byl ukončen.

Lidský genom se skládá z 3,2 miliard párů bází, které jsou rozděleny do 23 chromosomů. Lidé vlastní dvě sady těchto chromosomů, od každého z rodičů jednu. Ve skutečnosti tedy můžeme říci, že osobní genom jedince obsahuje šest miliard párů bází. Pouze dvě procenta této sekvence jsou transkripčně aktivní- jsou to geny kódující bílkoviny, kterých je 20- 25 tisíc [10]. Genem se nazývá úsek DNA se specifickou funkcí, který je schopen při dělení buňky utvářet svoje vlastní přesné kopie. Tyto kopie jsou přenášeny do dalších generací [27]. Geny kódují více než 300 tisíc různých molekul RNA a více než 1 milión druhů proteinů.

Zbývající část genomu nekóduje proteiny a nazývá se odpadní DNA (junk DNA). Tato část genomu není zatím dobře prozkoumána, ale již dnes je zřejmé, že určitý význam v organismu má. Obsahuje mezidruhově konzervované oblasti zanesené do genomu v průběhu evoluce a předpokládá se, že hraje důležitou roli v genové regulaci.

V sekvenci nukleotidů jsou zakódovány všechny instrukce pro vývoj a vlastnosti celého organismu. Lidské genomy se shodují v 99,9% bází.

Většina rozdílů je způsobena rozdíly pouze v jedné bázi (SNP, Single Nucleotide Polymorphisms). Tyto polymorfismy a jejich spojitost s častými onemocněními byly prozkoumány v projektu Hap Map<sup>2</sup>, který byl ukončen v roce 2006. SNP se vyskytují přibližně jedenkrát na 1000 nukleotidů, v lidském organismu jich

---

<sup>1</sup> [http://www.ornl.gov/sci/techresources/Human\\_Genome/project/about.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml)

<sup>2</sup> <http://www.hapmap.org>

jsou tedy přibližně 3 milióny. Nemají vždy přímý vliv na lidské zdraví, ale mohou ovlivnit odpověď organismu při vystavení zátěži prostředí.

Variabilita existuje také v počtu kopií jednotlivých genů (CNV, Copy Number Variations)<sup>1</sup>. V průzkumu genomů 270 lidí bylo odhaleno průměrně 70 míst CNV, což dělalo dohromady průměrně 250 000 nukleotidů. U všech zkoumaných osob bylo objeveno celkem 1447 různých míst CNV, což dohromady představuje zhruba 12% lidského genomu [16]. CNV jsou v organismu běžné, ale některé z nich mohou vést k různým postižením a nemocem.

V genomu může docházet k náhodným změnám, takzvaným mutacím. Mutace mohou být spontánní - vzniklé chybou v replikačním a reparačním mechanismu DNA, nebo indukované - vyvolané mutageny. Mutace mění genom na třech úrovních: změny genové, chromosomové a genomové. Při genových mutacích je gen měněn substitucí, inzercí, nebo delecí nukleotidů. Mutace chromosomové jsou způsobeny zlomy chromosomů a jejich ztrátami (deficiencí) či připojením k jinému chromosomu (translokace). Mutace genomové vznikají změnou počtu jednotlivých chromosomů (aneuploidie) nebo změnou počtu úplných chromosomových sad (euploidie). Vliv mutací na organismus závisí na konkrétním případě. Většina mutací je negativních, způsobujících genetické choroby a nádorové bujení, ale existují i mutace pozitivní, tedy daný organismus zvýhodňující.

Sekvenování je základní metodou detekce těchto genomových variabilit a je prvním krokem k porozumění podstatě onemocnění, které tyto variability vyvolávají.

Vzniká řada projektů mající za úkol urychlit vývoj sekvenovacích metod a prozkoumat důsledky plynoucí ze znalosti úplných genomových sekvencí.

Cílem Projektu 1000 genomů<sup>2</sup> je osekvenovat genomy tisíce dobrovolníků z celého světa a tyto informace zveřejnit ve veřejně přístupných databázích. Vznikne tak detailní mapa genetických variací, která umožní vědcům identifikaci spojitostí mezi genetickými variacemi a určitými nemocemi.

Projekt osobního genomu,<sup>3</sup> má za úkol prozkoumat možné přínosy a rizika znalosti lidských genomů. Cílem je vytvoření rozsáhlé, veřejně přístupné, databáze

---

<sup>1</sup> [http://www.sanger.ac.uk/humgen/cnv/data/cnv\\_data/display/](http://www.sanger.ac.uk/humgen/cnv/data/cnv_data/display/)

<sup>2</sup> [www.1000genomes.com](http://www.1000genomes.com)

<sup>3</sup> <http://www.personalgenomes.org>



kompletních genomových sekvencí, spolu se zdravotními údaji získanými od dobrovolníků.

Kalifornská nadace X Prize Foundation<sup>1</sup> vypsalala cenu deset miliónů dolarů pro vědecký tým, který jako první přečte genom stovky lidí během deseti dnů, s celkovými náklady nepřevyšující deset tisíc dolarů na jeden genom.

Dva nadační programy Národního Institutu Zdraví<sup>2</sup> (NIH) vyzývají vědce k přečtení lidského genomu za 100 000 dolarů do roku 2009 a za 1000 dolarů do roku 2014 [10]. Dosažení částky 1000 dolarů za genom vyžaduje 100 000x snížit cenu za osekvenovanou bázi a 10 000x zvýšit počet osekvenovaných bází za sekundu [19].

Genom za 1000 dolarů je meta, která umožní vznik personalizované medicíny a tím pádem možnost individuální a maximálně účinné léčby, zvolené na základě znalosti genetické informace jedince.

---

<sup>1</sup> <http://genomics.xprize.org>

<sup>2</sup> <http://www.nih.gov/>

## **2.4 Metody izolace a manipulace DNA**

DNA je možné izolovat z jakéhokoliv biologického materiálu. Nejčastěji se používá krev a sliny (bukální stěr). Izolovanou DNA můžeme štěpit, spojovat, zkracovat, prodlužovat, kopírovat do molekuly RNA či do nové molekuly DNA. Také můžeme modifikovat DNA přidáním či odebráním specifických chemických skupin. Tyto postupy nejčastěji vycházejí z biologických buněčných dějů a většinou jsou zajišťovány enzymy.

### **2.4.1 Izolace DNA**

Prvním krokem izolace DNA je narušení buněčné stěny a buněčné membrány. Buněčná stěna je zeslabena pomocí lysozomu, který rozpouští polymerní sloučeniny zpevňující buněčnou stěnu, nebo častěji pomocí ethylendiamin triacetátu (EDTA), který odstraňuje ionty hořčíku nutné k ochraně struktury buněčných obalů a inhibuje buněčné enzymy štěpící DNA. Obvykle se ke směsi přidává detergent, který odstraňuje molekuly lipidů, čímž způsobuje rozpuštění buněčných membrán. (např. SDS- *deodecyl sulfát sodný*). Vzniklá směs obsahuje nerozpuštěné buněčné zbytky, které jsou odstraněny centrifugací. Vzniklý buněčný extrakt obsahuje kromě DNA velké množství proteinů a RNA. Existuje řada metod, kterými lze z této směsi izolovat DNA.

Nejčastějším způsobem je fenol-chloroformová extrakce. Do směsi je přidán fenol s chloroformem a izoamylalkoholem. Chloroform je organické rozpouštědlo, které se nemísí s vodným roztokem buněčného extraktu, takže je směs rozdělena na dvě fáze - horní vodnou a dolní chloroformovou. Protřepáváním dochází k mísení fází a srážení proteinů působením fenolu. Izoamylalkohol zvyšuje rozpustnost fenolu v chloroformu, takže po skončení protřepávání projde fenol do chloroformové fáze. Centrifugací jsou odděleny jednotlivé vrstvy vzniklého roztoku. Na rozhraní mezi fázemi se objeví bílý prstenec koagulovaných proteinů. Vrchní vodný roztok nukleových kyselin je pipetou přenesen do čisté zkumavky. Pro dokonalé odstranění proteinů je nutno proces opakovat, dokud se nepřestane mezi fázemi objevovat bílá proteinová sraženina. Není vhodné provádět velké množství fenolových extrakcí protože při tomto postupu dochází k degradaci určitého množství DNA. Řešením může být přidání enzymu proteinázy (např. pronáza nebo proteináza K), který rozkládá polypeptidy na menší jednotky, které

lze mnohem snáze odstranit. Nakonec je nutné vzorek znovu extrahovat směsí chloroformu s izoamylalkoholem v poměru 24:1, aby došlo k odstranění fenolu.

Následuje zahuštění vzorku DNA přidáním etanolu, který způsobí vysrážení nukleových kyselin. Po centrifugaci vznikne na dně zkumavky bílá sraženina obsahující DNA i RNA. Po odsání ethanolu, je sraženina vysušena a rozpuštěna v pufru. V případě, že je potřeba získat čistou DNA, je nutné odstranit RNA přidáním enzymu ribonukleázy (RNAzy), který RNA rychle degraduje na ribonukleotidové podjednotky [3] [28].

## 2.4.2 Kontrola koncentrace a kvality DNA

Pro další manipulaci s nukleovými kyselinami je nutné znát jejich množství v roztoku, kvůli správnému dávkování látek přidávaných do reakcí. Koncentraci DNA lze změřit několika způsoby.

### Absorbční spektrofotometrie

Jedná se o nejčastější a nejjednodušší metodu. Malé množství vzorku je vloženo do přístroje (např. NanoDrop). Vzorek je prosvícen ultrafialovým zářením. Měří se množství absorbovaného záření, které je přímo úměrné množství DNA ve vzorku. Absorbance je obvykle měřena při 260nm a při této vlnové délce absorbance o hodnotě 1,0 odpovídá 50 µg dvouvláknové DNA (dsDNA, *double-stranded DNA*) na mililitr.

Tato metoda je také používána ke kontrole čistoty preparátu. Využívá faktu, že intenzita absorpce UV-záření nukleovými kyselinami a bílkovinami je odlišná při různých vlnových délkách. V čistém vzorku DNA při 260nm a 280nm je poměr absorbancí 1,8. Pokud je nižší, znamená to, že je vzorek kontaminován proteiny [3].

### Fluorimetrie

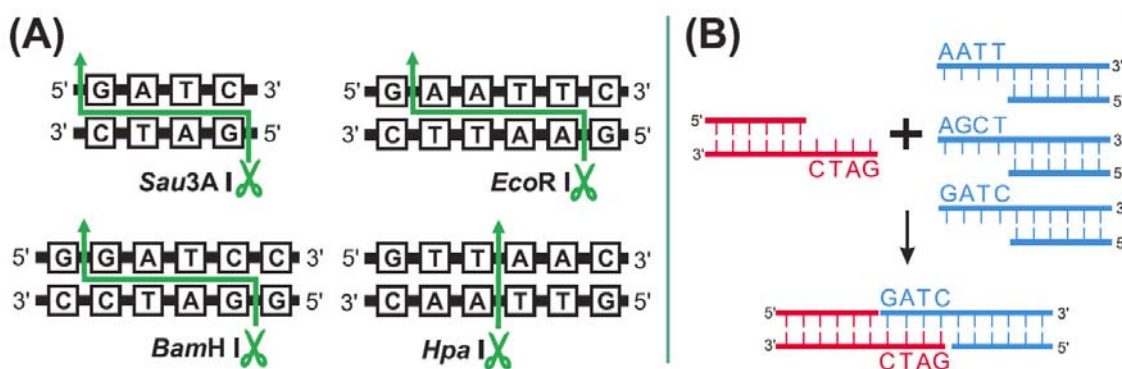
Tento způsob je založen na obarvení vzorku fluorescenčním barvivem. Ve fluorimetru lze pak změřit specifickou fluorescenci barviva. Intenzita fluorescence odpovídá množství navázaného fluorescenčního barviva. Toto množství odpovídá počtu molekul, na které se barvivo váže. Fluorimetrie umožňuje změřit správně koncentraci nukleových kyselin i ve směsi s proteiny, na druhé straně nás však na přítomnost kontaminujících proteinů neupozorní [28].

### 2.4.3 Štěpení DNA

Ke štěpení DNA se používá metoda enzymatická a metody mechanické.

Enzymatická metoda je založena na použití nukleáz. Nukleázy štěpí molekuly DNA v místě fosfodiesterových vazeb. Existují dva typy nukleáz. Endonukleázy, které štěpí vazby uvnitř molekuly DNA a exonukleázy, které odstraňují nukleotidy z konců molekuly DNA.

Dalším enzymem používaným pro štěpení DNA jsou restriční endonukleázy. Restriční endonukleázy jsou bakteriální enzymy štěpící dsDNA v místech se specifickou nukleotidovou sekvencí. Tato sekvence může být různě dlouhá, nejčastěji čtyři až osm nukleotidů. Některé nukleázy štěpí molekulu za vzniku tupých (souměrných) konců a jiné za vzniku konců kohezních (na každém konci výsledných fragmentů vzniká jednovláknový přesah). (Obr.4A)



Obr.4: Štěpení dsDNA restričními endonukleázami a spojování vzniklých fragmentů

Zdroj: <http://www.otvarena-veda.cz/ov/users/Image/default/C1Kurzy/Biolog/1kotrba.pdf>

Mechanickými metodami se rozumí sonikace (rozbití DNA na fragmenty pomocí ultrazvuku) a nebulizace (rozprášení DNA na fragmenty působením stlačeného vzduchu v přístroji zvaném nebulizér). Nevýhodou mechanického štěpení je, že je nespecifické, tedy nevíme, v jakých místech byla DNA rozštěpena [3].

### 2.4.5 Spojování DNA

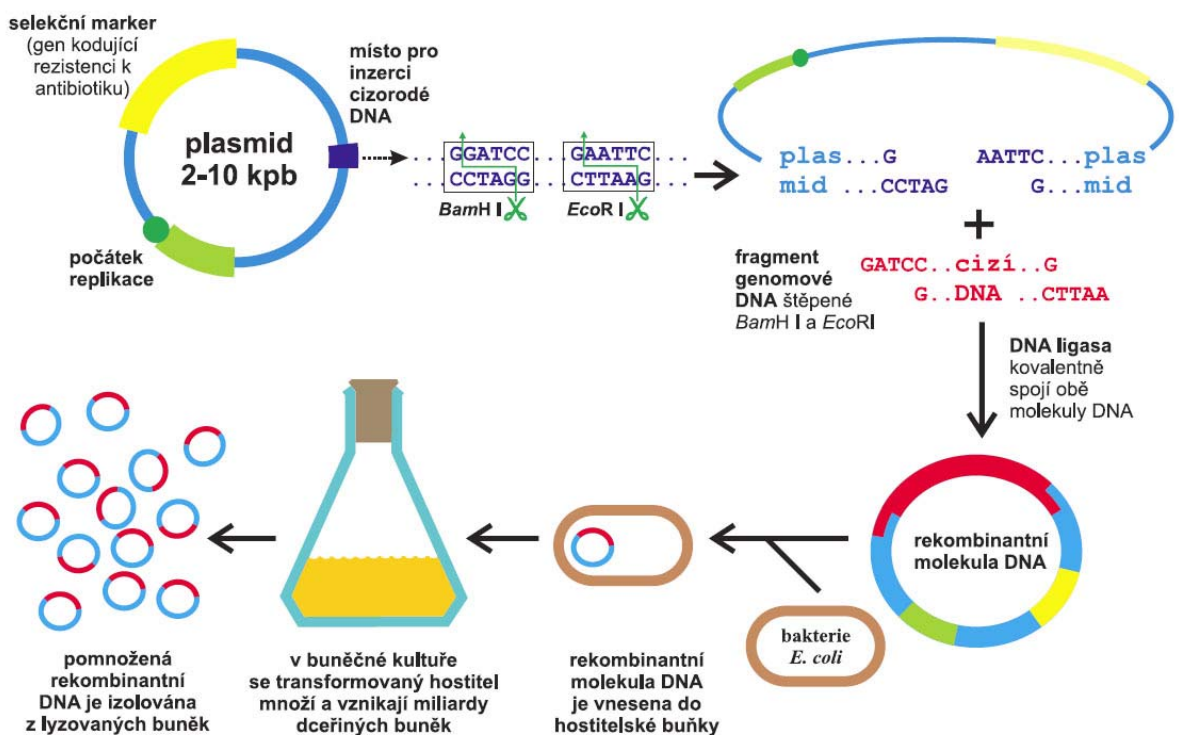
Spojování molekul zajišťuje enzymy zvané ligázy. V buňce slouží k opravování chyb vzniklých například během replikace DNA. Umožňují spojování fragmentů jednovláknových i dvouvláknových. Je možné spojovat tupé i kohezní konce. (Obr.4B)

Ligace kohezních konců je ve srovnání s ligací konců tupých mnohem efektivnější a stabilnější. Pro potřeby sekvenování je používána velice často ligace adaptérů na naštěpené fragmenty DNA. Adaptér je krátká sekvence dsDNA sloužící k výběru určitých fragmentů DNA nebo k jejich přichycení na podložku [3].

## 2.4.6 Kopírování DNA

### Klonování DNA

Izolovaná DNA je rozštěpena restriktivními endonukleázami a požadovaný úsek DNA je vložen do vektoru. Jako vektory jsou používány plazmidy, viry (bakteriofágy), a umělé chromosomy, umožňující klonovat úseky dlouhé až 3Mb (např. YAC, *yeast artificial chromosome*, nebo BAC, *bacterial artificial chromosome*). Vektor nese gen zajišťující rezistenci proti antibiotikům. Získáme rekombinantní molekulu DNA, kterou vložíme do hostitelské buňky, nejčastěji bakteriální. Bakterie jsou vloženy na živné médium obsahující antibiotikum. Dochází k pomnožení pouze bakterií, obsahujících gen zajišťující rezistenci proti antibiotikům, tedy bakterií, nesoucí požadovaný gen. (Obr.5)



Obr.5: Klonování DNA

Zdroj: <http://www.otvorena-veda.cz/ov/users/Image/default/C1Kurzy/Biolog/1kotrba.pdf>

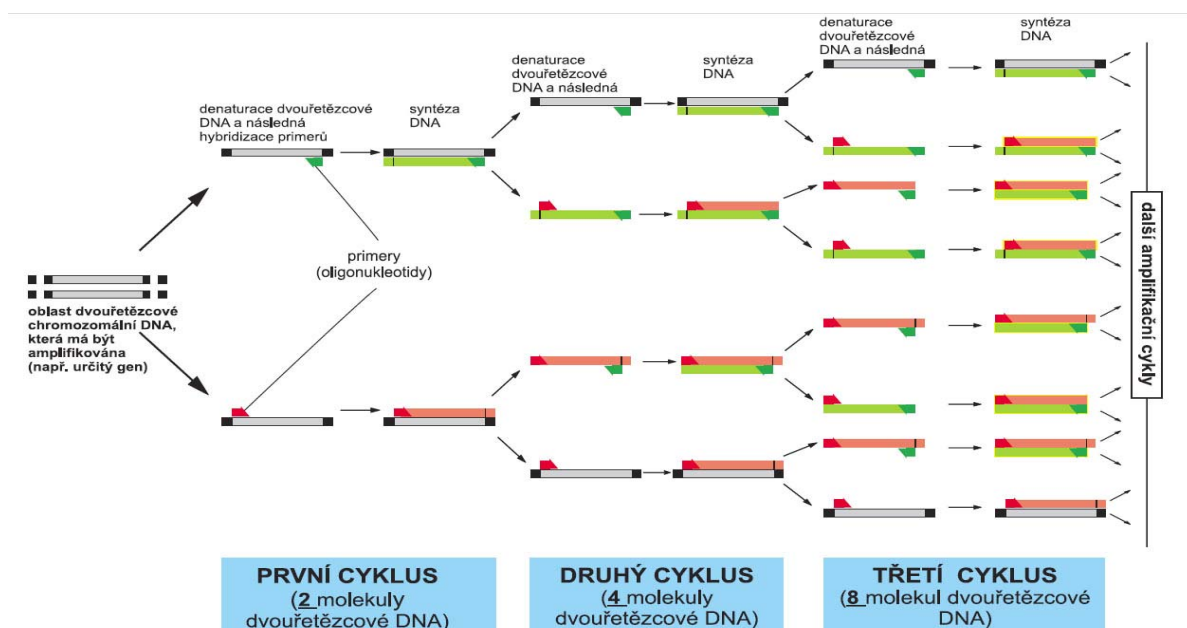
## PCR

V roce 1983 objevil Kally Muris novou metodu kopírování DNA, takzvanou polymerázovou řetězovou reakci (PCR, *polymerase chain reaction*). Tato metoda umožňuje v krátkém čase namnožit požadovaný úsek DNA bez použití buněk. Klíčovým enzymem je polymeráza, umožňující syntézu nového vlákna, komplementárního k teplátu. V případě PCR je používána termostabilní *Taq* polymeráza izolovaná z *Thermus aquaticus*, což je organismus žijící v horkých pramenech. PCR reakce sestává ze třech základních kroků, během nichž dochází k exponenciálnímu zvyšování množství DNA.

Prvním krokem je zahřátí směsi DNA, *Taq* polymerázy, množství deoxyribonukleosidtrifosfátů (dNTP) a primerů na teplotu 94- 96°C. Primery jsou krátké, synteticky vyrobené molekuly jednovláknové DNA (ssDNA, *single-stranded DNA*) komplementární k sekvenci ohraničující požadovaný úsek DNA. Teplem jsou porušeny vodíkové vazby a dojde k oddělení řetězců DNA, takzvané denaturaci.

Druhým krokem je ochlazení směsi na 50- 65°C. Ochlazení umožní navázání (hybridizaci) primerů na požadované úseky DNA podle komplementarity bází.

Posledním krokem je zahřátí směsi na teplotu 72°C. *Taq* polymeráza je připojena k jednomu konci každého primeru a dochází k syntéze nového vlákna. Tento cyklus je 30x opakován , což vede k syntéze až několika miliónů kopií [3]. (Obr.6)



Obr.6: PCR

Zdroj: <http://www.otevrena-veda.cz/ov/users/Image/default/C1Kurzy/Biolog/1kotrba.pdf>

## 3 KLASICKÉ METODY SEKVENOVÁNÍ

Sekvenováním se rozumí určení přesného sledu bází v molekule DNA. Mezníkem v sekvenaci se stal rok 1977, kdy nezávisle na sobě, byly publikovány dvě rozdílné metody sekvenace DNA.

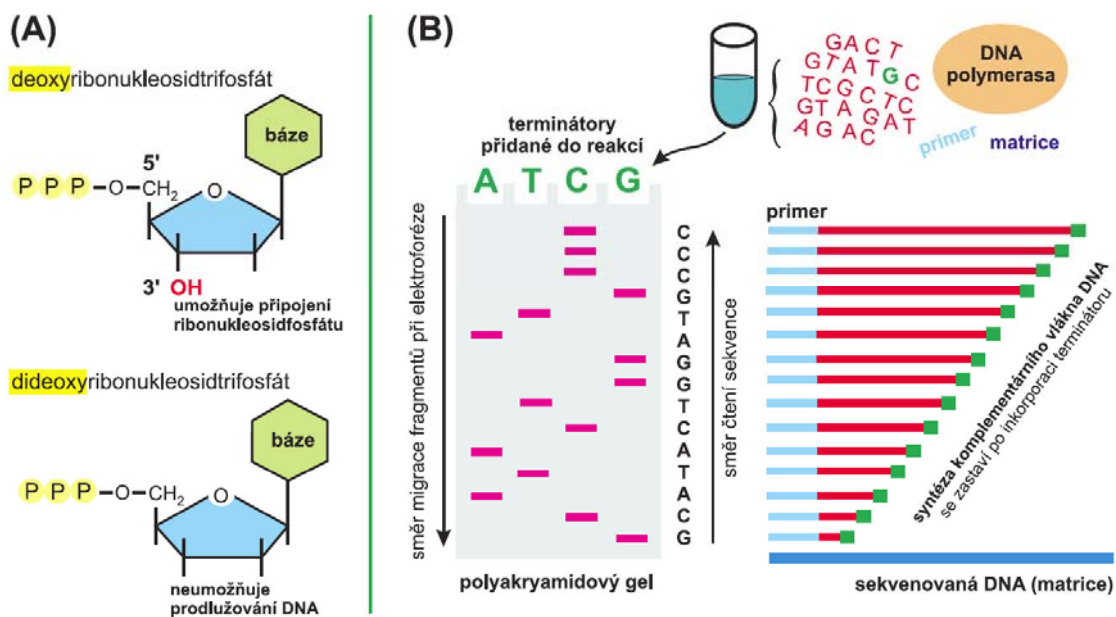
### 3.1 Maxam- Gilbertova metoda

Podstatou chemické metody sekvenování je specifické štěpení molekuly DNA chemickými činidly v místech báze určitého typu. Úsek genomu, který chceme sekvenovat, se rozdělí do čtyř zkumavek, kde jsou přítomna činidla, která jsou schopna chemicky modifikovat vždy jeden druh báze. Jiné chemické látky pak v pozměněných místech řetězec DNA rozštěpí. Směs různě dlouhých kousků DNA se rozdělí pomocí gelové elektroforézy. Delší úseky DNA se na gelu v elektrickém poli pohybují pomaleji než ty krátké. Po obarvení gelu tak získáme systém proužků, ze kterých lze vzájemným porovnáváním a kombinováním odvodit původní pořadí nukleotidů v sekvenci [31].

### 3.2 Sangerova metoda

Sangerova enzymatická metoda postupně vytlačila metodu Maxam-Gilbertovu a až do dneška je nejpoužívanější sekvenační metodou.

Vzorek DNA je denaturován a vzniklé jednovláknové úseky jsou vloženy do čtyř zkumavek. Každá zkumavka obsahuje primery, dNTP a jeden z fluorescenčně označených dideoxynukleotidtrifosfátů (ddNTP) v poměru 90:10. (Obr.7A) Po přidání DNA polymerázy do zkumavek začne syntéza nového vlákna na základě komplementarity bází. dNTP a ddNTP jsou připojovány na 3' OH skupinu ribózy. Ve chvíli, kdy je začleněn ddNTP, je ukončena syntéza řetězce, protože ddNTP obsahuje místo hydroxylové skupiny pouze  $H^+$  a následující dNTP se již nemá kam navázat. Ve zkumavce vzniknou různě dlouhé fragmenty zakončené ddNTP. Vzniklé směsi jsou podrobeny elektroforetickému vyšetření, ze kterého je odečtena sekvence vzorku. (Obr.7B)



Obr.7: A) rozdíl mezi dNTP a ddNTP B) Sekvenování DNA Sangerovou metodou  
Zdroj: <http://www.otevrena-veda.cz/ov/users/Image/default/C1Kurzy/Biolog/1kotrba.pdf>

Tato metoda byla dále vylepšována. Délka čtených úseků se zvýšila ze 100 bází na 400. Bylo to umožněno použitím gelu s až 48 oddělenými dráhami. Touto metodou bylo možno přečíst během jednoho dne 30kb dat.

Sangerova metoda je vhodná k paralelizaci a automatizaci a proto je s drobnými obměnami využívána v miniaturizované formě v automatických sekvenátorech.

V roce 1986 byla publikována zpráva o prvním automatickém sekvenátoru vyvinutém laboratoří Leroye Hooda z Caltech ve spolupráci s firmou Applied Biosystems. Tento přístroj používal upravenou Sangerovu metodu a elektroforézu probíhající v trubici. ddNTP byly označeny čtyřmi fluorescenčními barvivy. Vzorky s navázanými nukleotidy procházely postupně gelem naplněnou trubicí, srovnané podle délky fragmentů. Fluorescenční značky byly detekovány a zaznamenávány do počítače. Celý cyklus trval 13 hodin [8].

Tato metoda byla vylepšena vynálezem kapilární elektroforézy, kdy DNA fragmenty postupně procházejí místo trubicí, miniaturní kapilárou. Dnes jsou na trhu sekvenátory až s 96 kapilárami. (Applied Biosystems, ABI 3730xl DNA Analyzer).



## 4 NOVÉ METODY SEKVENOVÁNÍ

V současnosti dochází k velkému rozvoji sekvenovacích metod nové generace (NGS, *next-generation sequencing*). Některé z nich jsou ve stádiu vývoje, jiné už jsou komerčně dostupné. Ale čím se vlastně odlišují od klasické Sangerovy metody?

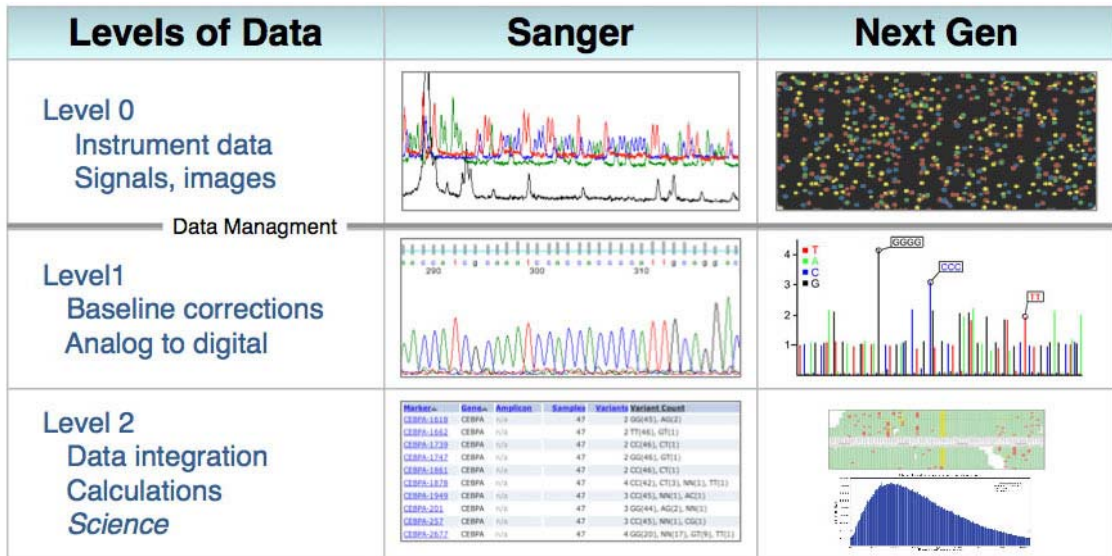
Existují dvě kategorie sekvenovacích metod nové generace. První z nich je sekvenování velkého množství kopií zkoumané sekvence a druhá je sekvenování jednomolekulové.

Jedním kritériem NGS je paralelizace, což je schopnost přístroje zpracovat milióny fragmentů najednou. Při této masivní výkonnosti bude k osekvenování genomu potřeba pouze jednoho běhu přístroje pro dokončení experimentu.

Metody nové generace nepoužívají k amplifikaci klasické klonování pomocí vektorů. Sekvence jsou vytvářeny z knihoven fragmentů. Příprava knihovny probíhá pomocí navázání speciálních adaptérů na oba konce fragmentů. Je možné sekvenování konců příslušných fragmentů DNA (*paired-end reads*), což umožňuje snadnější poskládání genomové sekvence. To je podstatné obzvláště v případě *de novo* sekvenování.

Oproti klasické metodě je potřeba menší množství vstupního materiálu. Důležitým rozdílem je čtení kratších úseků (32-300bp) oproti klasické Sangerově metodě (650-900bp) [12].

Průběh zpracování dat a jejich analýza je v případě NGS mírně odlišná od Sangerovy metody. V obou případech jsou data získávána a zpracovávána na třech úrovních (Levels). Úroveň 0 obsahuje hrubá data získaná přímo z přístroje. Přiřazením signálů k odpovídajícím bázím jsou získána hrubá sekvenační data (Level 1). Následuje analýza hrubých sekvenačních (např. přiřazení k referenční sekvenci, studium homologií apod. (Level 2) [30]. (Obr.8)



Obr.8: Level 0- 2

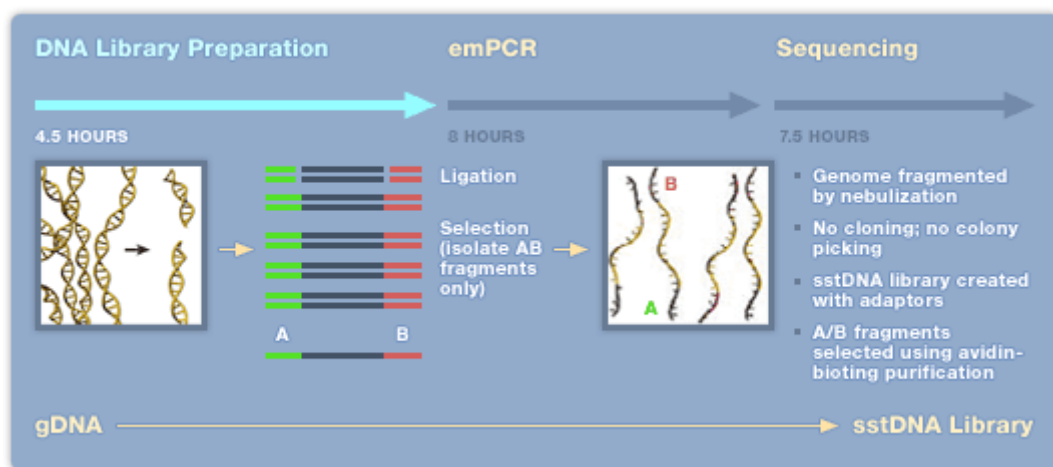
Zdroj: [http://www.geospiza.com/finchtalk/uploaded\\_images/levels-of-data-2-733545.png](http://www.geospiza.com/finchtalk/uploaded_images/levels-of-data-2-733545.png) (vlastní úprava)

## 4.1 Roche (454) GS FLX Sequencer

ROCHE FLX Sequencer byl představen v roce 2004, jako nástupce předchozího modelu, Genome Sequencer 20 System. Hlavním rozdílem je větší délka čtených úseků, (300bp oproti 100bp u předchozího modelu) a větší množství dostupných aplikací. Přístroj pracuje na principu pyrosekvenace.

### 4.1.1 Příprava knihovny

Vzorek DNA je nejprve nebulizován na fragmenty dlouhé 300- 800bp. Na konce fragmentů jsou ligovány dvouvláknové adaptéry. Jsou dvojího typu, adaptér A a B. Adaptér B se liší od adaptéru A tím, že obsahuje biotinovou značku, pomocí které jsou fragmenty přichyceny na streptavidinem obalenou magnetickou kuličku. dsDNA je denaturována hydroxydem sodným. Vzniklé jednovláknové fragmenty obsahující adaptéry A a B, tvoří DNA knihovnu. (Obr.9)



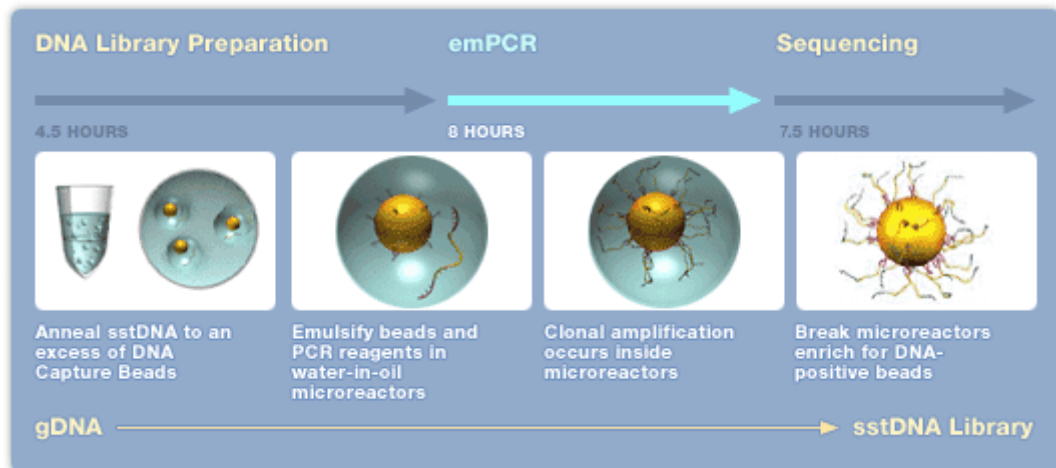
Obr.9: Příprava knihovny

Zdroj: <http://454.com/img/content/enabling-technology/process/figure7.gif>

### 4.1.2 Amplifikace DNA

Druhým krokem je PCR reakce probíhající v emulzi (emPCR, *emulsion PCR*). DNA knihovna je navázána na kuličky pomocí oligonukleotidů umístěných na jejím povrchu. Oligonukleotidy jsou komplementární k adaptérům navázaných na konce fragmentů v předchozím kroku. Díky kvantifikaci DNA knihovny je zajištěno, že na každé kuličce je navázán pouze jeden fragment. DNA kuličky jsou vloženy do zkumavky obsahující olej ve vodě a tato směs je protřepána. Okolo každé kuličky je

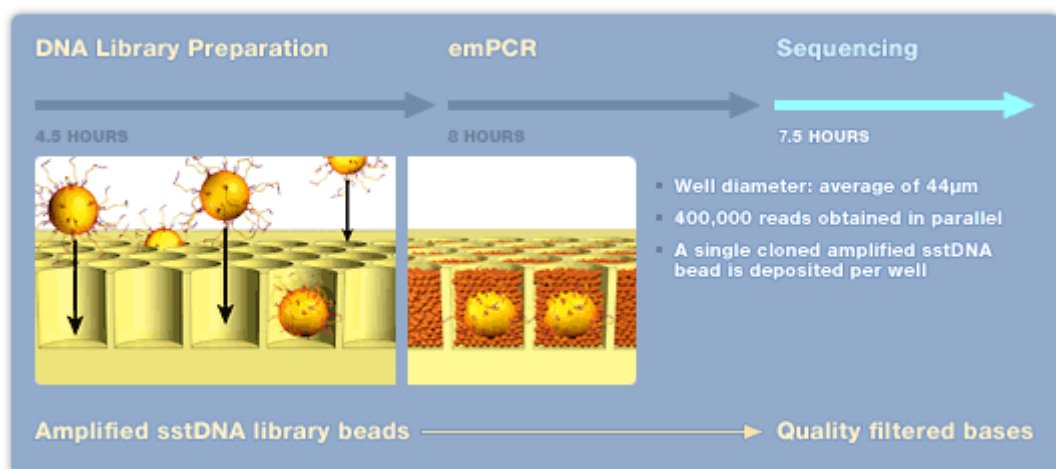
vytvořen speciální mikroreaktor, ve kterém dochází za střídání teplotních cyklů až k tisícinásobnému zmnožení DNA fragmentu. Následuje rozbití mikroreaktorů a vyjmutí DNA kuliček obsahujících klony vždy jednoho z původních fragmentů. (Obr.10)



Obr.10: Amplifikace pomocí emPCR  
Zdroj: <http://454.com/img/content/enabling-technology/process/figure8.gif>

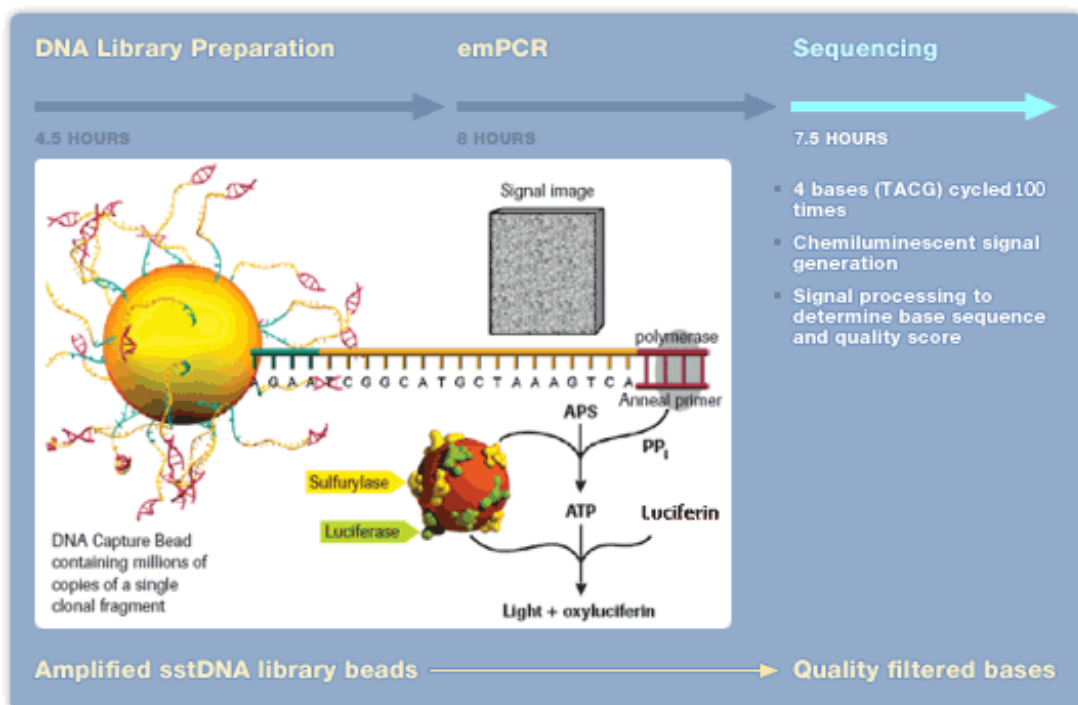
#### 4.1.3 Sekvence

DNA kuličky jsou umístěny na povrch titrační destičky (PTP- *PicoTiterPlate*), která obsahuje stovky tisíc jamek širokých 44 $\mu$ m. Průměr DNA kuličky je 26  $\mu$ m, čímž je zajištěno umístění pouze jedné DNA kuličky do jamky. DNA kulička je v jamce umístěna spolu s větším množstvím dalších dvou typů menších kuliček. První typ kuliček obsahuje enzymy potřebné pro pyrosekvenční reakci a druhý typ má za úkol DNA kuličky v jamkách imobilizovat. (Obr.11)



Obr.11: Umístění DNA knihovny do jamek titrační destičky  
Zdroj: <http://454.com/img/content/enabling-technology/process/figure9.gif>

Přes titrační destičku postupně protéká roztok obsahující DNA polymerázu a vždy jeden typ nukleotidů (A, T, C nebo G). Vždy, když je daný nukleotid komplementární k následující bázi, je polymerázou inkorporován do prodlužujícího se řetězce za uvolnění molekuly pyrofosfátu. Pyrofosfát je enzymem ATP sulforylázou přeměněn na ATP, které dodává energii luciferázové reakci, při které je luciferin oxidován enzymem luciferázou na oxyluciferin a světlo [18]. Světlo je detekováno CCD kamerou umístěnou na spodní části titrační destičky. Cyklus je opakován 100 krát pro každý nukleotid. (Obr.12)



Obr.12: Pyrosekvenační reakce

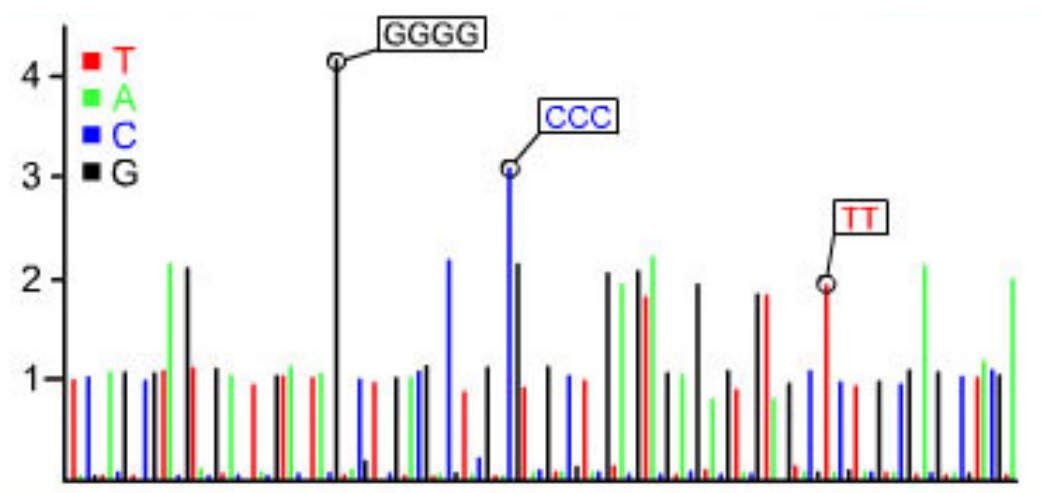
Zdroj: <http://454.com/img/content/enabling-technology/process/figure10.gif>

#### 4.1.4 Zpracování dat

Posledním krokem je zpracování dat, které probíhá částečně již během sekvenace a částečně až po jejím dokončení. Během jednoho chodu přístroje je zaznamenáno 100 Mb dat v průběhu sedmi hodin. Délka čtených úseků je 200 - 300bp a paralelně je čteno 400 tisíc vzorků.

Na začátku každého řetězce je umístěna sekvence čtyř nukleotidů, která slouží ke kalibraci systému (je součástí adaptéru). Detekovaný světelný signál je funkcí

inkorporovaných bází. V případě více stejných bází za sebou dochází k inkorporaci všech komplementárních nukleotidů a tedy i ke zvýšení intenzity detekovaného světla. Tato funkce však není lineární a z tohoto důvodu dochází k chybám v homopolymerních úsecích. Se vzrůstající délkou homopolymerních úseků se zvyšuje pravděpodobnost chyby. Sekvenci čteného úseku je možno přečíst z pořadí jednotlivých sloupců a z jejich výšky [21]. (Obr.13)



Obr.13: Level 1

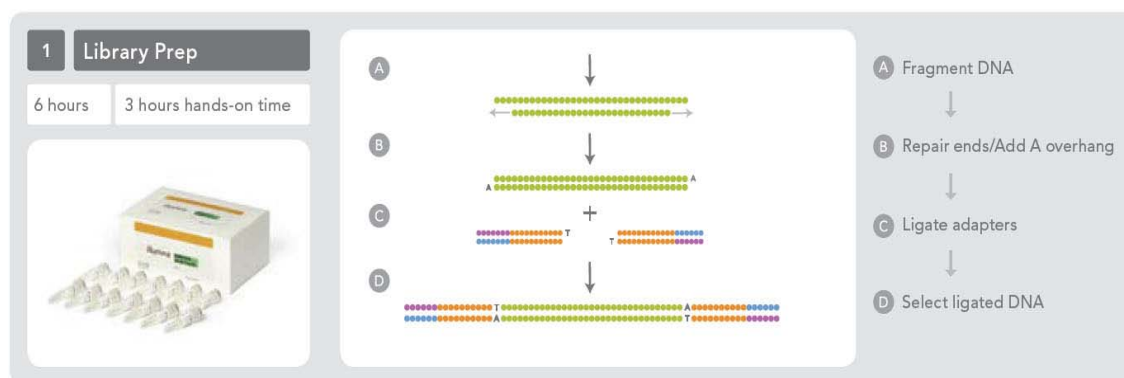
Zdroj: [https://www.roche-applied-science.com/publications/multimedia/genome\\_sequencer/flx\\_multimedia/wbt.htm](https://www.roche-applied-science.com/publications/multimedia/genome_sequencer/flx_multimedia/wbt.htm)

## 4.2 Illumina Genome Analyzer

Přístroj firmy Illumina (dříve Solexa) založený na sekvenování syntézou (SBS-sequencing by synthesis) byl představen v roce 2006. Čte úseky dlouhé 35- 50bp.

### 4.2.1 Příprava knihovny

Vzorek DNA je v přístroji zvaném nebulizér rozprášen působením stlačeného vzduchu na fragmenty dlouhé v průměru 800bp. Konce DNA jsou upraveny a na každý fragment jsou navázány dva rozdílné adaptéry. Fragmenty jsou vloženy do gelové elektroforézy a následně vyříznutím požadované části gelu extrahovány fragmenty dlouhé 150- 200bp. (Obr.14)

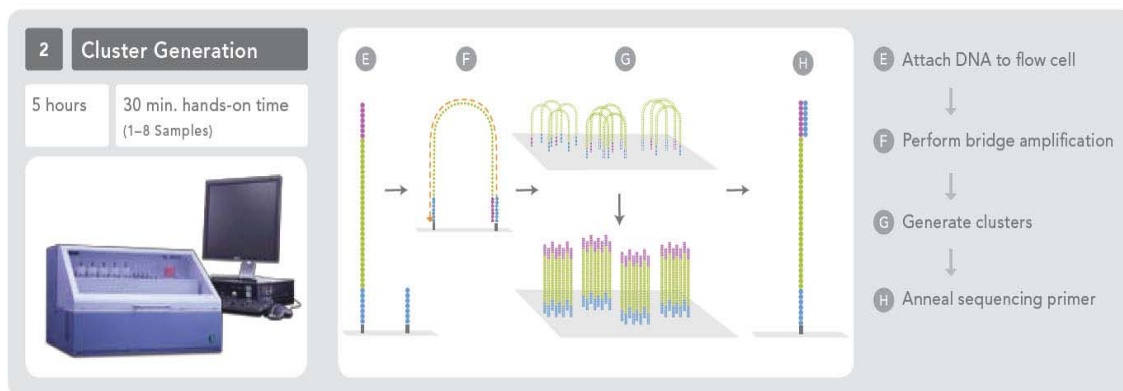


Obr.14: Příprava knihovny

Zdroj: [http://www.illumina.com/downloads/GenomeAnalyzer\\_Brochure.pdf](http://www.illumina.com/downloads/GenomeAnalyzer_Brochure.pdf)

### 4.2.2 Amplifikace

Fragmenty DNA jsou amplifikovány metodou, která se nazývá můstková amplifikace (*bridge amplification*). Probíhá na sklíčku v osmi oddělených liniích, jejichž povrch je pokryt oligonukleotidy komplementárními k oběma typům adaptérů, navázaných v průběhu přípravy knihovny. Jednovláknové fragmenty jsou hybridizovány každým koncem k jednomu z adaptérů tak, že vytvoří „můstek“. K fragmentu je dosyntetizováno druhé vlákno a poté jsou obě vzniklá vlákna oddělena denaturací. Výsledkem jsou dva identické fragmenty navázané na povrchu sklíčka. Tento proces je opakován, dokud nevzniknou tisíce kopií původního templátu umístěných ve shlucích, takzvaných clusterech. (Obr.15)

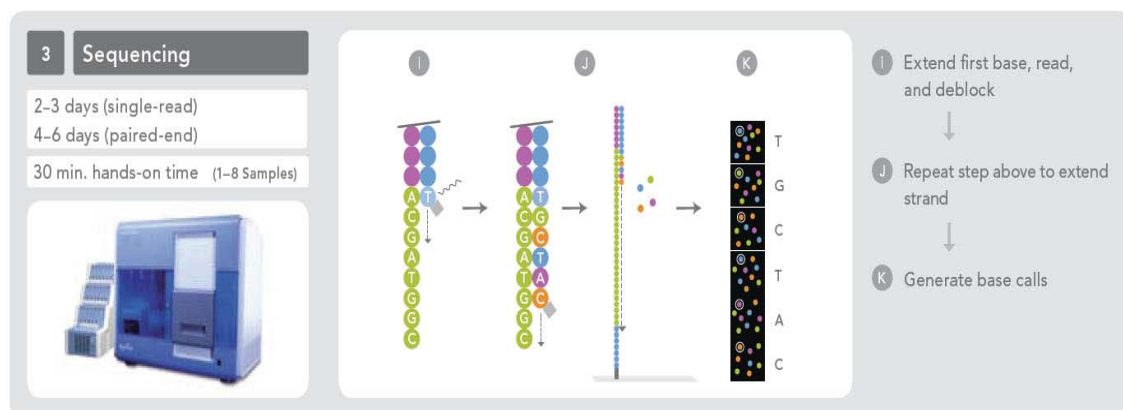


Obr.15: Můstková amplifikace

Zdroj: [http://www.illumina.com/downloads/GenomeAnalyzer\\_Brochure.pdf](http://www.illumina.com/downloads/GenomeAnalyzer_Brochure.pdf)

### 4.2.3 Sekvence

V každém cyklu jsou přidány čtyři typy odlišně fluorescenčně označených nukleotidů. Jejich 3'-OH konce jsou chemicky inaktivované, aby během cyklu nedošlo k inkorporaci více než jedné báze. Fluorescenční signál inkorporovaných nukleotidů je v jednotlivých clusterech detekován pomocí laserového paprsku. Následuje odstranění fluorescenční značky a odblokování 3' konce. Tento cyklus se opakuje [25]. (Obr.16)



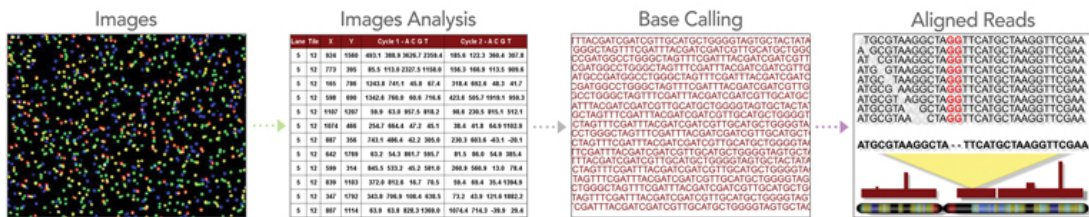
Obr.16: Sekvence

Zdroj: [http://www.illumina.com/downloads/GenomeAnalyzer\\_Brochure.pdf](http://www.illumina.com/downloads/GenomeAnalyzer_Brochure.pdf)



#### 4.2.4 Zpracování dat

Během každého cyklu dochází k začlenění pouze jedné báze. Po zpracování získaných obrázků Levelu 0 získáme data Levelu 1. (Obr.17) Z nich lze číst sekvenci jednotlivých fragmentů z pořadí inkorporovaných bází. K chybám dochází častěji na koncích fragmentů v důsledku zvyšování šumu pozadí zapříčiněném neúplným odstraněním fluorescenčních značek.



Obr.17: Fáze analýzy dat level 0-2

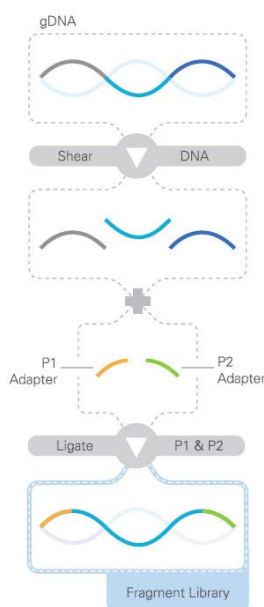
Zdroj: <http://www.illumina.com/images/pipelineWorkflowLr.jpg>

### 4.3 Applied Biosystems SOLiD™ System

Tento sekvenátor je komerčně dostupný od října 2007. Pracuje na principu ligace oligonukleotidů, o čemž vypovídá i název přístroje. (SOLiD, *Sequencing by Oligonucleotide Ligation and Detection*). Přístroj čte úseky dlouhé 35 bází.

#### 4.3.1 Příprava knihovny

Vzorek DNA je sonikován na fragmenty dlouhé 60-90bp. Na 5' konec fragmentu je ligován adaptér P1 a na 3' konec adaptér P2. Takto upravené fragmenty tvoří DNA knihovnu. (Obr.18)

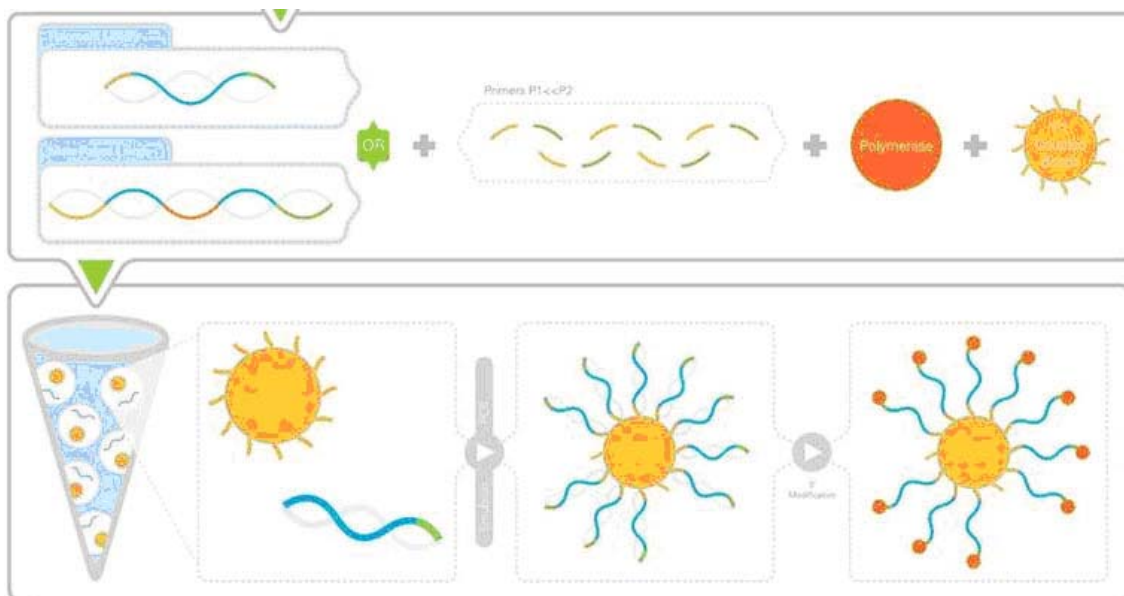


Obr.18: Příprava knihovny

Zdroj: [http://marketing.appliedbiosystems.com/images/Product\\_Microsites/Solid\\_Knowledge\\_MS/pdf/SOLiD\\_Brochure.pdf](http://marketing.appliedbiosystems.com/images/Product_Microsites/Solid_Knowledge_MS/pdf/SOLiD_Brochure.pdf)

#### 4.3.2 Amplifikace

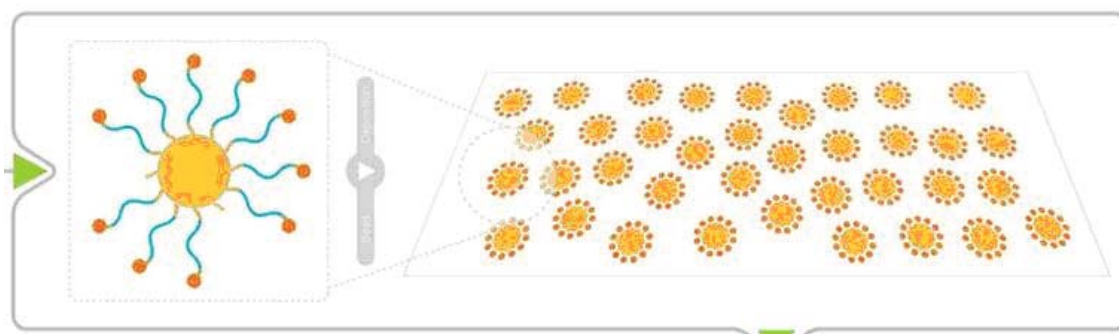
Vytvořená knihovna je amplifikovaná pomocí emPCR, stejně jako u sekvenátoru firmy Roche. DNA knihovna je vložena do zkumavky spolu s polymerázou, P1 a P2 primery a kuličkami obsahující oligonukleotidy o sekvenci komplementární k P1 adaptéru. Výsledkem jsou kuličky obsahující klony vždy jednoho z původních fragmentů. (Obr.19)



Obr.19: Amplifikace pomocí emPCR

Zdroj: [http://marketing.appliedbiosystems.com/images/Product\\_Microsites/Solid\\_Knowledge\\_MS/pdf/SOLiD\\_Brochure.pdf](http://marketing.appliedbiosystems.com/images/Product_Microsites/Solid_Knowledge_MS/pdf/SOLiD_Brochure.pdf)

3' konce fragmentů jsou modifikovány a kovalentně přichyceny na povrch sklíčka (*glass slide*), které je rozděleno až na osm segmentů. (Obr.20) V průběhu jednoho cyklu přístroje jsou zpracovávány dvě sklíčka najednou; na jednom sklíčku dochází k přijímání reagensů a na druhém dochází ke zpracování signálů.



Obr.20: Umístění DNA knihovny na povrch sklíčka

Zdroj: [http://marketing.appliedbiosystems.com/images/Product\\_Microsites/Solid\\_Knowledge\\_MS/pdf/SOLiD\\_Brochure.pdf](http://marketing.appliedbiosystems.com/images/Product_Microsites/Solid_Knowledge_MS/pdf/SOLiD_Brochure.pdf)

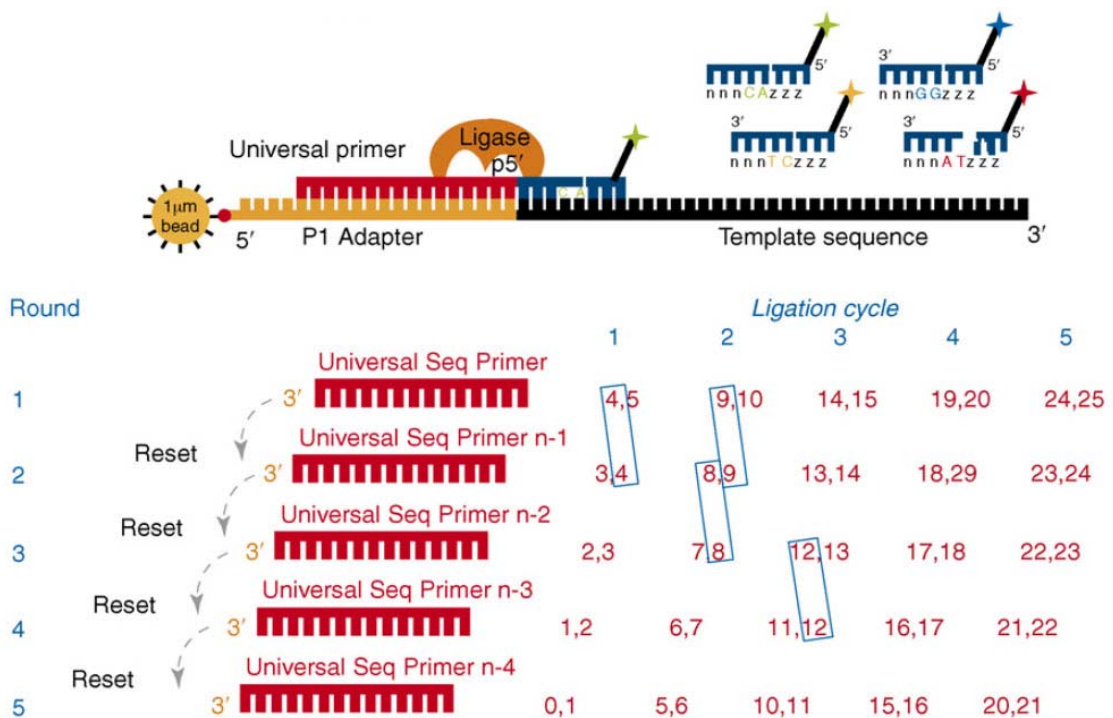
### 4.3.3 Sekvence

Na povrchu destičky (*flow cell*) je přichyceno tisíce kuliček obsahující identické kopie templátů, na jejichž koncích jsou navázány adaptéry. Na každé kuličce dochází k oddělené sekvenační reakci. Do tohoto kroku jsou si všechny NGS metody podobné. V následujících krocích používá SOLiD System velice odlišnou metodu sekvenování.

Nejprve jsou do reakce přidány primery komplementární k adaptéru navázanému v předchozích krocích.

Sekvence probíhá v pěti až sedmi cyklech, během nichž jsou přidávány fluorescenčně označené oktamerly, které jsou pomocí ligázy připojeny ke tvořícímu se řetězci. První tři pozice na 3' konci jsou degenerovány tak, aby se mohly navázat na kteroukoliv ze čtyřech bází. Na pozicích 4 a 5 (v pozdějších krocích 1 a 2) jsou dvě určité báze. Poslední tři pozice jsou universální a nesou fluorescenční značku, která charakterizuje dvě prostřední báze. V následujícím kroku dojde k odštěpení posledních třech nukleotidů spolu s fluorescenční značkou a nastává další cyklus. (Obr.21)

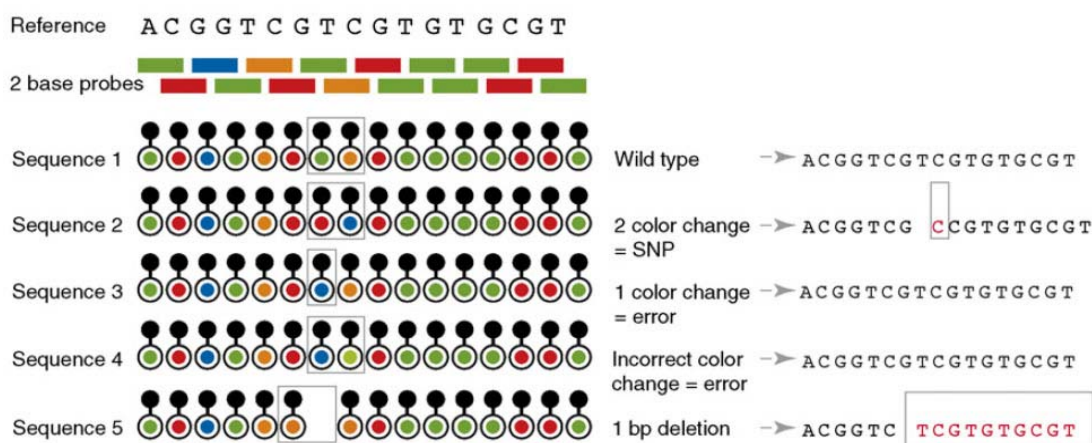
Po dokončení těchto pěti až sedmi cyklů dochází k resetování vlákna, tedy odstranění navázaných primerů i oktamerů. Dochází k navázání primeru o jeden nukleotid kratší a následuje další kolo ligace oktamerů.



Obr.21: Přehled sekvenačních cyklů

Zdroj: [http://marketing.appliedbiosystems.com/images/Product\\_Microsites/Solid\\_Knowledge\\_MS/pdf/SOLiD\\_Dibase\\_Sequencing\\_and\\_Color\\_Space\\_Analysis.pdf](http://marketing.appliedbiosystems.com/images/Product_Microsites/Solid_Knowledge_MS/pdf/SOLiD_Dibase_Sequencing_and_Color_Space_Analysis.pdf), příloha 139AP04-02

Tento systém se nazývá kódováním pomocí dvou bází. Máme 16 možných kombinací dinukleotidů a jen čtyři odlišné fluorescenční značky. Například zelenou značkou jsou určeny nukleotidy CA, AC, TG, and GT. (Obr.22) Abychom mohli identifikovat bázi, nestačí nám znát pouze jednu barvu. Musíme znát první bázi a pomocí té určíme bázi následující. Výhodou tohoto systému je, že dokážeme odlišit chybu systému (záměna jedné barvy oproti referenční sekvenci) od opravdového polymorfismu (záměna dvou barev oproti referenční sekvenci) [22].



Obr.22: Sekvenace a kódování pomocí dvou bází

Zdroj: [http://marketing.appliedbiosystems.com/images/Product\\_Microsites/Solid\\_Knowledge\\_MS/pdf/SOLiD\\_Brochure.pdf](http://marketing.appliedbiosystems.com/images/Product_Microsites/Solid_Knowledge_MS/pdf/SOLiD_Brochure.pdf), příloha 139AP04-02

#### 4.3.4 Zpracování dat

Každá báze je osekvenována dvakrát z důvodu snížení pravděpodobnosti chyby. Data Levelu 1 (Obr.23) se výrazně liší od dat předchozích platforem. Čtyři odlišné fluorescenční barvy jsou zapsány pomocí čísel 0, 1, 2, 3 [30].

```

Tue Oct 23 16:20:44 2007 /share/apps/corona/bin/filter_fasta.pl --
noduplicates --output=/data/results/AMELIA/AMELIA_20071010_1/
Yoruban_110_Frag_B/results.01/primary.m2_F3 --
name=AMELIA_20071010_1_Yoruban_110_Frag_B_ --tag=F3 --
minlength=50 --
mask=111111111111111111111111111111111111111111111111111 --
prefix=T /data/results/AMELIA/AMELIA_20071010_1/
Yoruban_110_Frag_B/jobs/postPrimerSetPrimary.m2_F3/rawseq
# Cwd: /home/pipeline
# Title: AMELIA_20071010_1_Yoruban_110_Frag_B_
>258_36_642_F3
T20222020210030010110032212323002333120323023221010
>258_36_704_F3
T01321312203302030212003223023332333120213113031012
>258_36_735_F3
T31203212213011131122001223131031321103203123233211

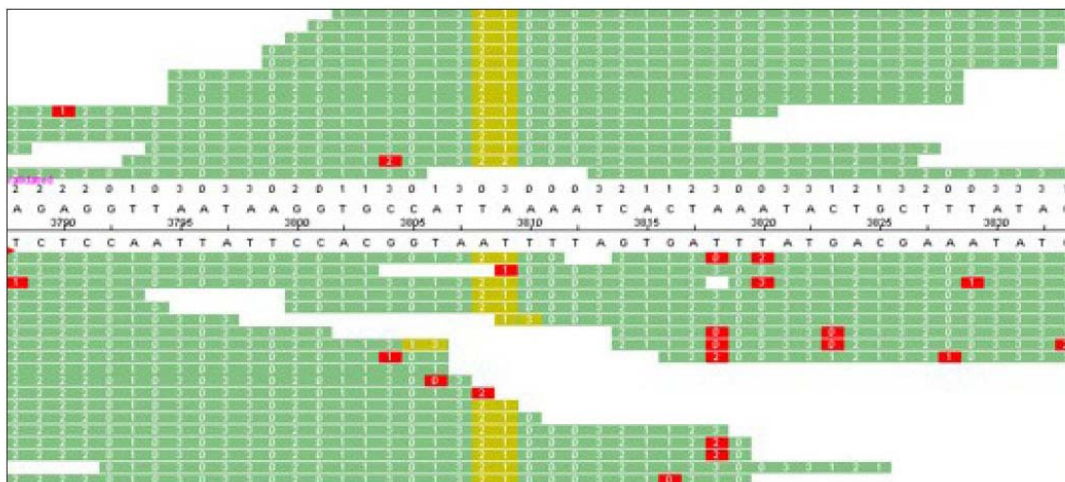
```

Obr.23: Level 1

Zdroj: [http://www.geospiza.com/finchtalk/uploaded\\_images/solid-seq-787621.png](http://www.geospiza.com/finchtalk/uploaded_images/solid-seq-787621.png)



Tento záznam je dekodován a přepsán do sekvence nukleotidů a přiřazen k referenční sekvenci (Level 2, Obr.24). K dekodování sekvence musíme znát první bázi, která pochází z adaptéru.



**Figure 3.** A snapshot of sequencing reads aligned in the SOLiD™ Alignment Browser (SAB). 0, 1, 2, and 3 represent the four colors used in 2-base encoding. Positions matching the color space reference are represented in green. Single color space mismatches are shown in red and adjacent color space mismatches are represented in yellow.

Obr.24: Level 2

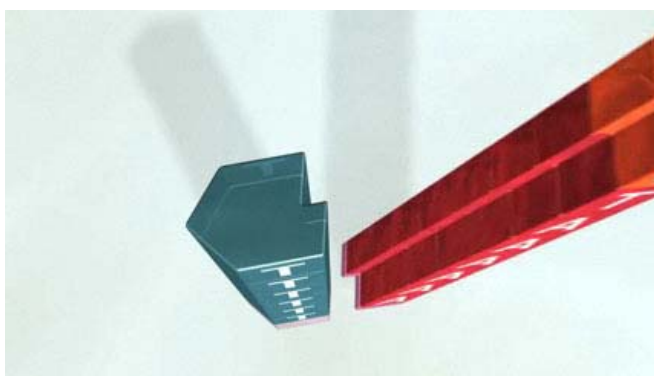
Zdroj: [http://marketing.appliedbiosystems.com/images/Product/Solid\\_Knowledge/PDF/SOLiD\\_Data\\_Management\\_AppNote\\_F2.pdf](http://marketing.appliedbiosystems.com/images/Product/Solid_Knowledge/PDF/SOLiD_Data_Management_AppNote_F2.pdf)

## 4.4 Helicos Heliscope

Nejnovější sekvenátor nové generace byl představen v únoru 2008. Pracuje na principu sekvenace syntézou a je to první přístroj, který dokáže číst jednotlivé molekuly DNA nebo RNA. To znamená, že odpadá krok amplifikace vzorků. Tento přístroj čte úseky dlouhé až 55 bází.

### 4.4.1 Příprava knihovny

Vzorek DNA je rozdělen na fragmenty dlouhé 100- 200 bází. Na 3' konec každého vlákna je navázán adaptér s fluorescenční značkou. Po přidání vzorku DNA na destičku (*flow cell*) hybridizují vlákna na oligonukleotidy umístěné na jejím povrchu, které jsou komplementární k navázaným adaptérům. (Obr.25) Díky tomu, že přístroj čte jednotlivé molekuly, jsou vlákna DNA rozložena na povrchu velice hustě. Na jednom čtverečním centimetru jich je 100 miliónu, což znamená že na celé destičce jich jsou miliardy. Sekvenovací destička je vložena do přístroje. Je možno sekvenovat dvě destičky najednou.



Obr.25: Hybridizace vláken na povrch destičky

Zdroj: <http://helicosbio.com/Technology/TrueSingleMoleculeSequencing/tabid/64/Default.aspx>

### 4.4.2 Amplifikace

Tento přístroj čte jednotlivé molekuly. Z toho důvodu zde amplifikační krok odpadá a po hybridizaci vláken na destičku následuje sekvenace.

### 4.4.3 Sekvenace

Laserem jsou detekovány fluorescenční značky, které nám po sejmutí CCD kamerou určí umístění jednotlivých vláken. Fluorescenční značka je poté odstraněna. Následuje první cyklus sekvenování. Je přidána DNA-polymeráza s jedním druhem

fluorescenčně značeného nukleotidu. Nukleotidy jsou inkorporovány do prodlužujícího se řetězce a poté je odstraněna polymeráza a všechny volné nukleotidy. Pomocí fluorescenčního signálu jsou určena a zaznamenána vlákna, na kterých došlo k začlenění daného nukleotidu. Fluorescenční značka je odstraněna a cyklus může začít znovu, postupně se všemi následujícími nukleotidy [24]. (Obr.26)

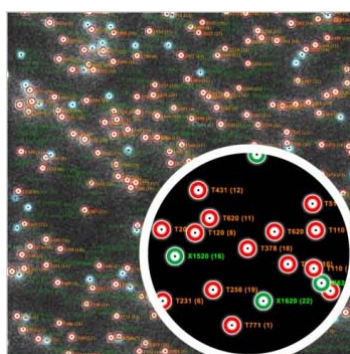


Obr.26: Sekvence

Zdroj: <http://helicosbio.com/Technology/TrueSingleMoleculeSequencing/tabid/64/Default.aspx>

#### 4.4.4 Zpracování dat

Během každého běhu přístroje je potřeba 14 terabytů volně paměti na disku. Takto velké množství paměti je potřebné z toho důvodu, že přístroj udělá 45 obrázků během jedné sekundy. Tato data Levelu 0 jsou zpracována do dat Levelu 1 a 2, která již tak velký objem nemají. (Obr.27)



Obr.27: Analýza obrazu

Zdroj: <http://helicosbio.com/Technology/DataPipeline/ImageAnalysis/tabid/81/Default.aspx>



## 5 POROVNÁNÍ

Parametry - cena, výkonnost, přesnost a úplnost spolu úzce souvisí. Snížení ceny projektu nebo zvýšení výkonnosti je často doprovázeno snížením přesnosti a úplnosti. Nové techniky sekvenování se snaží o zlepšení jedné nebo více zmíněných parametrů bez snížení těch zbývajících [1].

### 5.1 Cena

Cenu sekvenování je možné snížit paralelizací, která je umožněna miniaturizací, umožňující hustější uložení vzorků na sekvenovacím povrchu a tím pádem možnost testování mnoha vzorků naráz.

Snížení ceny je také možné docílit zredukováním množství použitých reagensů. Cena reagensů spotřebovaných během jednoho cyklu je přibližně stejná u všech tří NGS firem. Rozdílná je ale doba trvání cyklu a z toho plynoucí odlišná cena za osekvenovnou bázi [23].

Sekvenátor firmy 454 je s cenou \$84,39/Mb více než desetkrát levnější a Illumina (\$5,97/Mb) a ABI (\$5,81/Mb) jsou více než stokrát levnější, než Sangerova metoda (\$1000/Mb).

### 5.2 Výkonnost

Výkonnost sekvenátorů závisí na rychlosti detekce a stupni paralelizace [1]. Limitujícím faktorem pro metody nové generace není rychlost reakce, ale rychlost detekce produkovaných signálů [32].

Nejnižší výkonnost má klasická Sangerova metoda (0,002Mb/min), následuje sekvenátor firmy 454 (0,222Mb/min), Illumina (0,417Mb/min), ABI (0,463Mb/min) a jednomolekulový přístroj firmy Helicos s nejvyšší výkonností 1,389 Mb/min.

### 5.3 Délka čtených úseků a úplnost

Čím kratší jsou čtené úseky, tím těžší je jejich zpětné seskládání. Tento problém nastává hlavně v případě de novo sekvenování, kdy nemůžeme fragmenty přiřazovat k referenční sekvenci. Tento problém je řešen paired-end sekvenováním, kdy je mezi

jednotlivé fragmenty vložen adaptér o známé délce a tento složený fragment je sekvenován z obou stran. Pro de novo sekvenování stále zůstává dobrou volbou Sangerova metoda, kterou je možno přechíst úseky dlouhé až 900 bází. Z nových metod je nejvhodnější sekvenátor firmy 454, který dokáže číst fragmenty dlouhé až 300 bází a pro letošní rok 2008 je plánováno zvýšení délky čtených úseků na 400 bází.

Další komplikací je identifikace repetitivních sekvencí delších, než jsou čtené úseky.

## **5.4 Přesnost**

Důležitým aspektem NGS je přesnost. U sekvenátorů používajících Sangerovu metodu je přesnost určována pomocí algoritmů Phred vyjadřujících pravděpodobnost chybně inkorporovaného nukleotidu. Například skóre Q40 znamená jednu chybu na 10000 bází, což znamená přesnost 99,99%. Na první pohled se zdá být přesnost vysoká, ale v případě sekvenace diploidního lidského genomu umožňuje tato přesnost chybně identifikovat až 600 tisíc nukleotidů.

Nové metody sekvenování pracují na jiných principech a přesnost je porovnávána pomocí jiných algoritmů, které ale pracují na podobném principu. Přesnost je potřeba určit pro každý systém zvlášť a souvisí s pokrytím genomové sekvence. Vícenásobné pokrytí genomu umožňuje větší přesnost.

## **5.5 Doba cyklu**

Rozdílná je doba trvání cyklu jednotlivých přístrojů. To je důležité z důvodu plánování experimentů a plného vytížení sekvenátoru. Je důležité započítat dobu mezi jednotlivými cykly a dobu nutné obsluhy přístroje. V případě krátkých cyklů nebude přístroj pravděpodobně vytížen z důvodu nepřítomnosti obsluhy v nočních hodinách [1]. Přístroj ABI 3730xl pracující na principu Sangerovy metody má například dobu cyklu 180 minut, ale je připraven pro automatické zpracování více cyklů najednou. Přístroj firmy 454 běží 7,5 hodiny a u ostatních firem je doba cyklů v řádů dnů. Sekvenování jednotlivých úseků trvá u Illuminy 2,5 dne, u ABI 6 dní a u firmy Helicos 7 dní.

## **5.6 Počet a množství vzorků**

Sekvenátory se liší počtem testovaných vzorků v jednom cyklu přístroje. Osm vzorků naráz je možno osekvenovat přístrojem firmy Illumina, což je nejnižší možný počet u porovnávaných sekvenátorů. Následuje nový přístroj firmy Helicos s počtem 50 vzorků, 454 s počtem 192 a poté ABI s 256 vzorky. Nejvyšší počet, 384 vzorků naráz je možno osekvenovat Sangerovou metodou.

V některých případech, například při testování pravěkých vzorků, je k dispozici pouze malé množství vzorku DNA. Nejnižší množství je potřeba u firmy ABI (20ng) a 100ng u firmy Illumina a 454. Jsou vyvíjeny nové metody amplifikace, které se snaží snížit množství vstupní DNA. Jedná se například o kvantitativní PCR umožňující snížit množství potřebného vzorku až na pikogramy. Tato metoda je použitelná pro firmy 454, Illumina i ABI [14].

## **5.7 Možnosti rozšíření sekvenátoru**

Důležitá je možnost vylepšení stávajícího sekvenátoru. Určitou informaci o možnostech zvýšení výkonnosti sekvenátorů nám přibližují výrobci a určité možnosti plynou již z použitých principů.

Sekvenátor firmy 454 má v roce 2008 zvýšit výkonnost na 0,5Gb/den a zvýšit délku čtených úseků na 400bp.

Sekvenátor firmy 454 a ABI jsou si velice podobné metodou amplifikace, tedy emPCR na kuličkách. Důležitým rozdílem je průměr použitých kuliček (26 $\mu$ m versus 1 $\mu$ m) a jejich rozmístění (uspořádané v jamkách versus náhodné). Tyto rozdíly umožňují firmě ABI několikanásobně zvýšit výkonost systému hustějším rozmístěním DNA kuliček [8]. Pro rok 2008 je plánováno zvýšení výkonnosti na 9Gb/cyklus. Zákazníkem byla ale již prokázána výkonnost 17Gb/cyklus.

Helicos slibuje zvýšení výkonnosti až na magickou hranici 1000 dolarů.

## **5.8 Zpracování dat**

Průběh získávání a analýzy dat se v případě NGS výrazně odlišuje od Sangerovy metody. V obou případech můžeme průběh zpracování dat rozdělit do třech stupňů (Level 0-2).

V případě Sangerovy metody dochází v Levelu 0 k detekci fluorescence fragmentů seřazených podle délky pomocí laserového paprsku a převedení získaných dat do digitální podoby zobrazené v podobě křivek. Tato data jsou upravena a jednotlivým signálům jsou přiřazeny odpovídající báze, což tvoří data Levelu 1. Takto získané sekvence jsou pospojovány dohromady a případně přiřazeny k referenční sekvenci, což tvoří data Levelu 2.

V případě NGS metod nejsou fragmenty seřazeny podle délky, ale jsou sekvenovány na místě. Světelné signály jsou snímány pomocí CCD kamery a data Levelu 0 jsou ukládána v podobě obrázků. Těchto obrázků je během jednoho cyklu získáno značné množství o objemu až terabytů dat, což je náročné na paměť počítače a případné skladování. Tyto obrázky jsou analyzovány a převedeny na data Levelu 1, která jsou zobrazena v podobě grafů, nesoucích informace o sekvenci jednotlivých čtených úseků. Data Levelu 2 vzniknou stejně jako v případě Sangerovy metody.

Tato data mohou být zobrazena v různých formátech (Příloha 2) a zpracovávána v různých programech [30]. Z důvodu kratších čtených úseků by měly být ke zpracování dat používány nové programy, určené pro tento typ dat [15].

## 6 CO ZVÁŽIT PŘED NÁKUPEM SEKVENÁTORU

Rozvoj v oblasti celogenomového sekvenování láká k nákupu přístrojů nové generace. Kromě již zmíněných aspektů je ale potřeba zvážit i další věci.

První věcí, nad kterou je třeba uvažovat je cena. Cena samotného přístroje, ale také náklady spojené s přechodem na jinou techniku sekvenování. Je potřeba vzít na vědomí, kolik stojí provoz přístroje, s ervis, ale také úprava laboratoře. Některé přístroje mají speciální požadavky na teplotu a místo na kterém mohou stát. Další položkou se můžou stát skladovací zařízení na reagenty, jako jsou mrazáky a chladničky. Nezanedbatelnou položkou je úprava počítačové sítě a disků pro přenos a skladování obrovského množství dat, produkované přístroji nové generace. Důležité je mít zaměstnance, který bude umět tato data spravovat a analyzovat, který bude sledovat velice rychlý vývoj nových technologií a protokolů týkajících se přístroje. Po zakoupení přístroje je také třeba týdnů až měsíců k nastavení kontroly kvality, aby pracoval s minimálními chybami.

Všechny tyto aspekty je třeba před nákupem přístroje řádně zvážit. V rozhodování může pomoci konzultace s lidmi, kteří daný přístroj již používají [26]. Hodně cenných rad lze najít i na internetových fórech zabývajících se novými metodami sekvenování, například fórum SEQanswers<sup>1</sup>.

Také může hodně napovědět zaslání již osekvenovaného vzorku do laboratoře, která používá sekvenátor nové generace a porovnání obou získaných sekvencí [29].

---

<sup>1</sup> <http://seqanswers.com/>

## 7 DALŠÍ PERSPEKTIVNÍ METODY

Komerčně dostupné techniky sekvenování jsou nadále vylepšovány. Kromě těchto zmíněných technik existují další perspektivní metody, které jsou ve stádiu vývoje. Některé z nich pouze modifikují již používané postupy a jiné vyvíjejí úplně nové přístupy [11]. Existuje přehled vyvíjených technik a udělených grantů k dosažení genomu za 100 tisíc dolarů a za 1000 dolarů<sup>1</sup>.

Následuje přiblížení dvou metod, odlišných od technik stávajících.

### 7.1 Nanopórové sekvenování

Jedna z nejslibnějších technik, používající zcela odlišný způsob detekce bází, se nazývá nanopórové sekvenování. Principem je průchod molekul skrz nanopórovou membránu.

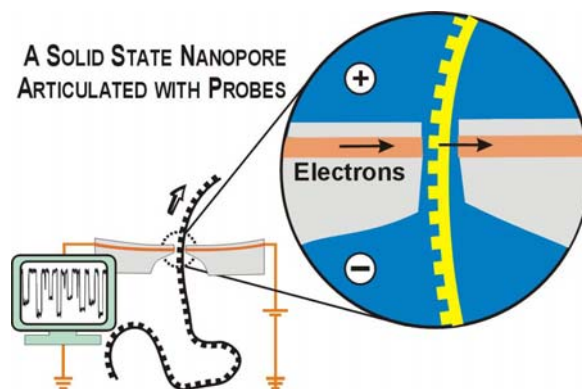
Tyto membrány se používají k oddělování jednovláknových molekul DNA od dvouvláknových, k oddělování DNA od RNA, rozdělování molekul podle jejich délky a co nás zajímá nejvíce, k identifikaci jednotlivých bází. Jsou testovány různé typy nanopórových membrán. Jako první byla testována membrána vyrobená z alfa-hemolysinu, což je protein, který je vložen do lipidové dvouvrstvy. Tento nanopórový materiál má ale své nevýhody a proto se testují materiály další. Snahou je minimalizace nežádoucích interakcí mezi molekulami DNA a póry membrány [17].

Molekula DNA prochází, podobně jako při elektroforéze pomocí elektrického pole, skrz póry membrány. Póry membrány mají průměr 1,5nm. Když vlákno prochází pórem, jeho nukleotidy vždy na chvíli zablokují ústí póru a změní elektrickou vodivost membrány, měřenou v pikoampérech (pA). Jednotlivé báze vyvolávají díky své fyzikální odlišnosti rozdílné výchylky vodivosti, pomocí nichž určíme pořadí procházejících bází [10]. (Obr.28)

Výhodami této metody je přímá a rychlá možnost sekvenování jednotlivých dlouhých molekul, pokles ceny díky nenáročnosti na reagentie a rychlost. Předpokládá se, že lidský genom by mohl být touto metodou osekvenován za necelý den [9].

---

<sup>1</sup> <http://www.genome.gov/25522229>



Obr.28: Nanopórové sekvenování

Zdroj: <http://www.mcb.harvard.edu/branton/projects-NanoporeSequencing.htm>

## 7.2 FRET (Fluorescent Resonance Energy Transfer)

Tuto metodu vyvíjí Susan Hardin pod firmou VisiGen Biotechnologies<sup>1</sup>. Jedná se o jednomolekulový přístup, při kterém je sledován fluorescenční rezonanční přenos energie. Upravená polymeráza obsahuje donor fluoroforu. Na každý ze čtyřech nukleotidů je navázán odlišně barevně označený fluoroforový akceptor. Ve chvíli kdy je inkorporován nukleotid, uvolní se FRET signál, který detekujeme. Výhodou tohoto systému je, že se jedná o přirozený proces syntézy DNA. Na rozdíl od ostatních přístupů sekvenování syntézou zde není třeba opakovat určité cyklické kroky. Na začátku jsou přidány všechny potřebné reagenty a poté už jen sledujeme inkorporaci jednotlivých nukleotidů tak, jako bychom sledovali video. Vědci z VisiGen Biotechnologies doufají v rychlost čtení miliónu bází za sekundu a jejich cílem je osekvenování lidského genomu během jediného dne [2].

<sup>1</sup> <http://www.visigenbio.com>

## 8 DĚDIČNÉ METABOLICKÉ PORUCHY

Dědičných metabolických poruch je v dnešní době známo 400- 500. Vyskytují se u jednoho člověka z tisíce. Jde o poruchy metabolických procesů, což jsou neustále probíhající řetězce přeměn, dokonale regulované a vzájemně propojené, ale i pečlivě oddělené. Hybnou silou jsou enzymy, bílkoviny s katalytickými funkcemi, které po spojení svého aktivního centra s určitou látkou bleskově provedou specifickou operaci (degradační, syntetickou, oxidoregulační).

Prvopočátek všech metabolických poruch je v genomu. Jsou způsobeny mutacemi genů. Nejvíce poruch je přenášeno autosomálně recesivně (90%). Následují poruchy vázané na X chromosom (6%), poruchy mitochondriální (2%) a autosomálně dominantní (2%). Příčinou projevu onemocnění může být deficit enzymu, aktivátoru, nebo transportního proteinu.

Tento deficit se může projevit nedostatkem, nebo naopak přebytkem produktu v příslušných metabolických cestách, což vede k systémovým, nebo tkáňovým specifickým projevům onemocnění.

Dědičné metabolické poruchy jsou nejčastěji diagnostikovány v novorozeneckém a kojeneckém věku, ale mohou se projevit v jakémkoliv věku. V České republice je prováděn novorozenecký screening na fenyلكetonurii a hypothyreózu. Diagnostika dědičných metabolických poruch probíhá na úrovni metabolitů, enzymů, DNA a tkání.

Léčba závisí na konkrétním onemocnění. Existují poruchy léčitelné, omezeně léčitelné, ale i neléčitelné. Cílem léčby je zlepšení kvality života. Metabolickou poruchu lze často ovlivnit úpravou diety vyřazující kritickou látku, jejíž metabolismus je defektní, nebo podáváním látek nedostatečně vytvářených.

Existuje okolo 500 dědičných metabolických poruch, ale již dnes je známo přes 4 tisíce enzymů<sup>1</sup>. Z toho je zřejmé, že řada dědičných metabolických poruch na své objevení teprve čeká. Techniky celogenomového sekvenování zcela jistě pomohou identifikovat mnoho dědičných metabolických poruch a umožní pochopit řetězec změn, který se od těchto poruch odvíjí a vede k dysfunkci buňky a orgánu. V budoucnu bude možné k léčbě použít genové terapie [5] [33].

---

<sup>1</sup> <http://www.expasy.ch/enzyme>



## 9 ETICKÉ PROBLÉMY

Techniky celogenomového sekvenování budou v nejbližší době k dispozici a masově se rozšíří. Jejich dostupnost a aplikace přinesou obrovské množství poznatků, ale také etických problémů.

### **Komerční firmy**

Genetické laboratoře mohou v současnosti existovat pouze na základě zápisu v obchodním rejstříku. Z tohoto důvodu není absolutně zaručena kvalita jejich práce a často porušují základní principy lékařské etiky. Mezi ně patří: beneficence (konání pro dobro a prospěch pacienta), nonmaleficence (nepoškozování pacienta), autonomie (právo na svobodné rozhodování) a spravedlnost.

Příkladem porušení autonomie jsou anonymní testy otcovství. Beneficence a nonmaleficence je porušována v případě vyšetření multifaktoriálních onemocnění. Některé firmy nabízejí prediktivní vyšetření na více než dvacet multifaktoriálních onemocnění. Jejich výpovědní hodnota je velice diskutabilní, neboť vliv genetických faktorů na tato onemocnění se pohybuje dle dnešních znalostí v řádu jednotek procent. To může mít za následek falešné uklidnění nebo zneklidnění pacienta a například změnu životního stylu k horšímu. Právě životní styl má jasný a obrovský vliv na tato onemocnění.

Velice důležitým krokem genetického testování je poradenství a správná interpretace výsledků. Tato služba u komerčních firem často chybí a je nechávána na praktických lékařích, kteří nemají dostatečné zkušenosti v lékařské genetice. To může mít za následek poškození pacienta.

Přehled genetických laboratoří podrobujících se externí kontrole kvality a pracujících na základě etických a medicínských principů je uveden v aktualizované databázi<sup>1</sup> [6].

### **Testování dětí**

Evropská společnost lékařské genetiky (ESHG) a Evropská komise doporučují odsunout prediktivní testování dětí do věku, kdy je pacient schopen pochopit všechny důsledky genetického testování a svobodně se rozhodnout, zda si testování přeje.

---

<sup>1</sup> <http://www.uhkt.cz/nrl/db>

Výjimkou jsou pouze onemocnění s nástupem v dětském věku nebo ta onemocnění, u kterých je znám vliv časného zahájení účinné prevence a léčby. Důvodem je možná diskriminace dětí ze strany rodičů a možné porušení zachování důvěrnosti dat [6].

### **Pojišťovny, diskriminace zaměstnavateli**

Znalost genomu přinese řadu informací o jedinci a o predispozicích k určitým onemocněním. O tyto informace by mohli mít zájem zdravotní pojišťovny či zaměstnavatelé, což by také mohlo vést k diskriminaci.

### **Povinnosti vůči příbuzným**

Získaná data nenesou informace pouze o osobě, která poskytla DNA k vyšetření, ale i o osobách příbuzných. Může se stát, že se blízký příbuzní dozví o svých zdravotních rizicích, aniž by chtěli [13].

## 10 ZÁVĚR

Od konce 70. let po současnost byla Sangerova metoda základní metodou získávání informací zapsaných v genomu.

V současnosti jsou k dispozici čtyři komerčně dostupné nové metody sekvenování genomu. Sekvenátor firmy 454 pracující na principu pyrosekvenace, sekvenátory firem Illumina a Helicos pracující na principu sekvenace syntézou a sekvenátor firmy ABI pracující na principu ligace oligonukleotidů. Sekvenátor firmy Helicos je prvním přístrojem, který čte jednotlivé molekuly bez nutnosti amplifikace.

Mnoho dalších perspektivních metod je ve stádiu vývoje a brzy můžeme očekávat jejich vstup na trh.

Předmětem této práce bylo popsání nových dostupných technik celogenomového sekvenování a jejich porovnání mezi sebou samými a srovnání s klasickou Sangerovou metodou.

Metody se liší v mnoha parametrech a z jejich porovnání nevychází jednoznačný vítěz. Volba nejlepší metody závisí na požadavcích a typech experimentů, které chceme na přístroji provádět. Kritérii, která musíme zvážit je délka čtených úseků, cena za Mb, výkonnost přístroje, trvání experimentu aj.

Techniky celogenomového sekvenování budou v nejbližší době k dispozici a masově se rozšíří. Brzy bude snížena cena sekvenování na magickou hranici 1000 dolarů. Již dnes se podařilo osekvenovat firmě AB lidský genom za 60 tisíc dolarů<sup>1</sup>, čímž byla pokořena první hranice stanovená NIH. Tato dostupnost a cenová přijatelnost přinese mnoho nových aplikací a poznatků o podstatě onemocnění, evoluci, populační genetice, mechanismech mutagenese a mnoho jiných. Zajímavou diskuzi o možnostech, významu a budoucnosti sekvenování je možné najít na stránkách časopisu Nature Genetics<sup>2</sup>.

---

<sup>1</sup> [http://marketing.appliedbiosystems.com/mk/get/SOLID\\_KNOWLEDGE\\_LANDING](http://marketing.appliedbiosystems.com/mk/get/SOLID_KNOWLEDGE_LANDING)

<sup>2</sup> <http://www.nature.com/ng/qoty/index.html>

## 11 SEZNAM POUŽITÉ LITERATURY

- [1] BENTLEY, David R. *Whole-genome re-sequencing. Current Opinion in Genetics & Development* [online]. 2006, vol. 16, is. 6 [cit. 2008-02-26], s. 545-552. Dostupný z WWW: <<http://www.sciencedirect.com/science/article/B6VSO-4M4TNKN-1/1/fb1195c884dbc2164262bcd352e98dfb>>. ISSN 0959-437X
- [2] BLOW, Nathan. *DNA sequencing: generation next-next. Nature Methods* [online]. 2008, vol. 5, no. 3 [cit. 2008-05-25], s. 267-274. Dostupný z WWW: <[www.nature.com/naturemethods](http://www.nature.com/naturemethods)>. ISSN 1548-7105
- [3] BROWN, T.A. *Klonování genů a analýza DNA : Úvod*. Výkonný redaktor prof. RNDr. Tomáš Opatrný, Dr.; RNDr. Martin Fellner, Ph.D.. 1. vyd. Olomouc : Univerzita Paleckého v Olomouci, 2007. 389 s. ISBN 978-80-244-1719-6.
- [4] DAVIES, Kevin. *Rozluštěný genom : Příběh největšího vědeckého objevu naší doby*. Prof. RNDr. Arnošt Kotyk, DrSc.. 1. vyd. Praha : Paseka, 2002. 304 s. Fénix; sv. 6. ISBN 880-7185-464-6.
- [5] ELLEDER, Milan. *O dědičných metabolických poruchách*. Sanquis [online]. 2001, č. 16 [cit. 2008-06-04], s. 12. Dostupný z WWW: <[http://www.sanquis.cz/clanek.php?id\\_clanek=113](http://www.sanquis.cz/clanek.php?id_clanek=113)>. ISSN 1212-6535.
- [6] GOETZ, Petr. *Hrozí diskreditace lékařské genetiky*. Medical Tribune [online]. 2008, č. 3 [cit. 2008-05-09], s. A14. Dostupný z WWW: <<http://www.medical-tribune.cz/archiv/mtr/191/5225>>. ISSN 1214-8911.
- [7] GROCHOVÁ, Ilga, GROCH, Ladislav. *Genetika v kardiologii Část I. Historie a evoluce moderní genetiky*. Cor et Vasa [online]. 2007, č. 5 [cit. 2008-06-04], s. 196-201. Dostupný z WWW: <<http://www.coretvasa.cz/pdf/3862.pdf>>. ISSN 01010-8650.
- [8] HUTCHINSON III, Clyde A. *DNA sequencing: bench to bedside and beyond*. Nucleic Acid Research [online]. 2007, vol. 35, no. 18 [cit. 2008-06-02], s. 6227-6237. Dostupný z WWW: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=17855400>. ISSN 1362-4962
- [9] CHAN, Eugene Y. *Advances in sequencing technology*. Mutation Research [online]. 2005, vol. 573, no. 1-2 [cit. 2008-06-04], s. 13-40. Dostupný z WWW: <[www.sciencedirect.com](http://www.sciencedirect.com)>. ISSN 0027-5107.
- [10] CHURCH, George M. *Genomy pro všechny*. Scientific American : České vydání. 2006, č. 2, s. 53-60. ISSN 1213-7723
- [11] LIN, Biaoyang, WANG, Jun, CHENG, Yin. *Recent patents and advances in the next-generation sequencing technologies*. Recent Patents on Biomedical Engineering [online]. 2008, no. 1 [cit. 2008-06-04], s. 60-67. Dostupný z WWW: <<http://www.bentham.org/biomeng/samples/biomeng1-1/Lin.pdf>>.
- [12] MARDIS, Elaine R. *The impact of next-generation sequencing technology on genetics*. Trends in Genetics [online]. 2008, vol. 24, no. 3 [cit. 2008-05-24], s. 133-141. Dostupný z WWW: <[www.sciencedirect.com](http://www.sciencedirect.com)>. ISSN 0168-9525
- [13] MCGUIRE, Amy L., CAULFIELD, Timothy, CHO, Mildred K. *Research ethics and the challenge of whole-genome sequencing*. Nature Reviews Genetics [online]. 2008, vol. 9, no. 2 [cit. 2008-06-04], s. 152-156. Dostupný z WWW: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2225443>>. ISSN 1471-0064.
- [14] MEYER, Matthias, et al. *From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing*. Nucleic Acids Research [online]. 2008, vol. 36, no. 1 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2248761>>. ISSN 1362-4962
- [15] POP, Mihai, SALZBERG, Steven L. *Bioinformatics challenges of new sequencing technology*. Trends in Genetics [online]. 2008, vol. 24, no. 3 [cit. 2008-06-04], s. 142-149. Dostupný z WWW: <[www.sciencedirect.com](http://www.sciencedirect.com)>. ISSN 0168-9525
- [16] REDON, R., et al. *Global variation in copy number in the human genome*. Nature [online]. 2006, vol. 444(7118) [cit. 2008-05-28], s. 444-454. Dostupný z WWW: <<http://www.nature.com/nature/journal/v444/n7118/abs/nature05329.html>>. ISSN 1476-4687
- [17] RHEE, Minsoung, BURNS, Mark A. *Nanopore sequencing technology: research trends and applications*. Trends in Biotechnology [online]. 2006, vol. 24, no. 12 [cit. 2008-06-04], s. 580-586. Dostupný z WWW: <[www.sciencedirect.com](http://www.sciencedirect.com)>. ISSN 0167-7799

- [18] RONAGHI, Mostafa. *Pyrosequencing Sheds Light on DNA Sequencing*. Genome Research [online]. 2001, vol. 11, no. 1 [cit. 2008-03-22], s. 3-11. Dostupný z WWW: <<http://www.genome.org/cgi/content/full/11/1/3#Article>>. ISSN 1088-9051/01.
- [19] RYAN, Declan, et al. *Toward nanoscale genome sequencing*. TRENDS in Biotechnology [online]. 2007, vol. 25, no. 9 [cit. 2008-06-02], s. 385-389. Dostupný z WWW: <[www.sciencedirect.com](http://www.sciencedirect.com)>. ISSN 0167-7799
- [20] WATSON, James D. *Tajemství DNA : Příběh jednoho z největších objevů 20. století*. Jitka Zykánová; František Sládeček. 1. vyd. Brno : Academia, 1995. 150 s. ISBN 80-200-0556-0.

## Seznam internetových zdrojů

- [21] 454 Life Sciences [online]. 2008 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.454.com/>>.
- [22] Applied Biosystems [online]. 2008 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.appliedbiosystems.com>>.
- [23] DAVIES, Kevin. *GATC Biotech Offers Increased Platforms, Productivity : German sequence contractor deploys three rival next-gen systems*. Bio-IT World [online]. 1.2.2008 [cit. 2008-05-25]. Dostupný z WWW: <<http://www.bio-itworld.com/issues/2008/feb/gatc-biotech/>>.
- [24] Helicos BioSciences [online]. 2008 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.helicosbio.com>>.
- [25] Illumina [online]. 2008 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.illumina.com/>>.
- [26] KAROW, Julia. *Users Weigh in on next-gen platforms; over half consider adding systems in '08*. In Sequence [online]. 2008, vol. 2, no. 1 [cit. 2008-03-21]. Dostupný z WWW: <[http://www.in-sequence.com/issues/2\\_1/webreprints/144249-1.html](http://www.in-sequence.com/issues/2_1/webreprints/144249-1.html)>.
- [27] Příspěvatelé Wikipedie. Gen [online]. Wikipedie: Otevřená encyklopedie, 2008 , 7. 05. 2008, 04:21 UTC [cit. 2008-06-04]. Dostupný z WWW: <<http://cs.wikipedia.org/w/index.php?title=Gen&oldid=2551223>>.
- [28] RACLAVSKÝ, Vladislav. *Metody molekulární genetiky* [online]. 2003 [cit. 2008-06-04]. Dostupný z WWW: <<http://biologie.upol.cz/metody/>>.
- [29] SALISBURY, Meredith. *Is next-gen technology right for your lab?*. Genome Technology [online]. 2008 [cit. 2008-06-04]. Dostupný z WWW: <[http://www.genome-technology.com/issues/2\\_13/markers/145963-1.html](http://www.genome-technology.com/issues/2_13/markers/145963-1.html)>.
- [30] SMITH, Todd. Color Space, Flow Space, *Sequence Space or Outer Space: Part II, Uncertainty in Next Gen Data* [online]. Geospiza, 24.3.2008 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.geospiza.com/finchtalk/labels/Next%20Generation%20Sequencing.html>>.
- [31] STAŇKOVÁ, Hana, KIZEK, René. *Jak se čte genom?* [online]. Český rozhlas, 4.5.2007 [cit. 2008-06-04]. Dostupný z WWW: <<http://www.rozhlas.cz/leonardo/priroda/zprava/343164>>.
- [32] TONER, Bernadette. *UCSD in discussion to license 'ordered array' tech for sequencing applications*. In Sequence [online]. 2008, vol. 2, no. 8 [cit. 2008-03-21]. Dostupný z WWW: <[http://www.in-sequence.com/issues/2\\_8/webreprints/145225-1.html](http://www.in-sequence.com/issues/2_8/webreprints/145225-1.html)>.
- [33] ÚSTAV DĚDIČNÝCH METABOLICKÝCH PORUCH VFN. *Metabolická příručka 2007*. Hlavní editor: Prim. MUDr. Sylvie Šťastná, CSc.. 2007. 72 s. Dostupný z WWW: <[http://www.udmp.cz/laborator/laborator\\_files/UDMP-2007.pdf](http://www.udmp.cz/laborator/laborator_files/UDMP-2007.pdf)>.

## 12 PŘÍLOHY

### Příloha 1- Porovnání sekvenátorů

	454	Illumina	SOLiD	Helicos	Sanger <sup>a</sup>
<b>Princip</b>	pyrosekvence	syntéza	ligace	syntéza	syntéza
<b>Metoda amplifikace</b>	emPCR	můstková	emPCR	neprobíhá	PCR
<b>Vstupní množství vzorku</b>	100 ng	0,1–1,0 µg	20 ng-20 µg	nezjištěno	5-30 µl
<b>Počet testovaných vzorků</b>	192	8	256	50	384
<b>Délka čtených úseků single read</b>	200-300bp	35-50bp	35bp	až 55bp	až 900 <sup>b</sup>
<b>Délka čtených úseků paired-end reads</b>	2x100bp	2x36bp	2x25bp	nezjištěno	-
<b>Výkonnost single read/ cyklus</b>	100 Mb	1500 Mb	3000-4000 Mb	nezjištěno	86,4 kb <sup>bg</sup>
<b>Výkonnost paired-end reads/ cyklus</b>	nezjištěno	3000 Mb	5000-6000 Mb	nezjištěno	-
<b>Výkonnost Mb/min</b>	0,222 <sup>g</sup>	0,417 <sup>g</sup>	0,463 <sup>g</sup>	1,389 <sup>g</sup>	0,002 <sup>cg</sup>
<b>Doba cyklu sinle read</b>	7,5h	2,5 dne	6 dní	7 dní <sup>f</sup>	180 min <sup>b</sup>
<b>Doba cyklu paired-end reads</b>	nezjištěno	5 dní	10 dní	nezjištěno	-
<b>Přesnost čtení báze</b>	99,50%	98.5%	99,94%	99,3- 99,4%	98,50%
<b>Přesnost</b>	>99,9%	>99,9% při 3x pokrytí	99,999% při 15x pokrytí	nezjištěno	99,7% při 6x pokrytí <sup>e</sup>
<b>Cena za cyklus</b>	\$8439 <sup>d</sup>	\$8950 <sup>d</sup>	\$17 447 <sup>d</sup>	\$18 000	\$86,40 <sup>g</sup>
<b>Cena za Mb</b>	\$84,39 <sup>d</sup>	\$5,97 <sup>d</sup>	\$5,81 <sup>d</sup>	\$1,285 <sup>g</sup>	\$1000 <sup>e</sup>

Zdroj: Webové stránky jednotlivých firem

a ABI 3730xl Applied Biosystems

b závisí na zvoleném režimu, údaje pro sekvenování extra dlouhých úseků

c závisí na zvoleném režimu, údaje pro značené resekvenování (TargetSeq™ Resequencing)

d Celkový součet ceny reagentů a spotřebního zboží, pracovní síly, amortizace přístroje a potřebný prostor na disku pro skladování dat

Zdroj: The impact of next- generation sequencing technology on genetics: Elaine R. Mardis

e Zdroj: Towards nanoscale genome sequencing

f [http://www.bio-itworld.com/BioIT\\_Article.aspx?id=73220](http://www.bio-itworld.com/BioIT_Article.aspx?id=73220)

g výpočet ze známých údajů



