

## 1-DAV-302: Princípy dátovej vedy, Zima 2023

### Domáca úloha č. 1

**Termín odovzdania: 24.10.2023, 22:00**

Kým koncom roku 2021 prevládal v Európskych krajinách variant vírusu SARS-CoV-2 označovaný ako Delta, začiatkom roku 2022 ho postupne začal vytláčať nový variant s označením Omicron a neskôr jeho subvariant, ktorý mal označenie BA.2. Na linke uvedenej nižšie nájdete prehľad registrovaných vzoriek v databáze GISAID z obdobia 1.12.2021-1.5.2022 z krajín Slovensko, Česká republika a Dánsko.

- a) Vytvorte prehľadnú vizualizáciu, ktorá demonštruje pre Slovensko, Českú republiku a Dánsko časový priebeh, ako variant Delta bol v populácii na Slovensku postupne nahrádzaný variantom Omicron a neskôr jeho subvariantom BA.2. Na vyhladenie dát použijete najskôr metódu pohyblivého okna a neskôr metódu lowess z prednášky. Ktorá metóda sa vám zdá na tento účel lepšia a prečo?  
(Na určenie dátumu odobratia vzorky použijete stĺpec "Collection date" a na určenie variantu použijete stĺpce "Variant" a "Pango lineage" pre identifikáciu subvariantu BA.2.)
- b) Pre každú z troch krajín odhadnite dátum, kedy podiel variantu Omicron prekročil 50%. (Takáto informácia je napríklad zaujímavá pre klinických lekárov, keďže niektoré lieky, ktoré fungovali na variant Delta nefungovali na variant Omicron.)
- c) Dňa 13.1.2022 spoločnosť Unilabs Slovensko osekvenovala plošne všetkých 346 pozitívnych vzoriek testovaných v ich laboratóriách, a z týchto vzoriek bolo 211 boli variant Omicron a zvyšné boli variant Delta. Tieto dáta nikdy neboli uložené do databázy GISAID. Zmenila by táto informácia výsledky vašich analýz pre Slovensko?
- d) Ak by ste pracovali len s dátami, ktoré boli k dispozícii na konci januára 2022, ovplyvnilo by to nejakým spôsobom výsledky vašich analýz? (Dátum pridania dát do databázy nájdete v stĺpci "Submission date".)
- e) Dáta z ktorej z krajiny by ste považovali za najinformatívnejšie / najspoľahlivejšie a prečo?

Váš kód a vaše odpovede odovzdajte do Google Classroom vo forme komentovaného notebooku.

Link na dataset: <https://compbio.fmph.uniba.sk/vyuka/pridav/data/gisaid-selection.tsv>