

Inzet van een centrale Callisto database voor de bestrijding van SARS-CoV-2 in NL

Auteurs: (10)(2e) (10)(2e) (10)(2e) (10)(2e) (10)(2e) (Centrum IDS)

De sequenties die wereldwijd worden bepaald van SARS-CoV-2 worden gedeeld in de GISAID database. In het breed gebruikte Nextstrain programma worden deze sequenties, inclusief een kleine set metadata, geanalyseerd en in zeer informatieve, interactieve visualisaties publiek gepresenteerd (<https://nextstrain.org/>). De data worden niet dagelijks bijgewerkt, de sequenties in deze database zijn niet allemaal even goed van kwaliteit en de metadata zijn beperkt. Hierdoor is de waarde van deze visualisaties voor de bestrijding van de uitbraak in NL beperkt. Zowel op het RIVM als op het EMC zijn virologen en epidemiologen met diverse tools de beschikbare sequenties aan het cureren en analyseren, om informatieve en betrouwbare visualisaties te leveren die gebruikt kunnen worden om het verloop van de SARS-CoV-2 uitbraak te duiden en te presenteren in bijvoorbeeld RT's, OMT's, publicaties en nieuwsberichten.

Bij IDS-vir is een eigen, besloten, versie van de Nextstrain ingericht, het Callisto platform, waarin voor verschillende virussen de Nederlandse sequenties worden geanalyseerd (ontwikkelaars: (10)(2e) en (10)(2e)). Dit platform is heel geschikt voor het hier boven beschreven doel. Met name in de volgende fase van de uitbraak (voorbij de 'piek'), kan dit ingezet worden in de opsporing van transmissies en nieuwe introducties in Nederland. Als voorwerk heeft (10)(2e) al een Callisto omgeving ingericht, waarbij hij de GISAID data van Nederland heeft ingevoerd (met dus nog de zeer beperkte set metadata die geupload wordt naar GISAID). Hieronder een aantal screenshots. Callisto is ook direct te bekijken via : <http://corona.callisto.rivm.nl/SARS-COV2> (N.B. uit beveiligings- en privacy-oogpunt is het alleen toegankelijk vanuit de CAMPUS werkplek; inloggen met de RIVM e-account). Daar is het interactieve karakter van het platform te zien.

De sequenties worden geanalyseerd via de zgn. maximum likelihood en de resulterende boom kan fylogenetische afstanden of afstanden in de tijd (sample date) aangeven. De takken en nodes kunnen worden voorzien van een kleur op basis van de gekoppelde metadata (in dit voorbeeld dus nog beperkt tot provincie en clade). Op de takken of nodes van de boom kunnen specifieke mutaties worden getoond. Verder kan er ingezoomd worden op specifieke clades of periodes etc. voor meer gedetailleerde informatie. Naast de boom worden kaartjes gegenereerd en plots met daarop geïdentificeerde mutatie-hotspots in de (Nederlandse) SARS-CoV-2 genomen.

Dankzij het besloten karakter van deze applicatie kan er meer gedetailleerde data in dan nu in het openbare GISAID gedeeld wordt. Daarmee kan sequentie analyse substantieel bijdragen aan het inperken van de uitbraak.

De meest noodzakelijke informatie die hiervoor gebruikt moet worden is op dit moment de plaatsbepaling. Voor toevoeging van 4-cijferige postcode of NUTS-3 code zouden we de stamnamen moeten koppelen met Osirisnummers en/of labnummers (lang niet alle cases waarvan sequenties zijn bepaald worden in Osiris gemeld). Toevoeging van Osirisnummer maakt ook dat EPI en LCI gemakkelijker de juiste case/sequentie kunnen herkennen.

Om dit platform centraal in te kunnen zetten in de SARS-CoV-2 bestrijding is het noodzakelijk om:

- 1 dagelijks geüpdatet, gecureerde alignment van de sequenties te uploaden
 - o Alle NL sequenties + een kleine geselecteerde set referenties vanuit de rest van de wereld
- Dagelijkse upload van de relevante metadata: monsternamedatum, postcode-4/NUTS-3, geboortedatum/leeftijd, geslacht, en de beschikbare identifiers (labnummer, osirisnummer)
- Toegang tot het besloten Callisto platform voor medewerkers van RIVM en EMC die betrokken zijn bij de bestrijding

Op dit moment wordt er alleen gesequenced door EMC en RIVM. De werkwijze zou nu dus als volgt kunnen zijn: RIVM stuurt nieuwe sequenties naar EMC. Die worden bij EMC samen met hun eigen sequenties toegevoegd aan de gecureerde alignment. Deze alignment wordt door (10)(2e) in Callisto gezet. EMC en RIVM sturen beiden csv's naar (10)(2e) met Fasta Id, + overige metadata (zie hierboven). Dagelijks wordt er een nieuwe analyse gemaakt in Callisto.

Op het moment dat er ook andere labs gaan sequencen zouden wij die data ook willen toevoegen aan de set. Waarschijnlijk is het wettelijk niet toegestaan om al deze labs ook toegang tot de SARS-CoV-2 Callisto te geven. Om te zorgen dat deze labs er toch iets aan hebben om data te delen: - dagelijkse nieuwsbrief met screenshots? Of een publieke NL Nextstrain met zeer beperkte set metadata?

