



HERA2

Molekulární surveillance a využití WGS **u bakteriálních vzdušných nákaz**

Z. Okonji, J. Kozáková, J. Zavadilová, L. Nováková, M. Honskus, S. Vohrnová, M. Musílek, P. Křížová

Oddělení bakteriálních vzdušných nákaz, Státní zdravotní ústav, Praha

Oddělení bakteriálních vzdušných nákaz (OBVN) zahrnuje:

NRL pro streptokokové nákazy (Dr. J. Kozáková, jana.kozakova@szu.cz)

NRL pro meningokokové nákazy (Mgr. Z. Okonji, zuzana.okonji@szu.cz)

NRL pro hemofilové nákazy (Mgr. L. Nováková, ludmila.novakova@szu.cz)

NRL pro difterii a pertusi (Mgr. J. Zavadilová, jana.zavadilova@szu.cz)

V projektu HERA2 = WP4 = VACPREV

CÍLE HERA2 projektu pro OBVN



- Cílem bylo **zavést sekvenování celého genomu (WGS) pro rutinní provádění** v Národních referenčních laboratořích (NRL) SZÚ, které zajišťují surveillance onemocněními způsobených meningokoky, pneumokoky, hemofily a *B. pertussis*.

Cíl je v souladu s dokumentem WHO „Defeating meningitis by 2030: a global road map“.

- **Proč zrovna WGS?** Jde o nejmodernější, nejpřesnější a celosvětově nejvíce využívaný přístup molekulární diagnostiky v mikrobiologických laboratořích. Metoda celogenomového sekvenování umožňuje stanovit kompletní DNA sekvenci genomu cílového organismu, což napomáhá rychlé identifikaci ohniska výskytu nálezů a umožňuje lepší porozumění způsobu šíření infekčních chorob.

- Je třeba **získávat kvalitní, mezinárodně porovnatelná data v reálném čase.**

Anonymizovaná data WGS jsou hlášena do mezinárodních databází (jako jsou PubMLST, TESSy, EMERT II, IRIS, SpIdNet, Pserenade, Pathogenwatch, Pasteur BIGSdb), ve kterých jsou využívána pro vyšetřování ohnisek, účely mezinárodní surveillance, a k monitorování přeshraničních hrozeb.

Implementace WGS protokolů pro *N. meningitidis*, *S. pneumoniae*, *H. influenzae* and *B. pertussis* v jednotlivých NRL



1. IZOLACE GENOMOVÉ DNA
2. KONTROLA KVALITY a KONCENTRACE DNA
3. FRAGMENTACE A NORMALIZACE DNA
4. PŘÍPRAVA A KONTROLA KNIHOVNY
5. SEKVENACE
6. ZÍSKÁNÍ, UKLÁDÁNÍ A ANALÝZA DAT

IZOLACE GENOMOVÉ DNA

- izolace genomové DNA je prováděna manuálně pomocí kitu QIA amp DNA (QIAGEN) nebo automaticky izolátorem MCA 1602 MAGCORE SUPER (KRD) dle protokolu MagCore Genomic DNA Bacterial Kit (RBC Bioscience) vždy včetně použití enzymu RNasa
- definitivní eluční objem rozdělit na dva aliquoty, jeden pro vlastní sekvenaci, druhý pro ověření kvality genomové DNA

KONTROLA KVALITY A KONCENTRACE DNA

- koncentrace genomové DNA je stanovena fluorometrem QUBIT s QubitTM dsDNA High Sensitivity 1xdsDN Assay Kit
- čistota genomové DNA je dána hodnotou poměrů absorbancí UV záření při vlnových délkách 260 nm, 280 nm a 230 nm na spektrofotometru NANODROP a to 260/280 a 260/230

FRAGMENTACE A NORMALIZACE DNA

- fragmentace genomové DNA je prováděna pomocí ultrasonikátoru COVARIS, fragmenty jsou štěpeny dle používané sekvenační chemie (fragmenty se střední hodnotou 300-350bp)
- koncentrace vzorků jsou měřeny na QUBITu
- fragmentace je kontrolována na platformě pro automatizovanou elektroforézu Agilent TAPESTATION
- ředění vzorků na výslednou koncentraci shodnou pro všechny vzorky

PŘÍPRAVA A KONTROLA KNIHOVNY

- amplifikační příprava knihovny kitem IDT XGen TM DNA Library Prep MC s následným měřením na QUBITu a kontrolou na TAPESTATION

SEKVENACE

- připraví se samplesheet („sekvenační návod“ pro sekvenované vzorky)
- příprava ekvimolární směsi knihoven vzorků („poolu“)
- denaturace a ředění poolu knihoven
- příprava sekvenátoru a sekvenační cartridge
- provedení vlastní sekvenace pomocí Illumina MiSeq

ZÍSKÁNÍ, UKLÁDÁNÍ A ANALÝZA DAT

- download raw dat z Illumina MiSeq a jejich uložení na WGS úložišti
- primární úprava dat pomocí základního bioinformatického softwaru (bowtie2)
- sestavení genomů principem assembly *de novo* (Velvet de novo Assembler)
- uložení sestavených genomů v mezinárodní databázi PubMLST
- typizace izolátů s využitím platformy BIGSdb (např. MLST, rMLST, cgMLST)
- sekundární analýzy WGS dat (např. PubMLST Genome Comparator)



Automatizace protokolu přípravy knihoven pro WGS v rámci OBVN

Kit XGen™ DNA Library Prep MC od firmy IDT

Pipetátor Biomek i5 (Beckmann Coulter)

Zatím optimalizováno pro agens *Streptococcus pneumoniae*

- srovnatelné koncentrace knihoven s manuálním postupem
- srovnatelné zastoupení kompletních fragmentů s manuálním postupem
- data vyhovují zavedeným parametrům kvality
- assembly *de novo* – nárůst počtu kontigů (40), odpovídá posunu průměrné velikosti fragmentů knihovny směrem k menším fragmentům

Očekávané výhody:

- úspora času (zpracování 96 vzorků v jednom běhu)
- snížené riziko kontaminace
- příprava knihoven různých agens v jednom běhu
- optimalizace spotřeby reagensů



NRL pro streptokokové nákazy – osekvenováno 165 izolátů *S. pneumoniae*

NRL pro meningokokové nákazy – osekvenováno 104 vzorků

(86x *N. meningitidis*, 3x *N. bergeri*, 1x *N. canis*, 1x *N. mucosa*, 6x *N. cinerea*, 5x *N. polysacharea*, 1x *N. subflava*, 1x *N. flavescens*)

NRL pro hemofilové nákazy – osekvenováno 67 vzorků

(56x *H. influenzae*, 2x *H. parainfluenzae*, 5x *H. haemolyticus*, 4x *Aggregatibacter segnis*)

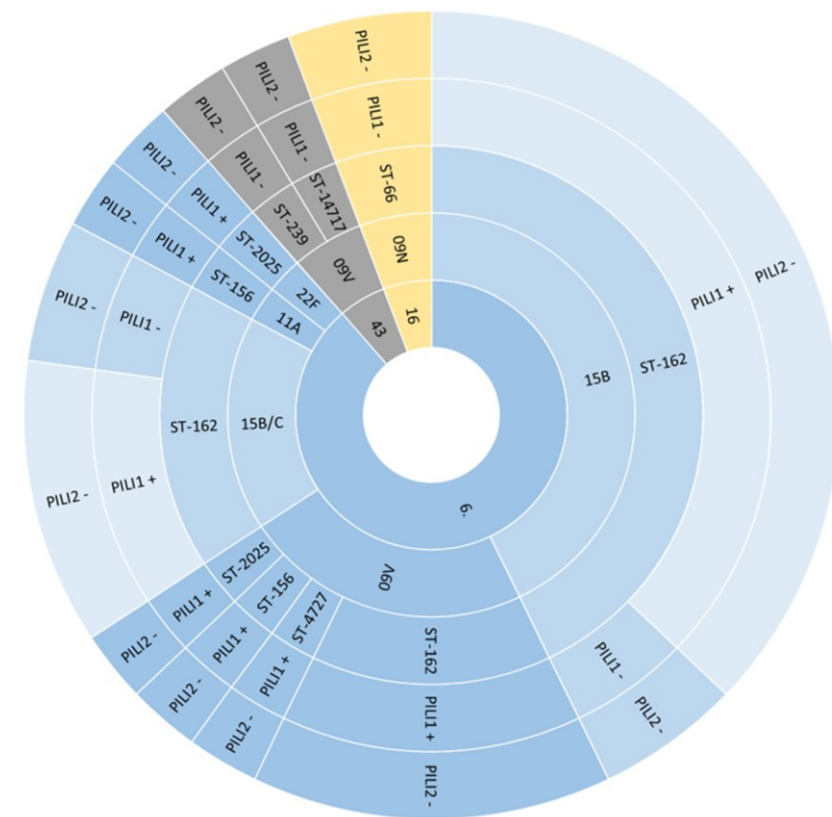
NRL pro difterii a pertusi – osekvenováno 51 izolátů *B. pertussis*

Streptococcus pneumoniae

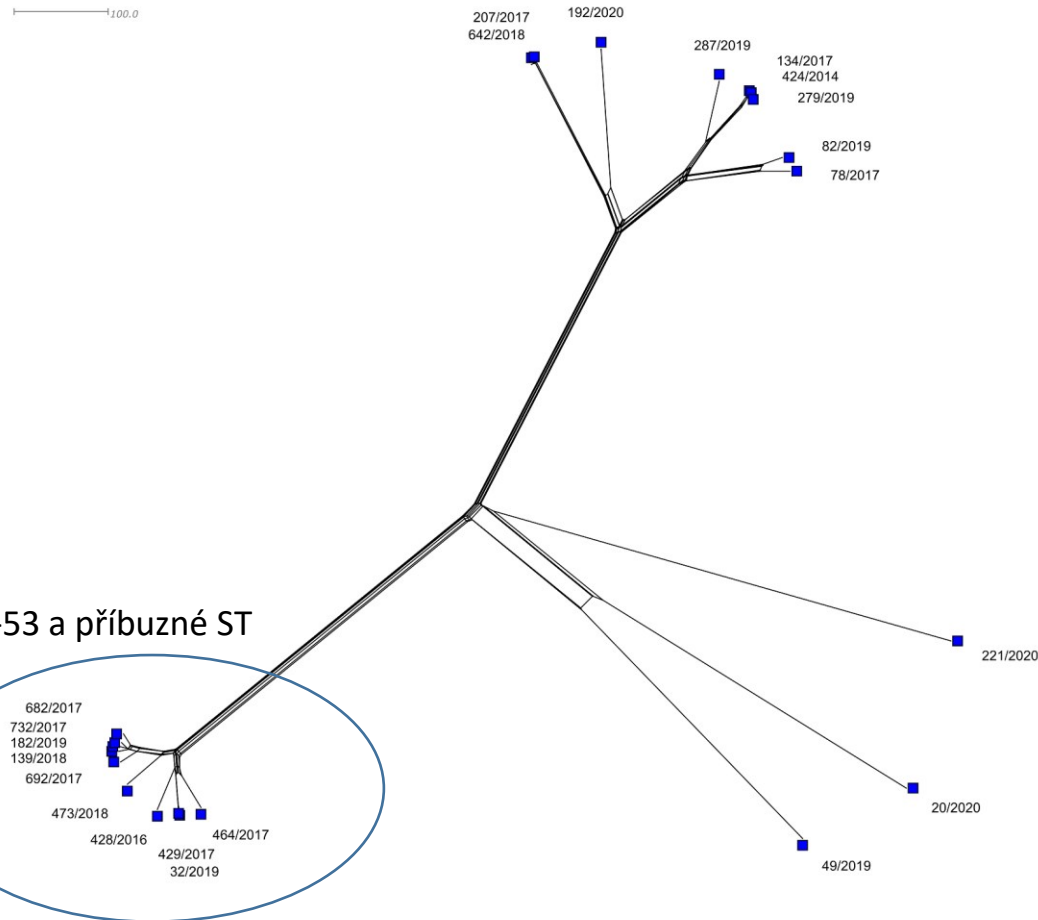


Celogenomová data vybraných izolátů *Streptococcus pneumoniae* byla analyzována pomocí **GPS Pipeline**, kdy byl u izolátů určen **GPSC (Global Pneumococcal Sequencing Cluster)**, sérotyp, sekvenční typ (ST) a dále byla zjišťována přítomnost genů pro Pilus-1 a Pilus-2, což jsou faktory virulence *S. pneumoniae* umožňující adhezi ke slizničnímu epitelu a potencující patogenezi onemocnění. GPSC je typizace používaná pro bakterie *S. pneumoniae* umožňující na rozdíl od sérotypu rozeznat skupiny izolátů, které v genomu nesou geny pro antibiotickou rezistenci, geny k tvorbě faktorů virulence či skupiny izolátů mající sklon k tzv. serotype replacementu.

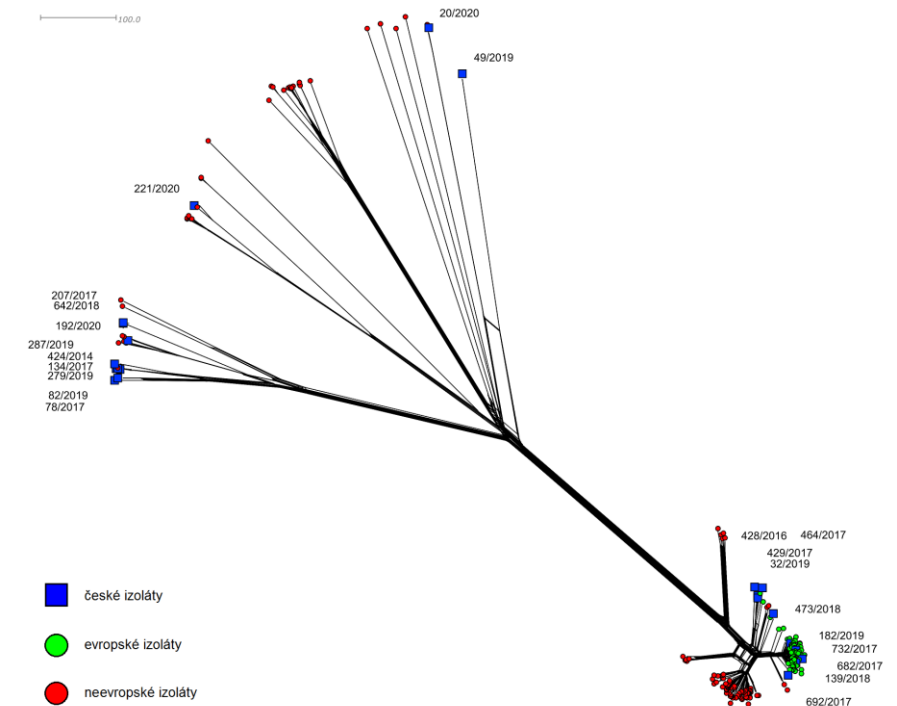
Byla provedena analýza celkem 35 izolátů *S. pneumoniae* patřící ke třem GPSC - 31 izolátů patřilo ke GPSC6, 2 izoláty ke GPSC16 a 2 izoláty ke GPSC43. Tyto tři GPSC obsahovaly 5 sérotypů – 2x 9N, 10x 9V, 1x 11A, 21x 15B z toho 6x 15B/C, 1x 22F. Na přiloženém grafu jsou zobrazeny GPSC a k nim přiřazené sérotypy, a dále výsledek ST a výsledek testování přítomnosti genů pro Pilus-1 a Pilus-2. Analyzováno bylo 21 izolátů sérotypu 15B. Sérotyp je v NRL STR určován pomocí kombinace PCR a Quellung reakce. Sérotyp 15B není pomocí metody Quellung rozlišitelný od sérotypu 15C, spolehlivou metodou rozlišující tyto sérotypy je celogenomová sekvenace. Analýzou celogenomových dat bylo zjištěno, že 6 izolátů určených Quellung reakcí jako 15B je co do genové výbavy sérotypu 15C. Všechny izoláty 15B (15B/C) patří ke GPSC6 a ST-162, všechny mají negativní gen pro Pilus-2, ale gen pro Pilus-1 je u části izolátů pozitivní a u části negativní. Skupina izolátů sérotypu 9V se řadí ke dvěma různým GPSC (GPSC6 a GPSC43) a šesti různým ST. U izolátů sérotypu 9V GPSC43 nebyly zjištěny geny pro Pilus-1 a Pilus-2. Skupina izolátů sérotypu 9V GPSC6 má pozitivní gen pro Pilus-1.



Fylogenetická analýza českých izolátů *S. pneumoniae* sérotypu 8 (n = 22) z let 2014 až 2020 metodou sekvenace celého genomu



Fylogenetická analýza českých a zahraničních izolátů *S. pneumoniae* sérotypu 8 z let 2014 až 2020 na základě dat z celogenomové sekvenace (zdroj genomů zahraničních izolátů – PubMLST DB)



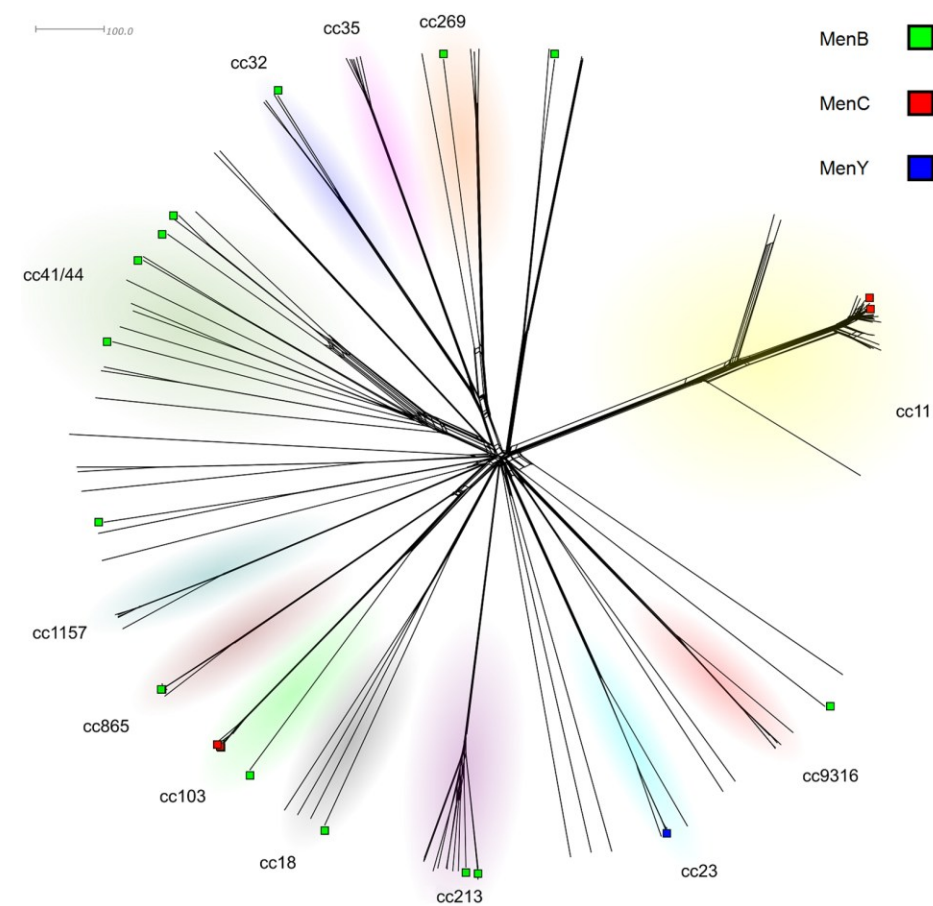
Neisseria meningitidis



V důsledku zavedení epidemiologických opatření proti pandemii onemocnění covid-19, došlo v letech 2020-2022 k poklesu počtu infekčních onemocnění přenášných vzdušnou cestou, včetně invazivního meningokokového onemocnění (IMO). Do fylogenetické analýzy s využitím WGS dat byly zařazeny všechny izoláty *Neisseria meningitidis* (celkem 126) z IMO z České republiky, které byly zaslány do NRL mezi lety 2018-2024.

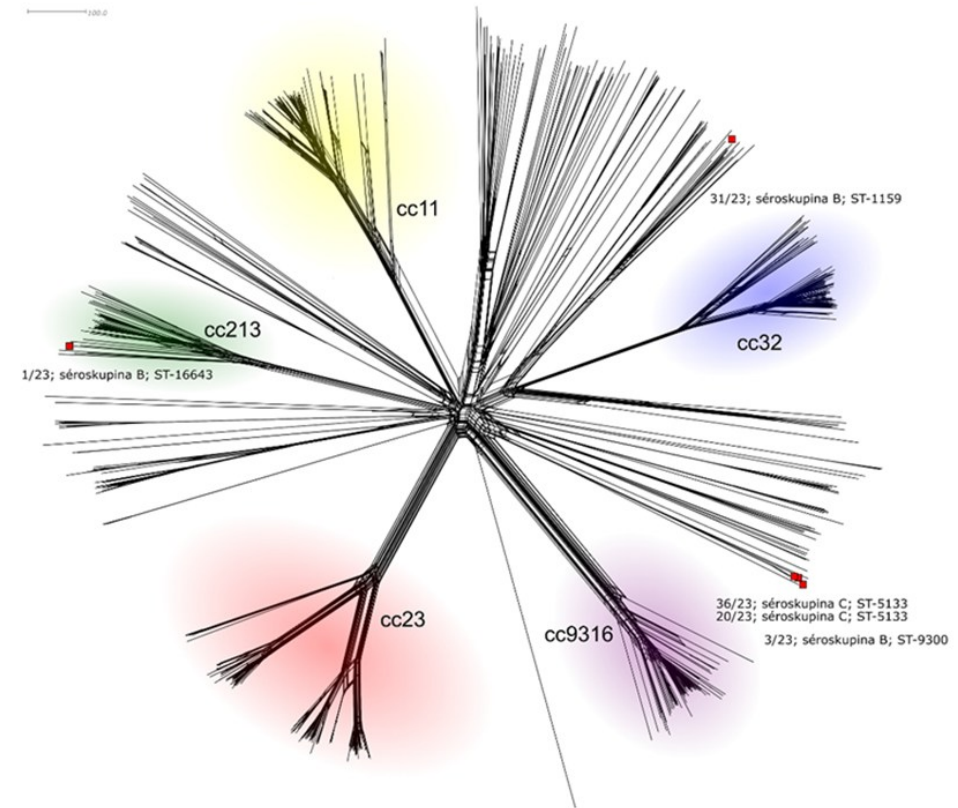
Analýza prokázala, že v průběhu covidového období (2020-2022) došlo ke změně populace meningokoků.

U izolátů MenACWY lze sledovat postupný a výrazný pokles celkové heterogenity – z 10 různých klonálních komplexů zachycených mezi lety 2018-2019 na pouhé 3 v post-covidových letech 2023-2024 (cc11, cc23 a cc103). Zároveň byla zjištěna významná redukce izolátů C; cc11, které v České republice v pre-covidových letech výrazně a dlouhodobě dominovaly. U izolátů MenB lze naopak pozorovat nárůst celkové heterogenity během období pandemie covid-19 a její opětovnou redukci na celkově nejnižší hodnoty v post-covidovém období. Toto snížení heterogenity výrazně změnilo strukturu populace izolátů MenB působících IMO v České republice. Izoláty *N. meningitidis* z post-covidových let (2023-2024) jsou na obrázku zvýrazněny a barevně odlišeny podle séro skupiny.



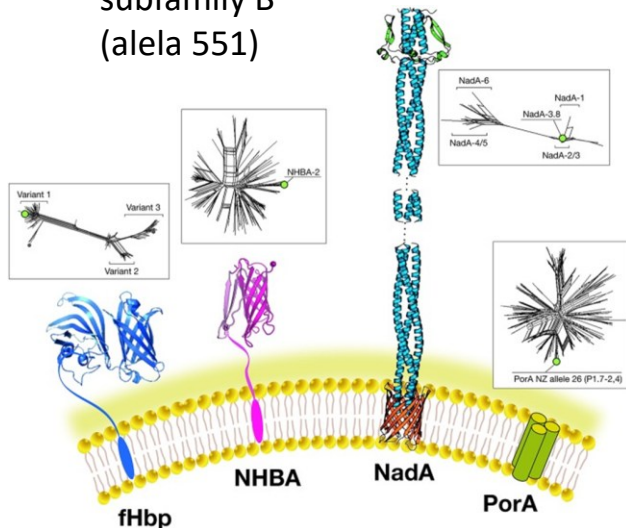
Evropská fylogenetická analýza izolátů *N. meningitidis* (n=811 izolátů) z let 2023 - 2024 metodou sekvenace celého genomu

Obsažené země v grafickém zobrazení: Belgie, ČR, Řecko, Itálie, Portugalsko, Slovinsko, Slovensko, Španělsko, Německo a Francie
Zvýrazněné izoláty *N. meningitidis* z ČR



MenB vakcíny

- | | |
|---|---|
| ➤ MenB-4C
(Bexsero) | ➤ MenB-FHbp
(Trumenba) |
| - multikomponent-
ní proteinová
vakcína | - bivalentní
proteinová
vakcína |
| - fHbp peptid
(alela 1) | - fHbp peptid
subfamily A
(alela 45) |
| - NHBA peptid
(alela 2) | - fHbp peptid
subfamily B
(alela 551) |
| - NadA peptid
(alela 8) | |
| - PorA variabilní
region VR2
(alela 4) | |



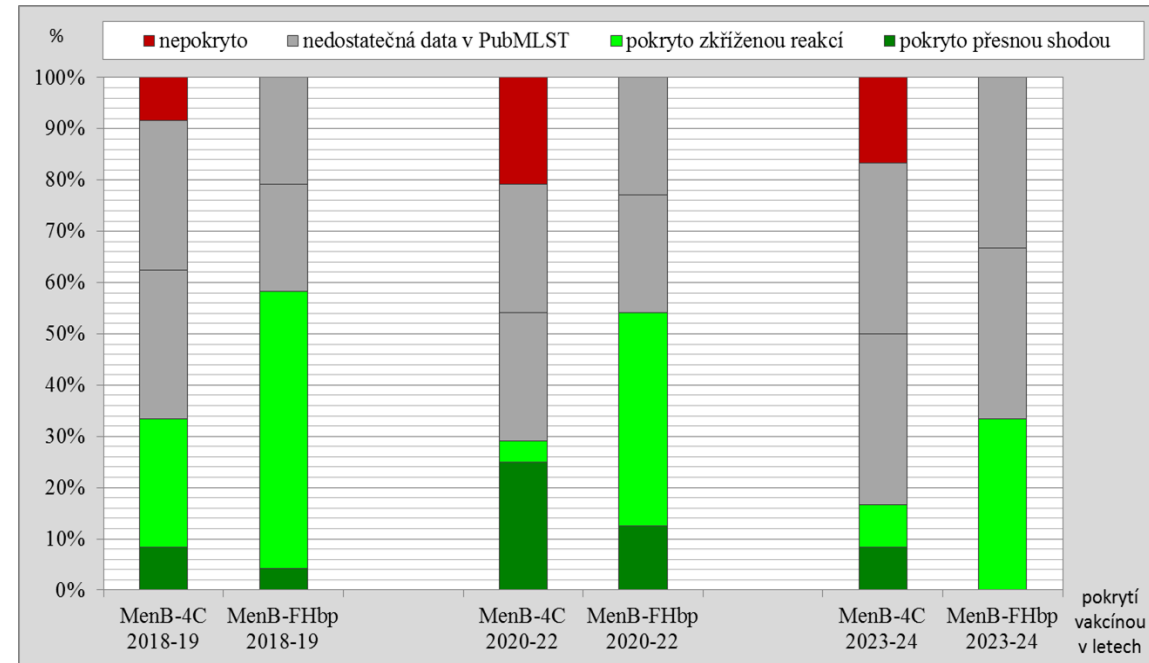
MenDeVAR

„meningococcal deduced vaccine antigen reactivity index“

- přítomnost alelických variant vakcinačních antigenů
- citlivost alelických variant na protilátky v baktericidním testu

The MenDeVAR Index:

- isolate contains ≥1 exact sequence match to antigenic variants found in the vaccine.
- isolate contains ≥1 antigenic variant deemed cross-reactive to vaccine variants through experimental studies.
- all the isolate's antigenic variants have been deemed not cross-reactive to vaccine variants through experimental studies.
- isolate contains antigens for which there is insufficient data from or are yet to be tested in experimental studies.



Teoretické pokrytí MenB izolátů má u obou MenB vakcín sestupný trend, u obou MenB vakcín je stále dostatečné.

- MenB-4C (Bexsero)
 - 2018-19: 62,5 %
 - 2020-22: 54,2 %
 - 2023-24: 50,0 %
- MenB-FHbp (Trumenba)
 - 2018-19: 79,2 %
 - 2020-22: 77,1 %
 - 2023-24: 66,7 %
- nárůst počtu izolátů, pro které nejsou dostupná data

Haemophilus influenzae



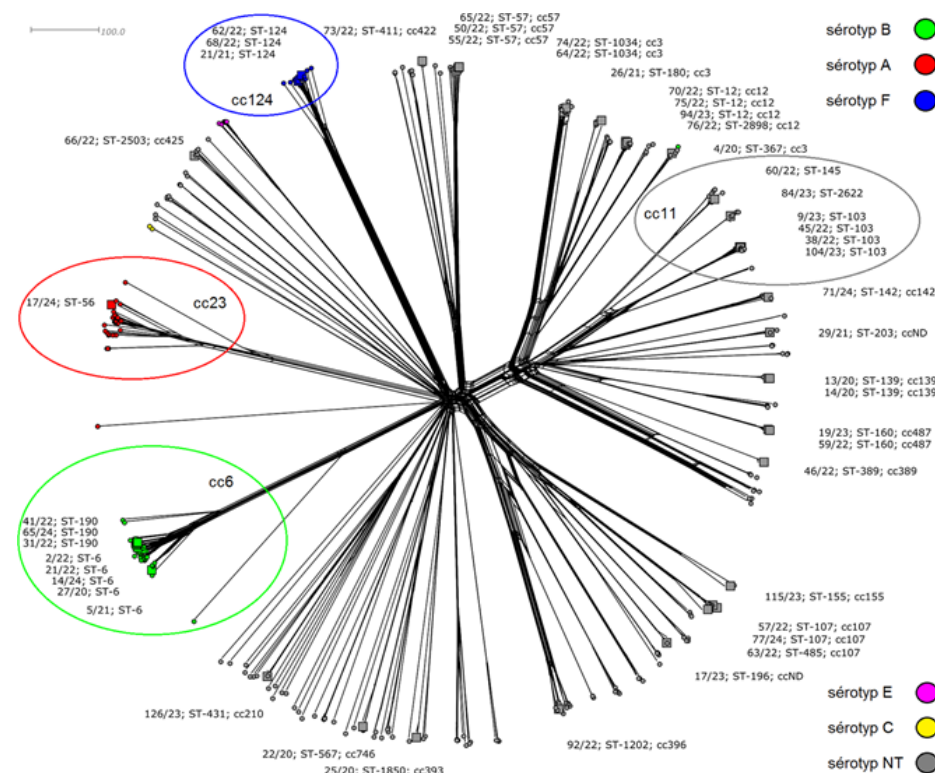
V rámci projektu HERA2CZ bylo celogenomově osekvenováno 56 izolátů *Haemophilus influenzae*.

Pro účely fylogenetické analýzy bylo vybráno 47 izolátů: 35x *H. influenzae* netypovatelný (NT), 8x *H. influenzae* sérotypu B, 3x *H. influenzae* sérotypu F, 1x *H. influenzae* sérotypu A. K českým izolátům bylo v rámci fylogenetické analýzy připojeno 340 invazivních izolátů *H. influenzae* z dalších evropských zemí (Portugalsko, n = 218; Francie, n = 105; Norsko, n = 9; Velká Británie, n = 5; Švédsko, n = 1; Dánsko, n = 1; Irsko, n = 1), které byly dostupné v databázi PubMLST. Všechny izoláty pocházejí z let 2020-2025.

Fylogenetická analýza na úrovni cgMLST (1037 lokusů) prokázala, že **opouzdřené izoláty *H. influenzae* tvoří samostatné, vysoce homogenní linie, které nejsou příbuzné k liniím neopouzdřených izolátů ani k sobě navzájem.**

S jedinou výjimkou patří všechny evropské izoláty *H. influenzae* A ke klonálnímu komplexu cc23, izoláty *H. influenzae* F patří výhradně k cc124 a pro izoláty *H. influenzae* B je typický cc6. Oproti tomu neopouzdřené izoláty patří k mnoha různým klonálním komplexům a na fylogenetické síti vytvářejí velké množství vysoce heterogenních genetických linií s různou mírou vzájemné příbuznosti.

Pomocí WGS byl recentně potvrzen první záchyt ***H. influenzae* A v ČR** (muž, možný export z Mexika).



Bordetella pertussis

V projektu HERA2CZ bylo celogenomově osekvenováno 51 kmenů *Bordetella pertussis*.

Analýza byla zatím provedena u 33 izolátů *Bordetella pertussis*, které byly izolovány v České republice v epidemickém roce 2024.

Všechny izoláty byly přiřazeny k sekvenčnímu typu ST-2, který patří do klonálního komplexu cc2.

Přítomnost genu pro pertaktin byla potvrzena u 29 izolátů, **u čtyř izolátů nebyl gen pro pertaktin nalezen**.

Celkem 27 izolátů neslo identický alelový profil ***ptxP3*, *ptxA1*, *ptxB1*, *ptxC4*, *ptxD1*, *ptxE4*, *fhaB1*, *fim2-1*, *fim3-1*** a bylo přiřazeno k antigennímu sekvenčnímu typu **BPagST-4**. Šest izolátů mělo odlišný BPagST profil v důsledku změny alelové varianty genu *fim3*. Pět izolátů neslo alelovou variantu *fim3-2* (BPagST-9) a jeden izolát variantu *fim3-26* (BPagST-93).

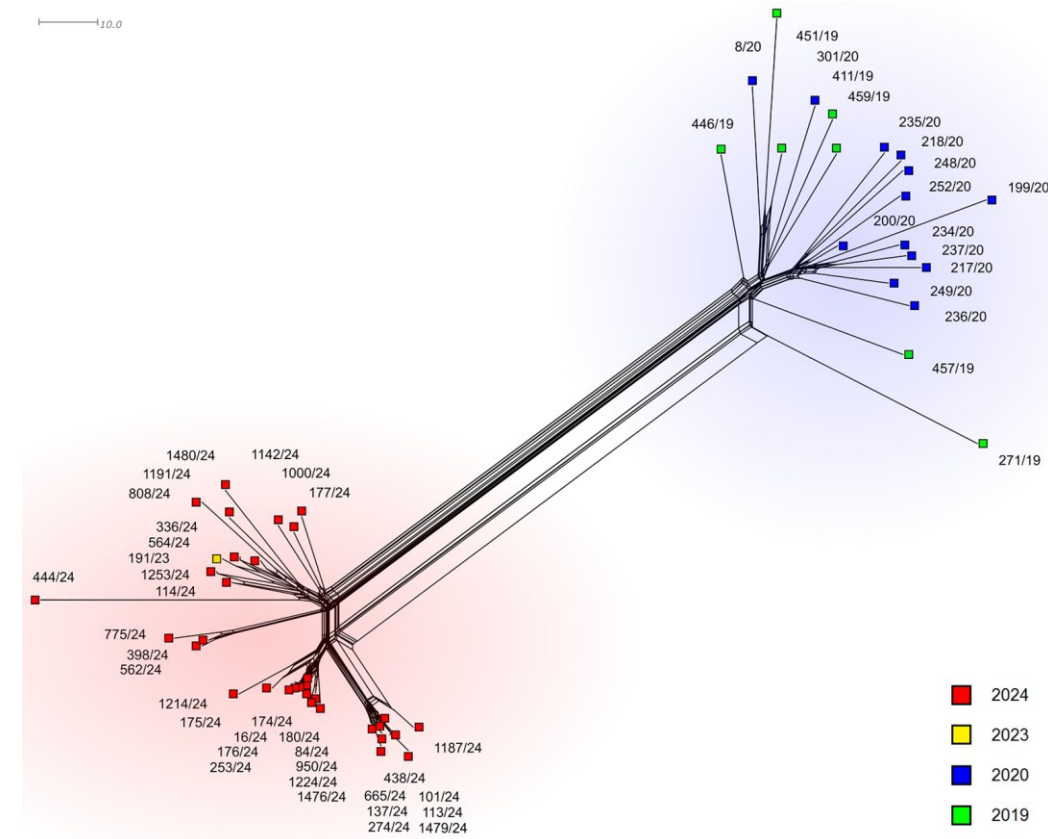
Během WGS testování byly nalezeny i **rezistentní kmeny** (3x ery, 2x coty) v NRL.

Izolát	Pertaktin	Pertusový toxin						Filamentózní hemaglutinin	Fimbrie		Antigenní sekvenční typ (BPagST)
	<i>prn</i> (5019)	<i>ptxP</i>	<i>ptxA</i>	<i>ptxB</i>	<i>ptxC</i>	<i>ptxD</i>	<i>ptxE</i>	<i>fhaB</i> 2400_5550	<i>fim2</i>	<i>fim3</i>	
0016/24	+	3	1	1	4	1	4	1	1	1	4
0084/24	+	3	1	1	4	1	4	1	1	1	4
0101/24	+	3	1	1	4	1	4	1	1	1	4
0113/24	+	3	1	1	4	1	4	1	1	1	4
0114/24	+	3	1	1	4	1	4	1	1	2	9
0137/24	+	3	1	1	4	1	4	1	1	1	4
0174/24	+	3	1	1	4	1	4	1	1	1	4
0175/24	+	3	1	1	4	1	4	1	1	1	4
0176/24	+	3	1	1	4	1	4	1	1	1	4
0177/24	+	3	1	1	4	1	4	1	1	1	4
0180/24	+	3	1	1	4	1	4	1	1	1	4
0253/24	+	3	1	1	4	1	4	1	1	1	4
0274/24	-	3	1	1	4	1	4	1	1	1	4
0336/24	+	3	1	1	4	1	4	1	1	2	9
0398/24	-	3	1	1	4	1	4	1	1	1	4
0438/24	+	3	1	1	4	1	4	1	1	1	4
0444/24	+	1	1	1	1	1	4	1	1	26	93
0562/24	-	3	1	1	4	1	4	1	1	1	4
0564/24	+	3	1	1	4	1	4	1	1	2	9
0665/24	+	3	1	1	4	1	4	1	1	1	4
0775/24	-	3	1	1	4	1	4	1	1	1	4
0808/24	+	3	1	1	4	1	4	103	1	2	136
0950/24	+	3	1	1	4	1	4	1	1	1	4
1000/24	+	3	1	1	4	1	4	1	1	1	4
1142/24	+	3	1	1	4	1	4	1	1	1	4
1187/24	+	3	1	1	4	1	4	1	1	1	4
1191/24	+	3	1	1	4	1	4	1	1	2	9
1214/24	+	3	1	1	4	1	4	1	1	1	4
1224/24	+	3	1	1	4	1	4	1	1	1	4
1253/24	+	3	1	1	4	1	4	1	1	2	9
1476/24	+	3	1	1	4	1	4	1	1	1	4
1479/24	+	3	1	1	4	1	4	1	1	1	4
1480/24	+	3	1	1	4	1	4	1	1	2	9

Fylogenetická analýza a genetická diverzita českých izolátů *B. pertussis* z let 2019 až 2024

Pro porovnání genetické diverzity epidemických izolátů z roku 2024 byly využity dříve osekvenované české izoláty *B. pertussis* z let 2019, 2020 a 2023, které patřily rovněž k ST-2.

Fylogenetická analýza (3592 lokusů) prokázala, že **izoláty formují dvě geneticky odlišné subpopulace**. Jednu subpopulaci tvoří všechny izoláty z roku 2024 ($n = 33$) a jeden izolát z roku 2023. Druhá subpopulace je tvořena izoláty z let 2019 ($n = 7$) a 2020 ($n = 13$). Pomocí WGS dat bylo možné rozlišit homogenní populaci českých izolátů *B. pertussis* ST-2 a prokázat, že **izoláty z roku 2024, které způsobily na našem území epidemii, jsou geneticky odlišné od izolátů *B. pertussis* z let 2019-2020.**



Deliverable WP4: D4.1 Implementation of WGS protocol for *N. meningitidis*, *S. pneumoniae*, *H. influenzae* and *B. pertussis* in NRLs



Metoda WGS byla realizována na platformě **Illumina MiSeq**. K základnímu zpracování WGS dat principem ***de novo* assembly** byl použit software **Velvet *de novo* Assembler**. Výsledné genomy bakteriálních izolátů byly nahrány do **mezinárodních databází**, které využívají platformy BIGSdb (PubMLST, BIGSdb-Pasteur) a ve kterých jsou veřejně dostupná. Díky systému automatického skenování jsou u každého genomu označeny známé alelové varianty druhově specifických genů, včetně genů multilokusové sekvenační typizace (MLST). Kombinace alelových variant MLST genů následně stanoví sekvenační typ izolátu a příslušnost ke klonálnímu komplexu (cc). Výše uvedené databáze zprostředkovávají také základní analýzu a **vzájemné porovnání genomů pomocí softwaru Genome Comparator**. Tato analýza může probíhat na širokém spektru rozlišení (MLST, rMLST, cgMLST) a umožňuje i srovnávání s izoláty z celého světa, které jsou v databázi PubMLST dostupné. Genome Comparator produkuje distanční matice, které jsou založené na počtu a alelické variabilitě všech analyzovaných lokusů u každého z analyzovaných izolátů. Na základě těchto matic jsou poté generovány **fylogenetické sítě** softwarem **SplitsTree4**.

Deliverable WP4: D4.2 Implementation of WGS based surveillance of invasive diseases caused by meningococci, pneumococci, haemophilus and *B. pertussis*

- Charakterizace agens (určování ST, klonálních komplexů, cgMLST)
- Charakterizace vakcinačních antigenů
- Geny rezistence
- Užití v molekulární surveillance in real-time
- Fylogenetické analýzy (porovnání českých izolátů či mezinárodních) včetně analýz v čase
- Informace při řešení epidemií
- Monitoring přeshraničních výzev

Souhrn



Přestože WGS je velice účinnou metodou, je náročnější na zdroje ve srovnání s jinými sekvenčními metodami.

Proces sekvenování celého genomu generuje velké množství dat, což představuje značné nároky na moderní technologickou výbavu a dostatečné výpočetní a analytické kapacity. Správa, sdílení a archivace WGS dat vyžaduje datové kapacity, pravidelnou údržbu systému a zabezpečení.

Projekt HERA2CZ se zaměřuje i na zvýšení celogenomové sekvenační kapacity v rámci jednotlivých NRL Státního zdravotního ústavu v Praze a rozšíření spektra WGS charakterizace na různé lidské patogeny, čímž nejen zvyšuje schopnosti monitorování infekčních chorob, ale také posiluje koordinační a reakční mechanismy systémů ochrany veřejného zdraví napříč Evropou. Zvýšením kapacity pro WGS sekvenaci a zpřesněním metodiky přispívá projekt HERA2CZ ke zlepšení systému ochrany veřejného zdraví a umožňuje detekci hrozby přeshraničního šíření infekčních onemocnění.

Budoucnost?



- Neustálý přístrojový **rozvoj technologií** celogenomového sekvenování, stejně tak jako vývoj chemie a vyhodnocovacích a hlásicích aplikací

Př. nové nástroje pro vizualizaci dat WGS se očekávají v rámci platformy ECDC EpiPulse, která byla zavedena v polovině roku 2024.

- **Zavedení systému WGS do surveillance** onemocnění způsobených meningokoky, pneumokoky, hemofily a *B. pertussis* a sdílení těchto údajů na národní i mezinárodní úrovni přispěje k rychlejšímu sdílení údajů a většímu propojení mezi evropskými státy.

- Oproti obrovskému mezinárodnímu sdílení WGS dat v současné době nejsou **údaje WGS sdíleny na národní úrovni**. SZÚ jako centrální místo pro celogenomové sekvenování vybraných patogenů v ČR vytvoří neomezený „WGS repozitář“, který umožní export dat do mezinárodních databází a integraci s mezinárodním systémem varování před hrozbami. Probíhají jednání o propojení „WGS repozitáře“ SZÚ s Informačním systémem infekčních nemocí v ČR (ISIN).

- **Udržitelnost** (nutnost snížení nákladů na WGS 1 vzorku na co nejnížší úroveň, automatizace, centralizace, podpora provádění WGS na národní úrovni)

- Nutnost vyřešit otázku zálohování získaných dat versus opakované sekvenování

- **Financování?!**



Děkuji za pozornost

Mgr. Zuzana Okonji

zuzana.okonji@szu.cz

Prezentovaný výzkum byl realizován týmem pracovníků Státního Zdravotního Ústavu díky projektu:

Genomická surveillance vybraných infekčních nemocí v České republice



Co-funded by the
European Union

Projekt „Genomická surveillance vybraných infekčních nemocí v České republice“ (HERA2CZ) je spolufinancován Evropskou unií. Prezentované názory a postoje však představují pouze názory autorů a nemusí nutně odrážet názory a postoje Evropské unie nebo Evropské výkonné agentury pro zdraví a digitální oblast (HaDEA). Ani Evropská unie, ani Evropská výkonná agentura pro zdraví a digitální oblast (HaDEA) za ně nemohou nést odpovědnost.