

VYUŽITÍ MUTANTŮ V GENOMICE

Metody genomiky a proteomiky

Genomika – lekce 6

Markéta Žďárská

Obsah

- Transgenní rostliny a funkční genomika
- Mutageneze, typy mutagenů
- Přístupy přímé genetiky
 - > EMS mutageneze
 - > Poziční klonování
 - > Sekvenování
- Přístupy reverzní genetiky
 - > Tilling

Transgenní organismy

- syntetické, modifikované nebo cizorodé geny mohou být vnášeny do živočichů a rostlin
- transgenní organismy slouží ke studiu funkcí genů

Funkční genomika

- cíl: hledání genů a určení jejich funkce v genomu
- 2 různé přístupy:
 - > **přímá genetik** „forward genetics“
 - fenotyp → gen
 - > **reverzní genetik** „reverse genetics“
 - sekvence DNA (gen) → fenotyp

Mutagenese

- mutanti – nástrojem obou přístupů
- různé typy mutagenů

1. chemické

2. fyzikální (záření)



klasické

- velký počet náhodných mutací, zasáhne všechny geny

3. biologické

- **moderní**
- inzerční mutagenese, menší počet mutací, zanechávají molekulární značku

Chemické mutageny

- způsobují bodové mutace

1. při replikaci DNA ale i v nereplikující se DNA

> **alkylační látky** (vnáší alkylovou sk. na báze; náhodné)

- **yperit** (hořtičný plyn; Ch. Auerbach)

- **ethylnitrosomočovina (ENU)**

- ethylová sk. ENU reaguje většinou s thyminem
- Bill Russell (1951) myší kmen („T-test stock) pro testování různých mutagenů



- **ethylmethansulfonát (EMS)**

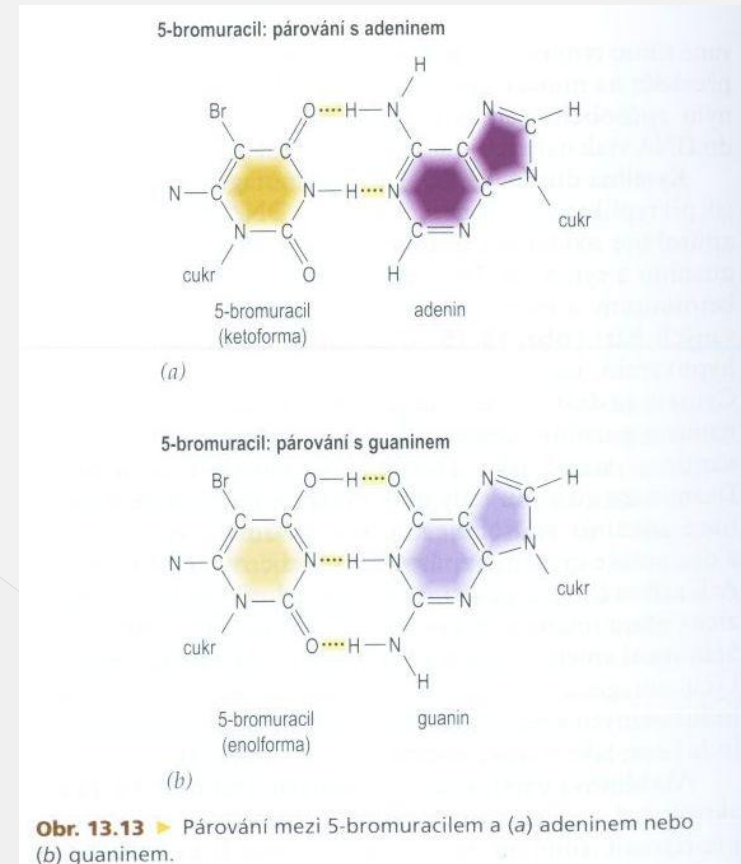
- ethylová sk. EMS reaguje s guaninem v DNA, vytvoří O-6-ethylguanin (místo G:C vznikají A:T)
- Maple J. & Moller SG, 2007 - Mutagenesis in *Arabidopsis*

> **HNO₂** (deaminace aminoskupin)

Chemické mutageny

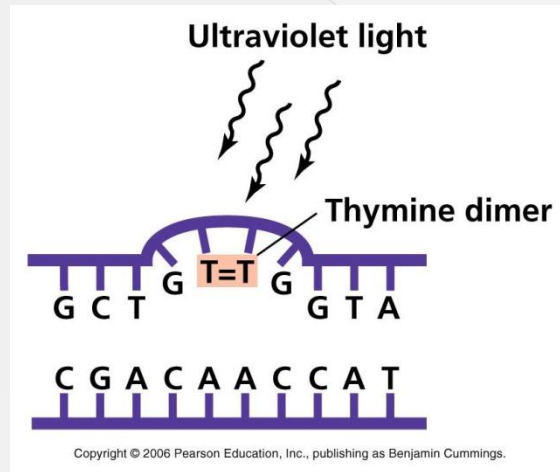
2. pouze při replikaci

- > **analogy bází:** 5-bromuracil, 2-aminopurin
 - způsobují chybné párování
- > **akridinová barviva:** proflavin, akridinová oranž
 - způsobují adice či delece 1-více bází → změna čtecího rámce → nefunkční genové produkty
- > **hydroxylamin**
 - specifická – indukuje tranzice pouze ve směru G:C → A:T



Fyzikální mutageny

- ionizující: RTG, gama záření, radioaktivní uhlík ^{14}C
 - > způsobují zlomy
- neionizující: UV záření 254nm je absorbováno bázemi → pyrimidinové dimery (naruší cukr-fosfátovou kostru)



- způsobují rozsáhlejší inserce a přestavby chromozomů
- mohou odstranit i více genů nebo vnést nové regulační sekvence
- nevhodné pro přesnou mutagenezi

Biologické mutageny

◉ inzerční mutageneze

1. T-DNA

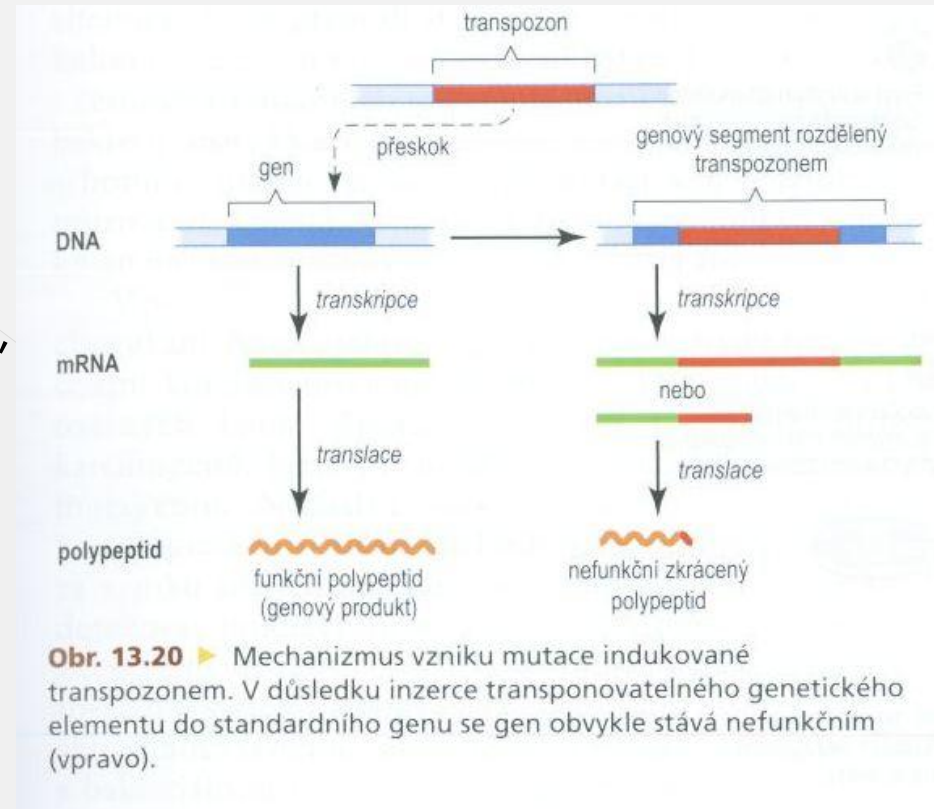
- > *Agrobacterium tumefaciens* je schopna vnést do rostlinného genomu část své DNA
- > důsledky začlenění T-DNA do genomu jsou různé, dáno povahou inzertu a místem, kde se integruje
- > způsobuje:
 - inaktivaci genu
 - aktivace (nese promotor, zesilovač)

2. Transpozony – transponovatelné genetické elementy

- > pohyblivé genetické elementy, nejsou tak stabilní jako T-DNA
- > mohou se přemístit z místa původní inserce → obnovení normálního fenotypu
- > v místě inserce zůstává stopa i po opuštění genomu, nevhodné pro rozsáhlou mutagenezi (Petersen et al, 2000)

T-DNA a transpozony

- ⊙ inzerční mutagenese
- ⊙ inzerce do:
 - > kódující oblasti
 - > nekódující oblasti – ovlivnění sestřihu intronů, genové exprese
- ⊙ 😊 výhody:
 - > reverzibilní mutace
 - > snadno se mapuje a region se lehce klonuje
- ⊙ ☹️ nevýhody:
 - > inzerce není náhodná



Jak obecně hledat mutaci?

1. menší mutace

- nové **jednonukleotidové polymorfismy** (SNP) ve vytipované oblasti
- změněný transkripční profil
- komplementace mutace transgenozí různých kandidárních genů ve WT formě

2. chromozomiální přestavby

- přestavby rozsáhlé → *in situ* hybridizace na metafázních chromozomech – rozlišení 5 Mbp
- přestavby menší → komparativní genomová hybridizace (CGN)
 - „DNA microarrays“ se sondami pokrývající různé části genomu
 - hybridizace s genomickou DNA – rozlišení desítky Mbp

Přímá „saturační“ genetika

- založena na saturační mutagenezi „saturation screen“
 - > působení mutagenu na organismus → analýza potomstva na určitý fenotyp
 - > identifikace mutantů → třídění do komplement. skupin
 - > mapování do obecných chromozomál. lokalizací pomocí známých markerů (značek) a klonování, sekvenování
- existuje mutace pro každý lokus → možnost určit skupinu genů odpovědných za daný znak
- cílem je dosažení saturačního bodu → odhalit všechny geny odpovědné za fenotyp
- mutageny (RTG, EMS, transpozony)
 - > Příklady:
 - > rostliny bez reakce na světlo
 - > neschopnost bakterií růst v přítomnosti určitých cukrů,...

Identifikace mutovaného genu u mutantní linie vybrané na základě fenotypového projevu

◎ na základě genetické mapy

1. mapování - (ko)segregační analýza

> nalezení přibližné pozice genu v genetické mapě, na základě genové vazby s **genetickými markery** (znaky, které vykazují polymorfismus = jsou rozdílné mezi rodičovskými genotypy)

2. nalezení konkrétní sekvence nesoucí mutaci

> **chromosom „walking“**

> **sekvenace**, srovnání s WT sekvencí

Typy genetických markerů

= znak se známou (snadno dohledatelnou) polohou v genetické mapě, který vykazuje polymorfismus (je rozdílný mezi rodičovskými genotypy)

1. Morfologické

2. Molekulární

- DNA markery – detekovatelné rozdíly v sekvenci
 - DNA se známou či určitelnou polohou v genomu
 - dobře detekovatelné lokusy se známou pozicí na chromozomu
 - jednonukleotidové polymorfismy (SNP)
 - ideálně co nejvíce a rovnoměrně rozmístěny

Přirozená morfologická variabilita *Arabidopsis* – ekotypy „accessions“

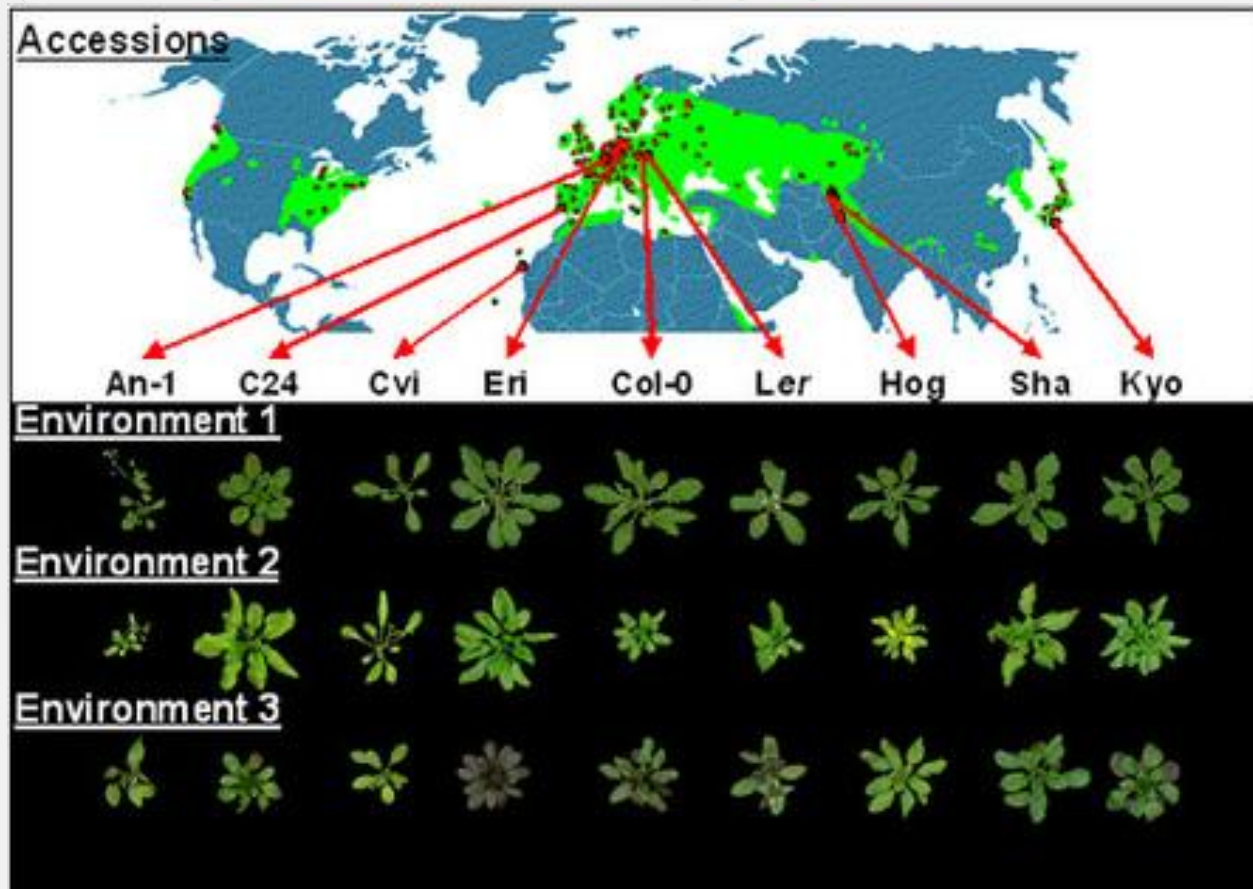


Fig. 1: Geographical distribution of *Arabidopsis thaliana* (green area on the world map) and overall phenotype of the rosette of a subset of accessions grown under 3 contrasted environment scenario.

<http://www.mpipz.mpg.de/102840/reymond>

DNA molekulární markery

(= proužek na elektroforéze či blotu)

- ◎ SSLP (Simple Sequence Length Polymorphism)
 - > délka genu (PCR produktů) amplif. pomocí spec. primerů
- ◎ RFLP (Restriction fragment length polymorphism) + Southern
 - > délka restričních fragmentů úseku genu, PCR po naštěpení genomové DNA a ligaci adaptorů
- ◎ RAPD (Random amplified polymorphism detection)
 - > délka náhodně amplif. úseků genu (krátké primery 8-10bp)
- ◎ AFLP (Amplified fragment length polymorphism)
 - > délka fragmentů genu, PCR po naštěpení genomové DNA a ligaci adaptorů

DNA molekulární markery

Arabidopsis thaliana

- křížení dvou jeho ekotypů: Columbia a Landsberg *erecta* (Col X Ler)
- rekombinační mapa obsahovala původně 67 markerů (Lister & Dean, 1993)
- dnes přes 1300 markerů (Hou et al, 2010)



Arabidopsis thaliana physical map with indication of the positions of the markers.

Poziční klonování

(positional cloning, map-based cloning)

- izolace genu pouze na základě znalosti jeho **pozice na mapě**
- nemusí být známa funkce genu (mechanismus působení, biochemická, podstata produktu)
- mapování genu na základě jeho **genetické vazby s molekulárním markerem** a následnou izolaci tohoto genu díky známé pozici na chromozomu
- potřeba standardní linie, která je zkřížena s mutantní linií, aby mohla být provedena rekombinační analýza s markery
- **cíl pozičního klonování :**
 - > gen s požadovanou mutací lokalizovaný v intervalu mezi dvěma nejbližšími markery
 - > úsek dostatečně malý → vybrat kandidátní gen, v němž je pak možné různými způsoby identifikovat mutaci
 - > obecně složité a časově náročné

Poziční klonování - schéma

1. Křížení mutantní linie se standardní linií odlišného, genetického pozadí a vytvoření **segregující generace** (většinou F2) → mapovací populace
2. Přibližná lokalizace mutantního lokusu rekombinační analýzou s genetickými markery (20-30 rostlin)
 - > „bulk segregation analysis“
 - a) analýza směsi DNA všech vzorků z rostlin určitého fenotypu najednou pro jednotlivé markery → snížení množství potřebných PCR reakcí (Lukowitz et al., 2000)
 - b) stanovení vazby je multiplex PCR s použitím fluorescenčně značených primerů (všechny markery najednou u jednotlivých vzorků) (Ponce et al, 1999)

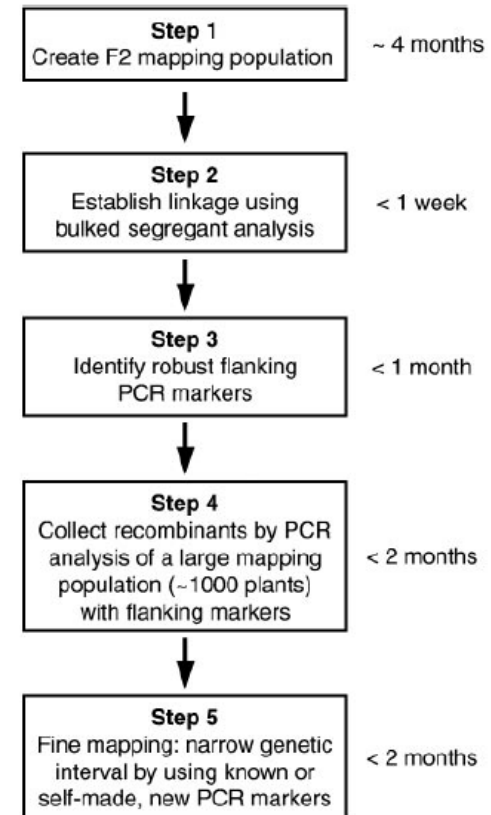
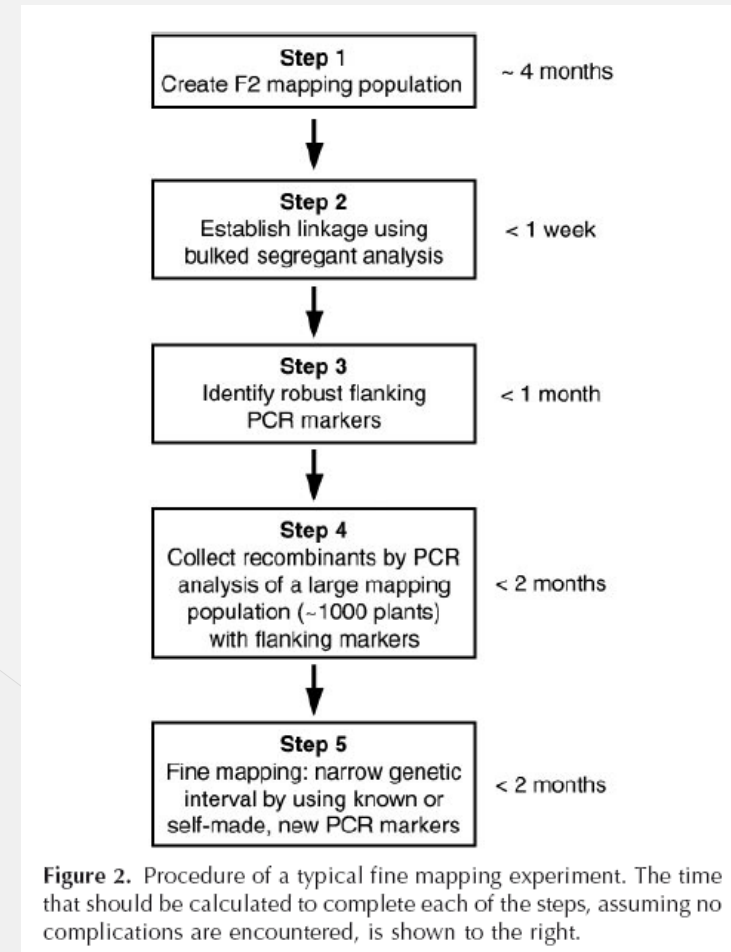


Figure 2. Procedure of a typical fine mapping experiment. The time that should be calculated to complete each of the steps, assuming no complications are encountered, is shown to the right.

Poziční klonování - schéma

3. identifikace **dvou markerů** vzdálených od sebe **<10%** rekombinace a definující genetický interval obsahující mutaci (cca 100 rostlin)
 4. **rozšířená analýza markerů** (1000 rostlin) a selekce rostlin, kde došlo k rekombinaci
 5. rekombinační analýzy ke zúžení definované oblasti na interval **< 1%** rekombinace (250 kb)
- 🕒 identifikace mutantního genu ve vymezené oblasti



Poziční klonování - aplikace

◎ *Arabidopsis*

- > mapa pozičně klonovaných genů obsahující 620 mutantních genů s fenotypovým projevem (Meinke et al, 2003)
- > počet genů *Arabidopsis* izolovaných metodou pozičního klonování každým rokem narůstá → objasnění dalších funkcí genů modelové rostliny

◎ **Pšenice** (*Tritium aestivum*)- zemědělské plodiny

- > molekulár.markery:
 - SSR (Simple Sequence Repeat)
 - SCAR (Sequence – Characterised Amplified Region)
- > izolování genů s markery komplikovanější – hexaploid
- > nedostatečná znalost genů kódujících zemědělsky užitečné znaky (zatím cca 20), např. plísně a rezistence k nim

◎ Geny chorob

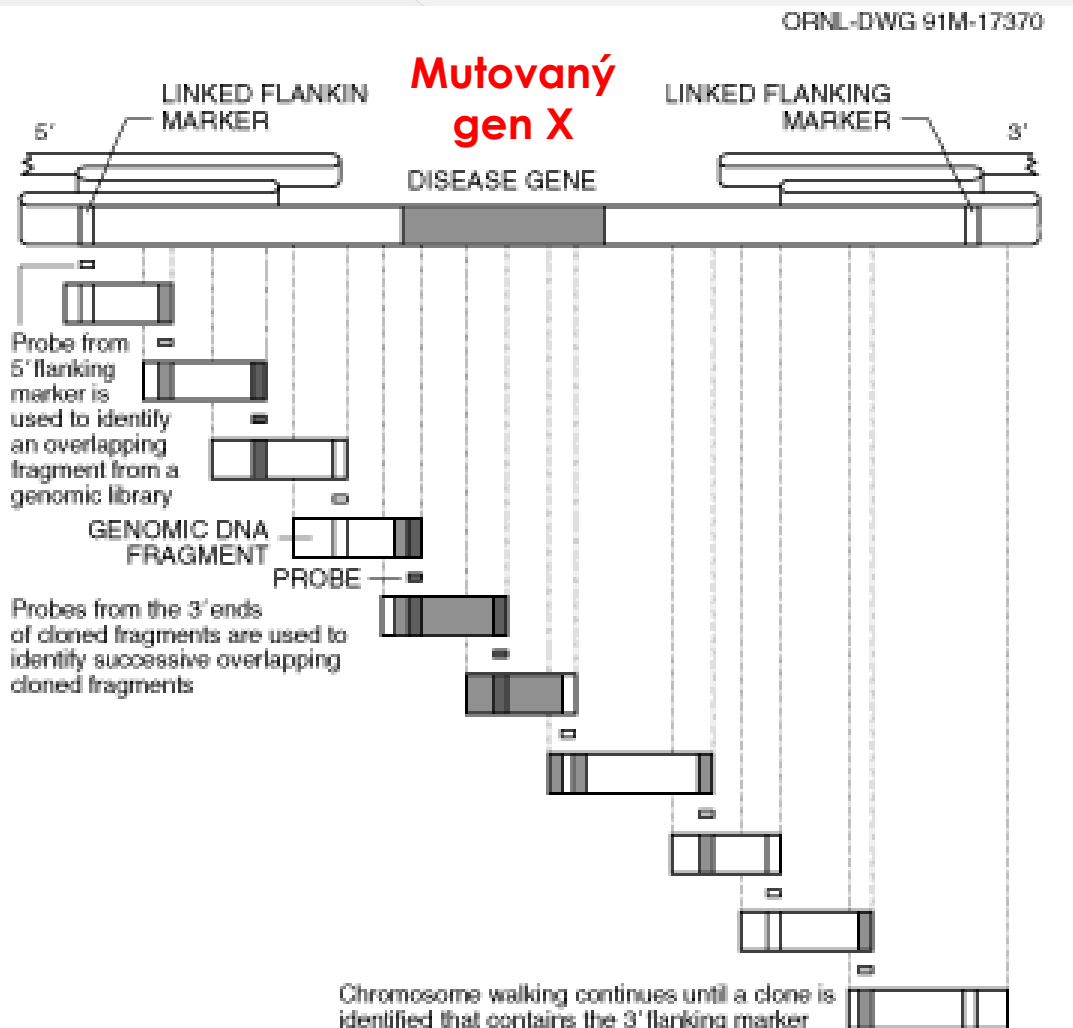
- > **Huntingtonova choroba** -je vzácné dědičné neurodegenerativní onemocnění mozku charakteristické nekoordinovanými trhavými pohyby těla a snížením mentálních schopností
- > **Cystická fibróza (CF)** je smrtelná lidská dědičná nemoc, která postihuje převážně dýchací a trávicí soustavu

Identifikace mutantního genu

- nalezení konkrétní sekvence nesoucí mutaci

- test na **alelismus**, tj. zkřížení testované mutantní linie s linií s „knock-out“ kandidátním genem a sledování, zda se zůstane mutantní fenotyp
- **molekulární komplementace** (u recesivních mutací)
 - > transformace mutantní rostliny sekvencemi standardní DNA z vymezeného úseku a určení, která obnoví standardní fenotyp
- **analýza celé sekvence DNA vymezeného genetického intervalu** → hledat změny, které způsobily mutaci
 - > SSCP (Single Stranded Conformational Polymorphism)
 - > HMA (Heteroduplex Mobility Assay)
 - > DGGE (Denaturing Gradient Gel Electrophoresis)
 - > dHPLC (denaturing High Performance Liquid Chromatography)
 - > hybridizace na čípech
 - > pyrosekvencování
 - > „chromosome walking“
- **sekvenování**

„Chromosome walking“



- nalezení optimálně dvou markerů obklopujících mutovaný gen
- knihovny velkých fragmentů genomové DNA (redundantní, náhodné):
 - > YACs, BACs
 - > = yeast (bacterial) arteficial chromosome, ~ 300 (100) kbp
- hledání překryvů na základě hybridizace koncové sekvence

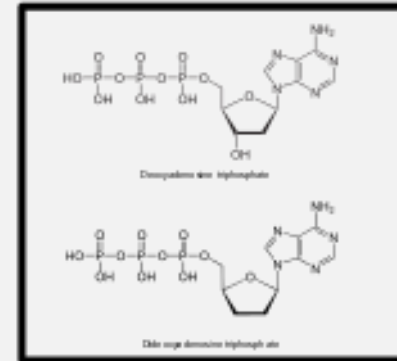
Sekvenování - klasické - Sanger

- enzymatická metoda je založena na sekvenaci pomocí **detekce ukončení prodlužujícího se vlákna DNA**
- sekvenaci předchází **příprava DNA knihovny**
 - > DNA sestřižena na kratší úseky
 - > naklonována do DNA vektorů
 - > amplifikována *in vivo* (v bakteriálních buňkách, ze kterých jsou plasmidy nesoucí klonované fragmenty následně extrahovány)
- sekvenace probíhá ve **čtyřech nezávislých reakcích**
- vznikají fragmenty DNA o různé délce zakončené značenými **dideoxynukleotidy**
- díky délce čtených úseků, které dosahují **700 až 800 bp** - stále **nejpoužívanější a nejpřesnější**

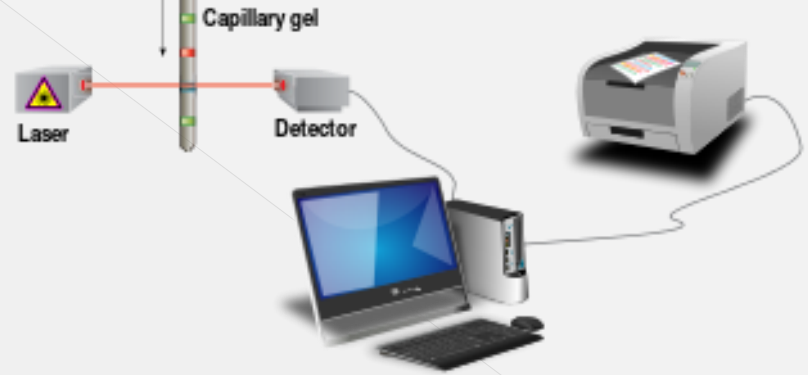
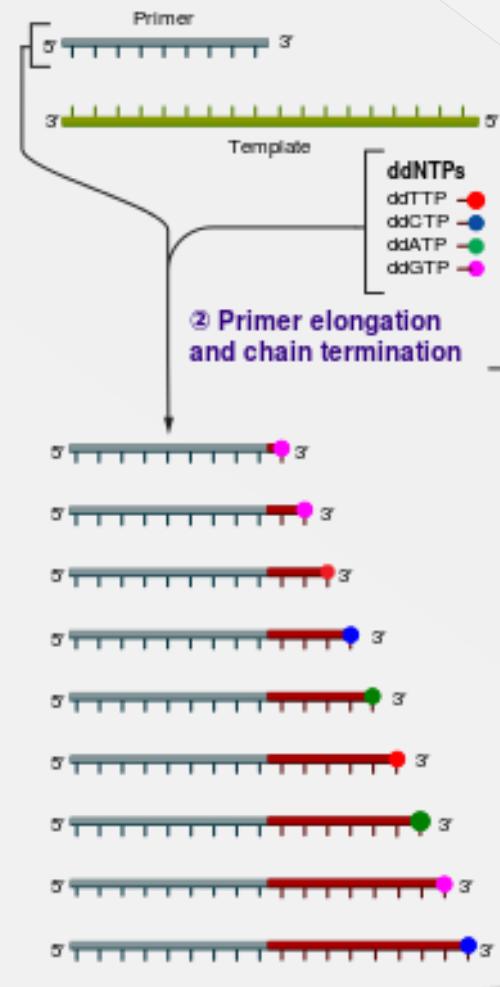
Sekvenování - klasické - Sanger

① Reaction mixture

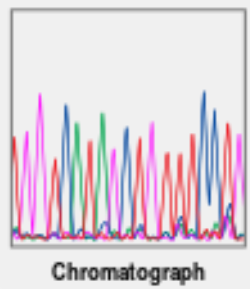
- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



③ Capillary gel electrophoresis separation of DNA fragments



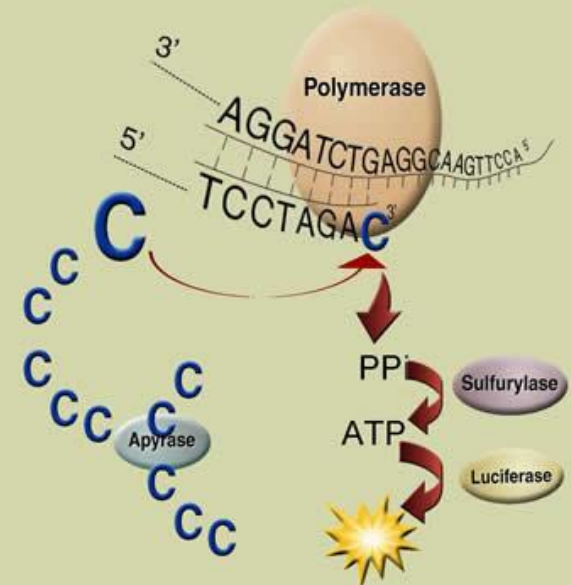
④ Laser detection of flouochromes and computational sequence analysis



Pyrosekvence - „Next-generation sequencing“

- na sekvenaci syntézou komplementární DNA
- liší se způsobem, jak je **detekováno začlenění daného nukleotidu** → nevyžaduje elektroforézu
- vyvinul v roce 1996 ve Stockholmu profesor **Pål Nyrén** se svým studentem **Mostafou Ronaghi**
- Reakční směs:
 - > DNA polymeráza, ATP sulfuryláza, luciferáza a apyráza;
 - > substráty adenosinfosulfát a luciferin
 - > do směsi jsou postupně vkládány nukleotidy různých typů (dATP, dGTP, dCTP, dTTP)
- světelné záření** → začlenění 1 nebo více nukleotidů tohoto typu vznikajícího řetězce začlenil jeden nebo více nukleotidů
 - > světlo vzniká v důsledku enzymatické reakce, na jejímž začátku je uvolnění pyrofosfátu z nově začleněného nukleotidu a na jejímž konci je spotřeba vzniklého ATP luciferázou k oxidaci luciferinu
- před přidáním dalšího nukleotidu je původní nukleotid (již obsažený v roztoku) rozložen

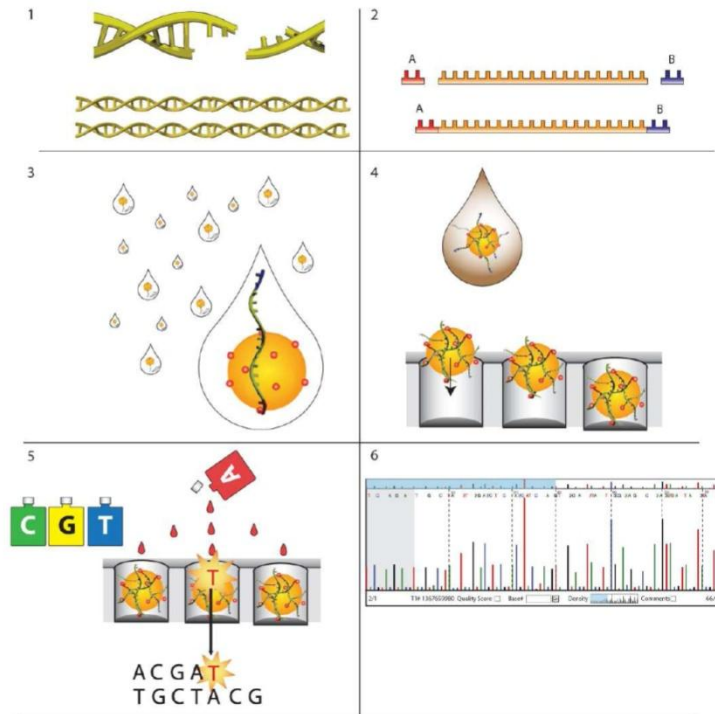
The Pyrosequencing Method



„Next-generation sequencing“

454 - 2005

- emulzní PCR
- pyrosekvenování



454 Genome Sequencers

FLX System

- 1 million of reads/run
- 400-650 bp/read
- 2 přístroje v ČR



GS Junior

- 0.1 millions of reads/run
- 400 bp/read

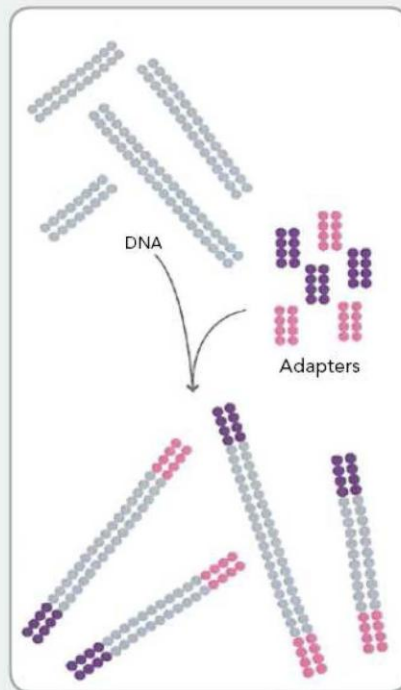
„Next-generation sequencing“

Solexa (Illumina) - 2007



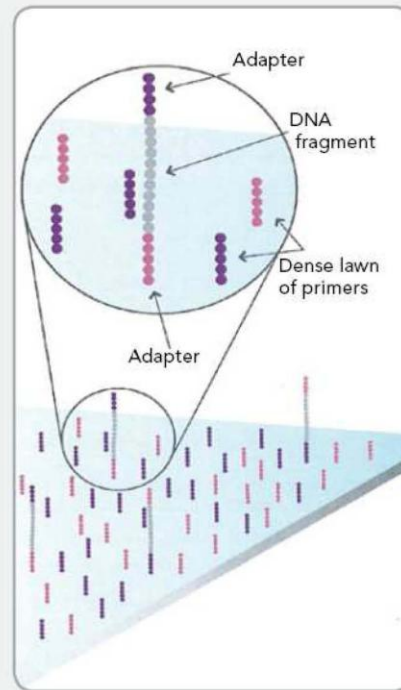
- mŭstková „bridge“ PCR
- sekvenovn pomocí DNA syntzy

1. PREPARE GENOMIC DNA SAMPLE



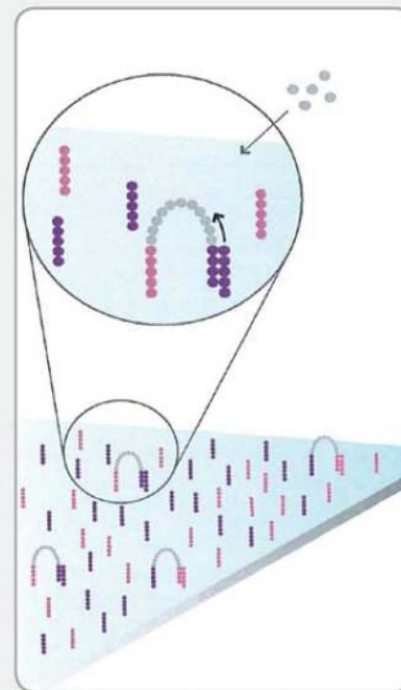
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



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

3. BRIDGE AMPLIFICATION



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

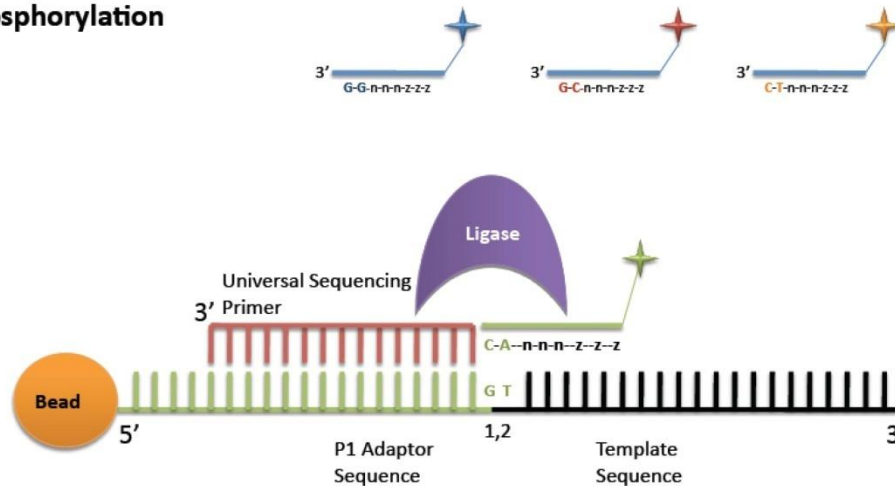
„Next-generation sequencing“

SOLiD (2008)



- emulzní PCR,
- sekvenování pomocí ligace

**First Cycle with primer 1-
Hybridization, Ligation, de-
Phosphorylation**



„Next-generation sequencing“

Přehled současných metod sekvenování nové generace

Platform	Year	Sequencing Method	Amplification	Detection	Features
454	2005	Pyro-sequencing	Emulsion PCR	Light	First NGS
Illumina	2007	Synthesis	Bridge PCR	Light	90% of Market
SOLiD	2008	Ligation	Emulsion PCR	Light	Lowest Error Rate
Ion Torrent	2010	Synthesis	Emulsion PCR	Hydrogen Ion	Semiconductor Chip
Pacific Biosciences	2010	Synthesis	None = Single Molecule	Light	Anchored Polymerases
Oxford Nanopore	2012	Nanopore	None = Single Molecule	Electrical Conductivity	“Run Until” Sequencing

Modified from T. C. Glenn. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11: 759-769.

„Next-generation sequencing“

Instrument	Run time	Millions of Reads/run	Bases / read	Yield MB/run
3730xl (capillary)	2 hrs	0.000096	650	0.06
PacBio RS	2 hrs	0.01	860 – 1,500	5-10
454 GS Jr. Titanium	10 hrs	0.1	400	50
Ion Torrent – 314 chip	2.5 hrs	0.25	200	50
454 FLX Titanium	10 hrs	1	400	400
454 FLX+	20 hrs	1	650	650
Ion Torrent – 316 chip	3 hrs	1.6	200	320
Illumina MiSeq	26 hrs	4	150+150	1200
Ion Torrent – 318 chip	4.5 hrs	4	200	800
Illumina GAIIx	14 days	300	150+150	96,000
SOLiD – 5500xl	8 days	>1,410 ^d	75+35	155,100
Illumina HiSeq 1000	8.5 days	≤1500	100+100	≤300,000
Illumina HiSeq 2000	11.5 days	≤3000	100+100	≤600,000

2012. NGS Field Guide (www.molecularecologist.com)

„Next-generation sequencing“

Chybovost jednotlivých sekvenačních metod

Platform	Primary Errors	Single-pass Error Rate (%)	Final Error Rate (%)
3730xl (capillary)	Substitution	0.1-1	0.1-1
454	Indel	1	1
Illumina	Substitution	~0.1 (85% of reads)	~0.1 (85% of reads)
SOLiD	A-T bias	~5	≤0.1
Ion Torrent	Indel	~1	~1
PacBio RS	CG deletions	~15	≤15
Oxford Nanopore	Deletions	≥4	4

2012. NGS Field Guide (www.molecularecologist.com)

Mnohonásobné prosekvenování jednoho úseku DNA kompenzuje vyšší chybovost sekvenačních metod nové generace.

Využití sekvenačních metod nové generace

◎ celé genomy

- > resekvenování → hledání sekvenčních variant v lidské populaci
- > *de novo* sekvenování → nemodelové organismy

◎ transkriptom (RNA-seq)

- > Identifikace dosud neznámých transkriptů
- > míra transkripce jednotlivých genů - přesnější než „microarrays“

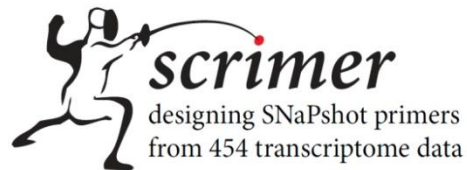
◎ cílené → určitá část genomu, skupiny genů

- > náhodné oblasti genomu → na základě délky po restričním štěpení genomové DNA (RAD-seq)
- > hybridizace k cca 100bp prábám se vyberou fragmenty DNA, kt. se sekvenují (Hyb-seq)

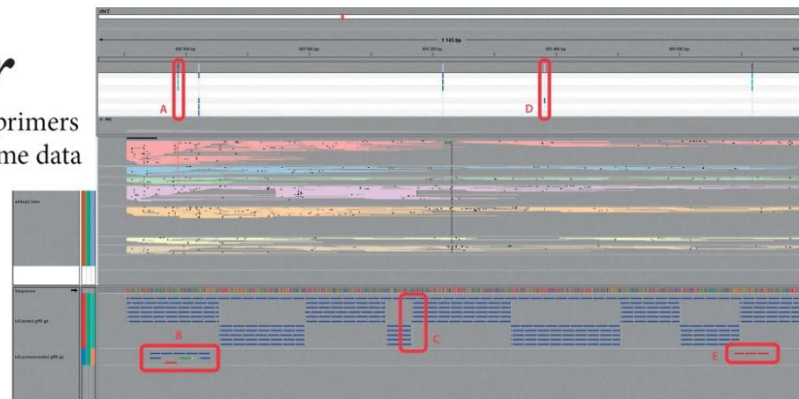
Využití sekvenačních metod nové generace

Identifikace SNPs (Single Nucleotide Polymorphisms)

- Sekvenování vzorků DNA pocházejících z mnoha jedinců.
- Vzorky lze individuálně tagovat - Multiplex Identifier Sequences (MIDs).
- Vhodné snížení komplexity
 - RNA-Seq (SNPs v kódujících oblastech)
 - RAD-Seq (SNPs v náhodných oblastech genomu, především nekódujících, větší variabilita)
 - Hyb-Seq (SNPs ve vybraných genech či genomové oblasti)



Libor Mořkovský,
Katedra Zoologie, PřF UK



Nepřímá „reverse“ genetika

- dnes → post-genomická éra
- známe geny (sekvence)
- neznáme
 - > funkce genů
 - většinou >50% predikovaných genů u eukaryot
 - > fenotypy, které způsobí mutace v těchto genech

Velkokapacitní náhodná mutageneze a „screening“

- ◉ systematická mutageneze –postupné
- ◉ EMS mutageneze
 - > místo hledání urč. fenotypu (přímá), se hledá gen zájmu se **změnami v nukleotidech**
 - > běžně „screen“ 1000 nebo 10000 jedinců
 - > využití PCR genu zájmu → hledání drobných změn v **migraci PCR produktu** na gelu či koloně
 - > všechny změny nevyřadí („knockout“) gen, některé se neprojeví -„silent“, na neesenciálních AMK pozicích
 - > metody běžně užívané:
 - **DHPLC** –“Denaturing High Performance Liquid Chromatography“
 - **DGGE** – „Denaturing Gradient Gel Electrophoresis“
 - **SSCP** – „Single-Stranded Conformation Polymorphism“

TILLING

(Targeting induced local lesions in genomes)

- získání (nalezení) bodových mutací ve vybraném genu
- bodové mutace
 - > potenciální změny regulace, interakce, ...
- představena na ***Arabidopsis thaliana*** (McCallum et al, 2000)
- Princip
 - > náhodná indukce bodových mutací (EMS)
 - > následné hledání linií s mutací v cílovém genu pomocí PCR a heteroduplexní analýzy

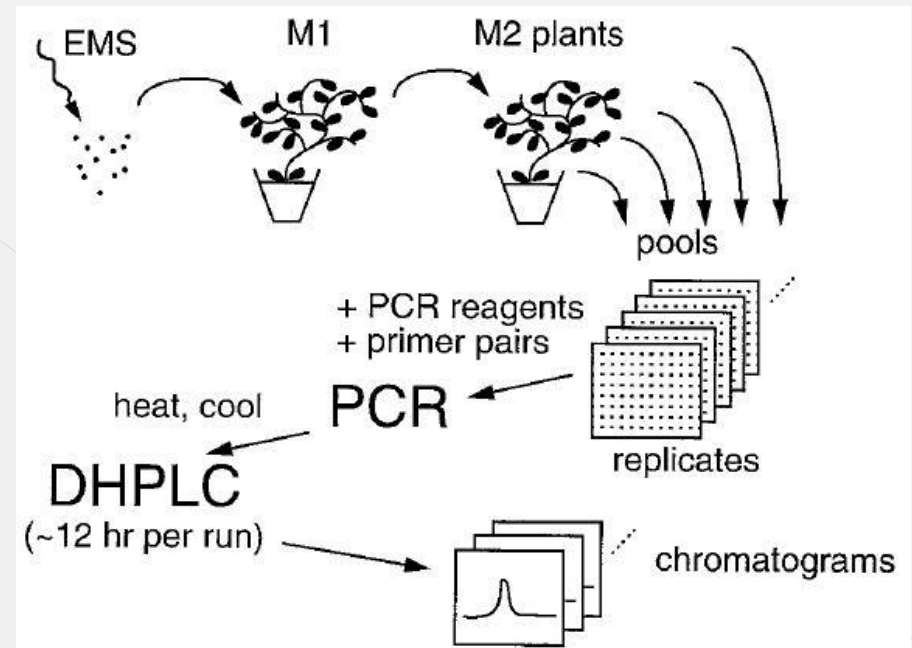


Figure 1. Schematic depicting the TILLING strategy applied to a plant such as *Arabidopsis*.

(McCallum et al, 2000)

TILLING – detekce mutací, strategie

1 Seeds are mutagenized to induce point mutations throughout the genome.



2 A founder population is grown from mutagenized seeds.



3 Founder population is self-fertilized to produce a crossed population.



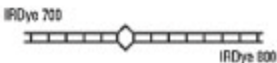
4 Seeds from the crossed population are stored and DNA samples are collected in 96-well plates.



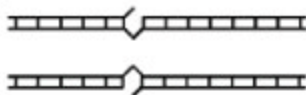
5 Up to eight 96-well plates are pooled into one and the samples (768) subjected to PCR with two gene-specific primers labeled with different IRDye® infrared dyes.



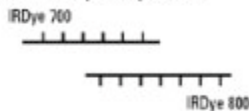
6 Resulting amplicons are heated and cooled, resulting in heteroduplexes between wild type and mutant samples.



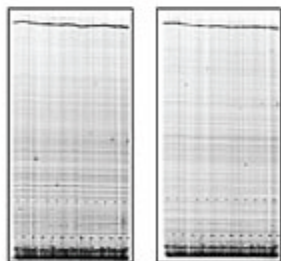
7 CEL I nuclease is used to cleave at base mismatches.



8 Samples are denatured and electrophoresed on a LI-COR 4300 DNA Analysis System.

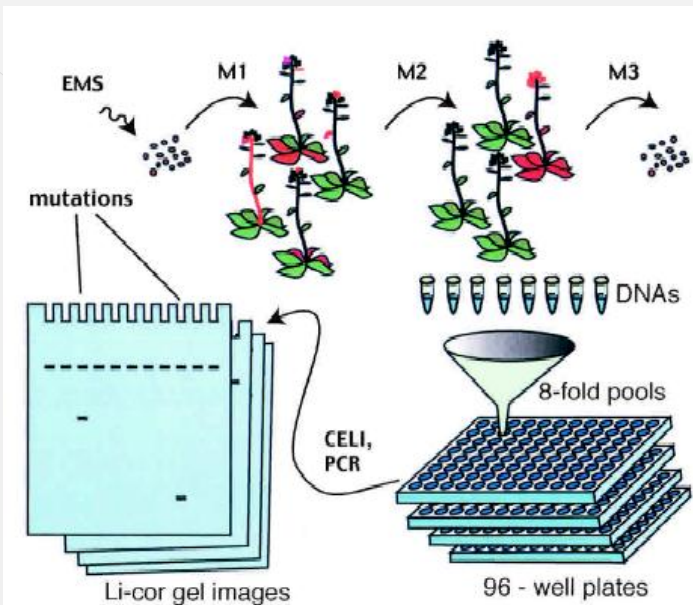


9 In lanes that have a mutation in the pool, a band will be visible below the wild type band on the IRDye® 700 infrared dye image. A counterpart band will be visible in the same lane on the IRDye® 800 infrared dye image. This band is the cleavage product labeled with IRDye® 800 infrared dye from the complementary DNA strand. The sum of the length of the two counterpart bands is equal to the size of the amplicon, which makes it easy to distinguish mutations from amplification artifacts.



10 After detection of a mutation in a pool (lane), the individual DNA samples in the pool are screened again to find out which of the eight pooled samples from the crossed population has the mutation.

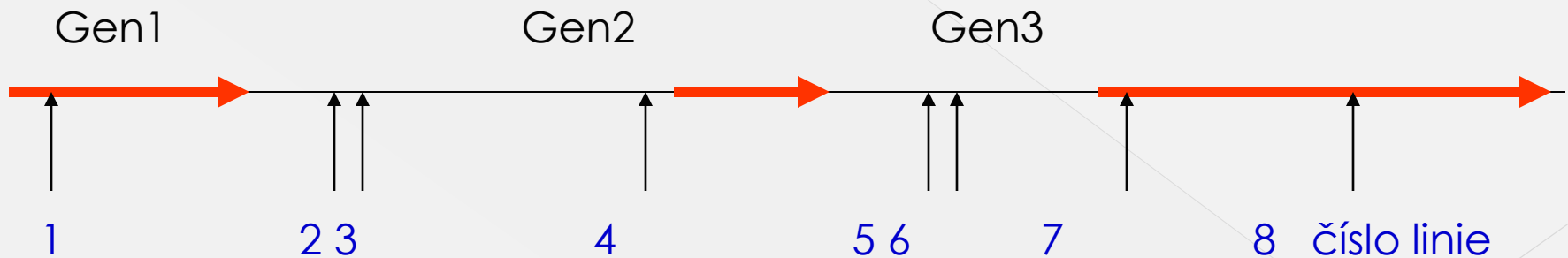
- amplifikace cílového fragmentu (koncově značené primery)
- reasociace s wt DNA
- štěpení heteroduplexu (ss nuclease-CEL I)
- elektroforéza (separace podle velikosti na různých platformách; LI-COR)
- vizualizace koncově značených fragmentů



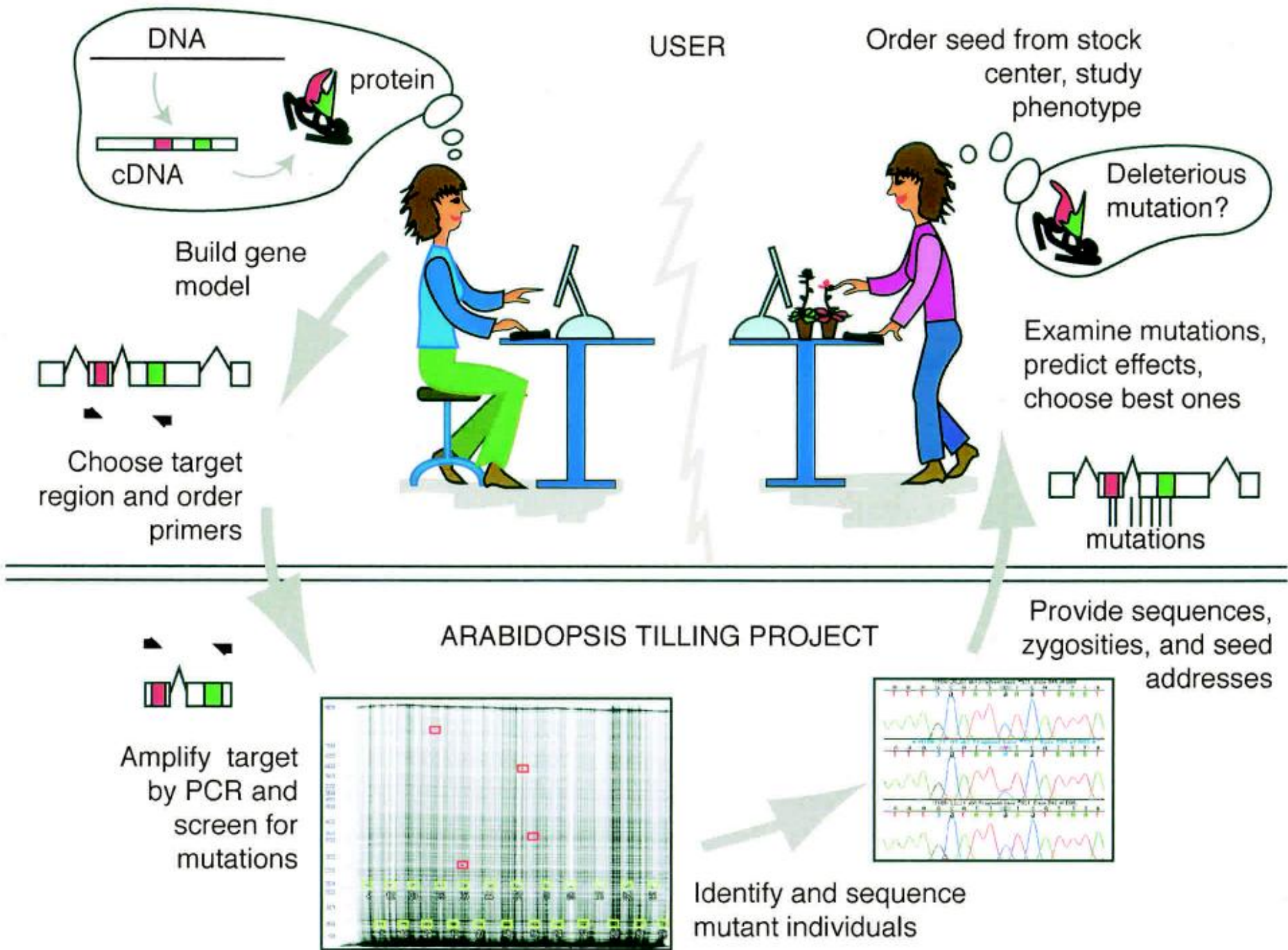
Metoda pracuje s amplifikovanými fragmenty o velikosti kolem 1 kb a mutaci je schopna analyzovat s přesností na 10 bp (Henikoff et al, 2004).

Rostliny s inzercí v určitém genu

- veřejně dostupné kolekce mutantů na různých místech genomu
- inzerce ve většině genů ***Arabidopsis thaliana***
- výběr *in silico*, objednání semen rostlin s inzercí v určitém genu



↑ = místa inzerce T-DNA v jednotlivých mutantních liniích



Děkuji za pozornost 😊