

SARS-CoV-2 har spridit sig och muterat. Bild från databasen GISAID (gisaid.org).

SARS-CoV-2

– undersökning med bioinformatik (lärrarhandledning)

Detta är en bioinformatikövning som byggs kring ett SARS-CoV-2-case där eleven får rollens om bioinformatiker på ett sjukhus. Uppgiften är att ta reda på så mycket som möjligt utifrån en sekvensering av patientens viurs.

Bioinformatikövningen kan kopplas till centralt innehåll i styrdokumentet för gymnasiekurserna Biologi 1 och Biologi 2 (eller Bioteknik) som exempelvis:

Biologi 1:

- Genetikens användningsområden. Möjligheter, risker och etiska frågor.
- Användning av genetiska data för studier av biologiska sammanhang.

Biologi 2:

- Mikroorganismer och deras betydelse för hälsa och sjukdom, och evolutionära processer.
- Användning av genetiska data för studier av biologiska sammanhang.

Elevinstruktionen inleds med bakgrundsinformation om patientens situation och läkarens frågor som kommit via epost (OBS! I Uppgift 6 ska eleven summera arbetet i ett epost-svar tillbaka till lä(k/r)aren):

1. Har viruset några särskilda förändringar i genen för spikeproteinet i den del av viruset som fäster till människans celler via ACE2-receptorn?
2. Är viruset av någon av de kända varianterna som är i omlopp? ("Brittiska", "Norska" eller "Sydafrikanska").
3. Kan vi se om patienten har smittats av kollegan som nyss varit i Italien?

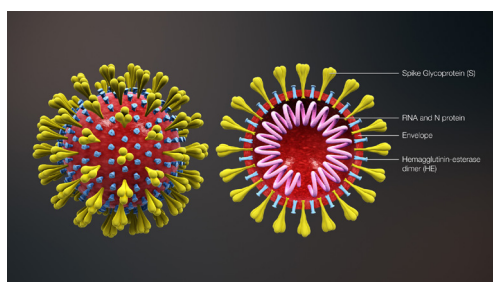
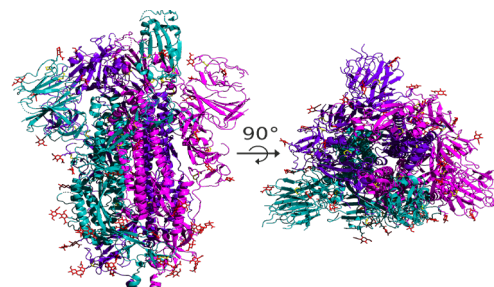


Illustration av SARS-CoV-2-viruset med S-proteinet i gul färg. Varje S-protein består av tre peptider.

Bild vänster: www.scientificanimations.com, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Co



I viruset finns en gen för S-proteinet. Tre S-peptider sätts ihop och bildar en trimer som här visas i grönt, rosa och lila. Dels från sidan och dels uppifrån.

Bild: Public domain (Wikimeida commons)



Spike-proteinet kallas också för surface glukoprotein, eftersom det sitter på virusets yta. I bilderna på första sidan visas olika illustrationer. Bilden till vänster visar att virusets yta är täckt av många utstickande S-proteiner (gula). Varje S-protein är sammansatt av tre peptidkedjor som sitter ihop i en trimer (tri = tre och mer = byggsten). Läs bildtexterna.

Viruset är ett RNA-virus, men sekvensen fås i form av en DNA-sekvens. Det finns två anledningar till att vi valde att göra bioinformatikövningen utifrån sekvensen för spikeproteinet (och inte hela virussekvensen):

- Nukleotid-sekvensen är totalt "bara" 3813 baser lång. Det får plats i en ruta på en A4 och är ganska enkelt att kopiera och klistra in (jämfört med om vi skulle haft närmare 30000 baser lång sekvens).
- Spike-proteinet är biologiskt intressant eftersom det påverkar receptorbindningen och därmed virusets förmåga att infektera värdceller.

Ett problem med att enbart jobba med en del av virusets genom är att vi har för lite information för att säkert kunna svara på frågan om geografiska ursprunget. Vi får helt enkelt för många träffar i databasen som matchar. Därför avslutas övningen med att man får ett accession number för den träff som matchar bäst när man kör hela genomet.

Översikt av övningens delar med facit och kommentarer och tips i förhållande till frågeställningarna som ges i caset

Uppgift 1 Fokus på spikeproteinet och hur muterat det är

Uppgift 2 Fokus på att försöka koppla smittan till Italien bara med spikeproteindata (går ej)

Uppgift 3 Ger en lösning på Italen-frågan utifrån hela virusgenomet

Uppgift 4-8 Värdering av metoden, sammanfattning av resultat samt reflektionsfrågor utan specifikt nya bioinformatik-analyser

1. Har viruset några särskilda förändringar i genen för spikeproteinet i den del av viruset som fäster till människans celler via ACE2-receptorn?

Uppgift 1 - Hur muterat är spikeproteinet?

Här används databasen gisaid.org och funktionen CoV-surver App där man tittar på 3D-bilder av spikeproteinet där mutationer/variationer visas som färgade bollar.

Några tips/varningar på vad man kan fastna på (fastna inte på detta!):

- Färgerna på bollarna har lite olika betydelser, men det räcker att veta att orange färg motsvarar en förhöjd risk för att en mutation/variation kan påverka virusets funktion (t ex förmåga att infektera).
- En variation/mutation är markerad med flera bollar som i en grupp. Tänk att det markerar en aminosyra där kolatomerna motsvarar kulorna i bollen (en kulmodell av molekyler kombineras med en enklare "kabelmodell" i grått för resten av aminosyrakedjan).
- Om man fastnar i ett inzoomat/utzoomat läge måste man ladda om webbsidan och börja om.
- När man placerar musen på en färgad boll eller en position i proteintrasslet så dyker det upp en beteckning där siffran motsvarar positionen i proteinet men det kan vara lite svårt ibland att förstå vad man ser/pekar på. Fastna inte på detta!



Svarförslag till deluppgifterna:

A. Hur många skillnader i aminosyror är det? SVAR: 10 st.

Stämmer det med antal färgade bollar i bilden? SVAR: Nej, det är fler färgade bollar än 10 och det beror på att det vi ser i 3D-bilden är trimerer dvs spikproteinet består av tre proteinkedjor som sitter ihop (jämför med bilderna på sida 1 i uppgiften). OBS! Antalet bollar stämmer inte riktigt (det borde vara 30 st, men vi ser bara 27, det beror förmodligen på att det finns två variationer som ligger väldigt nära varandra (H69del och V70del(69)).

B. Vilken beteckning har variationen närmast RBD (receptorbindande området)?

SVAR: N501Y. Alltså har vi ett svar på fråga 1 här: viruset har en mutation/Variation i spikproteinet nära det receptorbindande området. OBS! En kommentar här är att man kan ju även tänka sig att mutationer på andra positioner i proteinkedjan indirekt kan påverka formen och funktionen så att bindningen till receptorn påverkas. Det är alltså en förenkling att säga att enbart mutationer i området närmast RBD är intressanta. Om man tittar på bilden i gisaid så ser man att det finns ytterligare två orange-färgade variationer/mutationer i positionerna 69-70 (två deletioner) och 614. De kan alltså också påverka virusets funktion.

C. Skriv upp alla beteckningar för alla varianter som du fått fram för spikeproteinet. ar deluppgifterna. SVAR: H69del V70del(69) Y144del(143) N501Y A570D D614G P681H(674) T716I S982A D1118H. Kanske är det onödigt skriva ner (man har dem på skärmen)? De ska användas i nästa uppgift för att logiskt jämföra med en given tabell för de olika varianterna.

2. Är viruset av någon av de kända varianterna som är i omlopp? ("Brittiska", "Norska" eller "Sydafrikanska").

D. Vilken variant är patientens virus mest lik? Vilken kod har den?

Tabellen visar tre varianter av SARS-CoV-2 som ofta nämnts i mediareporteringen under 2021.

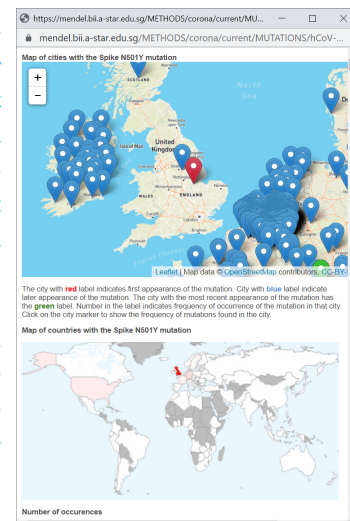
Virusvarianterna har olika koder. Några karaktäristiska variationer på aminosyranivå som förekommer hos de olika virusvarianterna bekrävs med beteckningar där siffrorna markerar en position i S-proteinet.

Namn	Kod	Karaktäristiska variationer på aminosyranivå
Brittiska varianten	B117	N501Y, 69-70del, P681H (* vissa med E484K?)
Sydafrikanska varianten	B1351	K417N, E484K, N501Y
Norska varianten	B1525	E484K, F888L

SVAR: N501Y finns i både den Brittiska och den Sydafrikanska. Men tittar vi närmare i vår lista från uppgiften C så ser vi att 69-70del och P681H finns i patientens virus och det matchar den Brittiska men saknas i Sydafrikanska. Norska utesluts då den saknar N501Y och/eller för att den ska ha mutationen F88L och E484K som saknas i vårt virus. Slutsats: vår patient bär på ett virus som kan kallas Brittiska varianten.

E. Var har aminosyrvaryationen N501Y uppkommit?

SVAR: I England. Se bild till höger. Man behöver zooma in kartan numera eftersom det är så tätt mellan fynden av denna mutation i världen. Den rödmarkerade visar var N501Y-variationen först upptäcktes och den gröna visar det senaste registrerade fallet med N501Y-variationen.





F. Hittills har analysen fokuserat helt på virusets arvs massa för S-proteinet. Jämförelserna har gjorts mot det ursprungliga viruset från Wuhan. Kan det finnas fler mutationer i virusets arvs massa som skiljer patientens virus från den ursprungliga varianten? Motivera ditt svar!

SVAR: Ja, det är högst troligt att det finns fler mutationer än det vi ser när vi enbart tittar på aminosyranivå i ett av virusets proteiner (Spike-proteinet). Det kan finnas mutationer som är tysta - dvs som inte märks som aminosyravarianter i spikeproteinet. Det kan också finnas mutationer i andra delar av virusets arvs massa som ju totalt var nästan 30000 bp, här har vi endast tittat på mindre än 4000 bp.

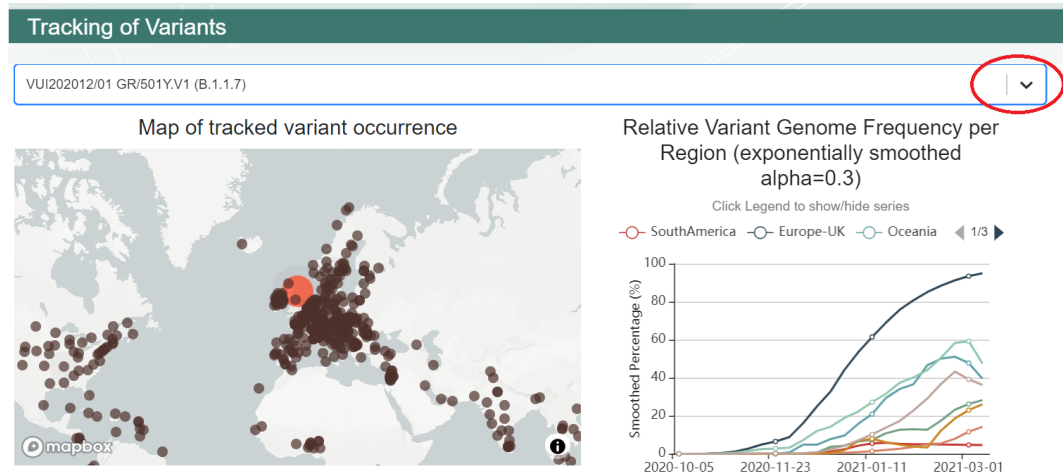
3. Kan vi se om patienten har smittats av kollegan som nyss varit i Italien?

För att komma åt denna fråga ska man pröva några olika verktyg - men eftersom det är svårt att svara på frågan kanske man upplever att de inte ger så mycket? Ett förslag kan vara att du som lärare demonstrerar t.ex. Uppgift 2 Metod I.

Uppgift 2 - Kommer smittan från Italien?

Metod I: Geografisk information om förekomst av kända virusvarianter

Använder gisaid.org men funktionen hCoV-19 Tracking of Variants. Man kan välja med rullgardinmenyn som är inringad i bilden nedan.



Kartans bruna cirklar markerar länder och/eller delstater i länder. När man för musen över punkterna ser man hur många registrerade fynd man gjort av den virusvariant som man valt att titta på.

A. Hur många olika virus följer man just nu i denna databas? **SVAR: 5 st (8 april 2021)**

B. Välj den virusvariant (kod) som du kom fram till stämmer bäst på patientens sekvens och studera kartan. Hur vanlig är denna variant just nu i Sverige där du bor? (finns den alls? i delar av landet eller i hela landet?) **SVAR: 8 april 2021 hittar vi över tio rapportplatser i Sverige. I Västernorrland exempelvis står att man registrerat 101 st med den brittiska varianten (B117). OBS! Siffrorna står alltså inte för olika varianter utan helt enkelt för antalet fynd av den varianten som man valde i rullgardinmenyn.**

C. Kan man utifrån kartans information avgöra om patienten kan ha fått viruset av sin kollega som rest i Italien? Ja eller Nej? Motivera ditt svar! **SVAR: Nej det är omöjligt att använda denna information för att säkert säga att det är kollegan som varit i Italien som smittat eftersom den brittiska varianten nu är så spridd i hela landet. Sannolikheten för att smittan kommit från kollegan som varit i Italien är inte högre än att det kan vara någon annan som smittat på grund av att denna variant blivit så vanlig. För att kunna avgöra om det är kollegan som smittat måste vi få mer information. T ex testa hela virusets sekvens och se om det säger något mer.**



Metod II: Studera sekvensen i detalj med en virusdatabas

Här används databasen www.ncbi.nlm.nih.gov/labs/virus/vssi och verktyget *Search by sequence*.

Träffbilden ger en tabell-vy (se bild nedan).

Coverage visar hur stor del av den sekvens vi matat in som har gått att matcha med träffarna i listan.

Identity visar med avrundat värde många % av nukleotiderna den inmatade sekvensen som är exakt lika mot databasens lagrade sekvenser.

Tabellen är väldigt bred. Ett sätt att få den mer hanterbar är att ta bort vissa kolumner. Via Select Columns (se röd ring vid siffran 4 i bilden nedan) kan du ta bort överflödiga kolumner så att tabellen blir mer lätthanterlig (OBS! man måste bekräfta med Apply längst ned i dialogrutan). Ta exempelvis bort Coverage, Species, Molecule type, Length, USA. Spara Geo Location som talar om var i världen som träffarna registrerats. Genom att klicka på rubrikerna på flikarna sorteras träffarna i stigande eller fallande ordning.

D. Titta igenom - är alla träffar från samma geografiska område? **SVAR: Nej, de är från olika områden. Många från USA t.ex. Man får max 500 träffar och de är förmodligen på 99% likhet alla som kommit med här. Alltså gav detta ingen tydlig signal om koppling till Italien. Vi måste titta vidare på om det finns några skillnader på detaljnivå i nukleotidsekvensen för att se om någon är mer lik än någon annan.**

NCBI Virus Sequences for discovery

Your search: nucleotide sequence
Length: 3813 nt

Database: Viral nucleotide collection
Search algorithm: BLASTN

Selected Results: 0

Download Align Build Phylogenetic Tree

5 Nucleotide (500) Protein (0) RefSeq Genome (0)

4 Select Columns

6 View BLAST Alignment of selected sequences

Accession	Coverage	Identity	Submitters	Release Date	Species	Molecule type	Len
MW699901	100%	99%	Sadri,N., et al.	2021-03-05	Severe acute respiratory s...	ssRNA(+)	29
MW699902	100%	99%	Sadri,N., et al.	2021-03-05	Severe acute respiratory s...	ssRNA(+)	29
MW699904	100%	99%	Sadri,N., et al.	2021-03-05	Severe acute respiratory s...	ssRNA(+)	29

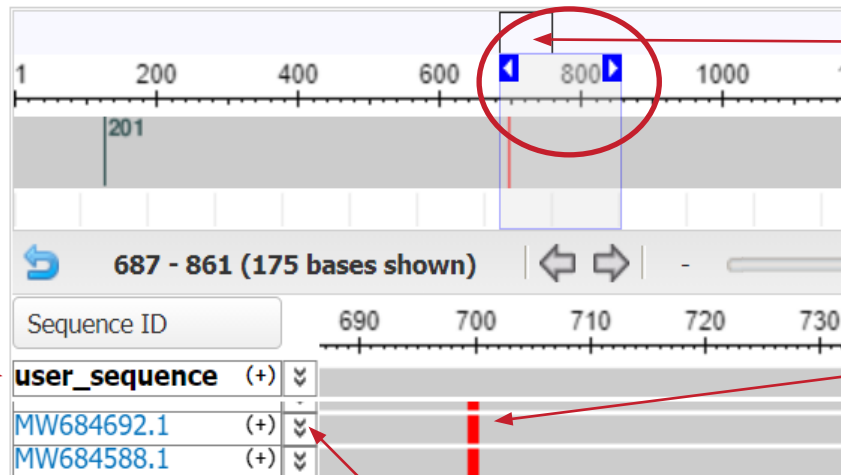
Nästa steg är att titta på hur sekvenserna ser ut i jämförelse. Det görs genom att man markerar alla träffar (vid röd ring numrerad med 5) och sedan BLAST Alignment (vid röd ring nr 6). Vår sekvens kallas "user_sequence" och skillnader markeras med röda streck.

E. Hur många skillnader finns det mellan patientens virus från träffarna i databasen? **SVAR: Ett rött streck, dvs det verkar bara finnas en mutation som skiljer vår sekvens från alla de träffar vi fått i databasen. Detta blir ett problem eftersom vi inte får något svar på om någon av alla träffar från olika delar av världen är mer lik än någon annan. Vi kan alltså inte svara på frågan ifall patientens virus kommer från Italien med denna metod heller. Men vi kan ju ändå stanna upp lite och se vad det är för typ av mutation som vår patient verkar ha som skiljer ut den från andra varianter.**

F. Vilken eller vilka nukleotider är muterade? Vilken position har de i sekvensen? **SVAR: I position 700 är det en nukleotid A i alla sekvenser vi fått fram, men vår patients virus har i denna position ett C istället.**



Graphic view of selected Blast results



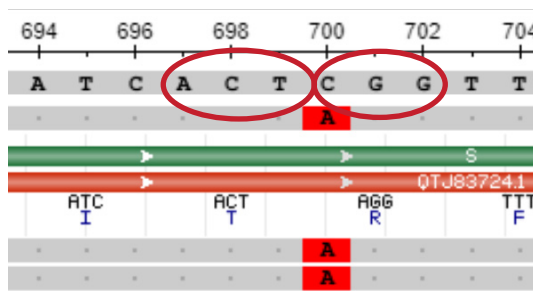
Flytta läsfönstret genom att peka och dra i rutan ovanför de blå pilarna.

Den sekvens du matat in kallas "user_sequence"

Röda streck markerar att här finns en skillnad mellan din sekvens och träffarna i databasen.

Klicka på dubbelpilen på den översta sekvensen i databasen. Då ser du kodoner:

G. Klicka på nedåtriktade dubbelsymbolen bredvid första träffen i listan. Hur har mutationen påverkat kodonet jämfört med träffarna i listan? **SVAR: Kodonet från kvävebaserna 697-699 är ACT och från 700-802 är det AGG i alla träffarna medan patientens virus har ett CGG som kodon där. Se bild nedan där dessa två kodon är inringade i patientens sekvens.**



		Second Base				
		U	C	A	G	
First Base	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA } UCG }	UAU } Tyr UAC } UAA } STOP UAG }	UGU } Cys UGC } UGA } STOP UGG } Trp	U C A G
	C	CUU } Leu CUC } CUA } CUG }	CCU } Pro CCC } CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } Arg CGC } CGA } CGG }	U C A G
	A	AUU } Ile AUC } AUA } AUG } Met or Start	ACU } Thr ACC } ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } Val GUC } GUA } GUG }	GCU } Ala GCC } GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } GGA } GGG }	U C A G

Genetiska koden av Sarah Greenwood, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Om man tittar i genetiska koden i tabellen till höger ser man att detta är en så kallad tyst mutation eftersom både AGG och CGG kodar för samma aminosyra (Arg = arginin).

H. Gav denna metod någon vägledning i varifrån patienten fått sin smitta? Varför/Varför inte? **SVAR: Tyvärr blev vi inte något klokare på ifall patienten kan ha fått viruset från sin kollega som varit i Italien. Vi har för många träffar som är lika (minst 500 st är 99% lika) och de är alla lika sinsemellan - ingen är mer lik vår sekvens.**

Här skulle man kunna återkoppla till proteinstrukturerna i 3D som vi tittade på i början - var är position 700 i förhållande till aminosyrorna? Men eftersom det är en tyst mutation kanske inte det ger så mycket?



Uppgift 3 - Analys av hela virusgenomet

Vi gjorde caset utifrån en specifik sekvens i databasen som vi vet är från Italien. Vi plockade ut sekvensen för S-proteinet och förde in den punktmutation som du hittade i position 700. Men vi har alltså inte delat hela virussekvensen på närmare 30000 baser. Det kan man ju göra i en separat fil. Som alternativ valde har vi här en liten genväg (ursäkta!). I uppgiften får eleven nämligen direkt information om det accessionsnummer som kom högst upp i träfflistan när hela genomet kördes: MW450666. Detta ID-nummer används i databasen GenBank: <https://www.ncbi.nlm.nih.gov>. Man klickar där på första träffen. Bilden nedan visar hur informationen är strukturerad när man kommer vidare till själva datat i GenBank.

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ITA/APU-POLBA01/2020, complete genome

GenBank: MW450666.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS	MW450666	29780 bp	RNA	linear	VRL 08-JAN-2021
DEFINITION	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ITA/APU-POLBA01/2020, complete genome.				
ACCESSION	MW450666				
VERSION	MW450666.1				
KEYWORDS	.				
SOURCE	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)				
ORGANISM	Severe acute respiratory syndrome coronavirus 2 Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.				
REFERENCE	1 (bases 1 to 29780)				
AUTHORS	Chironna,M., Sallustio,A., Loconsole,D., Accogli,M., Centrone,F., Capozzi,L., Del Sambro,L. and Parisi,A.				
TITLE	Genome Sequence of SARS-CoV-2 VUI 202012/01 strain identified from a patient coming back from London to Apulia region, Italy				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 29780)				
AUTHORS	Chironna,M., Sallustio,A., Loconsole,D., Accogli,M., Centrone,F., Capozzi,L., Del Sambro,L. and Parisi,A.				

A. Vad säger titeln på forskningsartikeln? **SVAR:** Inringat i rött är TITLE för den forskningsrapport som ligger bakom denna datarapportering. Det är alltså en studie som specifikt rör en patient som kommit från London till Italien.

B. Ger informationen svar på om patient X blivit smittad av sin kollega på Företaget som varit i Italien? **SVAR:** Eftersom vi fått en så nära träff med denna sekvens och då antar att den är mer lik än andra sekvenser (det står inget om det i uppgiften) så kan det tala för att det finns en koppling till Italien. Men eftersom den italienska patienten kommit från London så kan ju vår patient fått smittan via någon annan person som också varit i London... Det är inte helt lätt att dra slutsatser här. Och i nästa uppgift handlar det om att fundera vidare på just det - vad säger resultaten egentligen?

C. Hur ska man gå vidare om man vill smittspåra utifrån det vi vet nu? **SVAR:** Ett förslag är att be kollegan som varit i Italien att göra ett PCR-test med sekvensering av hela SARS-CoV-2 och jämföra eventuella resultat. Om kollegan inte verkar vara smittad ens så kan det spåret strykas. Klassisk smittspårning där personer som patienten varit i kontakt med kontaktas är viktigt att göra. Genom att PCR-testa även dessa och vid positivt utslag göra en helgenomsekvensering kan man eventuellt spåra den som smittat.



Uppgift 4 - Värdering av resultat

Diskutera resultatens användbarhet genom att ta ställning till påståendena nedan. Stämmer de, ja eller nej? Motivera!

- I. Stor likhet mellan en patients virus-sekvens och en träff i databasen pekar på att det är den som lämnat provet till databasen som smittat patienten.
- II. Stor likhet mellan patientens virus-sekvens och en träff i databasen pekar på att den patienten blivit smittad av kommer från samma geografiska område som den som lämnat prov till databasen.
- III. Flera virus-sekvenser från samma geografiska område som alla liknar patientens virus-sekvens stärker påstående II.
- IV. Om det var möjligt att få in data för viruset hos den person som de facto har smittat patienten skulle man få 100 procent överensstämmelse mellan virus-sekvenserna.

SVAR:

I. Här kan man motivera ett nej med att patienten inte har träffat precis denna italienska person vars virussekvens hamnat i databasen. Alla personer som har SARS-CoV-2 får inte sina virus sekvenserade i sin helhet så därför är det mer sannolikt att det är någon annan person än just denna som smittat. Man kan se på datum när datat skickats in också när man drar slutsatser om denna fråga.

II. Här kan man motivera ett nej med hänvisning till att vi idag har så global spridning av SARS-CoV-2-varianter att det är svårt att säkert koppla enskilda virustyper till ett visst geografiskt område. Geografiskt område säger en del, men inte allt då vi har global rörlighet idag.

III. Om vi skulle få ett sådant resultat - att flera mycket likartade virusekvenser från ett område tyder det på ett gemensamt ursprung för den gruppen virus. Och om det är många i samma geografiska område som bär på samma typ av virus så tyder det nog på smitta inom regionen, det skulle alltså kunna stärka påstående II.

IV. Här kan man argumentera för nej med hänvisning till ständiga mutationer och pågående evolution. Kanske kommer frågor om virusets mutationshastighet upp. Hos de RNA-virus som har de högsta mutationshastigheterna är sannolikheten för mutation 10^{-4} /bp/replikation (motsvarar 0.01% sannolikhet att en viss nukleotid mutaterar när viruset replikerar). Detta är så högt att det under en och samma infektion i en person hinner bildas en population av virus där det kan finnas en viss variation mellan enskilda virus.

Uppgift 5 - Mutationer och vacciner

- A. Alla mutationer påverkar inte hur bra vacciner fungerar. Vad tror du krävs för att en mutation hos viruset ska påverka effekten av vacciner? SVAR: Vacciner bygger på att någon/några delar av virusets strukturer tillsätts till kroppen för att få immunförsvaret att reagera på. För att vårt immunsvaret ska fungera mot ett virus sedan så måste viruset ha kvar någon/några av de delar (antigen) som vi använt vid vaccinationen. Så länge mutationer inte påverkar virusets byggbitar - t.ex. utseendet på spike-proteinet - borde vaccinen fungera. Det krävs mutationer som förändrar virusets utseende så pass mycket att de antikroppar som bildats av vaccineringen inte passar längre mot viruset.
- B. Vad kan öka sannolikheten för att en virusvariant som bär på vissa mutationer återfinns i databaserna? (och tvärtom: varför hittas vissa virusvarianter inte i databaserna?) SVAR: Om viruset ger så mycket (och kanske särskild allvarliga) symptom att en patient kommer till sjukvården ökar sannolikheten att man kommer sekvensera och skicka in data till databaserna. Virusvarianter som inte ger några symptom kanske aldrig upptäcks överhuvudtaget.



Uppgift 6 - Sammanfatta ett svar till läkaren

Vad kan du berätta för läkaren och patienten om vad du med bioinformatik kommit fram till angående det virus som smittat X?

Formulera ett kort e-postsvar och skicka till lä(k/r)aren! SVAR: Här ska svaren på frågorna sammanfattas:

1. **Har viruset några särskilda förändringar i genen för spikeproteinet i den del av viruset som fäster till människans celler via ACE2-receptorn?** SVAR: Ja, N501Y ligger nära det receptorbindande området. Den unika mutation som skiljer patientens virus från andra kända varianter är dock en tyst mutation som inte påverkar aminosyrasekvensen.
2. **Är viruset av någon av de kända varianterna som är i omlopp? ("Brittiska", "Norska" eller "Sydafrikanska").** SVAR: Den passar in på den brittiska varianten.
3. **Kan vi se om patienten har smittats av kollegan som nyss varit i Italien?** SVAR: Det är svårt att avgöra helt men det finns en viss koppling till Italienska fynd av mycket liknande virus. Det behöver göras fler tester, förslagsvis kontaktas kollegan som varit i Italien för att göra PCR-test och sekvensera vid positivt testresultat.

Uppgift 7 - Smittspårning och bioinformatik

Smittspårning går normalt ut på att en smittad person kontaktar alla dem som hen har varit i kontakt med. I artikeln intill menar man att smittspårningen av SARS-CoV-2 skulle bli mer effektiv om man använde bioinformatik.

- A. Varför gör man inte det redan? SVAR: Bioinformatik är inte en naturlig del av vården i alla lägen. Det är dyrare och tar längre tid att göra en sekvensering av hela virusets genom jämfört med enklare tester där man enbart vill påvisa om man har viruset eller inte.
- B. Vilka kostnader kan finnas med smittspårning med bioinformatik? SVAR: Dels kostnader för själva sekvenseringen. Men också för bioinformatikarbetet (lön till bioinformatiker). Det kommer in enorma mängder data som ska lagras vilket kräver servrar/lagringsutrymmen som också kostar både energi och arbetstimmar för de som serverar.
- C. Vad skulle motivera att samhället satsar pengar på detta? SVAR: Om en effektiv smittspårning på individnivå skulle göra att spridning av virus kan stoppas tidigt utan att samhället behöver vara stängt i så stor omfattning som det varit under pandemin. Kunna mota spridning av virus tidigare i förloppet. Om sjukdomen var väldigt dödlig också så skulle det nog motivera att alla metoder för att motverka det skulle prioriteras.

Uppgift 8 - Utvärdera övningen

- A. I övningen har du mött några olika databaser och bioinformatiska verktyg. Vad var lätt och vad var svårt? Berätta!
- B. Har arbetet väckt några frågor om biologi eller något annat? Berätta!

Vi vill tipsa om en sammanfattning av en didaktisk artikel om Bioinformatik-didaktik som vi tagit fram på Bioresurs. Där finns några allmänna tips som baseras på studier av elever som provat bioinformatikövningar. Det har bland annat visat sig att läraren är viktig för att stötta i att koppla görandet (i databaserna) med lärandet i/om det biologiska (vad är det vi tittar på nu, vad betyder det?). Tanken med några av frågorna i denna övning har varit att blanda uppgifter som fokuserar på att ta fram vissa uppgifter/fakta med frågor där man ska tänka till kring vad svaren säger oss (eller inte).



Gör ett eget case

När man förstått hur virusdatabasen fungerar så kan man konstruera egna "case" med tillhörande historier och problemställningar. För att ta fram en sekvens att utgå från (Patientens virus-sekvens) letar man först fram en riktig sekvens i databasen från något land som är intressant att använda för det "Case" man vill bygga. Sedan förändrar man sekvensen lite ("muterar") i en textredigerare (word) och låter den bli Patientens virus-sekvens.

1. Från virusdatabasens startsida, välj "Search by virus" (se bild ovan), blå ruta till höger.
2. Skriv in "SARS-CoV-2" i sökrutan, vänta och välj den formulering som dyker upp.
3. Träffbilderna i tabellen visar alla registrerade sekvenser. För att få fram exempelvis träffar från ett visst land, använd filtreringsverktygen som finns till vänster om tabellen. Klicka på "Geographic region" för att välja träffar efter region/land.
4. Klicka på ID-beteckningen (Accession) för en träff. En ruta med detaljer visas. Klicka igen på numret för träffen (accession number) och du kommer till själva sekvensdatat (se bild nedan).
5. Klicka på FASTA (se bild nere till höger) och kopiera den sekvens du får fram och klistra in i ett textdokument.
6. Redigera sekvensen (mutera) på valfritt sätt så att den skiljer sig lite åt från den du vill att den ska vara lik.
7. Formulera ditt "Case" med bakgrundsinformation och problem som ska lösas genom att utgå från den patientsekvens du tagit fram.

Samma metoder som beskrivits ovan fungerar för vilket virus som helst som finns i databasen. Det går att variera i det oändliga.

Accession	Release Date	Species	Nucleotide Details
LC435023	2019-05-24	Hepacivirus C	Hepatitis C virus subtype 6e N12
LC435024	2019-05-24	Hepacivirus C	Hepatitis C virus subtype 6e N12
LC435027	2019-05-24	Hepacivirus C	Hepatitis C virus subtype 6e N12
LC435028	2019-05-24	Hepacivirus C	Hepatitis C virus subtype 6e N12
EF424625	2007-04-26	Hepacivirus C	Hepatitis C virus subtype 6e N12

Skärmdump efter att du följt instruktionerna i punkt 4. Klicka på inringat ID-nummer för att komma vidare.

Hepatitis C virus subtype 6e

GenBank: LC435023.1

FASTA Graphics

Go to: ☺

LOCUS LC435023 9341 I
DEFINITION Hepatitis C virus subtype 6e I complete cds.
ACCESSION LC435023
VERSION LC435023.1
KEYWORDS .
SOURCE Hepatitis C virus subtype 6e
ORGANISM Hepatitis_C_virus_subtype_6e

Bild som visar en sökträff för sekvensdata för en hepatitis C av subtyp 6e. Ett klick på "FASTA" tar dig till nukleotidsekvensen som sedan kan kopieras.