



Analýza genomů mikrobiálních komunit

doc. RNDr. Milan Bartoš, Ph.D.

bartosm@vfu.cz

Přírodovědecká fakulta MU, 2012

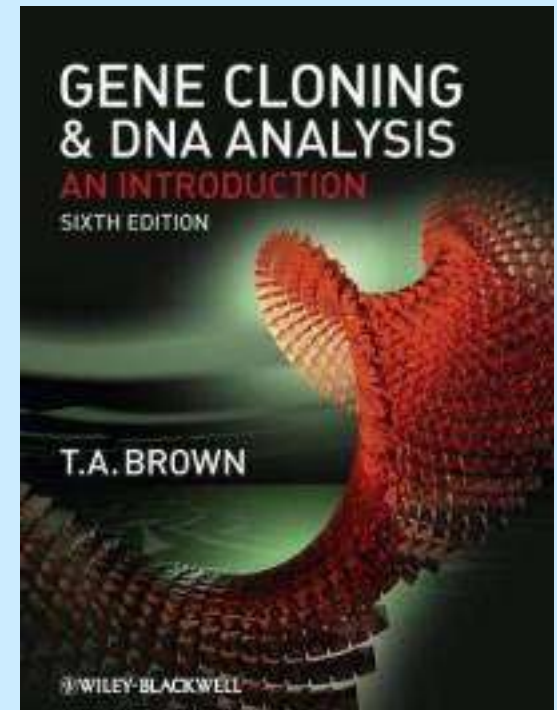
Obsah přednášky

- 1) Identifikace genů v genomu**
- 2) Určení funkce neznámého genu**
- 3) Studie transkriptomu a proteomu**
- 4) Interakce protein-protein**
- 5) Analýza mikrobiálních komunit**
- 6) Příklady aplikací a výsledků**



Doporučená literatura

Brown (2010): Gene Cloning & DNA Analysis. Wiley-Blackwell, Sixth edition



A odkazy na jednotlivé články

Historická poznámka

1975

- První genom, fág Φ X 174

1995

- První buněčný genom, *Haemophilus influenzae*

do roku 2001

- Kvasinky, *Drosophilla*, *Caenorhabditis*,
Arabidopsis, *Homo*

21. století

- Post-genomika neboli funkční genomika
- Bioinformatika = molekulární biologie *in silico*

O co jde?

Hledáme pozice a funkce všech genů
v genomu

1996 – *S. cerevisiae* má 6 000 genů,
funkce známa u 3 600 z nich

Zbytek odhalen *in silico*, ale ne u všech



Jak identifikovat geny v genomu

Znám sekvenci aminokyselin
u polypeptidu?

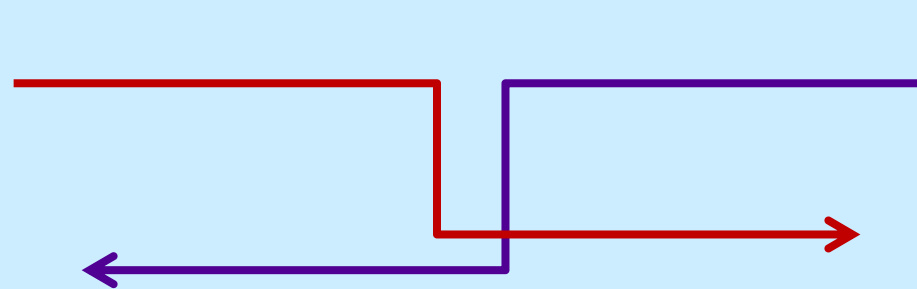
Znám sekvenci
cDNA?

↓
ANO

↓
NE

↓
Predikuji sekvenci
genu

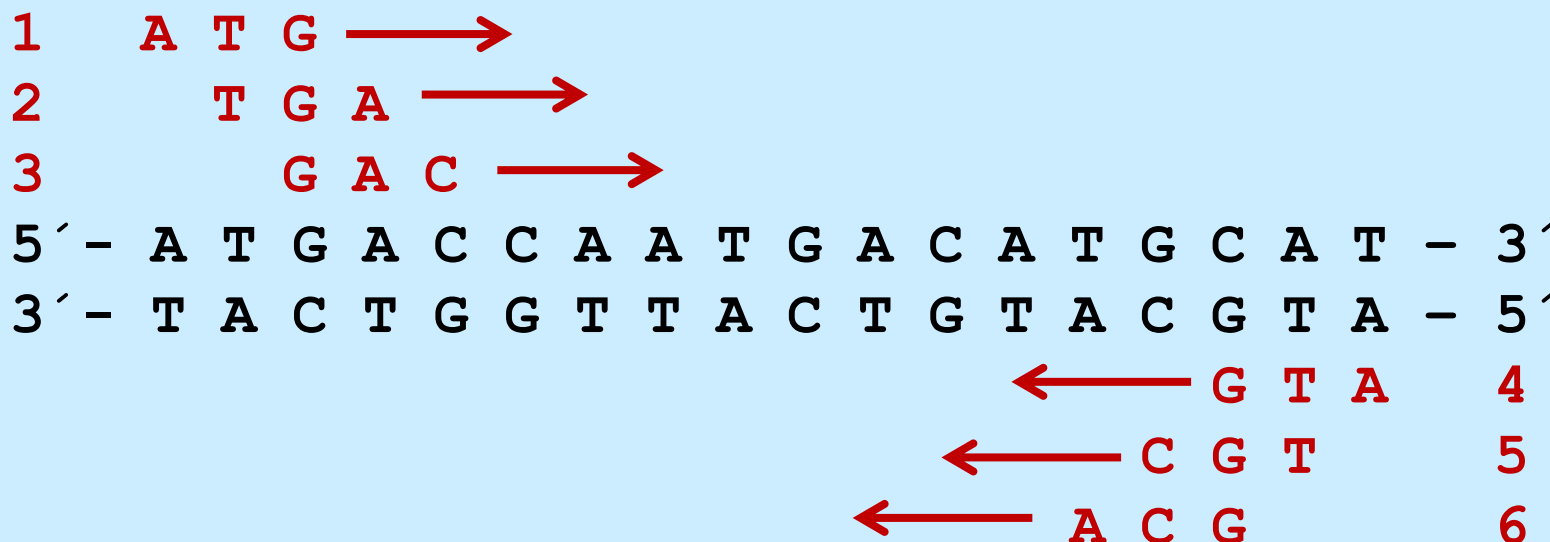
↓
Hledám ORF



Hledání ORF

ORF = open reading frame = otevřený čtecí rámec

- dlouhá sekvence kodónů vyznačená iniciačním (obvykle, ale ne vždy ATG) a nesmyslnými (TAA, TAG, TGA) kodony
- Prohledání je třeba udělat v 6 variantách!



Jak často se vyskytne v nahodilé sekvenci jeden z nesmyslných kodonů?



- 1) Kodonů je 4^3 , tj. 64
- 2) Kterýkoli z nesmyslných kodonů se vyskytne 1 x za 64 kodonů, tj. 192 nukleotidů



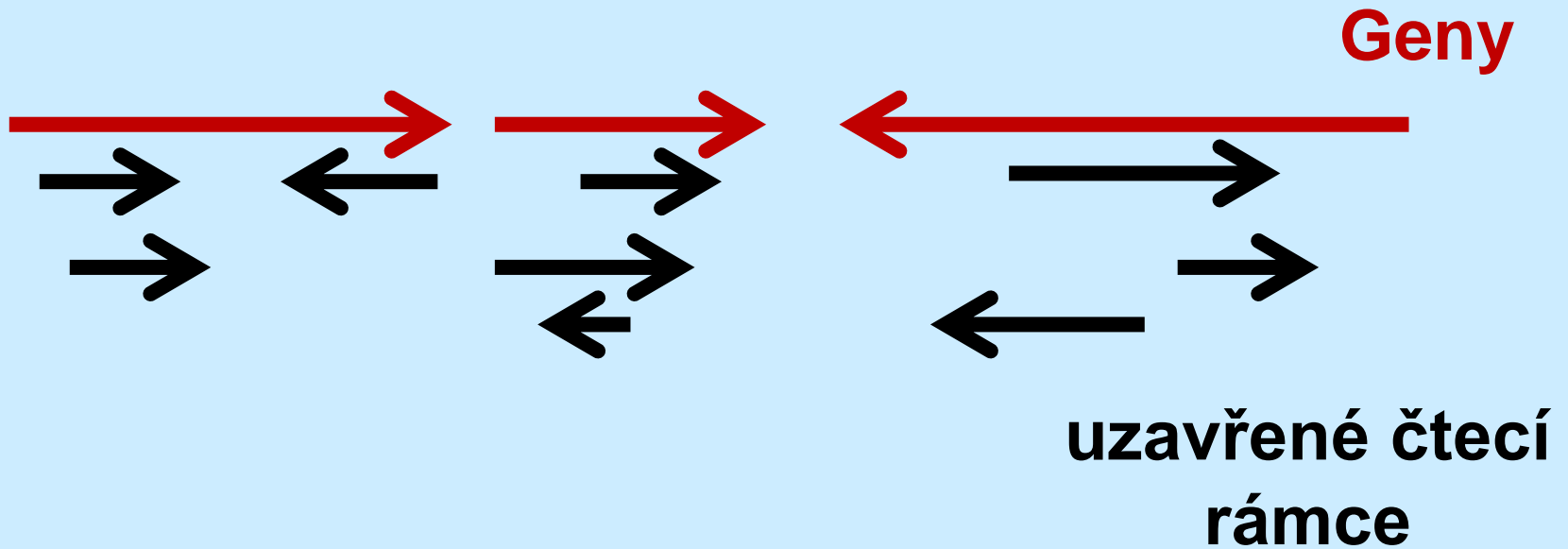
Délky ORF

V nahodilé sekvenci nebude ORF delší než 30-40 kodonů a ne všechny budou obsahovat ATG

- **Průměrná délka genu u *Escherichia coli* je 317 kodonů**
- **Průměrná délka genu u *S. cerevisiae* je 483 kodonů**
- **Průměrná délka genu u člověka je 450 kodonů**

Bakteriální ORF

Genomy bakterií jsou kompaktní, dlouhé ORF jsou zpravidla místy výskytu genů, krátké ORF (uzavřené čtecí rámce) nejsou geny



Eukaryotické ORF

- mnoho intergenových oblastí - více krátkých ORF
- exony, introny

Jak odlišit gen od ORF?

- Preference kodonů (codon bias)
- Hranice exon-intron
- Regulační sekvence proti směru transkripce
- Hledání homologických sekvencí
- Srovnání sekvencí s příbuznými genomy

Preference kodonů

- **Některé aminokyseliny jsou kódovány více kodony – kodonové rodiny**
- **Některé kodóny jsou ale využívány častěji než jiné, tzv. vzácné kodony**
- **Jestliže některý ORF obsahuje vzácný kodón častěji, pak tento ORF pravděpodobně není gen**

Tabulky s preferencemi kodonů

<http://www.kazusa.or.jp/codon/>



**Podívejte se na tabulku
preferenčních kodonů pro
*Mycobacterium tuberculosis H37Rv***

- 1) Jak často je využíván kodon AUG?**
- 2) Který kodon je využíván nejčastěji?**
- 3) Který kodon je využíván nejméně často?**

- 1) AUG = 18,4%**
- 2) GCC = 59,8%**
- 3) UAA = 0,5%**





***Podívejte se na tabulku
preferenčních kodonů pro
Mycobacterium tuberculosis H37Rv***

***Vypracujte tabulku četnosti využití
jednotlivých kodonů v kodonových
rodinách***





***Podívejte se na tabulku
preferenčních kodonů pro gen pro
katalázu u Mycobacterium
tuberculosis H37Rv***

- 1) Jak často je využíván kodon AUG?***
- 2) Který kodon je využíván nejčastěji?***
- 3) Který kodon je využíván nejméně často?***

- 1) AUG = 0,0%**
- 2) GCC = 67,6%**
- 3) Řada kodonů 0,0%, řada 6,8%**



Hranice exon-intron

- Tyto hranice se vyznačují přítomností tzv. konvenčních sekvencí

Hranice lze hledat prostřednictvím NCBI

- ***Saccharomyces cerevisiae* PHO5**
- Zadejte v kategorii „gene“
- Hranice hledejte v odkazu „Genomic context“ – „MapViewer“

Regulační sekvence proti směru transkripce

- **Konvenční sekvence jsou velmi variabilní, proto je použití tohoto nástroje poměrně problematické**

Hledání homologických sekvencí

Předpokládá se, že pokud dva geny (z různých organismů), které mají podobné funkce, mají podobné sekvence, pak mají společný původ.

S ohledem na degeneraci genetického kódu je vhodnější pracovat na úrovni sekvence aminokyselin



- **20 aminokyselin oproti 4 nukleotidům znamená, že když se vyskytnou dvě stejné aminokyseliny je méně pravděpodobné, že jde o náhodu**

Srovnání homologie DNA-protein

G A P G M W L R L A A G S F E H A G
GGTGCACCCGGTATGTGACTGCGATTAGCAGCGGGATCATTTCAGCATGCAGGG
* * ***** **** * ** * ** * ** * * ** * ** *
GATACACCCGTATTTGACAGCAATTTGCAGGGGGATGATTGCACCATGGAGCG
D T P R I W E E P A G G W L H H G A



Vypočítejte shodu (v %) pro sekvenci nukleotidů a sekvenci aminokyselin

- Nukleotidové sekvence jsou shodné ze 76% (41/54)
- Sekvence aminokyselin jsou shodné z 28% (5/18)



Je daný ORF genem?

Prohledáme databázi pomocí BLAST

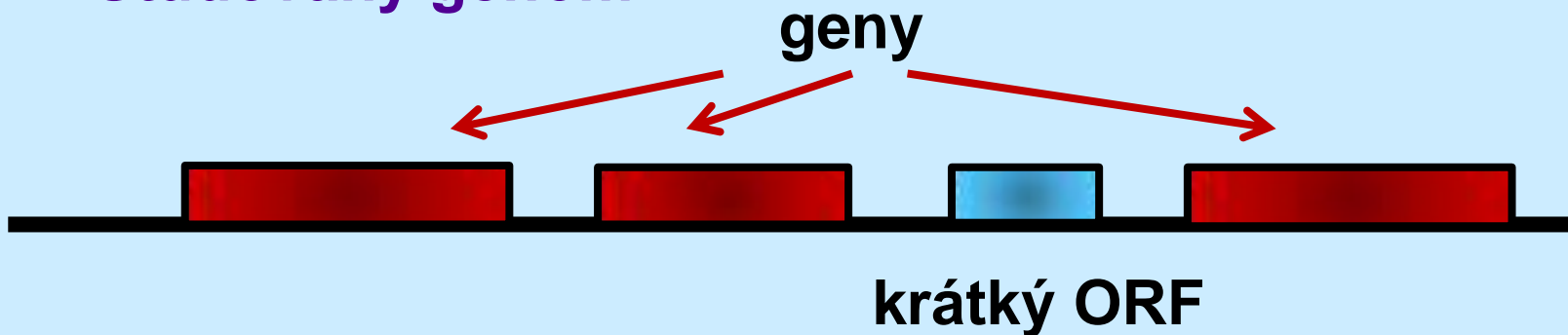
Pokud je sekvence delší než 200 AA a je z 30% a více identická s jednou ze sekvencí v databázi, jsou tyto sekvence téměř jistě homologické a zkoumaný ORF je skutečný gen



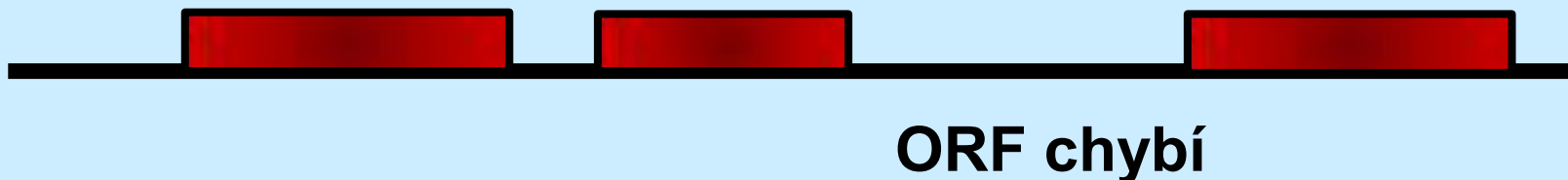
Srovnání sekvencí s příbuznými genomy

- U příbuzných organismů existují homologické geny
- ORF bez homologa zřejmě není gen, platí především pro krátké ORF

Studovaný genom



Příbuzný genom



Komparativní genomika

Výše popsáný způsob vyhledávání homologických sekvencí

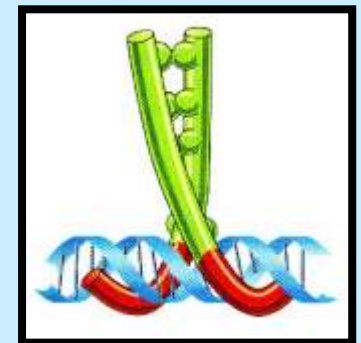
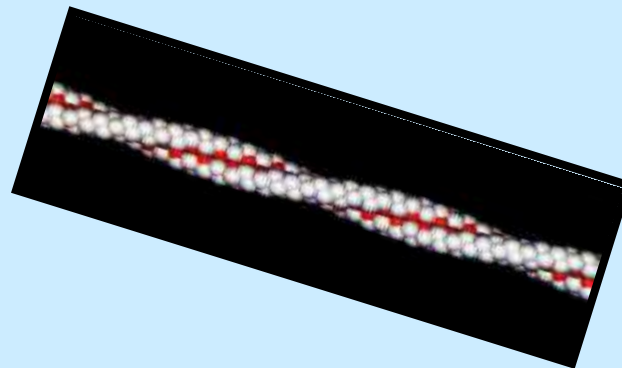
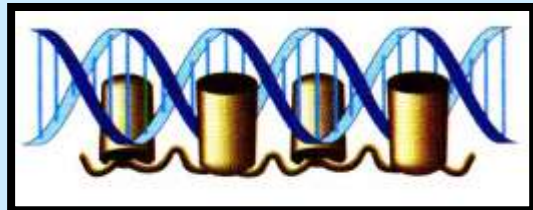
- **Byl použit k umístění genů v genomu *Saccharomyces cerevisiae***
- **Následně i u dalších kvasinek – *S. paradoxus*, *S. mikatae*, *S. bayanus***

Určení funkce neznámého genu

- **Určení homologie s genem, jehož funkce už je známa**
- **Bioinformatické studie**
- **Metodami reverzní genetiky**
- **Knock-outováním genu**

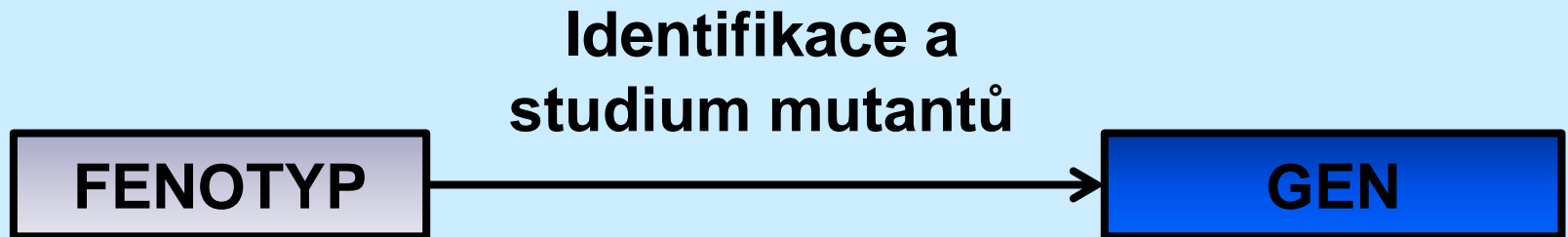
Bioinformatické studie

- **Zatím v počátcích**
- **Na základě sekvence nukleotidů nebo aminokyselin lze predikovat přítomnost α -šroubovic a β -struktur a odhadnout funkci**
- **Proteiny vázající se k membránám = α šroubovice**
- **Zinkové prsty, svinutý helix, leucinový zip, otočka-helix-otočka = regulační proteiny**

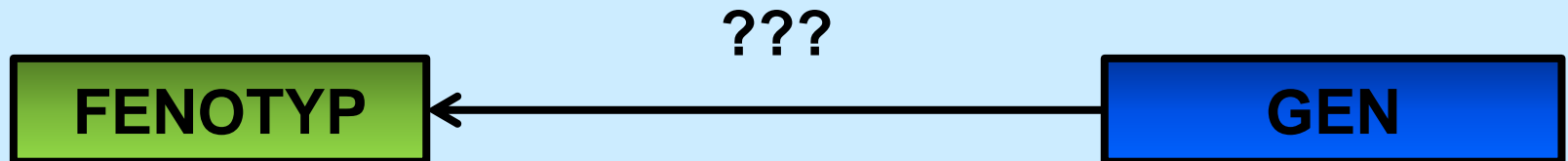


Reverzní genetika

Standardní genetika



Reverzní genetika



Reverzní genetika

Standardní genetika

- **Gen zodpovědný za fenotyp je identifikován určením, které geny zodpovědné za mutantní fenotyp jsou v organismu inaktivovány**
- **Získání mutantů (chemické mutageny, záření, přirozeně v populaci)**
- **Křížení a stanovení genetické mapy**
- **Bližší charakterizace molekulárně-biologicky**

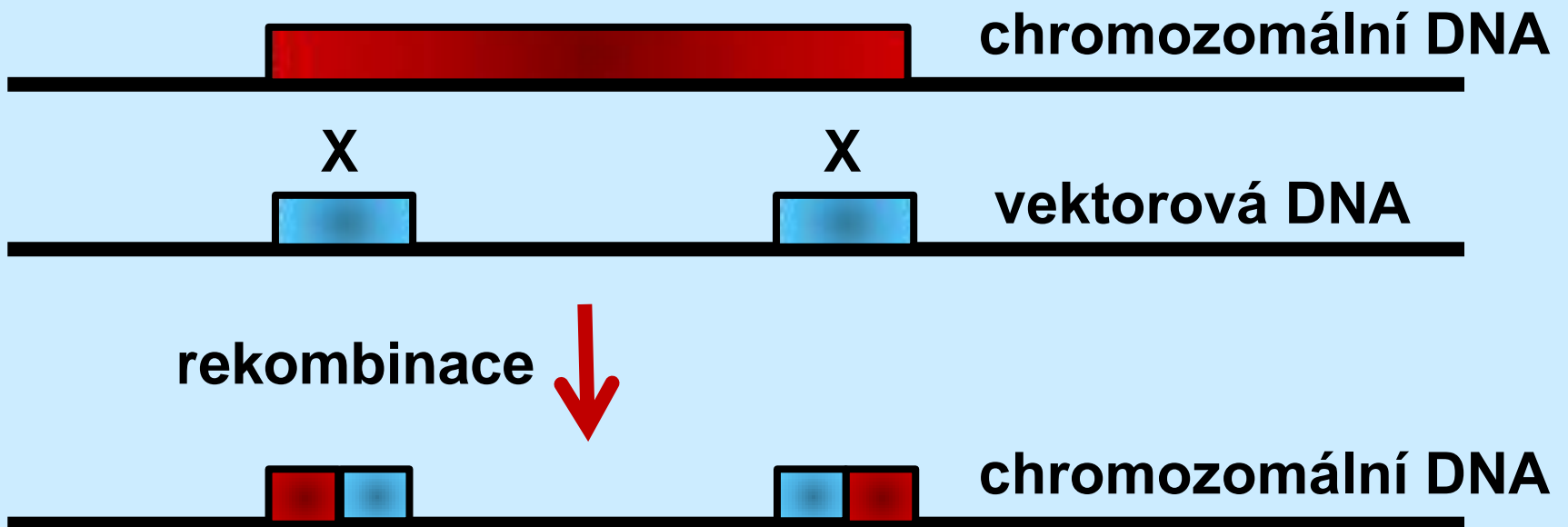
Reverzní genetika

- **Výchozím bodem není fenotyp, ale gen**
- **Vyvolání mutace v genu a identifikace výsledných fenotypových změn**

Knock-outování genu

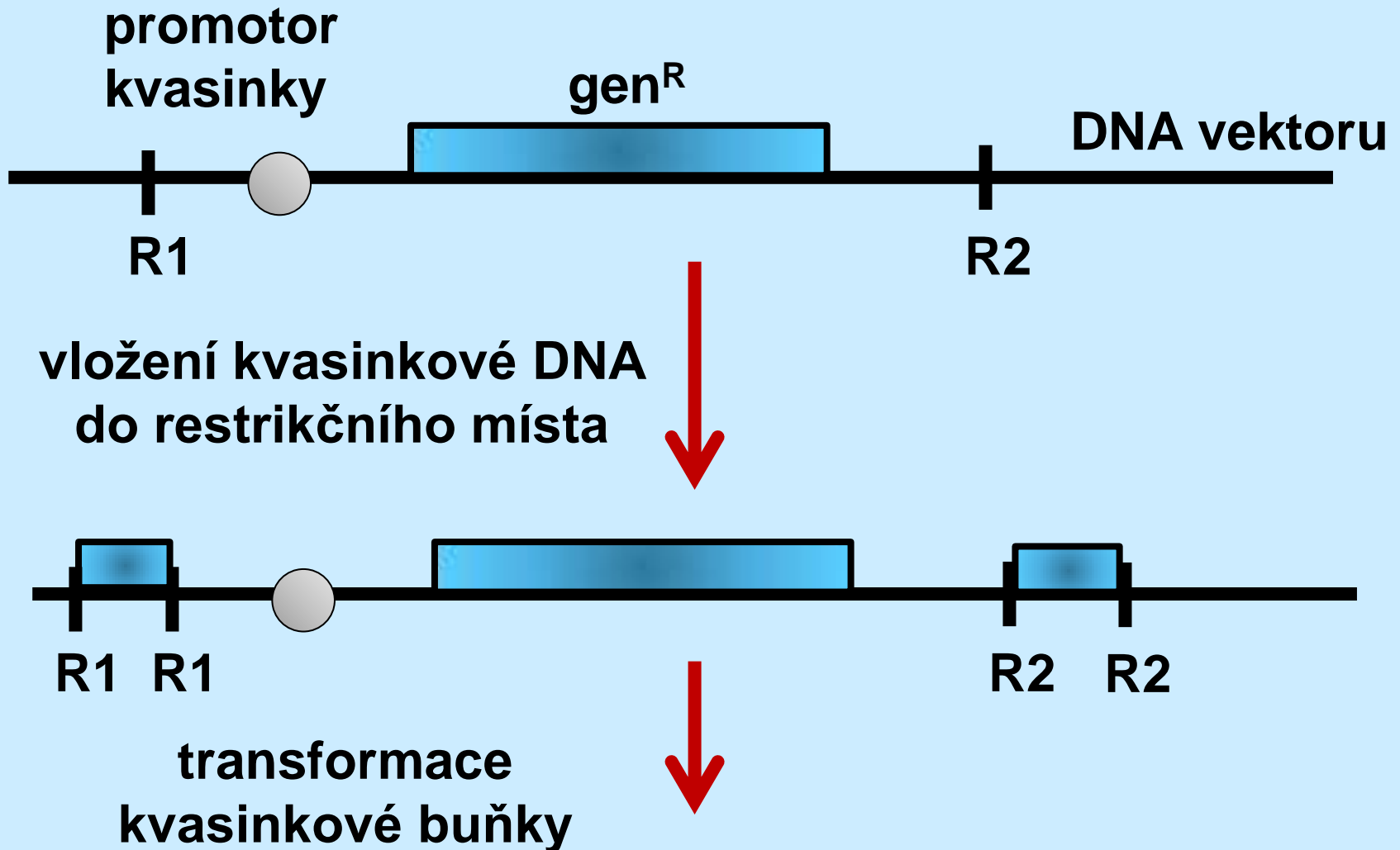
Deletovaná forma genu je využita k vyřazení funkční formy genu v organismu

- Prove se homologickou rekombinací
- Sleduje se změna fenotypu po knock-outu



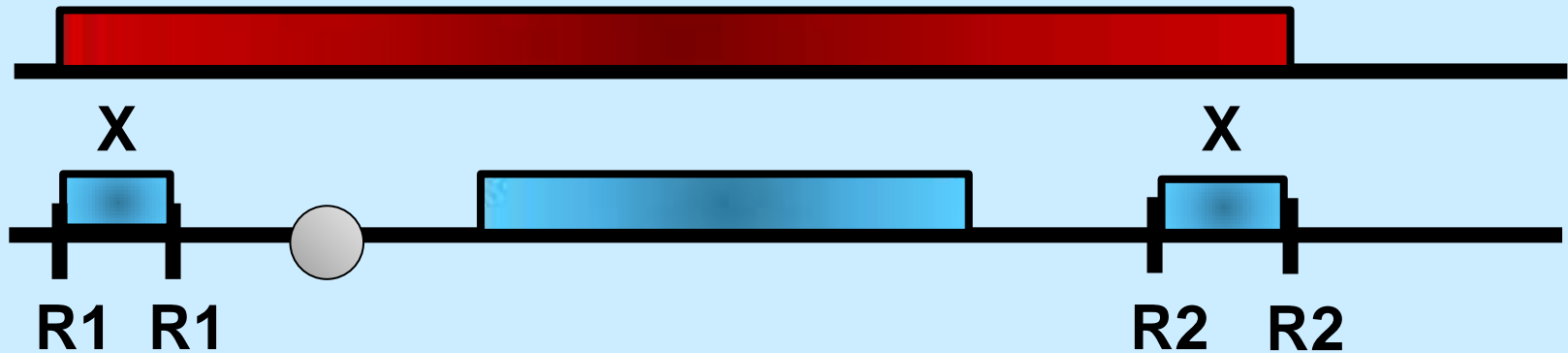
Knock-out u S. cerevisiae

Deleční kazety nesoucí gen rezistence k ATB



Knock-out u S. cerevisiae

chromozómová kopie cílového genu



homologickou rekombinací
dojde k přerušení cílového
genu kopií genu rezistence



exprese genu rezistence

Studium transkriptomu

**Transkriptom je veškerá mRNA buňky
a odráží celkový obraz genové
exprese v této buňce**



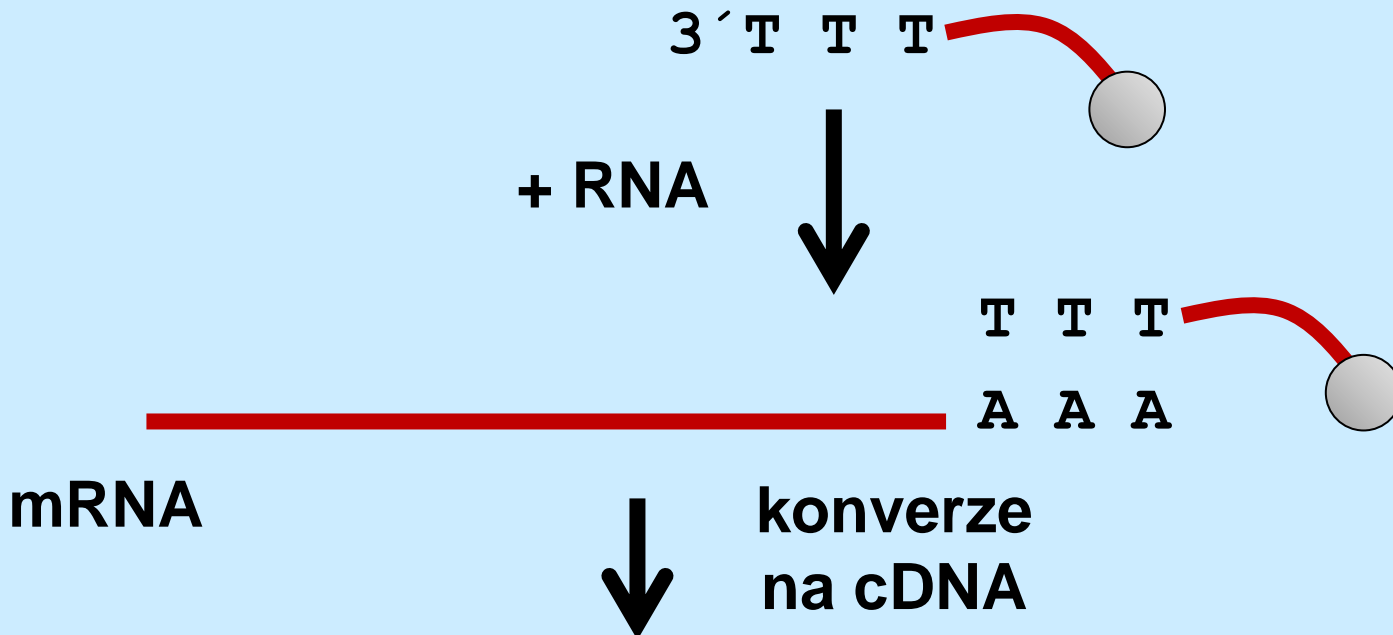
Studium transkriptomu

- **Transkriptom může být velmi složitý**
- **Může obsahovat stovky až tisíce různých mRNA**
- **Klasický postup zahrnoval přípravu cDNA a její srovnání s genomovou DNA**
- **Postup velmi zdlouhavý**
- **Analýza sekvenováním**
- **Analýza s využitím čipů**

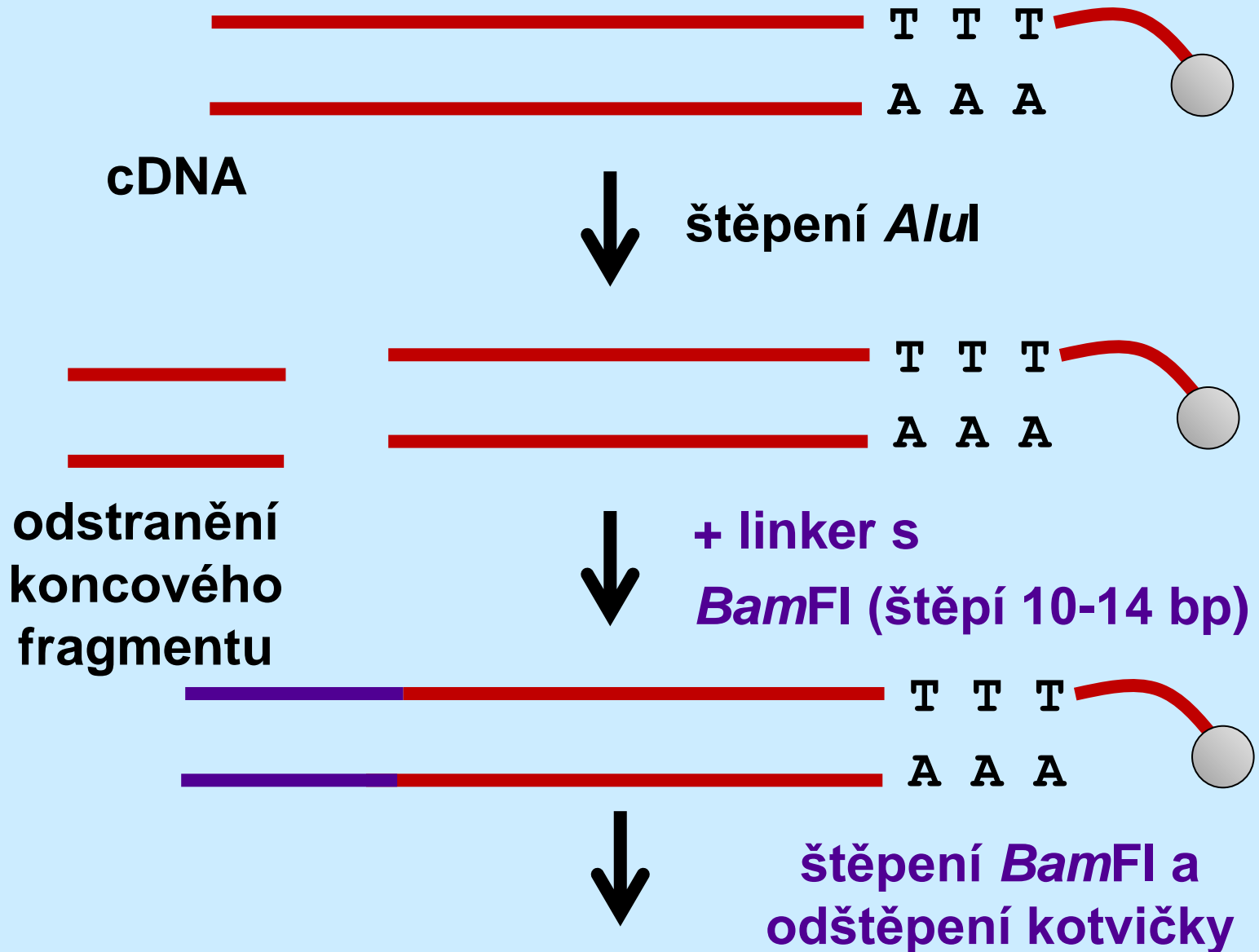
Metoda SAGE

serial analysis of gene expression

- Nevyužívá cDNA, ale krátké 12 bp sekvence
- mRNA jsou imobilizovány na celulózové kuličky s oligo (dT)



Metoda SAGE



Metoda SAGE



shromáždění fragmentů, ligace s dalšími fragmenty –
vznik katenanů a jejich sekvenování



Po sekvenování lze jednotlivé sekvence
odlišit, protože jsou odděleny místem *BsmFI*

<http://www.sagenet.org/findings/index.html>

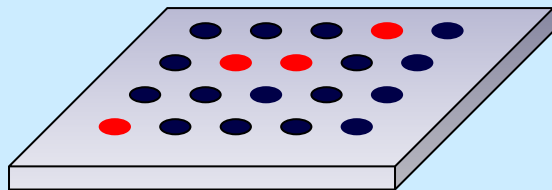
**Metoda SAGE byla použita ke studiu
transkriptomu *S. cerevisiae***

Cell 88, 243-251, 1997

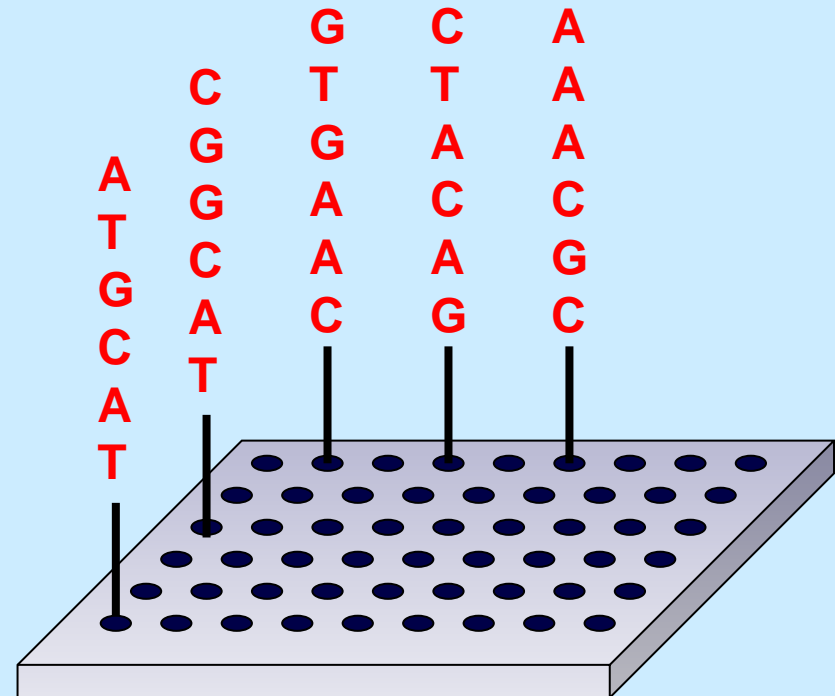


Nastupuje éra čipů

- 1) Všech 6 000 kvasinkových genů lze umístit na microarray 80 x 80
- 2) Alternativně lze využít DNA čipů na silikonu



sekvence
jednotlivých
genů

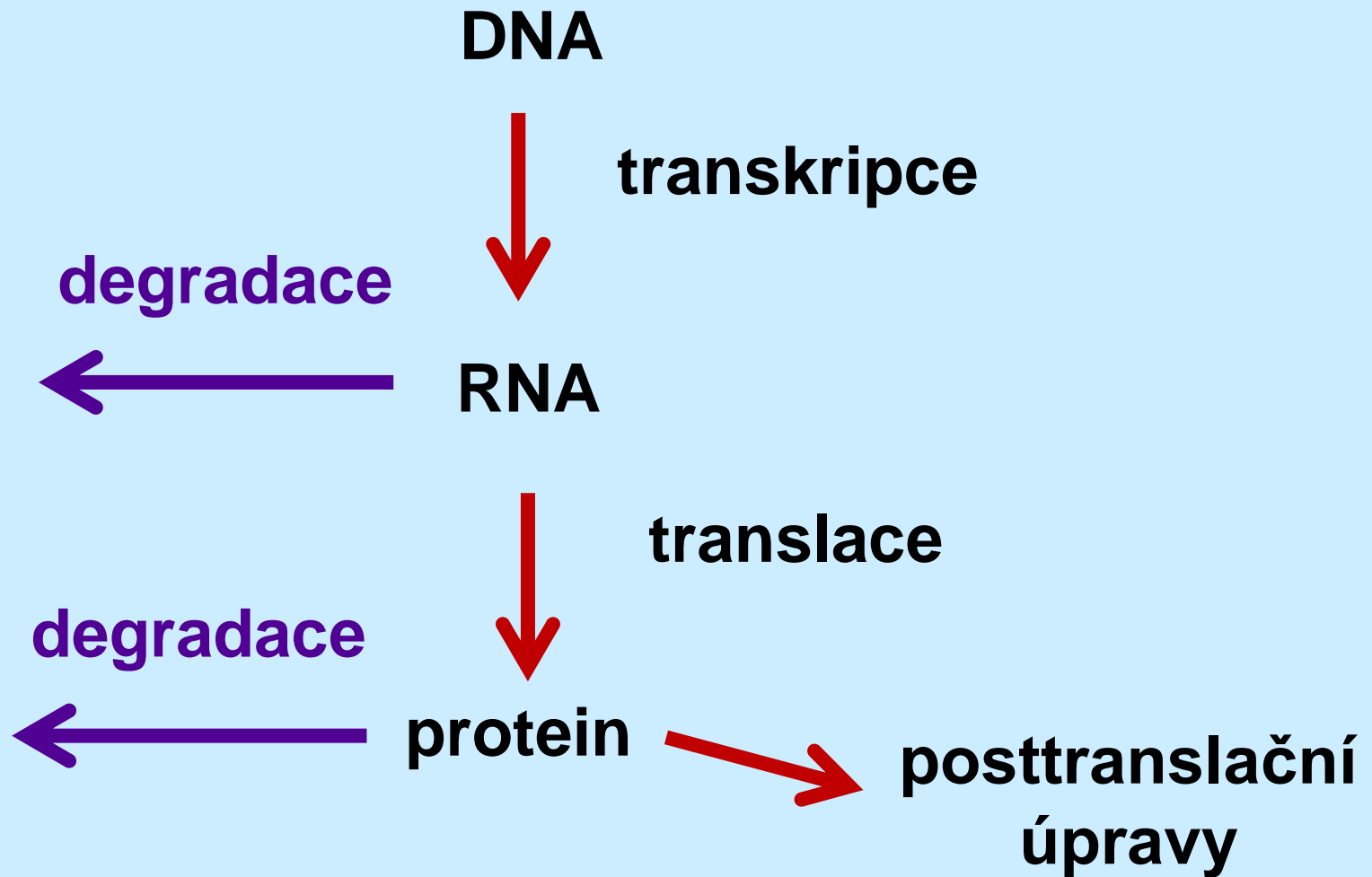


Studium proteomu

**Proteom jsou všechny proteiny buňky
a odrážejí její biochemickou kapacitu**



Proč studovat proteom?



Transkriptom neodráží celou genovou expresi

Jak studovat proteom?

Separace proteinu

- **PAGE**
- **Dvourozměrná gelová elektroforéza**
- **Izoelektrická fokusace**

Identifikace proteinů po separaci

- **Hmotnostní spektrofotometrie**
- **MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight)**
- **ICAT (isotope coded affinity tag)**

Studium interakcí protein-protein

Poskytují další informace o aktivitě genomu

Phage display

- Vystavuje proteiny na povrchu bakteriofága
- Testovaný protein je konfrontován s knihovnou takových fágů

Yeast two-hybrid system

- Genová exprese *S. cerevisiae* nastává po interakci dvou transkripčních faktorů
- Lze jím rozpoznat, jestli dva proteiny spolu interagují – pokud spustí expresi

Phage display – příklady aplikací

- **1994 Folgori – identifikace dvou epitopů na obalovém proteinu viru hepatitidy B – místo vhodné pro vývoj protilátek**
- **2008 Majumdar – vazebné vlastnosti membránového glykoproteinu gp41 viru HIV**
- **2002 Rosander – studium exprese extracelulárních bakteriálních proteinů – faktorů virulence - *Staphylococcus*, *Lactobacillus*, ...**

Yeast two-hybrid system - aplikace

- **Interakce obalových proteinů viru hepatitidy C s lidskými proteiny exprimovanými z cDNA knihovny (2009)**
- **Interakce viru SV40 s nádorovými antigeny (1993)**
- **Studium interakcí proteinů rodiny FemABX (rezistence k penicilinům) *Staphylococcus aureus* (2003)**
- **Studium proteinů zodpovědných za pohyblivost *Treponema pallidum* (2009)**

Analýza mikrobiálních komunit



Metagenomika

Studium genetického materiálu z organismů získaných přímo v nějakém prostředí aniž by bylo třeba kultivovat a izolovat jednotlivé druhy

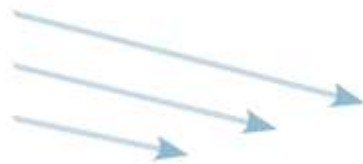
- Vzorky z půdy
- Mikrobiální společenstva trávicího traktu člověka

Základní premisa metagenomiky

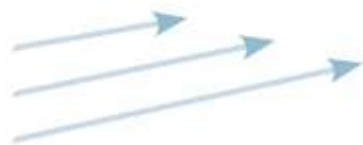
- Většina mikrobiální diverzity se v průběhu kultivace ztratí – většinu bakterií neumíme kultivovat

Koncept metagenomiky

THE METAGENOMICS PROCESS



Extract all DNA from
microbial community in
sampled environment



DETERMINE WHAT THE GENES ARE (Sequence-based metagenomics)

- Identify genes and metabolic pathways
- Compare to other communities
- and more...

DETERMINE WHAT THE GENES DO (Function-based metagenomics)

- Screen to identify functions of interest, such as vitamin or antibiotic production
- Find the genes that code for functions of interest
- and more...

Metody metagenomiky

- **Sangerova metoda sekvenování - shotgun**
- **Masívní paralelní pyrosekvenování a podobné techniky**

Pár historických poznámek I

1985

- Norman Pace započal se sekvenováním genů pro 16S rRNA (publikováno 1991)
- Nyní se sekvenují všechny geny
- Když se začalo se sekvenováním genů pro 16S rRNA, zjistilo se, že dokážeme kultivovat méně než 1% bakterií a archeí

1995

- Healy izoloval funkční geny z kolekce mikroorganismů ze senného nálevu

1996

- DeLong položil základy environmentální fylogeneze na bázi 16S rRNA, vzorky z mořské mikroflóry

Pár historických poznámek II

2002

- **Breitbart a Rohwer objevili 5000 nových druhů virů ve 200 l mořské vody – shotgun sekvenování**
- **Následně zjištěno více než tisíc druhů virů v lidské stolici, a zřejmě miliony druhů virů a fágů v 1kg mořských sedimentů**

2003

- **Craig Venter vedl Global Ocean Sampling Expedition (GOS)**

2004

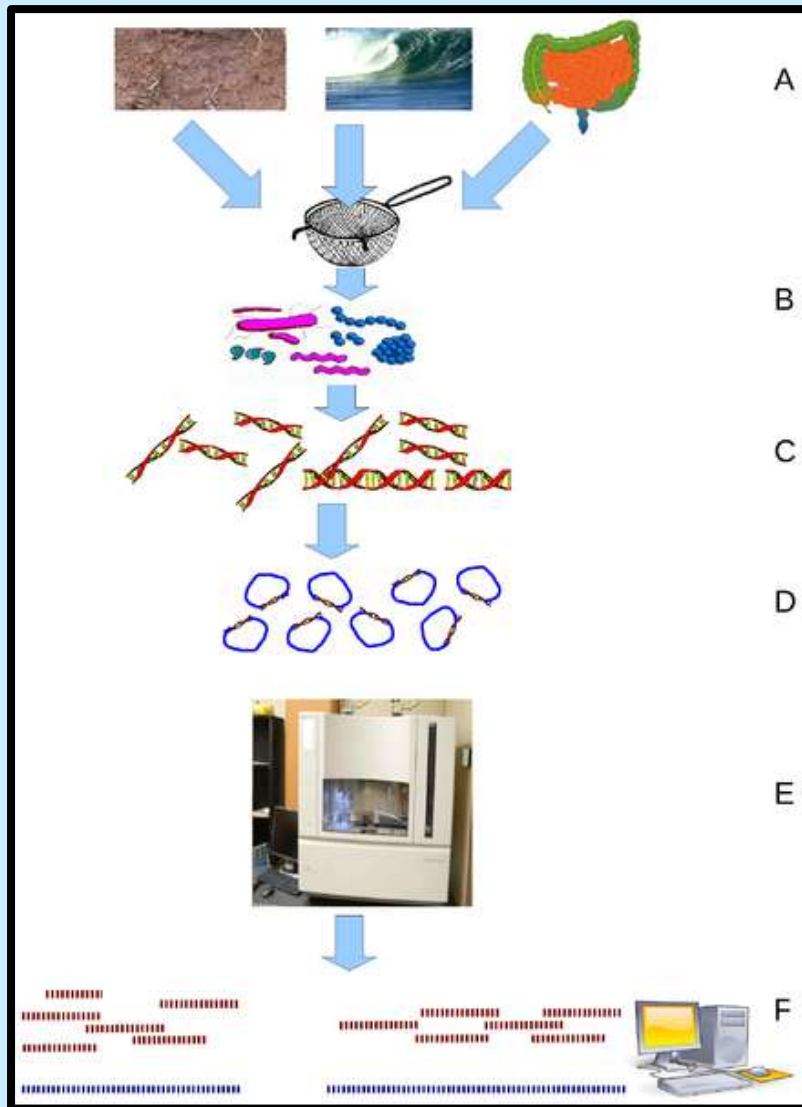
- **Tyson a Banfield sekvenovali DNA z drenážního systému dolů – nekultivovatelné bakterie a archae**

Pár historických poznámek III

2005

- **Schuster první sekvence získané pyrosequencováním z environmentálních vzorků**

Shotgun sekvenování



Odběr vzorků

Filtrování částic

**Lyze buněk a extrakce
DNA**

**Klonování a konstrukce
knihoven**

Sekvenování klonů

**Tvorba kontigů a
skládání sekvencí**

Masívní paralelní sekvenování

Pyrosekvenování

- **záznamy o délce 800 bp**
- **paralelnost vede k sumě 200-500 Mbp**

Illumina a SOLiD

- **záznamy o délce 25-75 bp**
- **paralelnost vede k sumě 20-50 Gbp**

Obrovské množství dat vyžaduje jejich specifické bioinformatické zpracování



Příklady

Metagenom kravského bachoru

- **279 Gbp**

Metagenom lidského střeva

- **567,7 Gbp**
- **3,3 milionu genů**

Problémy s kompletací výsledků

- **Repetitivní sekvence**
- **Sekvence více druhů mohou být poskládány do falešných contigů**

Používané programy

- **Phrap**
- **Celera Assembler**
- **Velvet assembler**

Co metagenomika sleduje

Predikce genů v mikrobiálních komunitách

- Podle vnějších příznaků – porovnáním s již identifikovanými geny (BLAST)
- *Ab initio* – nové geny identifikované podle regulačních oblastí a kódujících sekvencí

Sledování diverzity druhů

- Kolik druhů
- V jakém poměru
- Vyváženost množství jednotlivých druhů

Ke všemu byly vyvinuty specializované programy

Program GLIMMER

Využívá Markovových modelů

- **Aplikace na *Haemophilus influenzae*, *Helicobacter pylori* a další genomy**
- **Umožnil lokalizaci teoreticky všech genů v genomu**

<http://www.cbcb.umd.edu/software/glimmer/>

Salzberg et al. (1998): Nucleic Acids Research, 1998, Vol. 26, No. 2, 544-548

V co metagenomika vyústuje

Studium metabolismu mikrobiálních společenství

- **S detailním využitím DNA čipů a analýz proteomu**
- **Hledají se markery průběhu onemocnění**

Metatranskriptom

- **Sledování expresních profilů na úrovni mRNA**
- **Teprve v počátcích**

Studium viromů

- **Virový metagenom**

První příklad metatranskriptomu

- **2008**
- **Zemědělsky významný symbiont
*Sinorhizobium meliloti***
- **Použita metoda pyrosekvenování**

- **Identifikovali 20 nových transkriptů mRNA**

Mao, Ch. (2008): Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. BMC Microbiology 2008, 8:72

Aplikace metagenomiky I

Medicína

- **Studium bakteriálních společenstev v různých částech lidského těla**
- **Korelace složení mikrobiálních společenstev v těle a zdravotního stavu**

Biopaliva

- **Využití mikrobiálních společenstev při konverzi celulózy na etanol**
- **Produkce metanu a vodíku**
- **Vyhledávání účinnějších enzymů ve společenstvech**

Aplikace metagenomiky II

Remediace

- **Monitorování vlivu polutantů na mikrobiální společenstva**
- **Sledování mechanismů odbourávání kontaminantů**

Zemědělství

- **Interakce rostlina-mikroorganismus**
- **Využití znalostí o využívání nutrientů pro zlepšení výživy rostlin**
- **Sledování složení mikrobiálních společenstev v souvislosti se zdravím rostliny**

Aplikace metagenomiky III

Biotechnologie

- **Hledání nových léčiv – nové geny, nové enzymy**
- **Hledání nových chemikálií**
- **Hledání nových agrochemikálií**

Human Microbiome Project

<http://commonfund.nih.gov/hmp/>



- **Cílem je charakterizovat mikrobiální společenstva z několika míst lidského těla – dýchací cesty, ústní dutina, kůže, gastrointestinální trakt, urogenitální trakt**
- **Analyzovat úlohu těchto mikroorganismů v lidském zdraví a nemoci**

Individuální variabilita

<http://the-scientist.com/2012/06/13/microbial-menagerie/>

- **Od 242 zdravých dobrovolníků (18-40 let) odebrány vzorky z 18 tělních částí u žen a 15 u mužů**
- **Pro některé tělní části existují charakteristické druhy**
- **Některé druhy jsou „univerzálně“ přítomny všude**
- **Každý člověk si nese své individuální specifické kmeny jednotlivých druhů**
- **Složení komunit je dlouhodobě stabilní**
- **Nejrozmanitější komunity v ústech a střevě, nejchudší ve vagíně**

Individuální variabilita u CRISPR



Co je to CRISPR?

- **RNAi analogický systém určený k degradaci virových NA**
- **Popsaný v roce 2008**



Vlastnosti CRISPR

- **Využívá interní „antivirové“ sekvence začleněné v obrácených repetitích (CRISPR)**
- **CRISPR = clusters of regularly interspaced short palindromic repeats**
- **Po transkripci této sekvence dochází k jejich postupnému štěpení Cas proteiny**
- **Výsledné produkty interferují s nukleovou kyselinou vstupujícího viru**

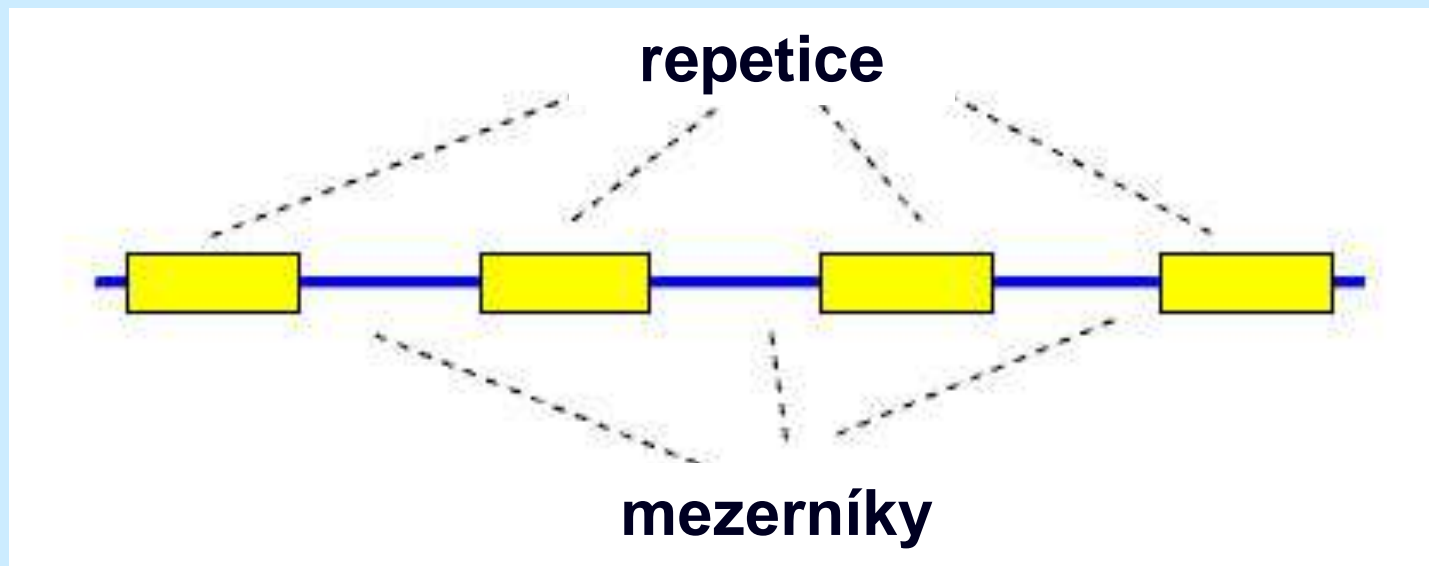
Brouns et al. (2008): Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes, Science 321, 960-964

Struktura CRISPR

Ke konci roku 2008 byly CRISPR popsány u asi 40% sekvenovaných eubakterií a téměř všech archeí

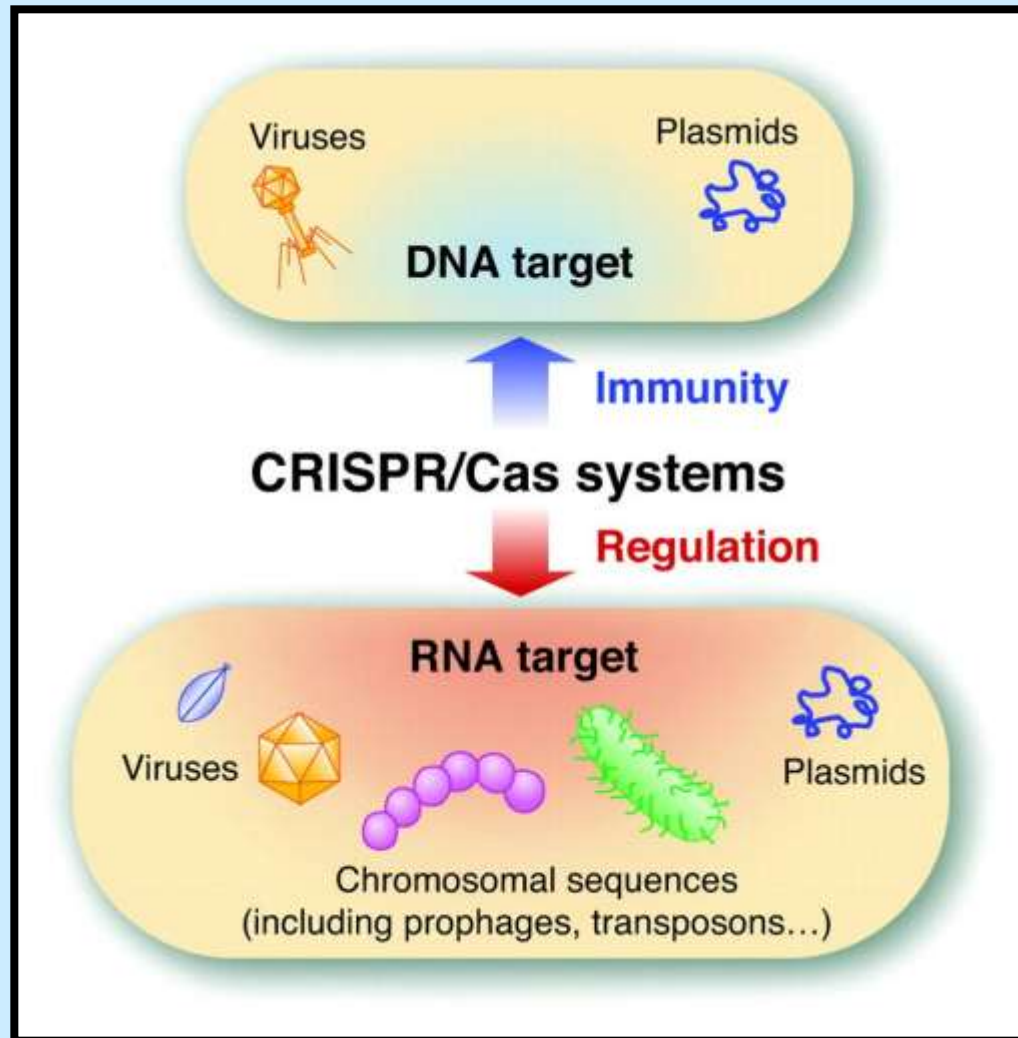
Všechny obsahují krátké repetice o délce 24 až 48 nukleotidů a mezerníky o přibližně stejné délce

Marraffini a Sontheimer (2008): Science 322, 1843 - 1845



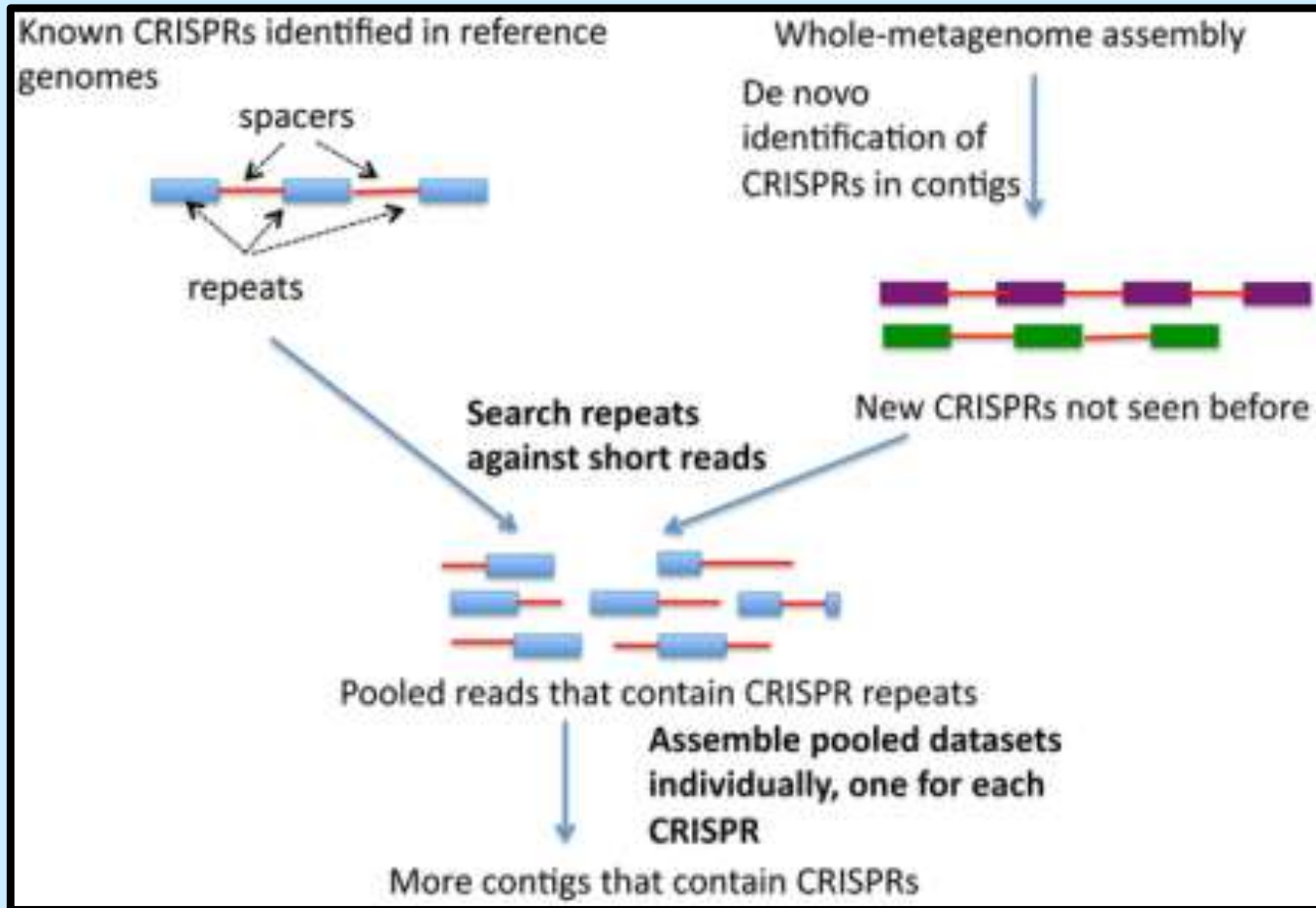
Edgar (2007): BMC Bioinformatics 8:18

CRISPR interference



P. Horvath et al., *Science* 327, 167-170 (2010)

Individuální variabilita u CRISPR



Rho et al. (2012): Diverse CRISPRs Evolving in Human Microbiomes. PLoS Genetics 8 (6), e1002441

Co zjistili

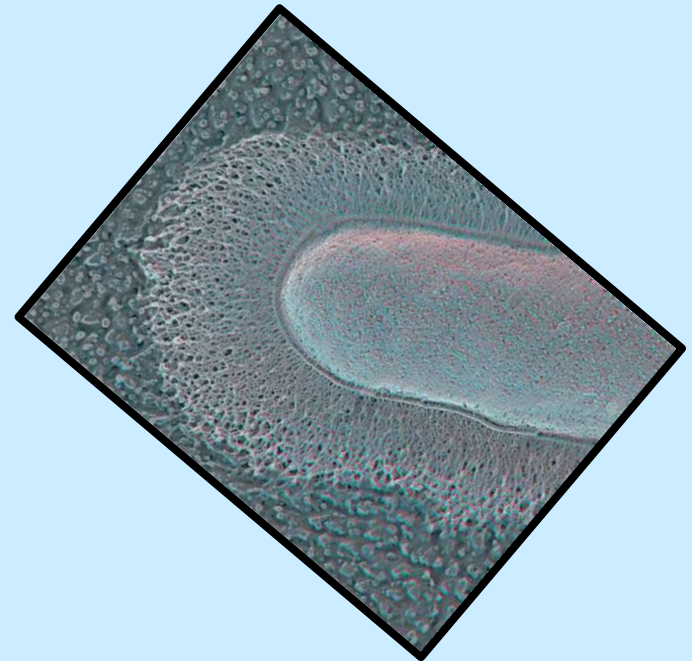
- **CRISPR pochopitelně odráží zastoupení mikroorganismů**
- **Sekvence mezerníků v CRISPR se mění podle toho s jakým infekčním agens se nositel potká**
- **Distribuce CRISPR je specifická**
- **Mezi jedinci se sekvence mezerníků v CRISPR liší**
- **Liší se i CRISPR z různých míst stejného jedince**

Výzkum bakteriálních společenstev osídlujících lidský organismus zásadním způsobem posouvá naše znalosti o významu mikroorganismů především ve vztahu mikroorganismus-lidské zdraví



Bacteroides thetaiotaomicron?

- **Obsahuje 260 enzymů schopných zpracovat rostlinný materiál**
- **Štěpí polysacharidy na glukózu a jiné lidským organismem zpracovatelné cukry**



Helicobacter pylori?

- **Reguluje aciditu žaludku**
- **Reguluje hladinu ghrelinu, hormonu hladovění**



málo Helicobacterů?

Bacteroides fragilis?

- polysacharid A (PSA) je rozpoznáván dendritickými buňkami, které jej prezentují T lymfocytům – vznikají regulační T lymfocyty
- regulační T lymfocyty tlumí prozánětlivé T lymfocyty



Další informace najdete

<http://dels-old.nas.edu/metagenomics/index.shtml>

- **populární informace o metagenomice**

<http://wiki.biomine.skelleftea.se/biomine/molecular/index.htm>

- **výukový popisný materiál o metodách analýzy mikrobiálních společenstev**

Scientific American, June 2012

- **populární článek o významu mikroorganismů sídlících v lidském těle**

Studium lidského mikrobiomu

<http://www.metahit.eu/>

- **Projekt 7.FP - Metagenomika lidského střevního traktu (2008-2012)**

<http://www.nature.com/nature/journal/v464/n7285/pdf/nature08821.pdf>

- **Informace o katalogu genů mikrobů z lidských střev**

<http://www.scientificamerican.com/article.cfm?id=microbiome-graphic-explore-human-microbiome>

- **Interaktivně o některých klíčových druzích mikroorganismů v našem těle**

Shrnutí

- 1) Identifikace genů v genomu**
- 2) Určení funkce neznámého genu**
- 3) Studie transkriptomu a proteomu**
- 4) Interakce protein-protein**
- 5) Analýza mikrobiálních komunit**
- 6) Příklady aplikací a výsledků**