

Inovace studia molekulární a buněčné biologie reg. č. CZ.1.07/2.2.00/07.0354

Investice do rozvoje vzdělávání



Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Genomika (KBB/GENOM)

Investice do rozvoje vzdělávání



Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Sekvenování genomů

Ing. Hana Šimková, CSc.

Investice do rozvoje vzdělávání



Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Cíl přednášky

- seznámení se strategiemi celogenomového sekvenování, s postupy používanými při celogenomovém sekvenování, anotacemi a sekvenačními technologiemi nové generace

Klíčová slova

- syntenie, metylační filtrace, whole-genome shotgun (WGS) sekvenování, hierarchické sekvenování, minimum tiling path, anotace, technologie 454, Illumina (Solexa), Solid

Investice do rozvoje vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

CELOGENOMOVÉ SEKVENAČNÍ PROJEKTY

ČÍM UMOŽNĚNY?

- automatizovanými sekvenačními systémy (fluorescenční barviva, kapiláry)
 - vyšší kapacita, nižší objem reakcí, větší délka čtené sekvence
 - nižší náklady
- vývojem programů pro uspořádávání sekvencí

PROČ SEKVENOVAT?

- objevování genů
- detekce SNP a jejich využití pro mapování
- komparativní genomika – studium evoluce

STRATEGIE SEKVENOVÁNÍ

A) Redukční přístupy zaměřené na geny

- a) sekvenování s nižším pokrytím
- b) sekvenování ESTů
- c) sekvenování založené na C_0t frakcionaci
- d) sekvenování založené na metylační filtraci

B) Celogenomové náhodné sekvenování

C) Hierarchické sekvenování

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Investice do rozvoje vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

A) Redukční přístupy zaměřené na geny

a) Sekvenování s nižším pokrytím (coverage)

- sekvenování 3 ekvivalentů genomu (BACu) zachytí 95% všech sekvencí.

Není vhodné pro velké genomy s hodně repeticemi. Může pomoci syntenie s již osekvenovanými genomy. (**Syntenie** = výskyt konzervovaného uspořádání genů v oblasti s rozsahem stovek kb, pozorované pro genomy blízce příbuzných druhů)

Speciální případ – sekvenování konců BACů (**BAC end sequencing**) - dražší než běžné sekvenování z plasmidů, ale sekvence z obou konců BACu vytvářejí kostru umožňující poskládání celých genomů.

Investice do rozvoje vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

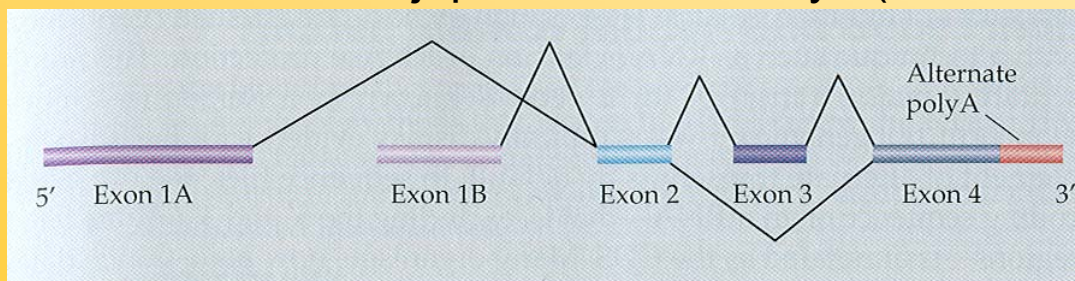
INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

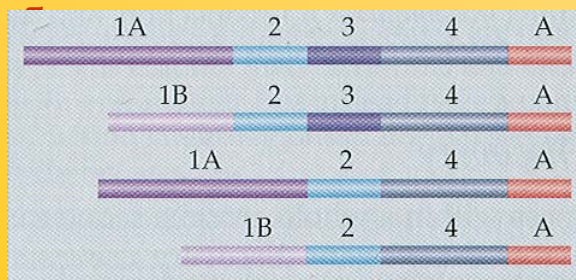
b) Sekvenování ESTů (koncových sekvencí z klonů cDNA)

😊 levnější způsob získávání informací o genech, vhodné k odvozování molekulárních markerů z kódujících sekvencí a k potvrzení genů vytipovaných na základě sekvence

☹ zachytí jen cca 60% genů, jsou to jen zlomky genů, neobsahují regulační sekvence, obsahují poměrně dost chyb (reverzní transkriptáza není přesná)



Struktura genu
– alternativní sestřih



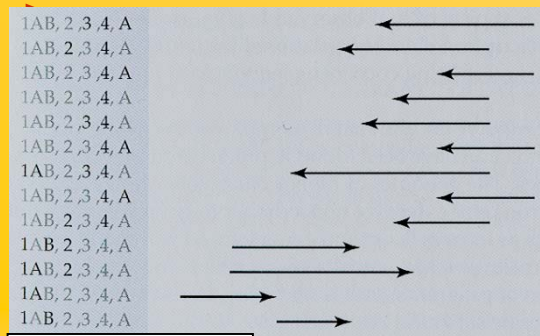
Různé
cDNA

Jeden gen bývá reprezentován několika ESTy.

Reprezentativní hlavní transkript = **RefSeq**

Pro mnoho organismů byly vytvořeny **databáze cDNA o plné délce** (full-length cDNA)

Člověk: www.h-invitational.jp



Různé ESTy

Gibson a Muse, 2004

— Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

B) Celogenomové náhodné sekvenování (whole-genome shotgun, WGS)

Celý genom se mechanicky rozbije na malé fragmenty obvykle tří velikostních frakcí (např. 2 kb, 10 kb, 50 kb), z kterých se vytvoří knihovny. Inzerty se sekvenují z obou stran – získání **párových konců**. Sekvenuje se nadbytek sekvencí (pokrytí 5-10x). Jednotlivé sekvence pospojovány pomocí počítače – 90% automaticky, zbytek ručně. Menší knihovny – k vlastnímu sekvenování, 50 kb knihovna – k ověření správnosti sekvenčních kontigů a zaplnění mezer.

☺ Jednodušší, rychlejší a levnější. Celkově jsou potřeba jen tři knihovny.

☹ Náročné na hardware a software – může dělat jen několik laboratoří na světě. Práce se hůř rozděluje mezi laboratoře. Neobejde se bez kvalitních fyzických a genetických map.

Investice do rozvoje vzdělávání

C) Hierarchické sekvenování (sekvenování klon po klonu, clone-by-clone sequencing, CBCS)

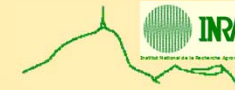
Vychází z knihovny dlouhých inzertů (např. BAC, PAC) a fyzické mapy. Sekvenují se klony z MTP (minimum tiling path).
Pro každý klon se konstruuje
1-2 knihovny krátkých inzertů a provede se jejich shotgun
sekvenování.

Investice do rozvoje vzdělávání

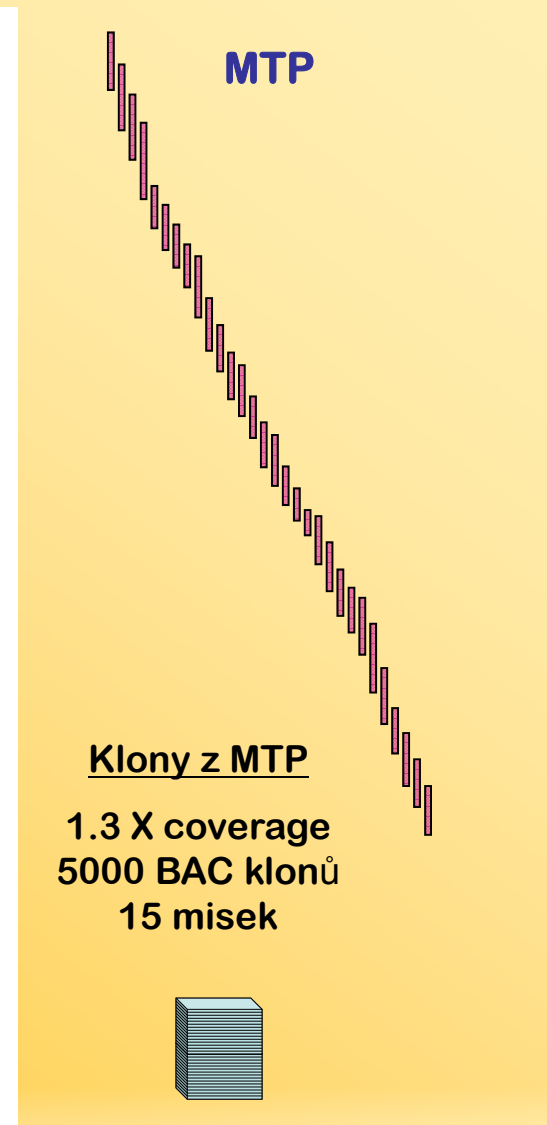
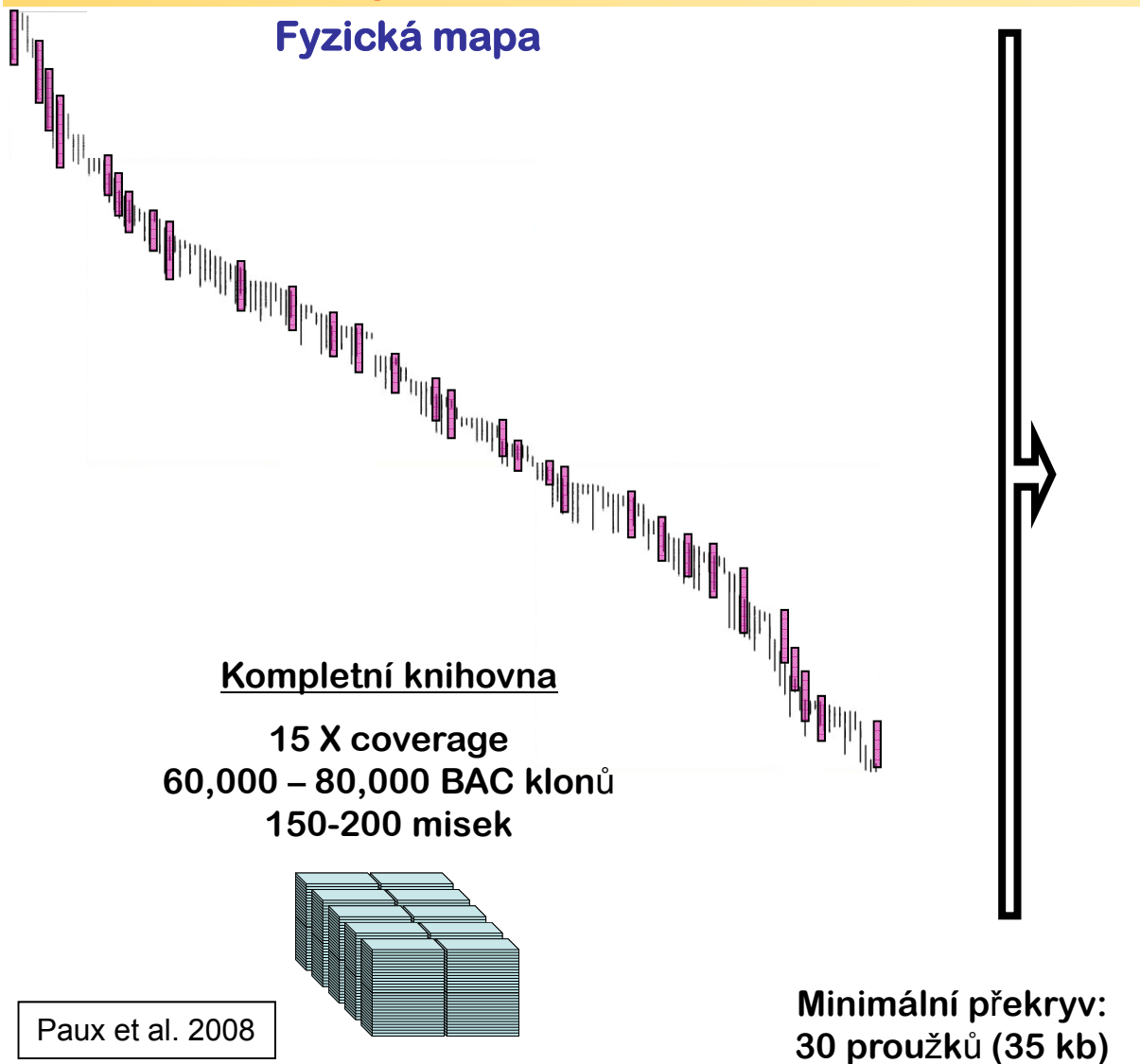


Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Hierarchické sekvenování - vytvoření MTP



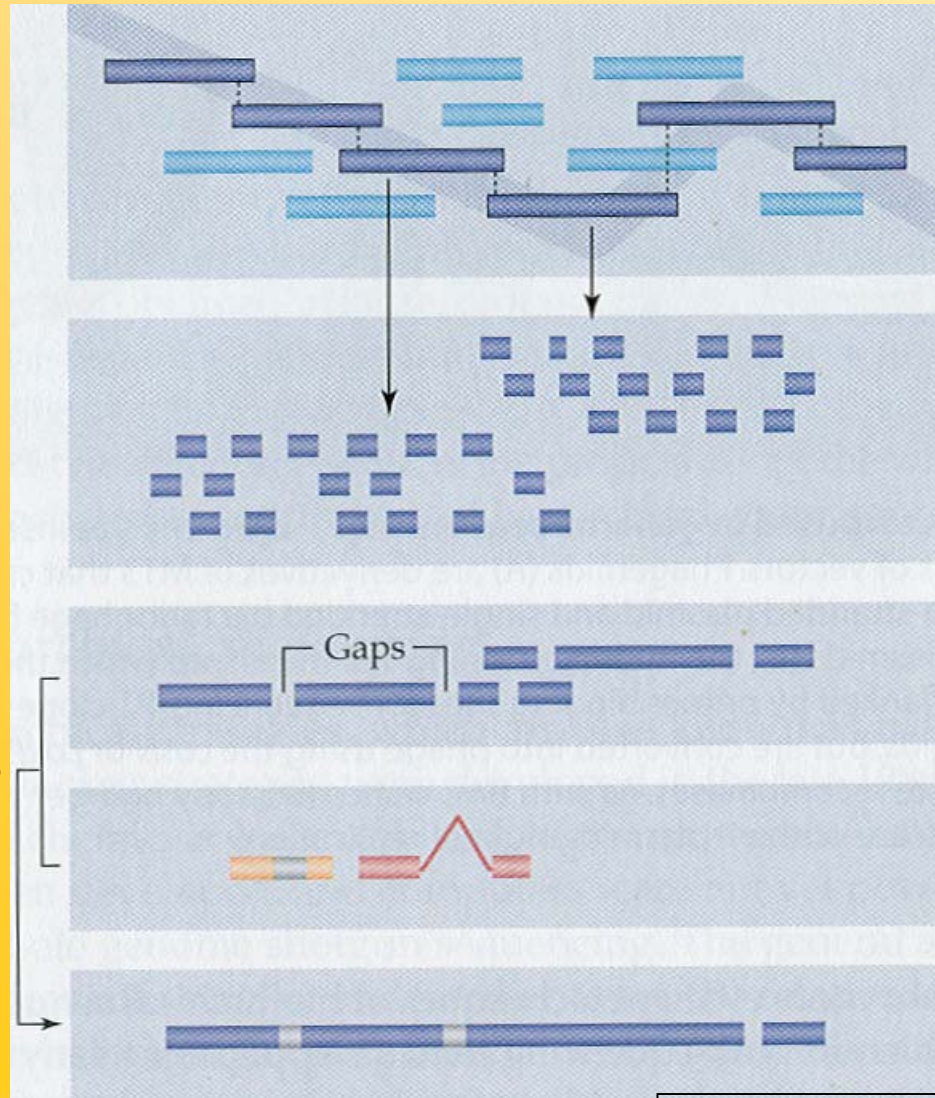
- minimální sestava klonů, které vytvářejí kontinuální sekvenci podél celého chromozómu (**program FPC**)



Paux et al. 2008

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Hierarchické sekvenování



Sestavení fyzické mapy, výběr „minimální trasy pro sekvenování“ (MTP)

Shotgun sekvenování vybraných BAC nebo jiných klonů, uspořádání sekvencí

Kontigy sekvenovaných klonů obsahují mezery

Přidání údajů z párových konců a cDNA

„Kostra“ (scaffold) z kontigu sekvencí

Gibson a Muse, 2004

Investice do rozvoje vzdělávání

Spojení

Hierarchické sekvenování

- ☹️ 1-2 knihovny z každého BAC klonu v MTP – pracnější
- 😊 práce se dá snadno rozdělit mezi laboratoře (nejlépe po chromozómech), méně problémů při uspořádávání, dá se přednostně zaměřit na genově bohaté euchromatinové oblasti

Investice do rozvoje vzdělávání

Postup celogenomového sekvenování

Genomická DNA (WGS)

BAC klony (CBCS)

- 1) Příprava knihoven malých inzertů – z náhodně našťípané DNA po velikostní selekci - obvykle 2 velikostní kategorie (2 kb, 10 kb), pro WGS ještě knihovna 50kb.
- 2) Příprava DNA pro sekvenování – založena na alkalické lyzi bakteriálních kultur (automatizováno)
- 3) Vlastní sekvenování – z knihoven 2 kb (případně 10 kb). Sekvenační reakce běží z primerů komplementárních k části vektoru. Sekvenování z obou konců inzertu – získání párových konců.
- 4) Zpracování sekvenačních dat: pro čtení sekvencí – program **phred** – převádí píky fluorescence na nukleotidy, ke každému udává míru spolehlivosti (pravděpodobnost chyby). Phred skóre nad 20 = spolehlivé. Očišťuje sekvenci od sekvence vektoru.

- 5) Uspořádání sekvencí (sestavování sekvenčních kontigů) – program **phrap** – přesah minimálně 40bp s homologií nad 97%. Pro určení konečné sekvence se kombinuje 5-10 překrývajících se čtení – odvodí se konsenzuální sekvence (nahrazuje manuální opravování).
K manuálnímu upravování a opravování – editor **consed** – k opravování špatných překryvů, dvojznačných nukleotidů, překlenování mezer (navrhne primery pro sekvenování do mezery).
- 6) Tvorba „kostry“ (scaffolding) – skupiny sestavených sekvencí (sekvenční kontigy) jsou spojovány a orientovány na základě údajů z párových konců (2 sekvence + vzdálenost mezi nimi) – zejména z knihoven dlouhých inzertů.
K ukotvování kostry na chromozóm se využívají genetické a fyzické markery, detekované v sestavách sekvencí, a syntenie.
- 7) Opravy sekvencí a zaplňování mezer – využívá údajů z genomových projektů jiných druhů, ze sekvenování cDNA, z párových konců 50 kb klonů. Pro kvalitní dokončenou sekvenci – většinou ručně, hlavně v oblastech repetit. Zaplnění mezer – většinou přesekvenováním těchto oblastí.

ANOTACE SEKVENCÍ

1. Identifikace genů

- na základě homologie se známými genovými sekvencemi v databázích
- ab initio - na základě
 - existence dlouhých ORF
 - blízkosti transkripčních a translačních motivů a polyA signálu
 - konsenzuální sekvence pro rozhraní exonu a intronu

Programy – pro bakterie a jednodušší eukaryota – např. **Genefinder**
- pro složitější eukaryota – využívají statistický přístup – **Hidden Markow models (HMM)**

K přijetí sekvence mezi geny nutné ještě **další důkazy**:

1. identita s nějakou cDNA v databázi
2. shoda s jedním nebo více ESTy ze stejného organismu
3. podobnost nukleotidové nebo AK sekvence se sekvencemi genů jiných organismů v databázích
4. proteínová struktura homologní s nějakou doménou v databázi
5. spojení s typickými promotorovými sekvencemi (TATA box), shluky vazebných míst pro transkripční faktory, u savců blízkost CpG ostrova
6. důkaz, že mutace v sekvenci vyvolává nějakou změnu fenotypu

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



2. Identifikace regulačních sekvencí

- jde o konzervované nekódující sekvence

Identifikace za pomoci **komparativní genomiky**.

Metody:

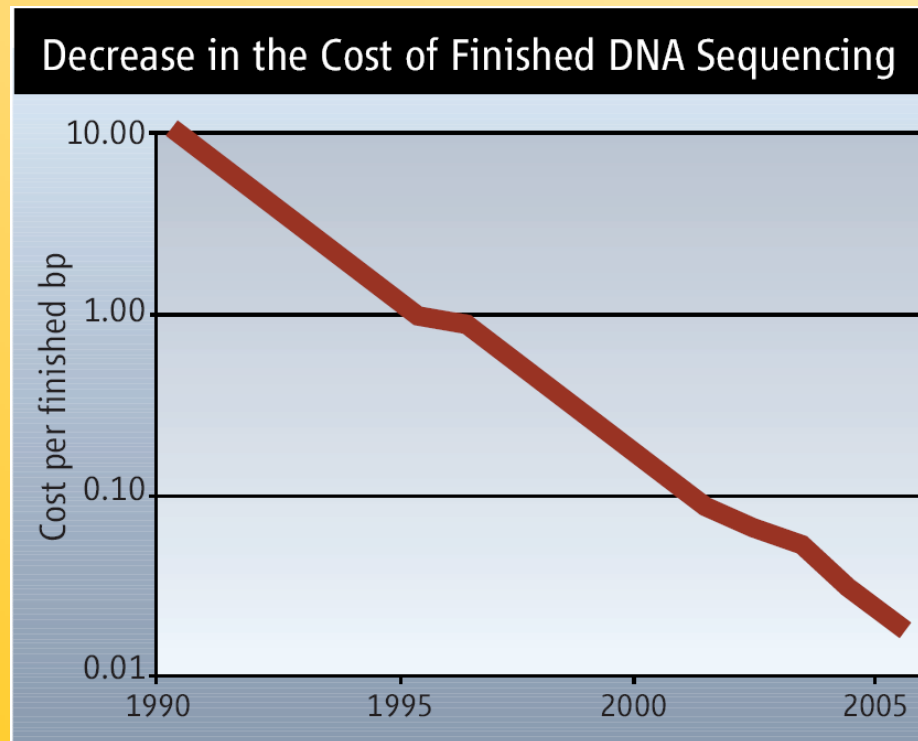
- phylogenetic footprinting – např. porovnáním lidské a myší DNA (programy **PipMaker**, **VISTA**)
- phylogenetic shadowing – porovnáním několika blízce příbuzných druhů (program **eShadow**)

Investice do rozvoje vzdělávání

SEKVENAČNÍ TECHNOLOGIE NOVÉ GENERACE

- cílem – urychlení a zlevnění celogenomového sekvenování (resekvenování)
 - předpoklad pro personalizovanou genomiku a farmakogenomiku

The Race for the \$1000 Genome (Science 311: 1544 – 1546, 2006)



Free fall. As with computer technology, the plunging cost of DNA sequencing has opened new applications in science and medicine.

Investice do rozvoje vzdělávání

Sekvenační technologie nové generace (= masivně paralelní sekvenování)

- místo konstrukce knihoven v bakteriích – „knihovny“ amplifikované DNA

	Platforma		
	454 (Roche)	Solexa (Illumina)	SOLiD (ABI)
Sekvenační princip	Pyrosekvenování	Polymerase-based sequencing by synthesis	Sekvenování založené na ligaci
Způsob amplifikace	PCR v emulzi	„Můstková“ amplifikace	PCR v emulzi
Pár. konce/vzdálenost	(Ano)/3 kb	Ano/200 bp	Ano/600 bp -10 kb
Mb/běh	500 Mb	1300 Mb (2 600 Mb)	30 Gb (120 Gb)
Čas/běh (pár. konce)	7,5 hod (10 hod)	4 dny	5 dnů
Délka čtené sekvence	400 bp (1000 bp)	35 bp (70 bp)	50 bp (100 bp)
Náklady na 1 Mb	84,39 \$ (250 bp)	5,97 \$ (35 bp)	5,81 \$ (35 bp)

Sangerovo sekvenování (ABI) – 1700 \$ /1Mb

■ perspektiva

Upraveno podle Mardis E. (Cell, 2008)

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Investice do rozvoje

454 sekvenování

1. Příprava „knihovny“ jednořetězcové DNA (sstDNA)
2. PCR amplifikace v emulzi (emPCR)
3. Sekvenování (pyrosekvenování)

Při jednom běhu přístroje čten současně až 1 milión sekvencí.

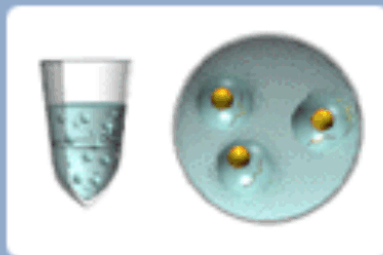
Investice do rozvoje vzdělávání



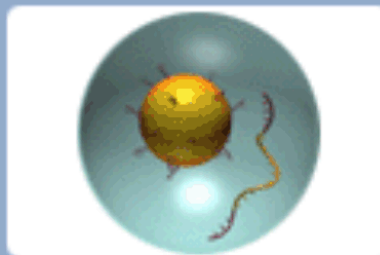
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Ad 2) PCR v emulzi

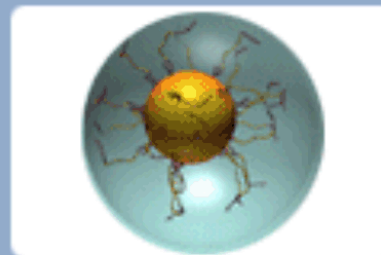
1. SstDNA „knihovna“ je navázána na agarózové kuličky s oligonukleotidy komplementárními k adaptérům. **Každá kulička - max. 1 sstDNA.**
2. Přidány reagencie pro PCR
3. Přidán olej a vytvořena emulze – každá kulička s DNA zachycena v kapičce reakčního roztoku obklopené olejem = mikroreaktor
4. PCR v mikroreaktorech
5. Rozbití mikroreaktorů a obohacení o kuličky s amplifikovanou DNA



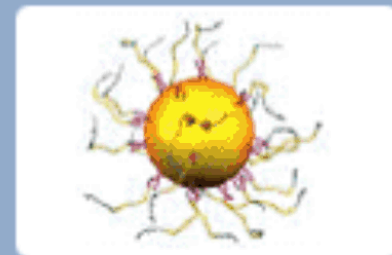
Anneal sstDNA to an excess of DNA Capture Beads



Emulsify beads and PCR reagents in water-in-oil microreactors



Clonal amplification occurs inside microreactors



Break microreactors enrich for DNA-positive beads

Investice do rozvoje vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Roche

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

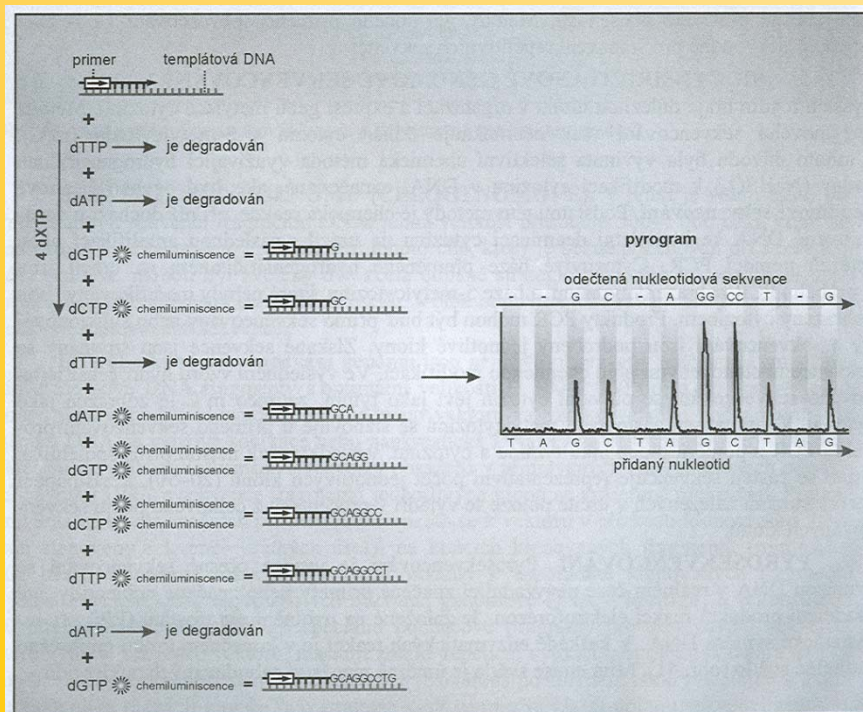
Ad 3) Sekvenování

- masivně paralelní pyrosekvenování – na pikotitrační destičce z optických mikrovláken

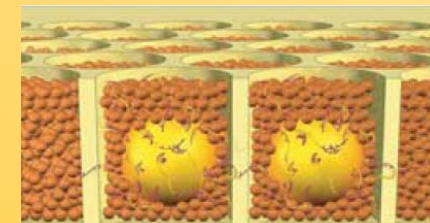
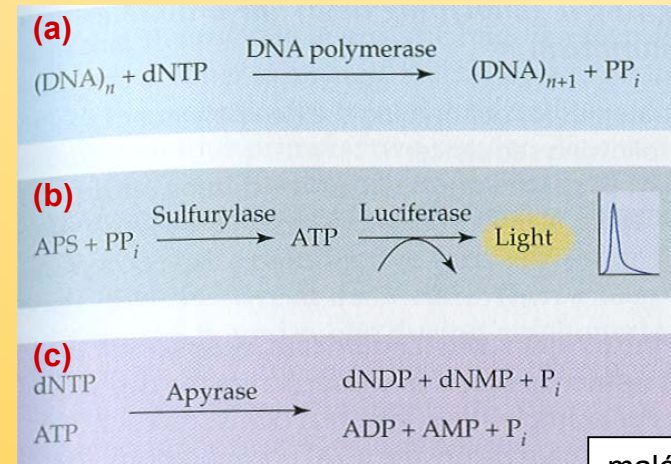
Pyrosekvenování

- založeno na - cyklických reakcích (přidávání T, A, G, C, T, A, ...)
- uvolnění pyrofosfátu v případě, že se inkorporovala specifická báze (a).

Následuje kaskáda reakcí, která na konci vede k emisi viditelného světla (b). Před dalším cyklem - degradace neinkorporovaného nukleotidu (c).



Obr. 42 Princip pyrosekvenování DNA



malé kuličky s enzymy

Roche

Investice do rozvoje vzdělávání



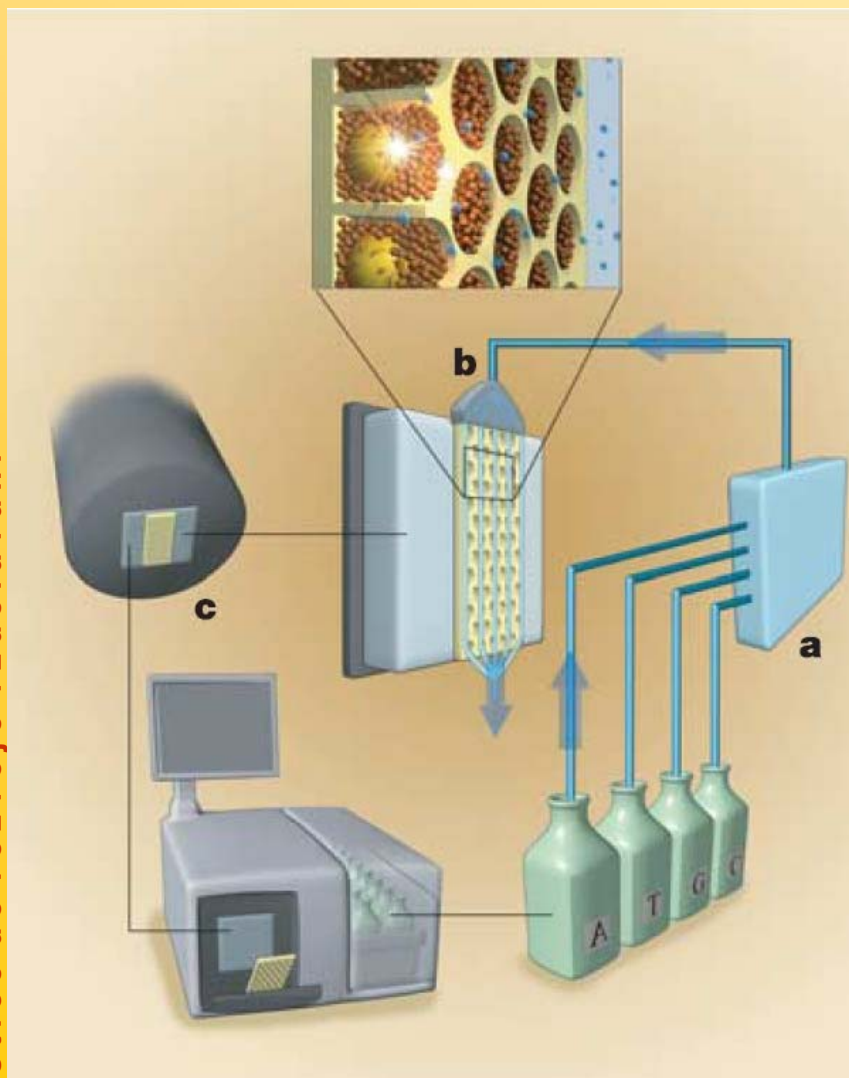
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Sekvenátor (Roche)



Sekvenační zařízení

se skládá z těchto základních subsystémů:

- Zařízení pro mikrofluidiku (a)
- Průtoková komůrka s pikotitrační destičkou (b)
- Zobrazovací zařízení s CCD kamerou (c)
- Počítač pro uživatelské rozhraní a kontrolu přístroje

Investice do rozvoje vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



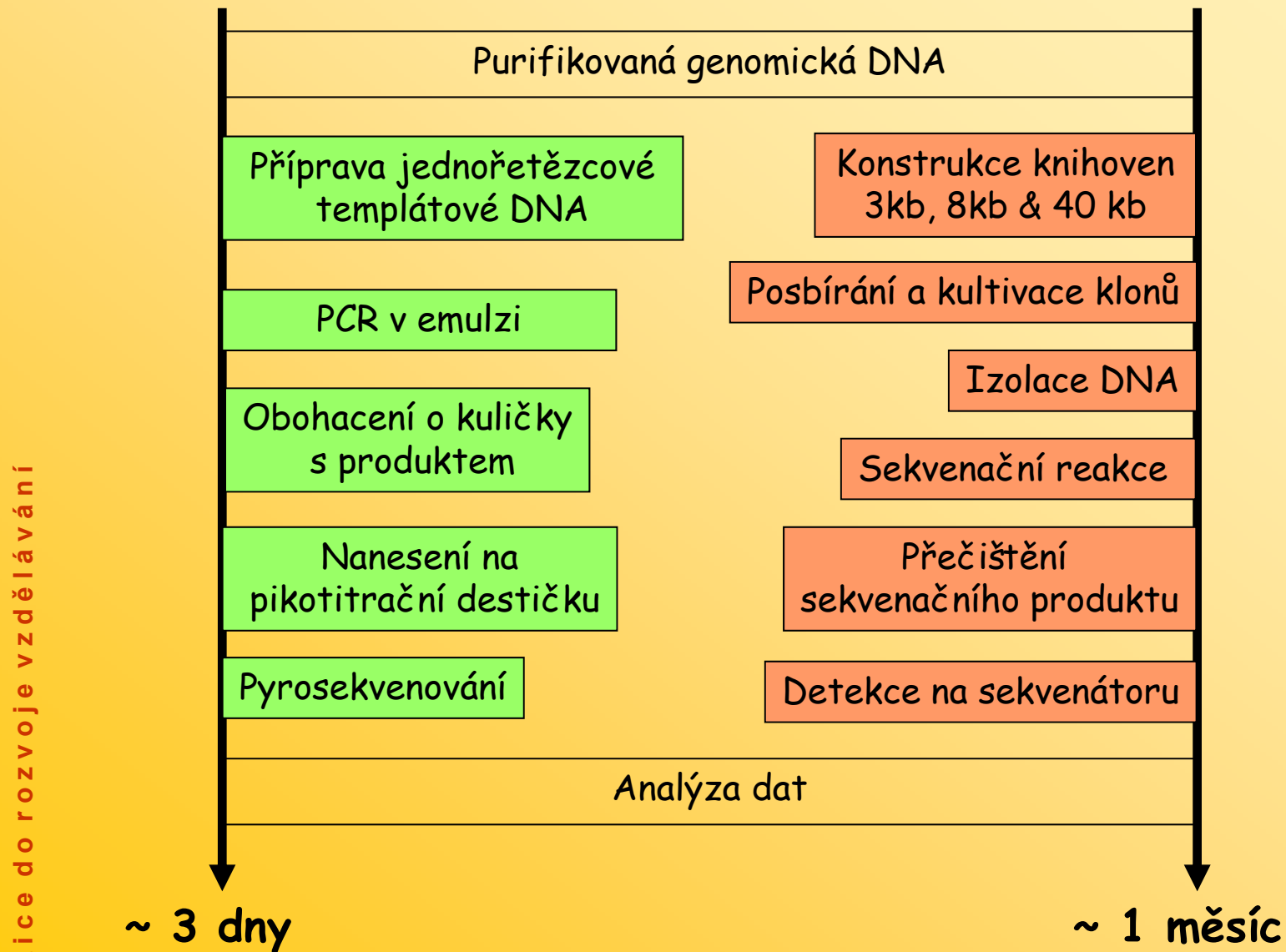
OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

454 sekvenování

Tradiční sekvenování



Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.