



Sekvenování nové generace a *in silico* analýza celogenomových dat

Předmět: Antibiotická rezistence

Mgr. Kristína Nešporová, h17005@vfu.cz

Doc. RNDr. Monika Dolejská, Ph.D.

Podpořeno: IVA VFU 2020, 2020FVHE/2150/30

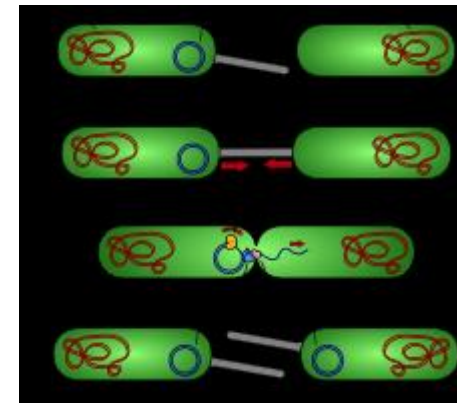
Antibiotická rezistence

Antibiotická rezistence

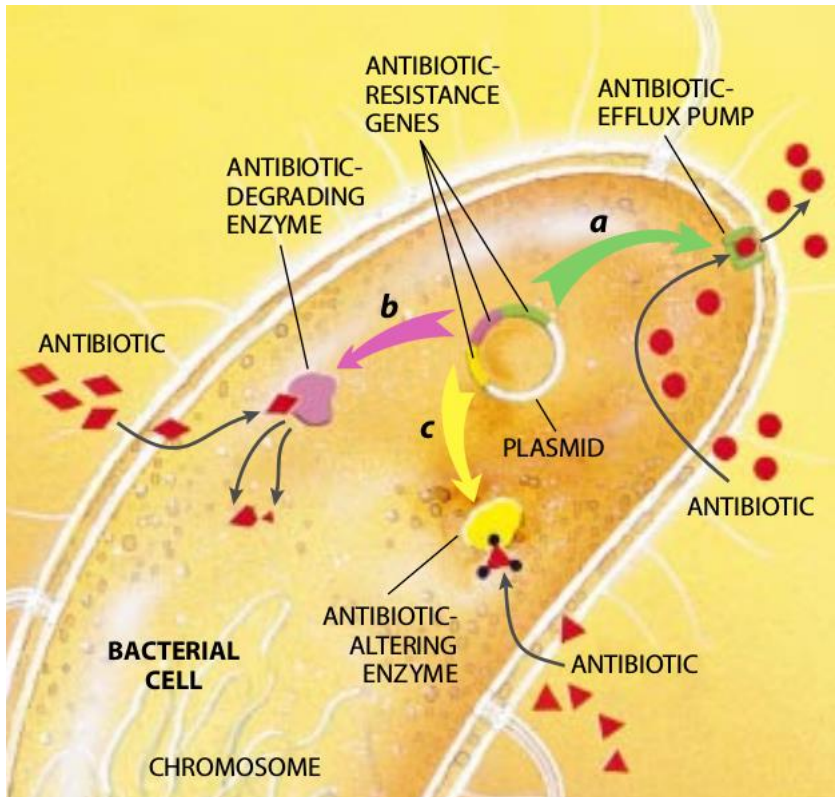
- Soustavně narůstající výskyt
- U lidí, zvířat, v prostředí – cesty přenosu nejsou komplexně popsány
- Problém pro medicínu – hrozí až postupný návrat do „preantibiotické éry“
- Předpovědi WHO – v roce 2050 bude umírat na komplikace spojené s léčbou infekcí rezistentních k antibiotikům více lidí ročně než na kardiovaskulární choroby a rakovinu dohromady
- Vývoj nových ATB – pomalý, ekonomicky nevýhodný

Původ ATB rezistence

- Rezistence k sloučeninám typu antibiotik se začala u bakterií tvořit již dávno díky jejich soužití s ostatními organismy (plísně, jiné bakterie), které tyto látky produkují
- Masivní šíření s používáním antibiotik – významný je horizontální přenos rezistence pomocí mobilních genetických elementů (plazmidy, transpozony, inzerční sekvence)



Mechanismy ATB rezistence



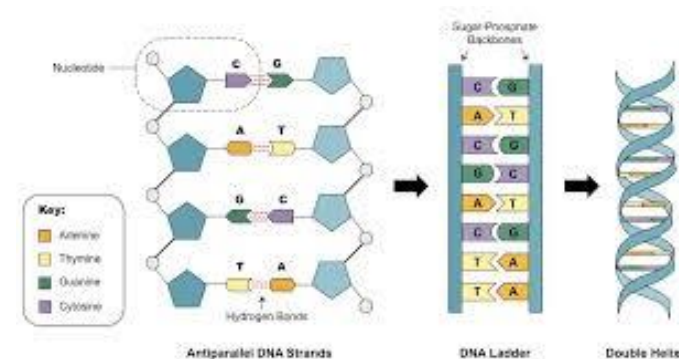
- Změna cílového místa
- Efluxní pumpy
- Bránění vstup ATB do buňky
- Enzymatický rozklad ATB
- Všechny mechanismy rezistence mají svůj genetický podklad kódovaný v DNA bakterií

Převzato z Levy, B. 1998. The Challenge of Antibiotic Resistance. Scientific American. vol. 3;

DNA a získání její sekvence

DNA – deoxyribonukleová kyselina

- 1869 - Miescher z lidských bílých krvinek poprvé izoloval DNA.
- 1944 - Avery ukázal, že genetická informace je uchovávána pravděpodobně molekulou DNA a nikoli proteiny, jak se do té doby věřilo.
- 1953 - Watson a Crick s na základě dat Franklinové a Wilkinse poprvé postulovali model sekundární struktury molekuly DNA
- 1966 - Nirenberg, Ochoa a Khorana rozluštili genetický kód.
- 1975 - 1977 - Sanger s Barrellem a Maxam s Gilbertem vyvinuli techniky sekvenování DNA
- 1985 - Mullis spolu s kolegy vynalezl PCR



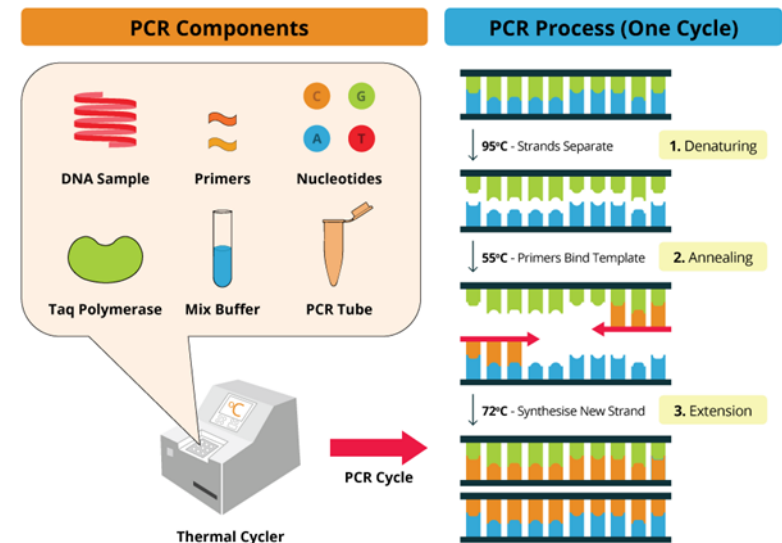
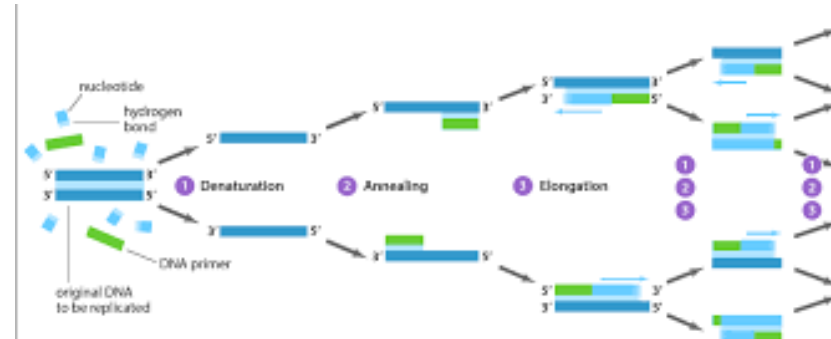
Funkce DNA

- Nese genetickou informaci
- Přepisování do RNA (ve chvíli potřeby)
- RNA je templát pro vytvoření proteinů
- Proteiny jsou základní stavební a funkční jednotky

		druhý nukleotid					
		U	C	A	G		
prvý nukleotid	U	UUU fenylalanín UUC UUA leucín UUG	UCU serín UCC UCA UCG	UAU tyrozín UAC UAA koniec UAG koniec	UGU cysteín UGC UGA koniec UGG tryptofán	U	třetí nukleotid
	C	CUU leucín CUC CUA CUG	CCU prolin CCC CCA CCG	CAU histidín CAC CAA glutamín CAG	CGU arginin CGC CGA CGG	C	
	A	AUU isoleucín AUC AUA AUG začiatok	ACU treonín ACC ACA ACG	AAU asparagín AAC AAA lyzín AAG	AGU serín AGC AGA arginin AGG	A	
	G	GUU valín GUC GUA GUG	GCU alanín GCC GCA GCG	GAU kys. asparagová GAC GAA kys GAG glutamová	GGU glycín GGC GGA GGG	G	

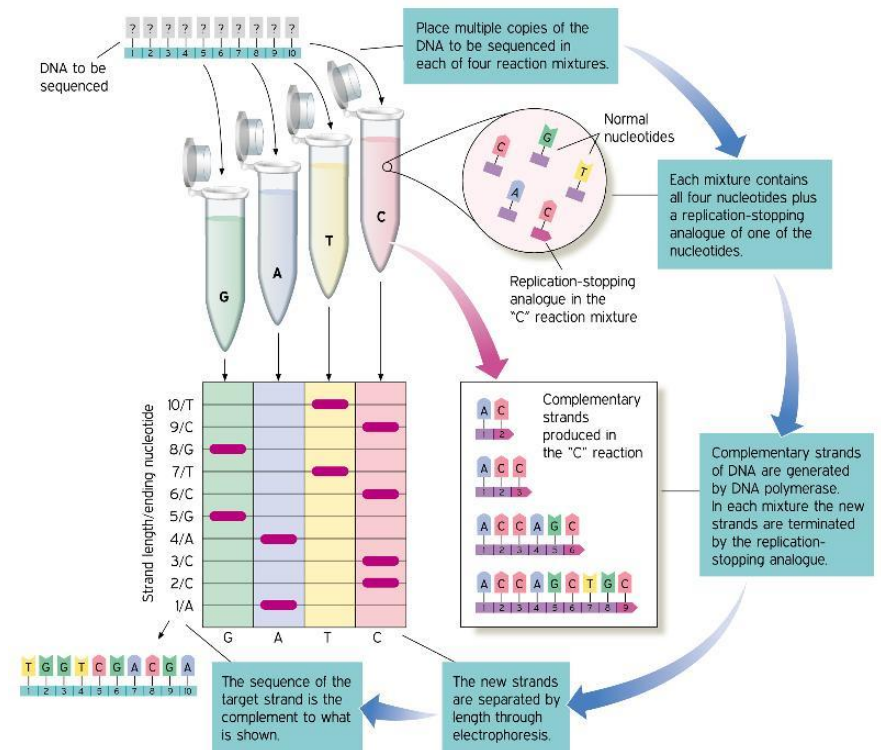
PCR – polymerázová řetězová reakce

- Polymerázová řetězová reakce
- Slouží k amplifikaci úseků DNA
- Při malém množství DNA je problém ji vizualizovat
- Znalost úseku, který chceme zmnožit
 - 1. Denaturace
 - 2. Nasednutí primerů
 - 3. Syntéza DNA – taq polymeráza
- Opakování cyklů – exponenciální amplifikace daného úseku



Sangerovo sekvenování

- DNA je amplifikovaná – namnožená do více kopií (dnes pomocí PCR, dříve izolace velkého množství DNA)
- Dvouvláknová molekula je pomocí tepla rozpojena – jednovláknová slouží jako templát
- Primery, DNA Polymeráza, Nukleotidy
- 4 zkumavky – každý doplněn o jeden typ modifikovaného „vadného“ nukleotidu, který je po čase začleněn a zastaví reakci
- Paralelně probíhá velké množství reakcí – „vadný“ nukleotid je začleněn na všechna místa, kam patří
- Srovnání hmotnosti fragmentů - elektroforéza



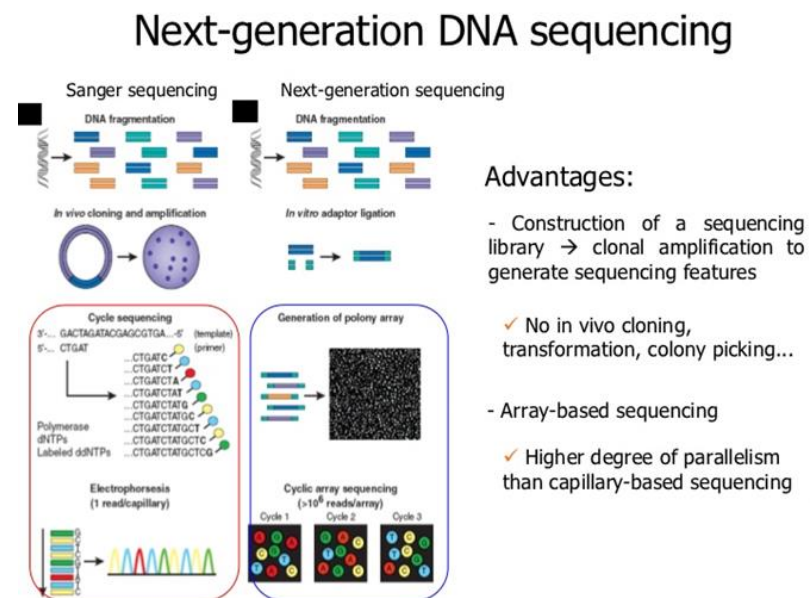
Problémy v získání kompletní informace

- PCR není vhodnou metodou pro získání sekvence určité DNA
 - Hodí se pro detekci genů, které očekáváme
- Sangerovo sekvenování je limitováno délkou úseků, které můžeme získat
 - Řádově maximálně tisíce páry bazí
 - DNA jednoduchých organismů jako bakterie mají řádově miliony bp
 - Časově a finančně náročné

Sekvenování nové generace

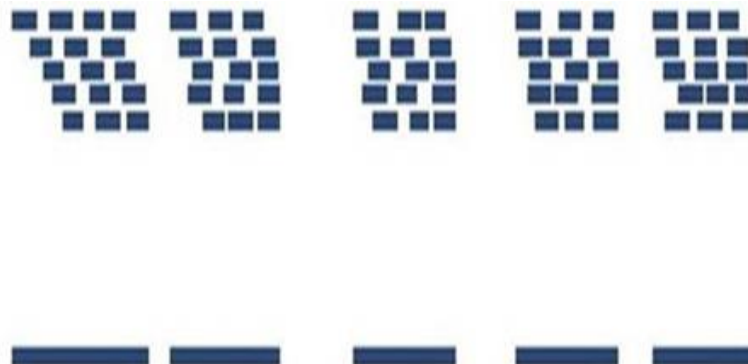
Sekvenování nové generace (Next-generation Sequencing)

- Vytvoření postupů, které umožňují rychlé a levnější určení sekvence DNA
- 1. Založené na paralelním sekvenování milionů fragmentů DNA – short read sequencing
- 2. Syntéza komplementárního vlákna jediné molekuly DNA a záznam použitých nukleotidů v čase – long read sequencing

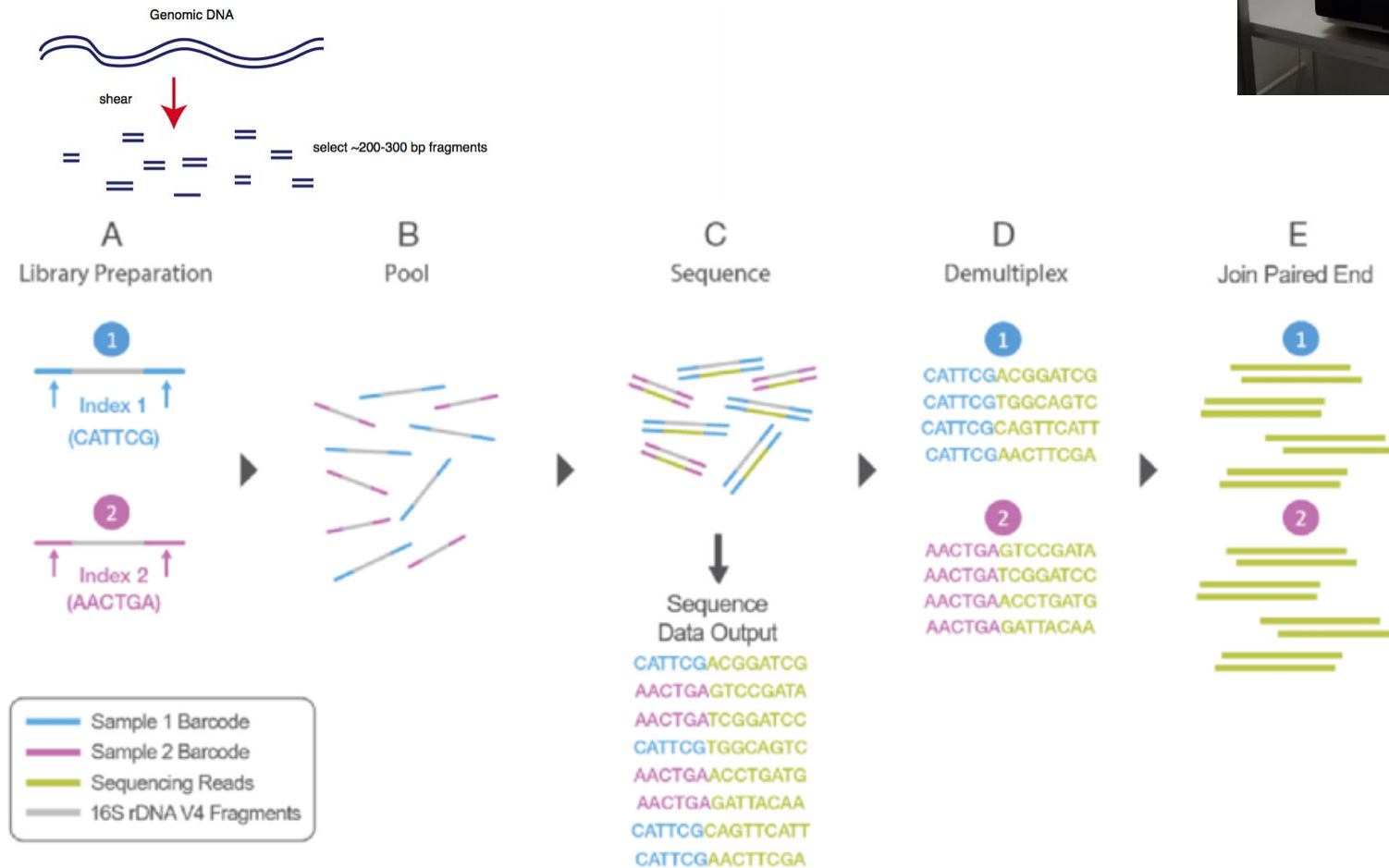


Sekvenování „short read“

- DNA pro sekvenování je fragmentovaná (cca 300 bp)
- Výsledkem jsou krátké sekvence
- Jedna pozice v genomu je v procesu sekvenování přečtena několikrát a je součástí více získaných krátkých sekvencí (typicky desítek), které mají různý posun vůči sobě – krátké sekvence lze skládat do delších úseků
- Výhody – při nastavení vysoké prosekvenovanosti vzorku je nízká chybovost
- Nevýhody – metoda je citlivá na místa, která se v genomu opakují (typicky inzerční sekvence) – není schopná navázat a udělat zlom, nedostaneme uzavřený genom a kompletní informaci o pozici genů
- Sekvenování technikou Illumina

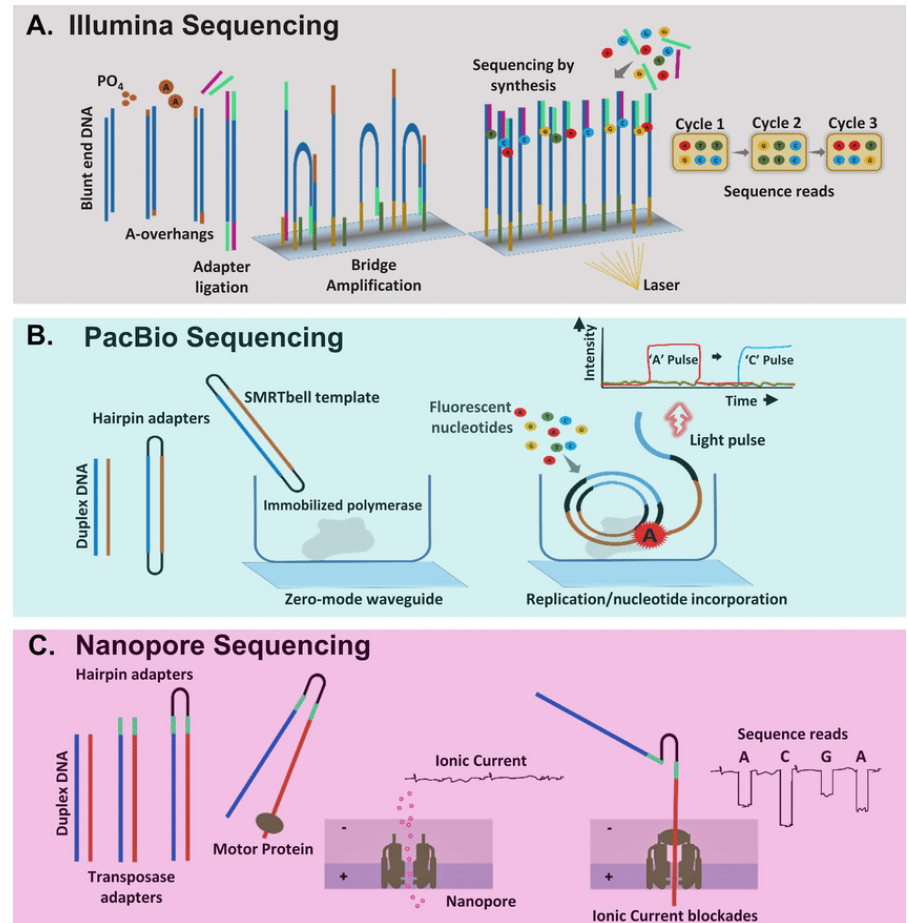


Illumina



Sekvenování „long read“

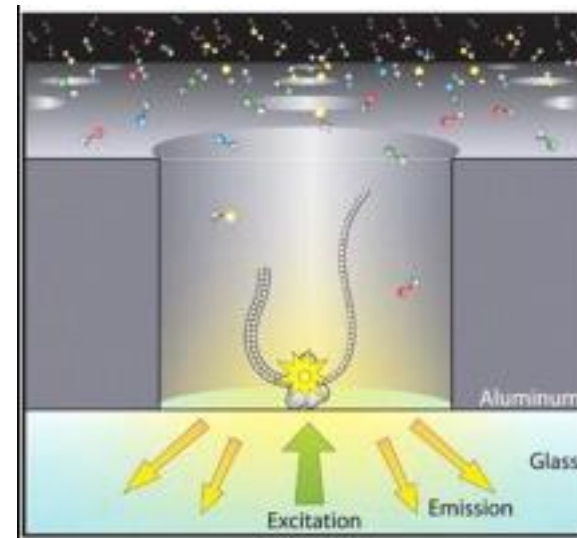
- Získány dlouhé úseky DNA
- Pouze v jedné/dvou kopiích – větší možnost začlenění chybného nukleotidu
- PacBio – přesnější ale drahé
- MinIon – cenově dostupný, vyšší chybovost



https://www.researchgate.net/publication/338048326_Current_challenges_and_best-practice_protocols_for_microbiome_analysis/figures?lo=1

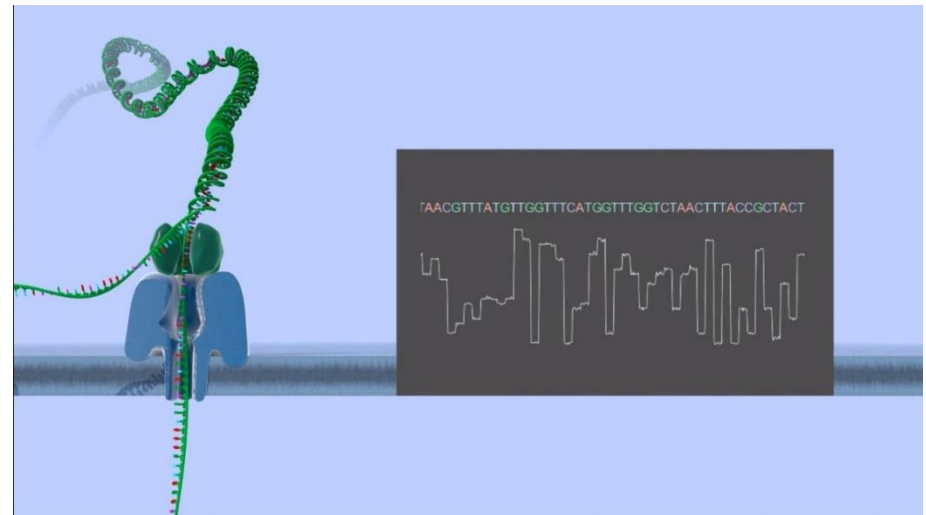
Pacific Biosciences („long ready“)

- SMRT (Single Molecule Real Time)
- Enzym DNA polymerázy v póru
- Prochází jediné vlákno DNA a syntetizuje se komplementární vlákno a je vyzářen flourescenční signál, kterým je příslušný nukleodit značen



MinION („long ready“)

- MinION
- Jediné vlákno DNA je elektroforeticky vedeno přes nanopór v elektricky nabitě membráně
- V průběhu průchodu je změřena změna napětí membrány
- Z této hodnoty se určí, která báze v danou chvíli pórem prošla



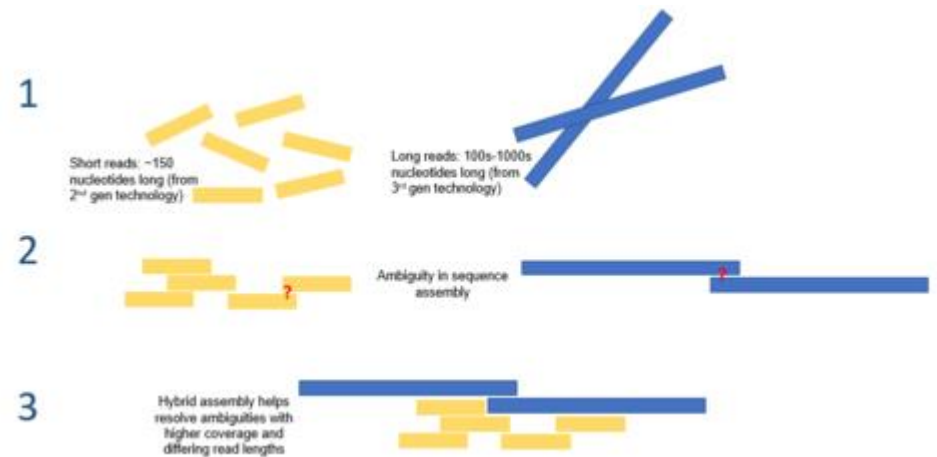
MinION („long ready“)

- Výhody
 - Velmi rychlé
 - Skladné zařízení
 - Velikost flash disku
 - Lze vzít do terénu
 - Sekvenování ve vesmíru
- Nevýhody
 - Větší chybovost
 - Firma vyvíjí novější přístupy pro zpřesnění



Hybridní sekvenační data

- Pokud chceme získat uzavřené sekvence a zároveň mít vysokou přesnost získaných dat (vhodné například pro referenční genom u fylogenetických studií), je vhodné využít tzv. hybridní assembly, které kombinuje data z „long“ a „short read“ sekvenování
- Dlouhé „reads“ umožňují získat představu o tom, jaké je pořadí jednotlivých úseků genomu
- Kombinace se krátkými „reads“ zvyšuje přesnost



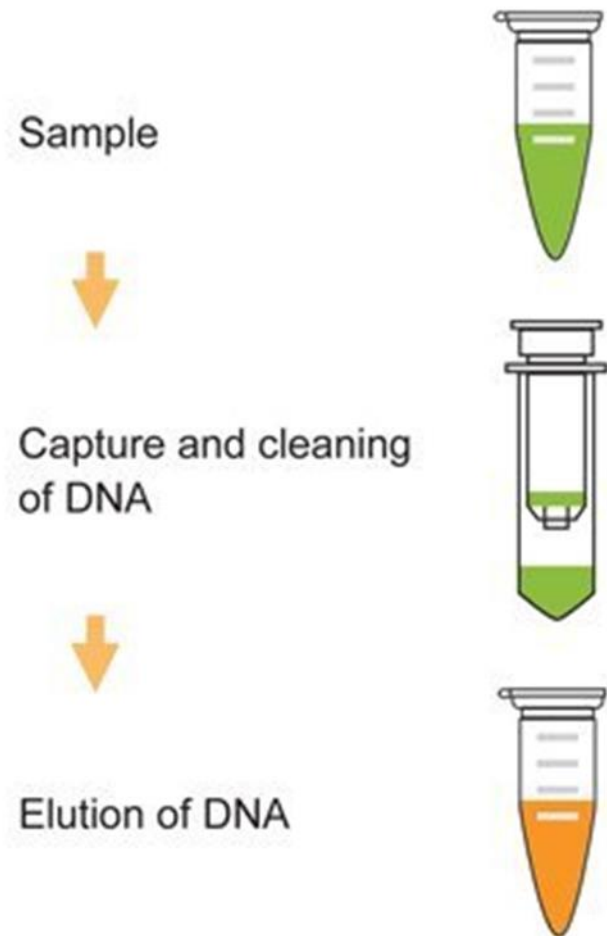
Výběr sekvenační metody

- Finance, dostupnost přístroje
- Typ analýzy
- V tomto kurzu se seznámíme s prací s krátkými ready získanými technologií Illumina (přístroj Miseq)

llumina – postup sekvenování

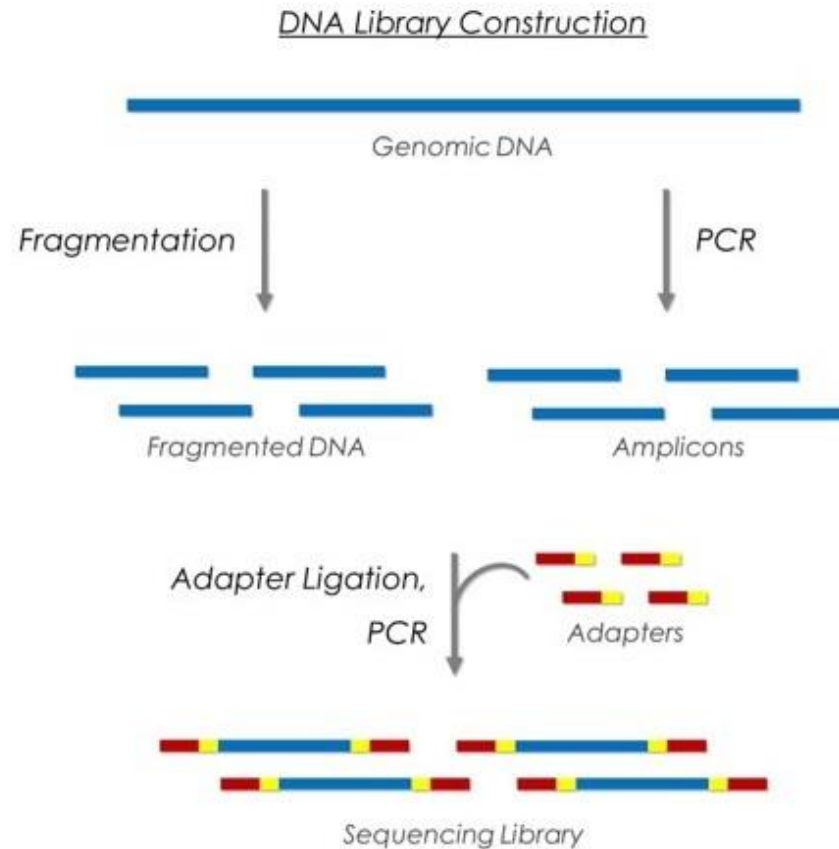
Izolace DNA

- Cíl – získání celogenomové DNA
- Problémy
 - Znečištění proteiny, organickými látkami
 - Fragmentace DNA
- Správná volba kitu/postupu je důležitá pro dosažení dostatečné kvality sekvenačních dat



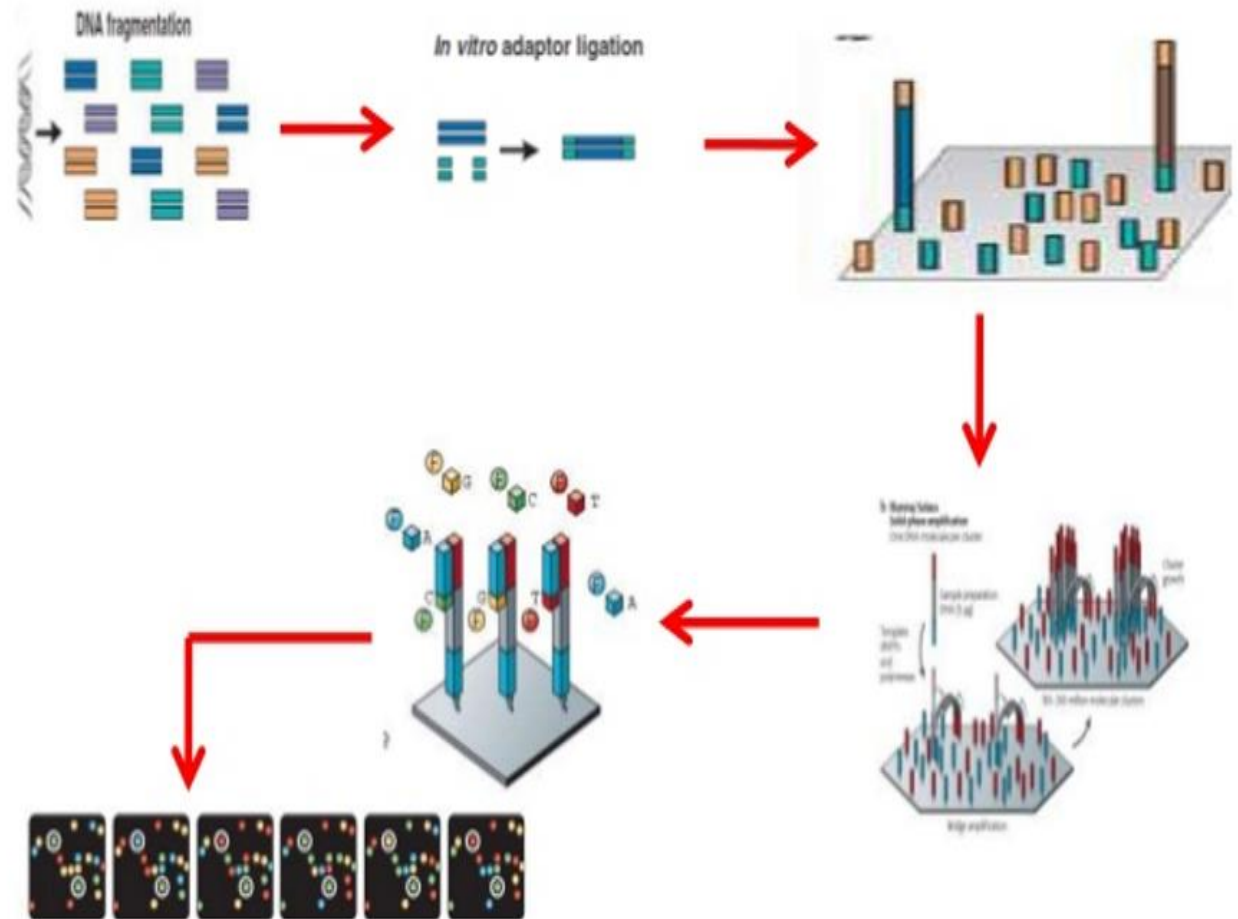
Příprava knihoven

- Fragmentace DNA
- Adaptery – indexy, úseky pro navázání primerů
- Smíchání DNA všech vzorků



Sekvenační proces

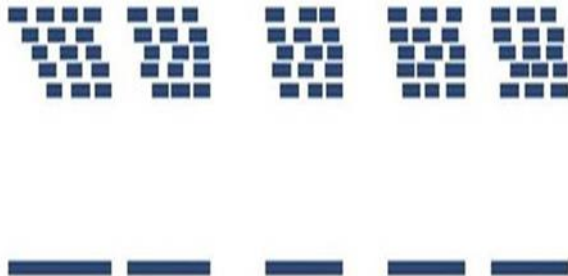
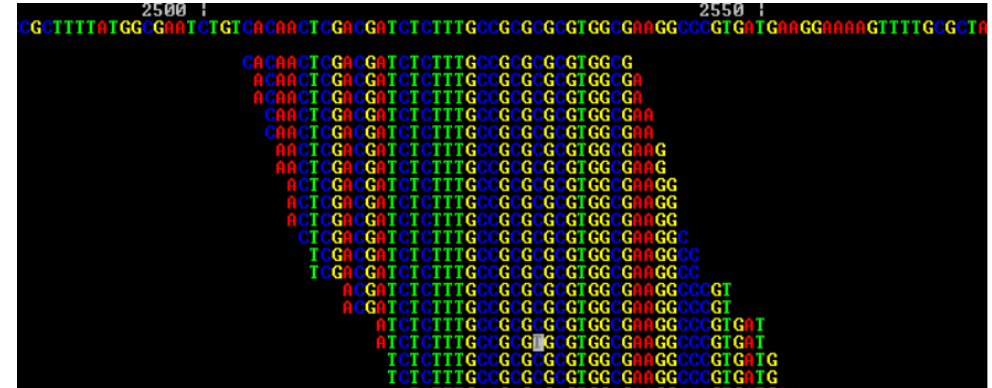
- Navázání fragmentů
- Můstková PCR
- Sekvenování pomocí DNA syntézy
- Záznam signálu jednotlivých nukleotidů



Analýza sekvenačních dat u bakteriálních izolátů

Analýza NGS dat („short ready“)

- Zpracování raw dat
 - Odstranění adapterů
 - Odfiltrování dat s nízkou kvalitou
- Assembly
 - Skládání získaných úseků DNA do větších celků (contigů)
 - De novo x mapování



Quality assessment

23 July 2018, Monday, 12:28:55

[View in Icarus contig browser](#)

[Download report](#)

Text, TSV and LaTeX versions of the table, plots in PDF. Additionally, detailed contigs and genome statistics.

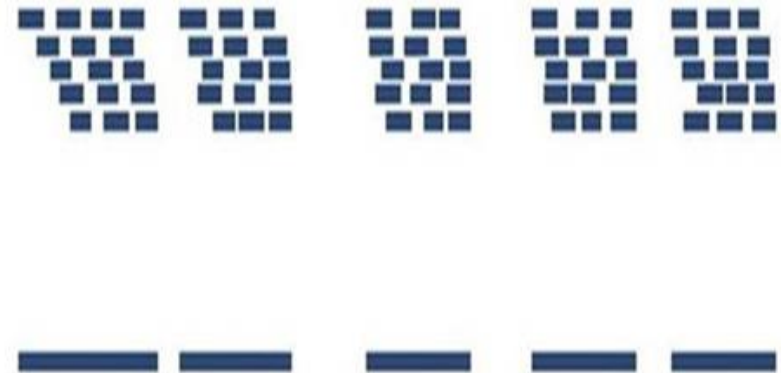
All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Worst Median Best Show heatmap

Statistics without reference	11_1466_R4	12_1471_R4	21_1473_R4	22_1474_R4	23_1483_R4	24_1494_R4	25_1507_R4	3_1489_R5	5_1523_R5	7_1484_R5
# contigs	107	475	429	420	675	336	288	150	196	284
# contigs (≥ 0 bp)	158	520	568	528	923	389	331	233	302	412
# contigs (≥ 1000 bp)	96	435	351	323	522	288	233	123	158	220
# contigs (≥ 5000 bp)	63	302	220	189	280	170	159	69	96	131
# contigs (≥ 10000 bp)	55	186	150	135	195	136	125	58	79	105
# contigs (≥ 25000 bp)	45	57	78	62	54	72	72	47	53	68
# contigs (≥ 50000 bp)	28	3	19	25	7	23	26	30	32	38
Largest contig	492,356	83,995	120,034	140,869	71,886	91,321	197,905	380,734	275,730	222,895
Total length	5,557,760	5,378,449	5,448,262	5,553,640	5,538,566	4,857,961	5,161,958	5,610,359	5,376,084	5,647,476
Total length (≥ 0 bp)	5,570,029	5,396,709	5,486,882	5,586,443	5,601,526	4,874,455	5,175,253	5,635,934	5,410,729	5,685,399
Total length (≥ 1000 bp)	5,549,724	5,349,417	5,391,918	5,285,742	5,428,833	4,823,854	5,122,559	5,592,239	5,349,358	5,602,812
Total length (≥ 5000 bp)	5,475,862	4,987,046	5,066,736	4,977,910	4,846,352	4,549,939	4,933,309	5,485,254	5,203,674	5,401,571
Total length (≥ 10000 bp)	5,417,971	4,158,300	4,574,181	4,574,876	4,255,601	4,205,071	4,672,045	5,410,443	5,074,557	5,217,945
Total length (≥ 25000 bp)	5,238,857	2,077,786	3,425,598	3,288,599	2,069,987	3,304,369	3,785,405	5,217,177	4,697,961	4,856,346
Total length (≥ 50000 bp)	4,613,582	2,116,233	3,136,189	2,017,680	432,183	1,612,085	2,174,531	4,617,947	3,960,573	3,651,600
NS0	199,536	19,815	33,023	33,859	18,268	37,313	43,658	169,781	114,178	69,548
NS7	77,936	11,020	15,551	18,681	10,418	17,534	24,004	67,322	46,531	33,423
L50	9	84	54	42	86	42	35	11	16	25
L75	21	175	112	186	89	76	23	34	54	54
GC (%)	57.01	57.34	50.72	50.61	50.77	50.82	50.62	50.18	50.46	50.4
Mismatches										
# N's	0	0	0	0	0	0	0	0	0	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0	0

Assembly

- Ze souboru .fastq, který obsahuje všechna čtení (tzv. „raw data“) a informaci o kvalitě získáme pomocí assembly .fasta soubor, který obsahuje pouze konsenzus dané sekvence ve formě jednotlivých contigů



- Contig – sekvence DNA poskládána z krátkých readů, až do místa zlomu, kde už nebylo možné dále navázat

- Kvalitu assembly lze rozpoznat

- Nízkého počet contigů – čím méně zlomů, tím kvalitnější máme
- Délky nejdelšího contigu
- N50 – délka contigu, který je uprostřed, pokud bychom dost 101 contigů pro jednu sekvenci, pak by šlo o délku 51. contigu
- Celkové velikosti genomu – pokud je příliš malá, data nám chybí, pokud je významně větší, došlo ke kontaminaci jinou bakterií
- Program QUAST a jiné

Quality assessment

23 July 2018, Monday, 12:28:55

[View in Icarus contig browser](#)

[Download report](#)

Text, TSV and Latex versions of the table, plots in PDF. Additionally, detailed contigs and genome statistics.

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Worst Median Best Show heatmap

Statistics without reference	11_1466_R4	12_1471_R4	21_1473_R4	22_1474_R4	23_1483_R4	24_1494_R4	25_1507_R4	3_1489_R5	5_1523_R5	7_1484_R5
# contigs	107	475	429	420	675	336	288	150	196	284
# contigs (≥ 0 bp)	158	520	568	528	923	389	331	233	302	412
# contigs (≥ 1000 bp)	96	435	351	323	522	288	233	123	158	220
# contigs (≥ 5000 bp)	63	302	220	189	280	170	159	69	96	131
# contigs (≥ 10000 bp)	55	186	150	135	195	136	125	58	79	105
# contigs (≥ 25000 bp)	45	57	78	62	54	72	47	53	68	68
# contigs (≥ 50000 bp)	28	3	19	25	7	23	26	30	32	38
Largest contig	492 356	83 995	128 034	140 869	71 886	91 321	197 905	360 734	275 730	222 895
Total length	5 557 760	5 378 449	5 448 262	5 353 640	5 538 566	4 857 961	5 161 958	5 610 359	5 376 084	5 647 476
Total length (≥ 0 bp)	5 570 029	5 396 709	5 486 882	5 386 443	5 601 526	4 874 455	5 175 253	5 635 934	5 410 729	5 685 399
Total length (≥ 1000 bp)	5 549 724	5 349 417	5 391 918	5 285 742	5 428 833	4 823 854	5 122 599	5 592 239	5 349 358	5 602 812
Total length (≥ 5000 bp)	5 475 862	4 987 046	5 066 736	4 977 910	4 846 352	4 549 919	4 933 309	5 485 254	5 203 674	5 401 571
Total length (≥ 10000 bp)	5 417 971	4 158 300	4 574 181	4 574 876	4 255 601	4 295 071	4 672 045	5 410 443	5 074 557	5 217 945
Total length (≥ 25000 bp)	5 238 857	2 077 786	3 425 598	3 288 599	2 069 987	3 304 369	3 785 405	5 217 177	4 667 961	4 656 346
Total length (≥ 50000 bp)	4 613 582	2 116 23	1 361 892	2 017 680	432 183	1 612 085	2 174 531	4 617 947	3 960 573	3 651 600
N50	199 536	19 815	33 023	33 859	18 268	37 313	43 658	169 781	114 178	69 548
N75	77 936	11 020	15 851	18 681	10 418	17 534	24 004	67 322	46 531	33 423
L50	9	64	54	42	86	43	35	11	16	25
L75	21	175	112	96	186	89	76	23	34	54
GC (%)	57.01	57.34	50.72	50.61	50.77	50.82	50.62	50.18	50.46	50.4
Mismatches										
# Ns	0	0	0	0	0	0	0	0	0	0
# Ns per 100 kbp	0	0	0	0	0	0	0	0	0	0

Hodnocení NGS dat u bakterií

Centrum genomové epidemiologie (Dánská technická univerzita)

- Využití databází a programů pro mapování

- SpeciesFinder
- ResFinder
- VirulenceFinder
- PlasmidFinder
- SerotypeFinder
- MLST
- pMLST

```
kristina@kristina: ~/Lab/DTU/assembled
76e.fasta -o MLST_out/9_R12_ec_1476e -s ecoll -x
sudo docker run --rm -it -v /home/kristina/Lab/DTU/assembled:/
90e.fasta -o MLST_out/9_R13_ec_1500e -s ecoll -x
kristina@kristina:~/Lab/DTU/assembled$ for runID in *_ec_*.fasta; do sudo docker run --rm -it -v $MLST_db:/database
1 $runID -o MLST_out/$(basename $runID) -s ecoll -x; done
[sudo] password for kristina:
{"mist": {"results": {"allele_profile": {"adk": {"align_len": 536,
"allele": "6a",
"allele_name": "adk_6",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 536},
"func": {"align_len": 469,
"allele": "31",
"allele_name": "func_31",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 469},
"gyrB": {"align_len": 460,
"allele": "5",
"allele_name": "gyrB_5",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 460},
"lcd": {"align_len": 510,
"allele": "28",
"allele_name": "lcd_28",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 510},
"mdh": {"align_len": 452,
"allele": "1",
"allele_name": "mdh_1",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 452},
"purA": {"align_len": 478,
"allele": "1",
"coverage": 100.0,
"identity": 100.0,
"sbj_len": 478}}}}}
```

- online tooly x příkazová řádka
- Nutné sledovat verzi databáze (geny jsou průběžně přidávány)
- <https://cge.cbs.dtu.dk/services/>
- Video 1. CGE úvod

Center for Genomic Epidemiology

Home Services Instructions Output Overview of genes Article abstract

ResFinder-3.1 Server - Results

Input Files: [5_R5_ec_1523e.fasta](#)

[Show Acquired antimicrobial resistance results](#)

Acquired antimicrobial resistance gene - Results

Aminoglycoside						
Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
aph(3)-Ib	99.88	804/804	NODE_89_length_12494_cov_72.8769	155..959		AF391500
aph(6)-Id	100.00	837/837	NODE_89_length_12494_cov_72.8769	959..1795		AF391500

Beta-lactam						
Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
blaTEM-1A	100.00	881/881	NODE_145_length_1351_cov_89.4453	217..1077	Beta-lactam resistance	U01745908
blaCTX-M-8	100.00	878/878	NODE_125_length_2379_cov_47.2527	230..1105	Beta-lactam resistance	AF198171

Colistin						
Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
mcr-5.1	100.00	1644/1644	NODE_88_length_7282_cov_46.138	1294..2937		KY601501

Fluoroquinolone
No resistance genes found.

Fosfomicin
No resistance genes found.

Fusidic Acid
No resistance genes found.

Glycopeptide
No resistance genes found.

MLS - Macrolide, Lincosamide and Streptogramin B						
Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
mdf(A)	98.22	1233/1233	NODE_27_length_60922_cov_25.0956	27103..28335		Y08743

Nitroimidazole
No resistance genes found.

Oxazolidinone
No resistance genes found.

Phenicol
No resistance genes found.

Rifampicin
No resistance genes found.

Sulphonamide						
Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
su2	100.00	816/816	NODE_109_length_862_cov_112.932	33..848	Sulphonamide resistance	AY044118

Tetracycline
No resistance genes found.



Hodnocení NGS dat u bakterií

Centrum genomové epidemiologie (Dánská technická univerzita)

Tresholdy pro mapování dat

- Identita genu
 - Možnost nastavit filtr před samotným spuštěním toolu
 - Výchozí hodnota 90 %
 - Nastavené tresholdy by měly být udané v jakýchkoliv výstupech
 - Pro reportování genu s menší než 100 % ID lze použít doplnění „like“ např. *bla*_{TEM-1}-like
 - Funkce genu může a nemusí být obdobná (ale ani 100 % shodný gen se v jiném organismu stejného druhu nemusí chovat stejně – jak se gen projeví ovlivňuje širší okolí genu a další interakce)
- Query coverage

Hodnocení NGS dat u bakterií

Centrum genomové epidemiologie (Dánská technická univerzita)

Tresholdy pro mapování dat

- Query coverage
 - Již popsané geny mají známou velikost (počet párů bazí – bp)
 - Query coverage udává, poměr (procento) velikosti úseky, který byl v našich datech skutečně detekován ku očekávané velikosti daného genu
 - Výchozí pro cge tooly obvykle 60 %
 - Coverage pod 100 % nutně neznamená, že gen není v organismu přítomen celý, data mohou být nedostatečně prosekvenovaná, gen může být rozdělen včleněním inserční sekvence, je proto lepší očekávat jeho nefunkčnost, nemá-li další důkazy potvrzující opak

Hodnocení NGS dat u bakterií

- Antibiotická rezistence
- Existují geny a chromozomální mutace, na jejichž podkladu rezistence k ATB vzniká
- Popsané geny a mutace jsou ukládány v databázích a na základě jejich přítomnosti je možné predikovat, k jakým ATB může být bakterie rezistentní
- Přítomnost genu ve NGS datech nezaručuje rezistentní fenotyp (ale shoda je dle studií nad 90 %)
- Nepřítomnost genu rezistence k dané ATB skupině v NGS datech nezaručuje, že bakterie nebude rezistentní (existují a dále se vyvíjejí další nepopsané geny a mechanismy)

Hodnocení NGS dat u bakterií

- ResFinder
- Video 2. ResFinder
- <https://cge.cbs.dtu.dk/services/ResFinder/>

Center for Genomic Epidemiology

Home Services Instructions Output Overview of genes Article abstract

ResFinder 4.1

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.

The database is curated by Frank Møller Aarestrup (click to contact).

Updates

ResFinder and PointFinder software: (2020-10-21)
ResFinder database: (2020-11-11)
PointFinder database: (2019-07-20)

Chromosomal point mutations

Select threshold for %ID: 66 %
Select minimum length: 66 %
 Show unknown mutations, not found in the database

Acquired antimicrobial resistance genes

Select Antimicrobial configuration: Beta-lactam
Select threshold for %ID: 66 %
Select minimum length: 66 %
 Acquired disinfectant resistance genes

Center for Genomic Epidemiology

Home Services Instructions Output Overview of genes Article abstract

ResFinder-4.1 Server - Results

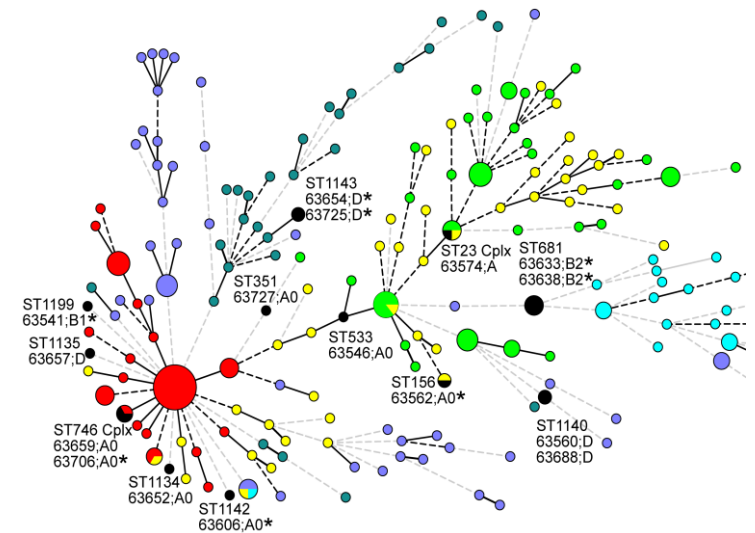
Input Files: *ecoll1.fa*

Warning:
One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	Resistant	blaCTX-M-8 (blaCTX-M-8_AF189721)
chloramphenicol	phenicol	Resistant	catA1 (catA1_V00622)
piperacillin-tazobactam	beta-lactam	No resistance	
cefotaxim	beta-lactam	No resistance	
ampicillin	beta-lactam	Resistant	blaCTX-M-8 (blaCTX-M-8_AF189721)
ampicillin-clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	Resistant	blaCTX-M-8 (blaCTX-M-8_AF189721)
ciprofloxacin	fluoroquinolone	Resistant	gyrA (p_S83L)
colistin	polymyxin	Resistant	mcr-5.1 (mcr-5_1_KY807921)
sulfamethoxazole	folate pathway antagonist	Resistant	su1 (su1_U12338)
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	fluoroquinolone	Resistant	gyrA (p_S83L)
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	Resistant	tet(B) (tet(B)_AF326777)
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	Resistant	blaCTX-M-8 (blaCTX-M-8_AF189721)

Hodnocení sekvenačních dat u bakterií

- MLST (z anglického Multi-locus sequence typing)
 - Typizace bakterií na základě alelických variant vybraných genů
 - Původně laboratorní technika – určeno na základě PCR a Sangerova sekvenování
 - Pro *E. coli* schéma se sedmi geny
 - Znalost sekvenčních typů může pomoci predikovat jejich vlastnosti (druhovú klasifikace pro to není dostačující)



<https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0005958.g002>

Hodnocení sekvenčních dat u bakterií

- MLST

- <https://cge.cbs.dtu.dk/services/MLST/>
- Video 3. MLST

Center for Genomic Epidemiology

Home Services Instructions Output Article abstract

MLST 2.0 (Multi-Locus Sequence Typing)

Software version: 2.0.4 (2019-05-08)
Database version: 2.0.0 (2020-10-05)

Select MLST configuration

- Escherichia coli#1
- Campylobacter upsaliensis
- Candida albicans
- Candida glabrata
- Candida krusei
- Candida tropicalis
- Carnobacterium maltaromaticum
- Chlamydiales spp.
- Citrobacter freundii
- Clonorchis sinensis
- Clostridium botulinum
- Clostridium difficile
- Clostridium septicum
- Corynebacterium diphtheriae
- Cronobacter spp.
- Dichelobacter nodosus
- Edwardsiella spp.
- Enterobacter cloacae
- Enterococcus faecalis
- Enterococcus faecium
- Escherichia coli#1

Isolate File

Name	Size	Progress	Status
------	------	----------	--------

Upload Remove

Center for Genomic Epidemiology

Home Services Instructions Output

MLST-2.0 Server - Results

mlst Profile: *ecoli*

Organism: *Escherichia coli*#1

Sequence Type: 57

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
adk	100	100	536	536	0	adk_6
fumC	100	100	469	469	0	fumC_31
gyrB	100	100	460	460	0	gyrB_5
icd	100	100	518	518	0	icd_28
mdh	100	100	452	452	0	mdh_1
purA	100	100	478	478	0	purA_1
recA	100	100	510	510	0	recA_2

extended output

Input Files: *ecoli1.fa*

Please download your results using the buttons below.

Results as text Results tsv Hit in genome sequences MLST allele sequences

CITATIONS

For publication of results, please cite:

Hodnocení sekvenačních dat u bakterií

- Virulence
 - Individuální vlastnost bakterie vyjadřující patogenitu určitého kmene ve srovnání s ostatními stejného druhu
 - Faktory virulence
 - adheziny – adheze na povrchy v organismu hostitele,
 - invaziny – pomáhají při pronikání bakterie do buněk a tkání v hostiteli,
 - toxiny
 - systémy zajišťující získání potřebných látek z prostředí/hostitele (často především záchyt železa)

Hodnocení sekvenačních dat u bakterií

- VirulenceFinder
 - Video 6. VirulenceFinder
 - <https://cge.cbs.dtu.dk/services/VirulenceFinder/>

VirulenceFinder 2.0

Software version: (2020-05-21)
Database version: (2020-05-29)

The database is curated by:
Flemming Scheutz, SSI
(click to contact)

Select species
Listeria
S. aureus
Escherichia coli
Enterococcus

Select threshold for %ID
90 %

Select minimum length
60 %

Select type of your reads
Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KIMA will be used for mapping. KIMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.
Assembled or Draft Genome/Contigs* (fasta)

VirulenceFinder-2.0 Server - Results

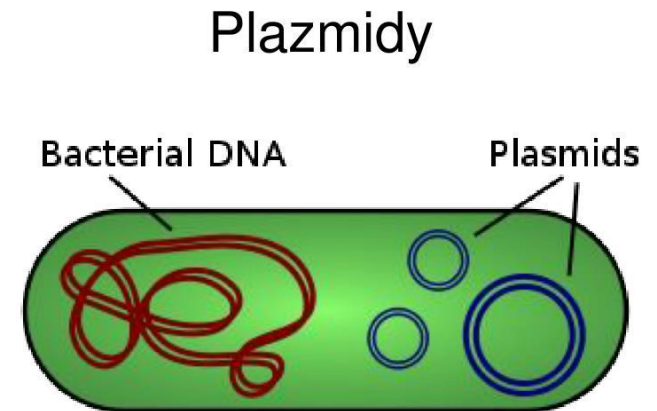
Organism(s): *Escherichia coli*

Shiga-toxin genes						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
No hit found						

Virulence genes for <i>Escherichia coli</i>						
Virulence factor	Identity	Query / Template length	Contig	Position in contig	Protein function	Accession number
chuA	99.93	1413 / 1983	NODE_125_length_7605_cov_20.128243	6075..7487	Outer membrane hemin receptor	UGDB01000003
cvaC	100	312 / 312	NODE_26_length_61255_cov_42.361095	40031..40342	Microcin C	X57525
gad	99.36	1401 / 1401	NODE_18_length_78229_cov_21.463125	10119..11519	Glutamate decarboxylase	BA000007
gad	100	1401 / 1401	NODE_23_length_62020_cov_21.996219	11045..12445	Glutamate decarboxylase	CP001846
hra	99.87	741 / 741	NODE_31_length_55707_cov_22.565131	54283..55023	Heat-resistant agglutinin	CP040456
hra	93.01	701 / 741	NODE_94_length_16077_cov_15.428966	121..809	Heat-resistant agglutinin	CP040456
hra	99.6	747 / 747	NODE_117_length_9719_cov_18.266472	7686..8432	Heat-resistant agglutinin	JF808724
tha	99.95	2088 /	NODE_142_length_3949_cov_20.727630	1143..3230	Adherence	CP000970

Hodnocení sekvenačních dat u bakterií

- Plazmidy
 - Malé kruhové molekuly DNA
 - Nacházejí se mimo chromozom
 - Mohou se předávat horizontálně (i mezidruhově)
 - Hrají důležitou roli jako vektory rezistence k antibiotikům, dezinfekčním látkám a těžkým kovům
 - Dělí se do skupin inkompatibility – neschopnost plazmidů ze stejné skupiny koexistovat v téže buňce



Autor: Spaully , Název: Plasmid (english).svg
Zdroj: [http://cs.wikipedia.org/w/index.php?title=Soubor:Plasmid_\(english\).svg&page=1](http://cs.wikipedia.org/w/index.php?title=Soubor:Plasmid_(english).svg&page=1)
Licence: <http://creativecommons.org/licenses/by-sa/3.0/>

<https://www.slideserve.com/oleg/mimojadern-d-di-nost-d-di-nost-kvantitativn-ch-znak>

Hodnocení sekvenačních dat u bakterií

- PlasmidFinder
 - Video 4. PlasmidFinder
 - <https://cge.cbs.dtu.dk/services/PlasmidFinder/>

Center for Genomic Epidemiology

Home Services Instructions Output

PlasmidFinder-2.0 Server - Results

Organism(s): *Enterobacteriaceae*

Enterobacteriaceae, <i>Aerobacter baumannii</i>						
Plasmid	Identity	Query / Template length	Contig	Position in contig	Note	Accession number
IncFIA(HI1)	100	388 / 388	NODE_158_length_1850_cov_23.759141	912..1299		AF250075
IncFIB(AP001918)	98.39	682 / 682	NODE_59_length_30681_cov_28.309943	1463..2144		AP001918
IncFIC(FII)	95.79	499 / 499	NODE_19_length_77206_cov_42.448358	6452..6948		AP001918
IncHI1A	99.76	420 / 420	NODE_22_length_67243_cov_25.894869	38464..38883		AF250076
IncHI1B(R27)	100	540 / 540	NODE_22_length_67243_cov_25.894869	24848..25387	R27	AF250076
Inc1-l(Gamma)	99.3	142 / 142	NODE_61_length_30173_cov_36.126972	11639..11780		AP005147

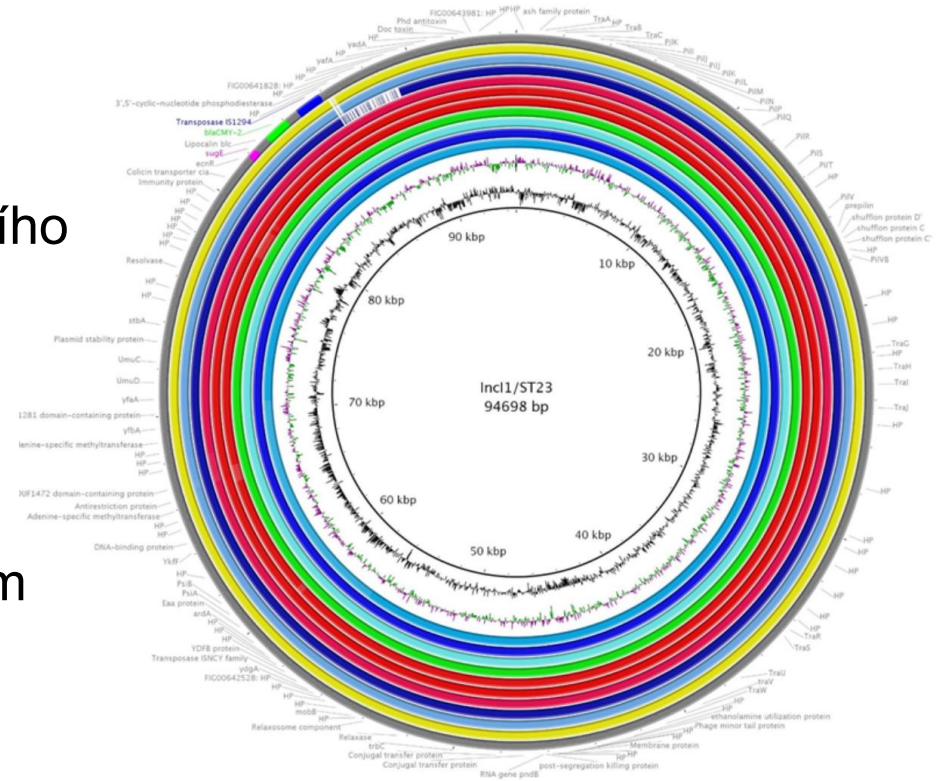
[extended output](#)

Input Files: *ecoli1.fa*

[Results as text](#) [Results tsv](#) [Hits in genome segs](#) [Plasmid sequences](#)

Hodnocení sekvenačních dat u bakterií

- Příbuznost plazmidů
 - Typizace vybraných genů
 - Určení plazmidového sekvenčního typu
 - Některé linie plazmidů jsou epidemické (vyskytují se často, někdy i v různých prostředích)
 - Často jsou spojené s konkrétním genem rezistence, který nesou



Hodnocení sekvenačních dat u bakterií

- pMLST
- Video č. 5 pMLST
- <https://cge.cbs.dtu.dk/services/pMLST/>

cge.cbs.dtu.dk/cgi-bin/webface.cgi?jobid=3FAU67EB0000148FAU54FUC5

Microsoft Word - Pr...

Center for Genomic Epidemiology

Home Services Instructions Output

pMLST-2.0 Server - Results

pMLST profile: *IncF RST*

Sequence Type: [F18:A8:B1]

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
FIA	100.0	100.0	383	383	0	FIA_8
FIB	100.0	100.0	373	373	0	FIB_1
FIC	100.0	100.0	200	200	0	FIC_4
FII	100.0	100.0	155	155	0	FII_18
FIIK						No hit found
FIIS						No hit found
FIIY						No hit found

extended output

Hodnocení sekvenačních dat u bakterií

- F antigeny
- Bílkoviny fimbrií
- FimTyper
 - Video 7. FimTyper
 - <https://cge.cbs.dtu.dk/services/FimTyper/>

cge.cbs.dtu.dk/cgi-bin/webface.fcgi?jobid=5FAD7AAE000343EB09FFFD1

Microsoft Word - Pr...

Center for Genomic Epidemiology

Home Services Instructions Output

FimTyper-1.0 Server - Results

FimH type: *fimH54*

FimH type - E.coli					
Fimtype	%Identity	Query/HSP length	Contig	Position in contig	Accession number
<i>fimH54</i>	100.00	489 / 489	NODE_67_length_26431_cov_22_163549	6483..6971	

extended output

Results as text Results tab separated Hit in genome sequences Fimtype sequences

Selected %ID threshold: 95 %

Input Files: *ecoli1.fa*

Support Scientific problems Technical problems

Copyright DTU 2011 / All rights reserved
Center for Genomic Epidemiology, DTU, Kemitorvet, Building 204, 2800 Kgs. Lyngby, Denmark

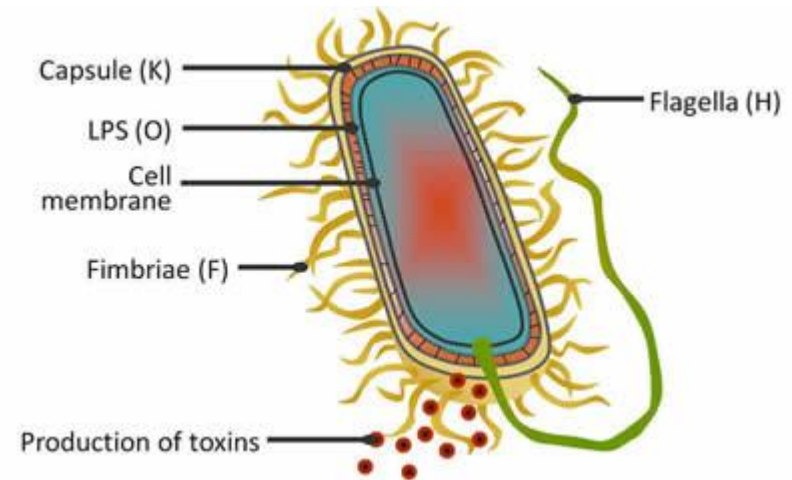
Hodnocení sekvenačních dat u bakterií

- Sérotypy

- Taxonomické členění *E. coli* na subdruhové úrovni
- Dle antigenních struktur
- O antigen (lipopolysacharid)
- H antigen (flagelární proteiny)

- SerotypeFinder

- <https://cge.cbs.dtu.dk/services/SerotypeFinder/>

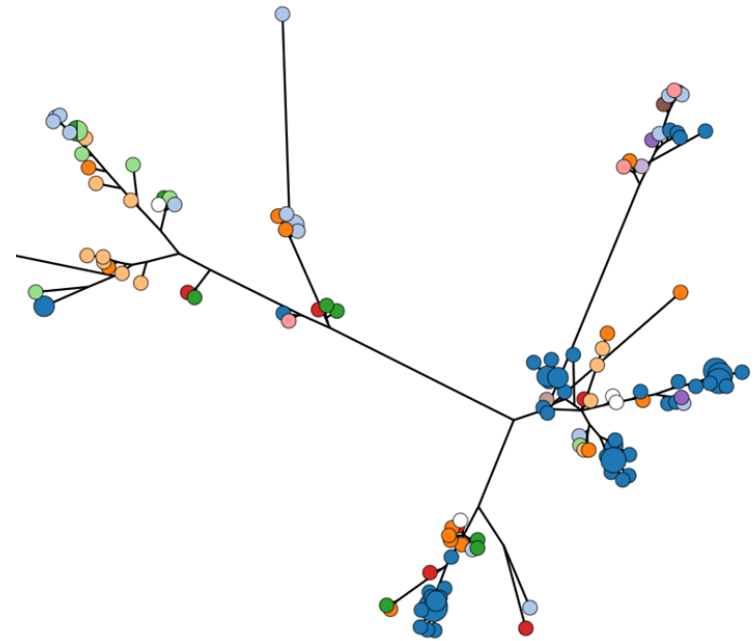


<http://www.ecl-lab.ca/en/ecoli/index.asp>

Hodnocení sekvenačních dat u bakterií

Fylogenetická analýza

- Různé programy pro analýzu příbuznosti bakteriálních izolátů
- Analýza „core genomu“ (geny, co mají dané izoláty společné) – počet SNP (odlišných míst, SNP – single locus variant)
- Výsledek – fylogenetický strom
 - CSI phylogeny
 - RAxML
 - ParSNP



Hodnocení sekvenačních dat u bakterií

- CSI Phylogeny
 - Video 8. CSI Phylogeny
 - <https://cge.cbs.dtu.dk/services/CSI-Phylogeny/>

Tree scale: 0.01

Origin

- Waste water
- Dog
- Bird
- Human



Hodnocení sekvenačních dat u bakterií

Grafické zpracování

- iTOL
 - Typy datasetů - Video č. 11. iTOL datasety úvod
 - Tvorba datasetů přes webové rozhraní - Video č. 12 iTOL datasety web interface
 - Dataset typu Binary – Video 13., 14. a 15. iTOL dataset binary
 - Dataset typu text – Video 16. iTOL text dataset
 - Export dat – Video 17. iTOL export dat

Hodnocení sekvenačních dat u bakterií

Alignment

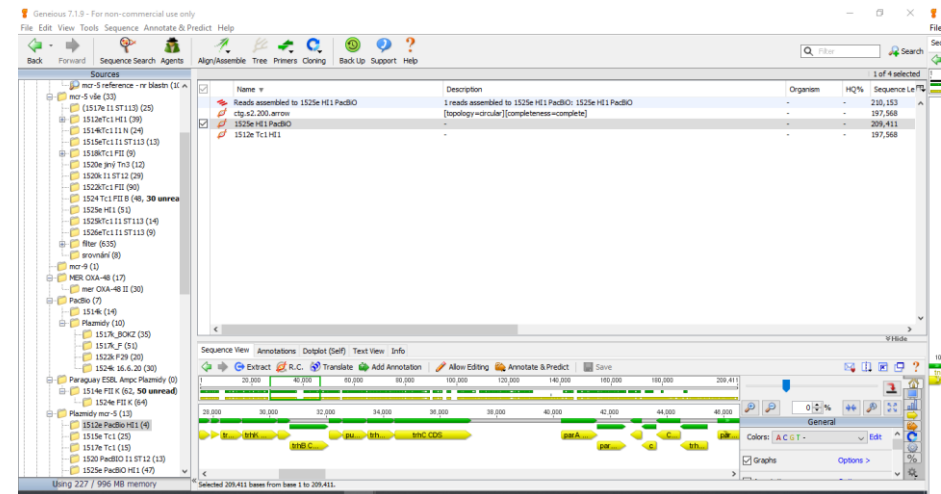
- Alignment
 - Seřazení minimálně dvou sekvencí pod sebe, tak aby odpovídající si úseky byly přímo pod sebou
 - Cílem je srovnat sekvence a z jejich podobnosti/rozdílnosti určit předpokládanou míru jejich možné příbuznosti
 - Program Geneious
 - Video 18. Geneious



Hodnocení sekvenačních dat u bakterií

Anotace

- Anotace
 - Na základě shody s dříve popsanými úseky lze popsat přítomné geny a předpokládat, jakou by mohly mít funkci
 - Shodné úseky s dříve popsanými genomy lze hledat přes BLAST
 - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
 - Program Geneious
 - Video 18. Geneious





Děkuji za pozornost

Dotazy?

h17005@vfu.cz