



Folkhälsomyndigheten

Helgenomsekvensering av svenska SARS-CoV-2 som orsakar covid-19

Delrapport 2, 2020-07-07



Denna titel kan laddas ner från: www.folkhalsomyndigheten.se/publicerat-material/. En del av våra titlar går även att beställa som ett tryckt exemplar från Folkhälsomyndighetens publikationsservice, publikationsservice@folkhalsomyndigheten.se.

Citera gärna Folkhälsomyndighetens texter, men glöm inte att uppge källan. Bilder, fotografier och illustrationer är skyddade av upphovsrätten. Det innebär att du måste ha upphovsmannens tillstånd att använda dem.

© Folkhälsomyndigheten, 2020.

Artikelnummer: 20106

Om publikationen

En central del i Folkhälsomyndighetens strategi för covid-19 innefattar övervakning av pandemins utveckling i Sverige och internationellt för att anpassa åtgärder utifrån aktuellt kunskapsläge och sjuklighet. En del i övervakningen innefattar förståelse för virusets spridning nationellt och internationellt liksom förändringar av virusets arvsmassa som föranleder viktiga strategiska åtgärder i hanteringen av pandemin. Detta är delrapport två med analys av helgenomsekvenseringsdata från prov av virus från svenska fall av covid-19. Ytterligare delrapporter kommer att publiceras allt eftersom nya sekvenser och analyser tillkommer.

Folkhälsomyndigheten

Karin Tegmark Wisell

Avdelningschef, avdelningen för mikrobiologi

Innehåll

Om publikationen	3
Ordlista	5
Sammanfattning	7
Summary	8
Bakgrund	9
Syfte	11
Resultat	12
Översikt av genetisk grupp tillhörighet hos de svenska sekvenserna	12
SNP-analyser av grupperna B.1/G, B.1/GH samt B.1.1/GR	13
Diskussion	17
Pågående och framtida analyser	18
Metod	19
Provinsamling	19
Urval	19
Sekvensering	19
Bioinformatiska analyser	19
Sekvensanalys och variationsbestämning	19
Fylogenetiska analyser	20
Bilaga 1	21
Klassifikationssystem använda i rapporten	21
GISAID klassifikationssystem	21
Nextstrains klassifikationssystem	21
PANGOLINs klassifikationssystem	22
Källor	23

Ordlista

Covid-19	Eng. Coronavirus disease, coronavirussjukdom och året, 2019, då den diagnostiserades första gången.
Ct-värde	Antal temperaturcykler som krävs för att detektera en signal i kvantitativ PCR, polymerase chain reaction (se även qRT-PCR nedan)
Fylogenetik	Evolutionsbiologisk terminologi för jämförande studier av släktskap mellan organismer.
Genetisk grupp	De grupper som bildas av jämförande analyser mellan olika virus arvs massa.
GISAID	Eng. Global Initiative on Sharing All Influenza Data, en internationell databas med möjlighet att snabbt dela sekvenseringsdata mellan länder och forskargrupper.
Helgenomsekvensering	Karakterisering av en organisms hela arvs massa.
Nextstrain	Ett mjukvarusystem bestående av två delar: en analytisk del som jämför arvs massa från bakterier och virus, (Augur) och en del som visualiserar data (Auspice).
PANGOLIN	Eng. Phylogenetic Assignment of Named Global Outbreak LINEages är ett mjukvaruverktyg för att klassificera sekvenser från SARS-CoV-2 mot de internationellt tillgängliga sekvenserna och därmed bestämma släktskap.
Provtagningsindikation	Den kliniska och epidemiologiska bild som Folkhälsomyndigheten publicerar som stöd till sjukvården vid provtagning av misstänkta fall av covid-19. I en situation av resursbrist avseende provtagnings- och laboratoriekapacitet finns också en prioritetsordning.
qRT-PCR	Quantitative reverse-transcriptase polymerase chain reaction. En metod som kan detektera RNA-virus i ett prov genom att ett visst genfragment amplifieras i temperaturcykler och producerar en signal som mäts i Ct-värde.
SARS-CoV-2	Eng. Severe acute respiratory syndrome coronavirus 2, på svenska svår akut respiratorisk sjukdom

	coronavirus 2. Är det virus som ger upphov till sjukdomen covid-19.
Sentinelövervakning	Sentinelövervakning innebär att ett urval av läkare rapporterar eller provtar ett urval av patienter som söker till läkarmottagningar som deltar i Folkhälsomyndighetens övervakning för influensa med anledning av symtom på luftvägsinfektion. Genom sentinelövervakningen kan man fastställa hur många personer med influensasyntom som faktiskt har influensa eller covid-19.
SNP	Eng. Single Nucleotide Polymorphism är en positionsbestämd förändring i arvsmassan som berör en enskild nukleotid. Den kan, men behöver inte orsaka funktionella förändringar i aminosyrasekvensen.
SNP-analys	Processen att ta SNP-positioner i en arvs massa. Sker genom att man matchar sekvensdata mot en referens för att studera var enskilda förändringar skett.
SNP-profil	Den unika profilen med förändringar som uppstår när en arvs massa förändras och dessa förändringar sedan ärvs nedströms vid kopiering av arvs massan.
WHO	World Health Organisation

Sammanfattning

Helgenomsekvensering av SARS-CoV-2 har etablerats vid Folkhälsomyndigheten och utförs genom samarbete med landets kliniska mikrobiologiska laboratorier. Delrapport 2 beskriver, liksom den första rapporten, den genetiska karakteriseringen av SARS-CoV-2 med fokus på genetiska grupper som etablerats i Sverige. Denna delrapport inkluderar ett utökad urval av sekvenser under perioden januari fram till den 30 april för att representera utbrottets etablering i landet. Liksom den första rapporten innehåller den även data från sentinelövervakningen som nu innefattar prov till och med juni.

Analyserna i det utökade urvalet bekräftar de slutsatser som framkom i första rapporten att flera olika genetiska grupper av viruset har introducerats i Sverige. Flera fall av covid-19 har konstaterats efter resa till Italien, och två av de huvudsakliga genetiska grupperna som konstaterats i sentinelövervakningen tillhör denna grupp. Efter resa till Österrike såg man initialt ett stort antal fall av en tredje genetisk grupp. Vi ser att en av de tre genetiska grupperna minskar. Genetiska grupper associerade med Kina, Iran och Sydkorea är i minoritet i de hittills tillgängliga sekvensdata.

Genom SNP-analys kan skillnader påvisas från fall under perioden januari till april med smittland Italien och Österrike, jämfört med SARS-CoV-2 tillhörande samma genetiska grupper från fall under februari till juni med bekräftad smitta i Sverige. Då samtliga dessa genetiska grupper även cirkulerade i andra länder i Europa under perioden februari/mars är det troligt att åtskilliga introduktioner av samma genetiska grupper som de som cirkulerade i Italien och Österrike skedde från återvändande resande från länder där ingen misstanke om smittrisk förelåg.

Internationellt har över 60000 SARS-CoV-2 sekvenser deponerats i GISAID (1). Trots detta är sekvensdata över utbrottet fortfarande ofullständigt från många länder. Riktad provtagningsindikation för SARS-CoV-2 samt internationellt sett begränsad diagnostisk kapacitet för att identifiera tidiga introduktioner har medfört att underlag i form av tillgång till tidiga virusprov och därmed möjligheten till sekvenser från dessa är ofullständiga från tiden mellan förmodad tidpunkt för introduktion av smittan i olika länder och de tidigaste provtagna fallen i dessa länder. Baserat på de tillkommande analyser som genomförts inom ramen för delrapport 2 stärks bilden att importerade fall från länder som inte var inkluderade i den svenska provtagningsindikationen troligen bidragit till utbrottet i Sverige. Då sentinelövervakningen innefattar endast ett stickprov av landets fall med influensaliknande symtom varför dessa sekvenser inte innefattar samtliga genetiska grupper av virus. Sekvenser från aktuella cirkulerande genetiska grupperna kommer fortsatt att erhållas genom det etablerade nationella mikrobiella övervakningsprogrammet för SARS-CoV-2. Sekvensdata kommer löpande att genereras och publiceras i den internationella databasen GISAID, med påföljande analys av den fylogenetiska förändringen över tid i utbrottet.

Summary

Whole-genome sequencing of SARS-CoV-2 has been established at the Public Health Agency of Sweden (PHAS) as a tool for following the phylogenetic changes of the viral populations circulating within Sweden. This is the second report in a series covering the genetic characterisation of the SARS-CoV-2 outbreak within Sweden. Our report focuses on the continuing spread of the virus within Sweden. Additionally, the report adds complementary data from the period of mid-March to the end of April. Samples included are selected based on a representation of the regions within Sweden (for samples belonging to the initial outbreak). Additionally, samples with a SARS-CoV-2 positive result from the sentinel program up until the 11:th of June are also included.

We identify three major genetic groups introduced into Sweden during the initial outbreak, confirming the conclusions from the first report. All three groups were common within Europe at the time of introduction, based on available sequencing data. All of these genetic groups have continued to spread within Sweden. However, a trend is that one of the three groups is decreasing over time.

An in-depth analysis of these three groups shows minor changes within the sampled population with Sweden as a country of origin for the infection, compared to import cases from Italy and Austria. This indicates that the primary spread of the virus might not have been driven by the travellers from countries identified by World Health Organisation (WHO) with community spread in the early part of the outbreak, namely China, South Korea, Iran, Italy, and Austria.

The Public Health Agency of Sweden continues to monitor the SARS-CoV-2 situation by whole-genome characterisation. It will continue to provide data for the international community to track the spread and evolution of the virus within Sweden. Finally, a separate national surveillance program has been established for SARS-CoV-2, aiming at continuously providing sequence data as support for decision making.

Bakgrund

Covid-19 pandemin, som orsakas av viruset SARS-CoV-2, har i dagsläget över 71000 diagnostiserade sjukdomsfall i Sverige och över 11 miljoner smittade i världen. Folkhälsomyndigheten beslutade tidigt att utveckla en sekvenseringsbaserad metodik för att följa virusets utveckling i Sverige.

I en första delrapport beskrev Folkhälsomyndigheten vilka genetiska grupper av SARS-CoV-2 som etablerade sig i Sverige (artikel 20089 Helgenomsekvensering svenska isolat av SARS-CoV-2 som orsakar covid-19). Denna rapport visade att tre huvudsakliga genetiska grupper cirkulerade i Sverige efter etableringen fram till den 15 mars då samhällssmitta konstaterades. Dessa grupper konstaterades också vara väl spridda i Europa och i viss mån i USA under motsvarande period.

Analysen i denna delrapport är baserad på helgenomssekvensering av SARS-CoV-2, både från kliniska prover skickade till Folkhälsomyndigheten fram till sista april samt från sentinelövervakningen av influensa och covid-19. Sentinelproverna ger en bra bild av spridningen i Sverige under perioden mars till juni då provtagningen inte baserades på resehistorik. Samtliga erhållna sekvenser har jämförts med internationellt publicerade sekvenser genom databasen GISAID (1). En del av dataanalysen är att identifiera nära släktskap mellan virus påvisade i Sverige och internationellt, vilket ger en indikation på att en i tiden nära gemensam smittkälla finns. För närvarande används i huvudsak tre sätt att klassificera släktskap mellan olika genetiska varianter av viruset: GISAID, Nextstrain och PANGOLIN (2, 3). Metodiken finns närmare beskriven i bilaga 1. Slutsatser i delrapport 2 är baserade på en kombination av GISAID och PANGOLINs system, en metodik fastställd i delrapport 1.

Vi redovisar även högupplösta analyser av variationer i virusets arvs massa på enskild nukleotidnivå - så kallade Single-Nucleotide Polymorphisms (SNPs). SNP-data möjliggör urskiljning av små genetiska skillnader även inom genetiska grupper. Metoden används i denna rapport främst för att jämföra Sveriges sekvenser mot varandra, för att avgöra om en skillnad i fortsatt cirkulerande genetiska grupper, och introducerade sådana, kan ses. Jämförande analys görs även om dessa kommer från lokala eller från den stora nationella introduktionen, eller om det är nya introduktioner.

Denna rapport inkluderar ett urval av prover med påvisad SARS-CoV-2 som inkommit till Folkhälsomyndigheten från landets kliniska mikrobiologiska laboratorier fram till den 30 april, där urvalet skett för att inkludera prover från så många regioner som möjligt (ingår delvis även i delrapport 1). Prover från sentinelövervakningen under en period som sträcker sig från mars fram till och med den 11:e juni, där SARS-CoV-2 påvisats, har också inkluderats. Totalt sett ingår 521 SARS-CoV-2-sekvenser från svenska fall med covid-19. Dessa sekvenser representerar endast en andel av alla fall av covid-10 som diagnostiserats i landet. Antalet diagnostiserade fall utgör vidare endast en andel av landets

faktiska fall. De analyserade sekvenserna utgör därmed endast ett stickprov men att urvalet har gjorts för att erhålla en så hög grad av geografisk representativitet som möjligt. Rapporten syftar till att karakterisera de olika genetiska gruppernas spridning i Sverige, vilket ger en överblick över det initiala utbrottet samt följer utbrottet över tid. De dominerande genetiska gruppernas arvs massa har dessutom undersökts på SNP-nivå.

Syfte

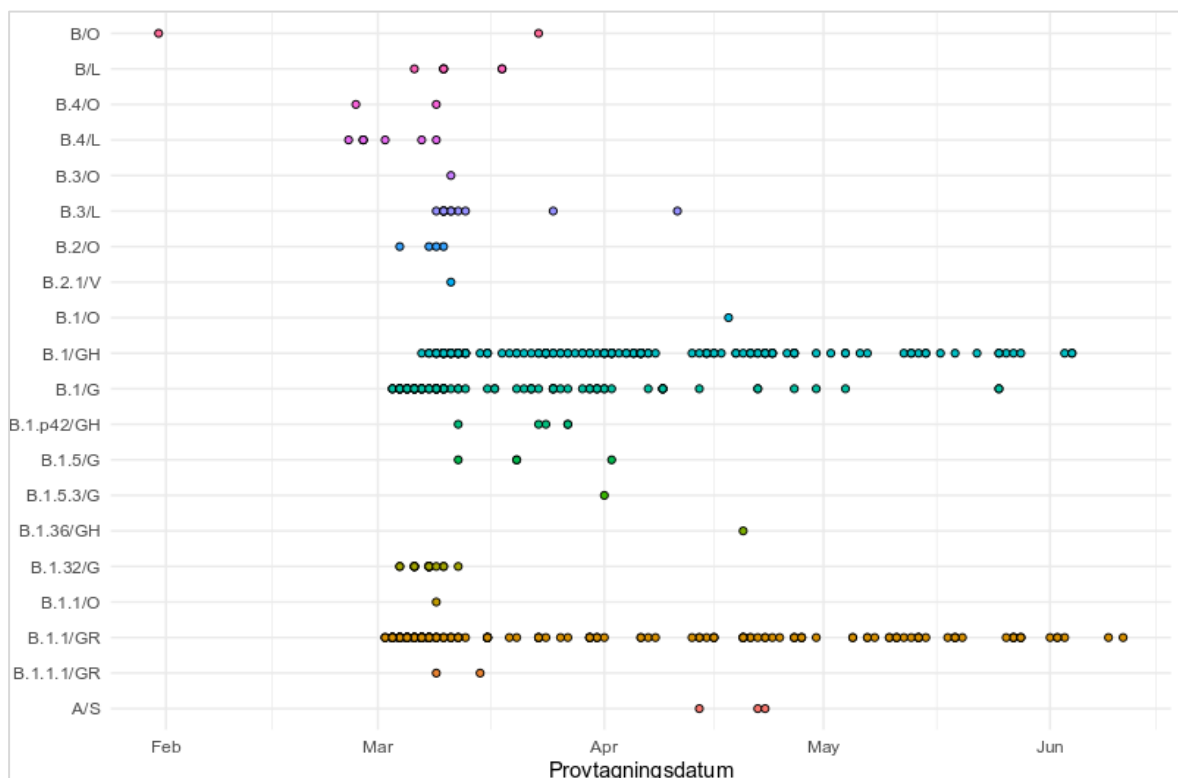
Genom att analysera prover från svenska fall där SARS-CoV-2 påvisats, möjliggörs förståelse för hur covid-19 introducerades i Sverige liksom hur smittspridningen i Sverige sett ut över tid. Analyser kan också påvisa genetiska förändringar som kan vara av relevans för att anpassa de diagnostiska verktyg som finns etablerade. Slutligen kan dessa analyser potentiellt upptäcka förändringar i viruset som är relevanta för immunitet vid en eventuell förnyad exponering, liksom känsligheten för olika läkemedel och på sikt även effekten av ett vaccin.

Resultat

Översikt av genetisk grupptillhörighet hos de svenska sekvenserna

De 521 prover som hittills har sekvenserats vid Folkhälsomyndigheten tillhör huvudsakligen tre genetiska grupper som fastställdes i delrapport 1 baserat på den kombinerade analysen av olika typningssystem: B.1/G, B.1/GH och B.1.1/GR. En sammanfattning av hur sekvenserna fördelas över de genetiska grupperna visas i figur 1.

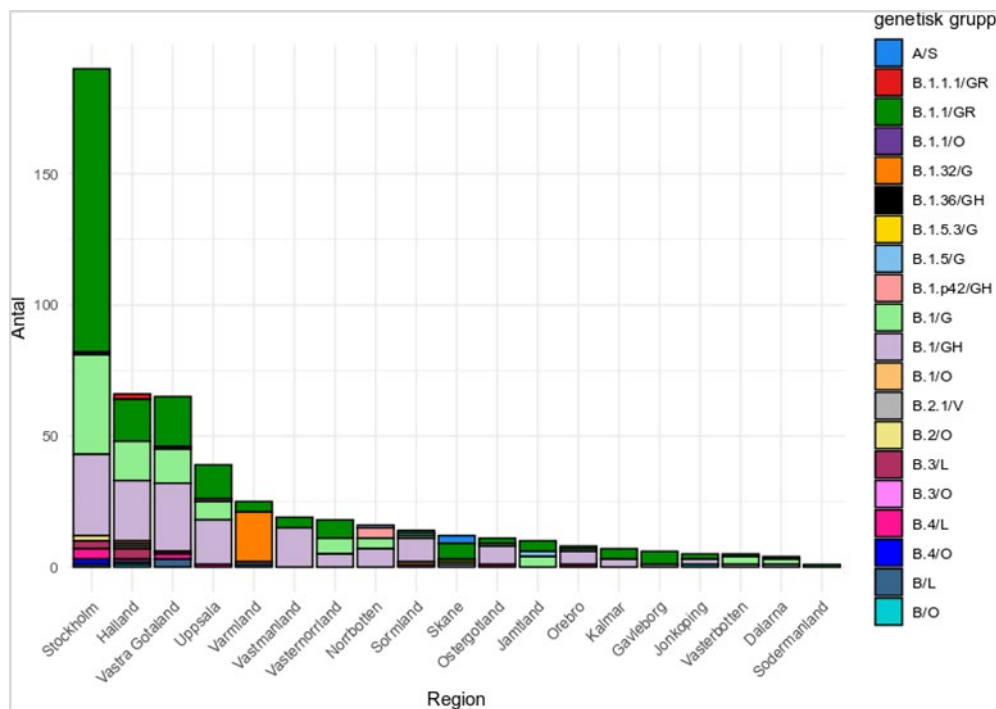
Figur 1. Fördelning av de svenska sekvenserna över tid, enligt den kombinerade klassifikationen av genetiska grupper. Y-axeln visar genetisk grupp och X-axeln visar provtagningsdatum.



Figur 1 visar att ett flertal genetiska grupper introducerades till Sverige under januari till maj, av dessa är det dock bara ett fåtal som uppnått någon högre frekvens i spridningen över tid. B.1.32/G kvarstår som en unik svensk grupp i den tillgängliga internationella data som finns i GISAID. B.1.32/G härstammar från importsmitta från Italien, vilket har stöd i smittspårningen under provtagningen och är genetisk mycket lik gruppen B.1/G. De grupper som tillkommit sedan första rapporten är B.2/O och B.4/L men nya fall som tillhör dessa grupper har inte upptäckts efter mars. Grupper med härstamning från grupp A samt tidiga B-grupper (B/O samt A/S och B/L) är ovanliga i dessa prover, och i dagsläget syns ingen tydlig indikation på cirkulation av dessa grupper i Sverige. Sentinelproverna stödjer en fortsatt spridning av de genetiska grupperna B.1/GH samt B.1.1/GR,

med enbart enstaka fall av B.1/G och B.3/L i underlaget mellan maj och juni månad.

Figur 2. Fördelningen av sekvenserade prover över regionerna, färgkodning efter genetisk grupp.



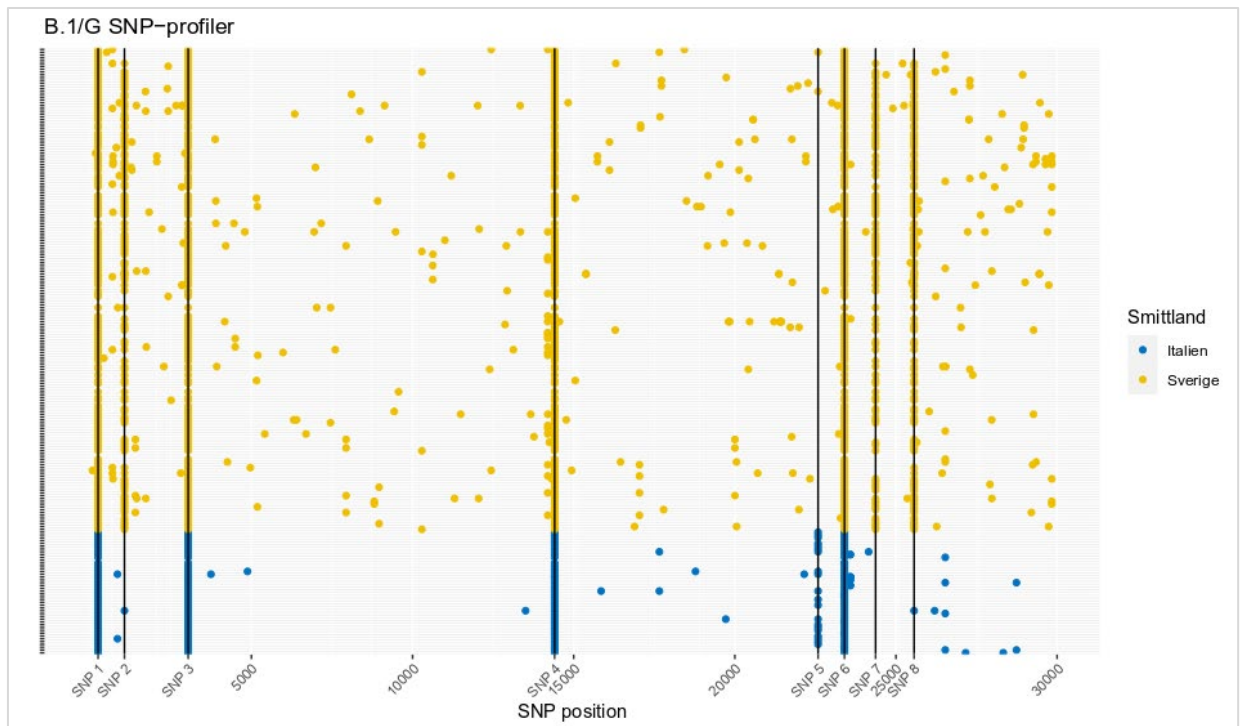
I figur 2 redovisas fördelningen av provmängd över regioner. Trots ett riktat urval i de prov som sekvenserats för att bättre representera samtliga regioner kvarstår det faktum att provtagningen på återvändande semesterresenärer under slutet av februari och början av mars påverkar fördelningen och vissa regioner dominerar därmed provmängden. Framtida analyser, med ytterligare data från det inrättade övervakningsprogrammet, kommer förhoppningsvis ge en bättre geografisk fördelningen av prover. Det är av stor vikt att kartläggning av de genetiska grupperna i landet fortgår när nationellt och globalt resande återupptas, för att säkerställa att spridningsmönster snabbt kan kartläggas. Fördelningen av de genetiska grupperna ser snarlik ut som i den första rapporten, vilket indikerar att de prover som nu lagts till analysen inte nämnvärt påverkat fördelningen av genetiska grupper av SARS-CoV-2 som cirkulerar i Sverige.

SNP-analyser av grupperna B.1/G, B.1/GH samt B.1.1/GR

I denna delrapport gör vi en utökad analys av mutationer som särskiljer sekvenser med smittland Sverige från de med smittland Italien eller Österrike, genom SNP-analyser av de tre dominerande genetiska grupperna. Genom att studera variationer mot virusets referenssekvens, kan vi följa virusets utveckling över tid. En majoritet av dessa förändringar i virusets arvs massa är oviktiga för virusets funktion, men ger ändå en möjlighet att med hög precision följa virusets utveckling vilket möjliggör studier av möjliga smittvägar. I figur 3, 4, 5 visas de genetiska profilerna

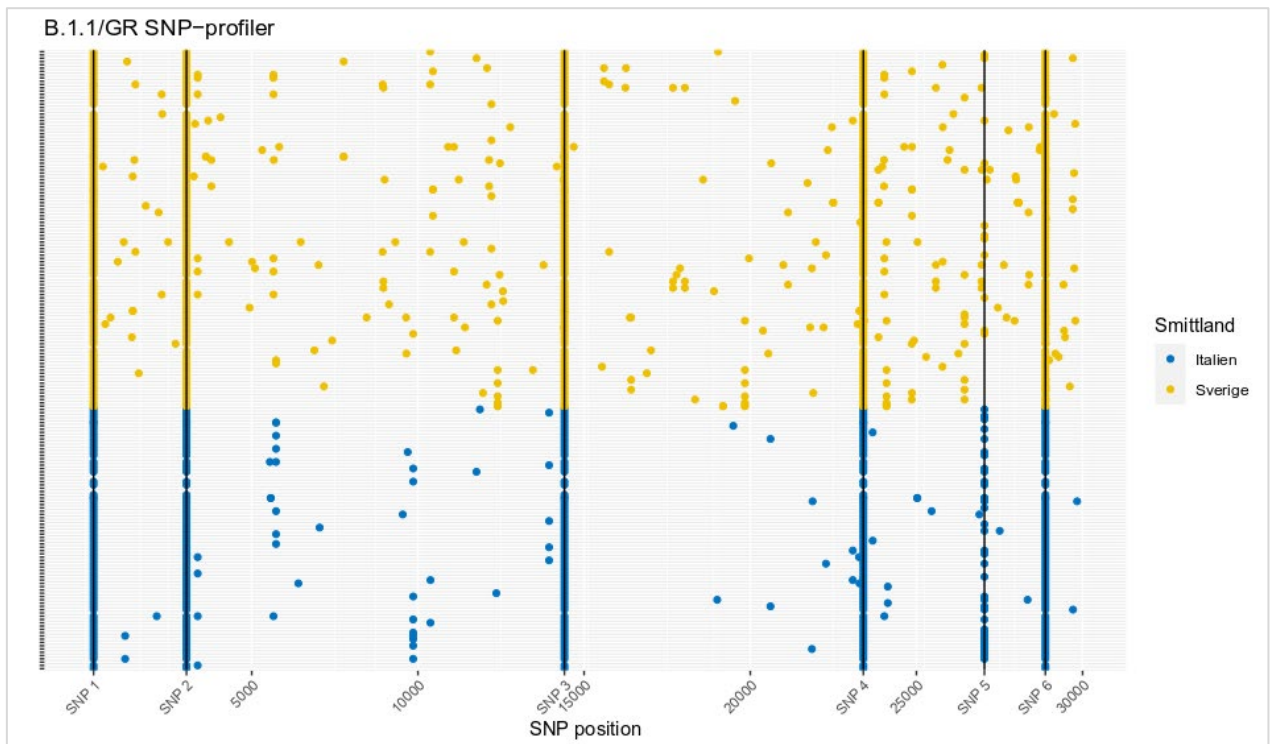
för de tre dominerande genetiska grupperna, fördelat på fall med bekräftat smittland Sverige eller Österrike/Italien, de länder som finns väl representerade i den initiala provtagningen.

Figur 3. SNP-profiler för den genetiska gruppen B.1/G, med fokus på Italien/Sverige. Notera positionerna 241, 1059, 3037, 14408, 22583, 23403, 24368, 25563 namngivna SNP 1-8 nedan.



SNP-profilerna inom den genetiska gruppen B.1/G uppvisar en klar skillnad mellan prov från fall med bekräftat smittland Sverige jämfört med fall från bekräftat smittland Italien. Sekvenserna uppvisar åtta SNP i jämförelse med referenssekvensen, fyra konserverade och fyra som skiljer sig mellan smittland Italien och Sverige. De huvudsakliga förändringar i virusets arvsmassa som skiljer sekvenserna från fall med smittland Italien jämfört med smittland Sverige finns i positionerna 1059, 23583, 24368 och 25563 (SNP 2, 5, 7 och 8). Position 24368 (SNP 7) saknas i samtliga prover med bekräftat smittland Italien. Även 1059 (SNP 2) och 25563 (SNP 8) sticker ut och har enbart ett prov som är bekräftat smittland Italien. Position 22583 (SNP 5) har enbart två prover med bekräftat smittland Sverige men är mycket vanlig i bekräftat smittland Italien. De svenska profilerna uppvisar även en större variation, vilket tyder på en bredare genetisk mångfald. Det är osannolikt att så många separata händelser av utveckling av virusets arvsmassa skulle ske inom en så kort tidsram, vilket tyder på att de prover som bidrar till spridningen i Sverige av B.1/G troligen inte kan förklaras med importsmitta från Italien. Framtida analyser kommer fokusera på att spåra de möjliga källorna till de olika SNP-profilerna inom gruppen B.1/G som antas vara möjligt när ytterligare nationella och internationella sekvenser finns tillgängliga för jämförelse.

Figur 4. Analys av SNP-profiler för den genetiska gruppen B.1.1/GR, med fokus på Italien/Sverige.

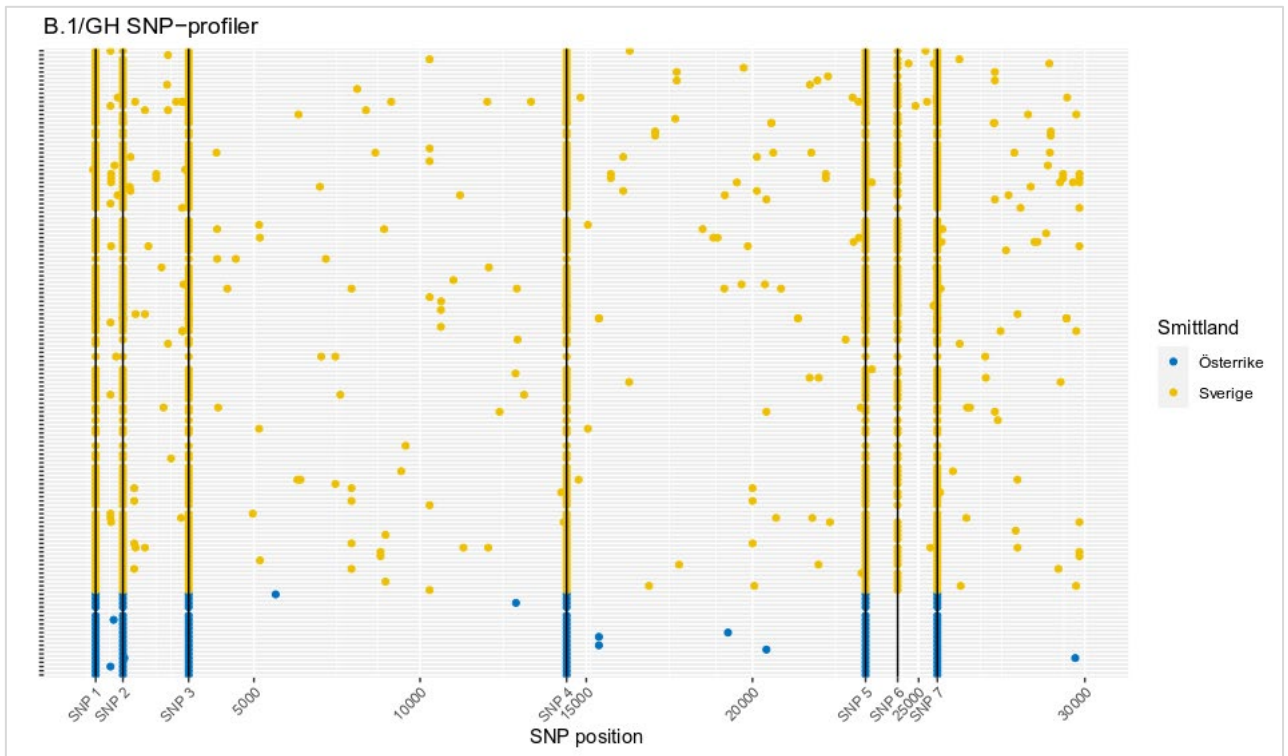


Liknande profilerna i Figur 3, varierar de prover med bekräftat smittland Sverige betydligt i jämförelse med prover från bekräftad smitta i Italien. Notera 241, 3037, 14408, 23403, 27046, 28881 namngivna SNP 1-6 nedan. Position 27046 (SNP 5) är vanligare bland proverna från Italien men saknas ofta bland de Svenska proverna.

I gruppen B.1.1/GR, som är senare i virusets utveckling än B.1/G, ser vi färre skillnader i prov med bekräftat smittland Italien jämfört Sverige. Enbart position 27046 (SNP 5) utskiljer sig konsekvent mellan smittland Sverige jämfört med Italien i alla sekvenser som ingår i jämförelsen.

Av de två genetiska grupper som primärt importerades med resande från Italien, finns utifrån denna analys begränsat stöd att just de varianter som konstateras i proverna med bekräftat smittland Italien var de som etablerats i Sverige. I gruppen B.1/G påvisas fyra SNP-skillnader från prov med bekräftad smitta i Italien, jämfört med de som smittats i Sverige. Ett antal separata förändringar av virusets genom gör det osannolikt att den genetiska gruppen utvecklades från den virusgrupp som konstaterades hos återvändande resenärer från Italien. Några omfattande slutsatser kring B.1.1/GR går inte att göra baserat på denna analys. Det finns ett antal skillnader, men mönstret är inte lika tydligt som i fallet med B.1/G. Variationen i B.1/G och B.1.1/GR är högre i prover med bekräftat smittland Sverige, något som inte speglas av de prover som kommer från bekräftad smitta i Italien.

Figur 5. Analys av SNP-profiler för den genetiska gruppen B.1/GH, med fokus på Österrike/Sverige. Notera positionerna 241, 1059, 3037, 14408, 23403, 24368, 25563 nedan namngivna SNP 1-7. Notera speciellt SNP 6 i de prover med bekräftat smittland Sverige som inte finns hos de med bekräftat smittland Österrike.



I fallet med B.1/GH är den konserverade händelsen i position 24368 (SNP 6) helt frånvarande i de prover som kommer från bekräftad smitta i Österrike. I en studie har Bluhm *et. al.* konstaterat att mutationen i position 24368, G24368T, är överrepresenterad i de svenska proverna, och dess frånvaro från prover med ursprung i Österrike, skedde importen troligen inte från Österrike (4). I studien av Bluhm *et. al.* konstaterades även att en majoritet av de sekvenser som bär den varianten kommer från England. En förklaring till skillnaderna i sekvensdata mellan de varianter som kommer direkt från Österrike och de som cirkulerar i Sverige, är att viruset spred sig tidigare från Österrike, till länder i Europa, och därifrån vidare till Sverige. Det är därför troligt att den stora spridningen av B.1/GH i Sverige inte kommer från resenärer från Österrike under perioden slutet av februari och början av mars.

SNP-analys är ett kraftfullt verktyg för att öka upplösningen i lokala utbrott, men begränsas av tillgången på internationellt sekvenser att jämföra mot. Analysen har även en förmåga att snabbt bli svårtolkad vid större grupper av prover. I denna studie har vi enbart fokuserat på ett urval av de SNP-profiler som vi haft tillgång till, där en specifik frågeställning fanns. Framtida rapporter kommer utveckla detta och ge en djupare överblick av SNP-varianter i de svenska proverna, och även studera de möjliga funktionella förändringar som kan utvecklas av underliggande SNP förändringar

Diskussion

I denna delrapport ser vi att de genetiska grupper som etablerades tidigt i landet har fortsatt att cirkulera men att varianter inom grupperna går att urskilja utifrån SNP-analys och att spridningsmönstret för dessa olika varianter skiljer sig åt inom de genetiska grupperna. Genom sentinelövervakningen kan vissa slutsatser om spridningens fortsatta utveckling antas, där en av dessa tre genetiska grupper till synes minskar över tid.

En djupare SNP-analys, Figur 3-5, visar att det är möjligt att inom de i Sverige dominerande genetiska grupperna urskilja ytterligare genetiska varianter. Baserat på data med smittland är det då även möjligt att bestämma om direkta importfall har samma sekvensvariant som de prover som är tagna i ett senare skede eller med bekräftad smitta i Sverige. Det är osannolikt att sådana förändringar är en effekt enbart av spridning inom landets gränser, och indikerar snarare att spridningen startades av flertalet större introduktioner inom en och samma genetiska grupp. Följsamheten mot de åtgärder och den kommunikation om smittläget som utgick från myndigheterna vid smittans utbrott gav resultat, då flertalet av introduktionerna stannat vid sporadiska fall. Slutsatserna i denna rapport tyder på att fall av covid-19 tidigt introducerades till Sverige från flera olika länder, inkluderande länder där landet själva och/eller WHO vid den tiden inte konstaterat samhällssmitta. I huvudsak tre genetiska grupper med variationer inom grupperna som skiljer dem från länder i den då inkluderade i provtagningsindikationen kan konstateras. Liknande analyser är även genomförda i främst Storbritannien, där ett stort antal introduktioner sågs i början av utbrottet för att sedan snabbt klinga av och ge plats åt ett fåtal dominerande genetiska grupper (5).

I delrapport 1 konstaterades att tre genetiska grupper (B.1/G, B.1/GH samt B.1.1/GR) dominerade, i denna delrapport visar vi att dessa har fortsatt att cirkulera. Det är främst två av de tre mest förekommande genetiska grupperna som kvarstår över tid (B.1/GH samt B.1.1/GR) medan B.1/G minskar, trots det stora antalet initiala introduktioner. Data från denna rapport indikerar att ett antal introduktioner av olika genetiska varianter skett till Sverige. Dessa introduktioner tillhör genetiskt närbesläktade virus men har särskiljande variationer i sin arvs massa som utesluter att de genetiska grupperna enbart introducerades genom resande till Österrike och Italien. Nya introduktioner har helt stannat av, mest troligt på grund av de åtgärder som genomförts på samhällsnivå för att minska smittspridningen, samt de internationella reserestriktionerna. Fortsatt provtagning och sekvensering inom det nationella övervakningsprogrammet såväl som sentinelövervakningen är nödvändig för att säkerställa en fortsatt hög beredskap inför ny import av annan genetisk grupp, såväl som för att kartlägga utvecklingen nationellt.

Pågående och framtida analyser

Folkhälsomyndigheten avser att fortsätta sitt arbete med genetisk karakterisering av SARS-CoV-2 inom landets gränser, samt att nogsamt karakterisera de möjliga importfall som kan bli effekten av ökat resande inom Europa och globalt. Utökade analyser av fylogenetiskt släktskap samt analyser av enskilda skillnader i arvsmassan hos de dominerande grupperna kommer även utföras för att säkerställa, i den mån det är möjligt, virusets ursprung i Sverige.

Metod

Provinsamling

Rapporten inkluderar prover från två separata insamlingar; insamling till följd av provtagningsindikation och insamling till följd av sentinelövervakningen. Prover till följd av provtagningsindikation samlades in hos Folkhälsomyndigheten mellan januari och slutet av april månad, under smittspridningens initiala fas.

Sentinelprover är insamlade från och med vecka nio och därefter löpande över tid.

Urval

Av sentinelproverna är samtliga prover med påvisad SARS-CoV-2, antingen med qRT-PCR för E-genen eller RdRp-genen, och med ett Ct-värde under 35 inkluderade. Av prover som samlats in till följd av provtagningsindikationen som då var riktad för resenärer och deras kontakter från Kina, Sydkorea, Iran, Italien och Österrike är merparten av proverna mellan januari och 15 mars är inkluderade. Förutom detta, är ett urval gjort på senare prover (mellan 15 mars och sista april) för att täcka så många geografiska regioner som möjligt. Även dessa prover har haft ett krav på Ct-värde under 35. Totalt utgör dessa urval 521 prover.

Sekvensering

Ion Torrent targeted Next Generation Sequencing (NGS), även kallat AmpliSeq, användes för helgenomsekvensering av SARS-CoV-2 RNA som isolerats från prover med påvisad covid-19. Metoden bygger på en uppsättning av specifika primrar, ”Ion AmpliSeq SARS-CoV-2 Research Panel” (Thermo Fisher Scientific), som i kombination med AmpliSeq-reagenser konverterar viralt RNA till DNA och amplifierar materialet för vidare sekvensering på ett GeneStudio S5-instrument (Thermo Fisher Scientific). Rådata efter sekvensering består av amplifierade regioner av viruset, så kallade amplicon, som processas och analyseras för att sammanfogas till ett komplett viralt genom.

Bioinformatiska analyser

Sekvensanalys och variationsbestämning

Sekvensanalys genomfördes med hjälp av Torrent Mapping Alignment Program (TMAP) och plug-ins i Torrent Suite (Torrent Variant Caller och Coverage Analysis) mot referensen Wuhan-Hu-1 NCBI Reference Genome (Accession number: MN908947.3,)(6). Endast sekvenser med minst 20x täckning på samtliga amplicon godkändes. Av dessa producerades så kallade konsensussekvenser inkluderande de förändringar som gavs stöd i sekvensanalysen. Varianter i form av enstaka nukleotidförändringar godkändes utan manuell inspektion medan varianter i form av infogningar eller borttagningar inspekterades manuellt, för att säkerställa att inga sekvenser med tekniska fel godkändes.

Fylogenetiska analyser

Fylogener genomfördes i enlighet med samtliga existerande metodiker: GISAID, Nextstrain och PANGOLIN. Klassifikation enligt GISAID genomfördes efter uppladdning till GISAID, likaså klassifikation enligt PANGOLIN (1, 3). Nexstrain-klassifikation skedde med hjälp av det integrerade verktyget Augur (2). Visualisering av data skedde i R, med hjälp av paketen ape, ggplot2, ggtree, dplyr, ggrepel, reshape2, tidyr, och treeio (7, 8, 9, 10).

Bilaga 1

Klassifikationssystem använda i rapporten

GISAID fokuserar på aminosyra förändringar i viruset, och har ett schema för klassifikation där konserverade mutationer bildar genetiska grupper. Nextstrain har ett schema för klassifikation som bygger på verktyget Augur, i korthet; sekvenser klassificeras löpande och namnges efter år (19, 20) och subgrupp (A, B, C). Detta sker när tillräckligt hög diversitet uppstått i en tillräckligt hög frekvens av genomen. PANGOLIN baserar sin analys på en dynamisk klassifikation som delar in sekvenserna i grupper baserat på ett antal formella kriterier, dvs att den har en eller fler förändringar från referensen, att det finns fem sekvenser med samma förändring och att de har starkt stöd i dataanalysen. I korthet ger GISAID och Nextstrain en grövre uppskattning om sekvensens släktskap och PANGOLIN ger en mer precis uppskattning. PANGOLIN själv har idag över 90 grupper, medan Nextstrain har fem och GISAID har sex.

GISAID klassifikationssystem

GISAID har ett system som bygger på identifikation av konserverade förändringar i genomet över tid och indelning i genetiska typer utifrån dessa förändringar. Förändringarna kan vara av både kodande typ, dvs att de förändrar aminosyran i det proteinet, eller av icke kodande typ, då man enbart ser en förändring på nukleotid nivå. I tabell 1 listas de nuvarande förändringar som leder till klassifikation i enlighet med GISAIIDS system.

Bifogad tabell 1. Klassifikationsschema för GISAID

Genetisk-typ	Genetisk variant	Proteinvariant
S	C8782T, T28144C	NS8-L84
L	C241, C3037, A23403, C8782, G11083, G25563, G26144, T28144, G28882	Referens
V	G11083T, G26144T	NSP6-L37F + NS3-G251V
G	C241T, C3037T, A23403G	S-D614G
GH	C241T, C3037T, A23403G, G26563T	S-D614G + NS3-Q57H
GR	C241T, C3037T, A23403G, G28882A	S-D614G + N-G204R

Nextstrains klassifikationssystem

I korthet, genetiska grupper av viruset kommer namnges löpande, när sådana grupper uppnår tillräckligt hög spridning med tillräckligt tydlig signal (i form av genetisk förändring). Grupper delas även in i huvudgrupper och undergrupper, där

en huvudgrupp är definierad som en grupp som upprätthåller 20 % av den globala spridningen. Det krävs även att den skiljer sig två mutationer från sin källa. På detta sätt definieras de första två huvudgrupperna som 19A och 19B, för året de uppstod och för att de skiljer sig med två mutationer. Inom dessa huvudgrupper kan även mindre grupperingar uppstå, som följs över tid och om de definieras som en huvudgrupp kommer de döpas om och utgöra en ny del av klassifikationen

PANGOLINs klassifikationssystem

- a. Fylogenetiskt bevis för att härstamma ur en tidigare grupp genom att uppfylla alla fyra kriterier 1) uppvisar en eller fler nukleinsyra skillnader från härstamningen 2) Innehåller åtminstone fem genom 3) inom dessa fem 22 delar de åtminstone en skillnad från härstamningen 4) ett bootstrap värde på minst 70% vid skiljepunkten för sin gren mot härstamningen.
- b. Släkten som definieras i a kan i sin tur vara ursprung till nya släkten (A kan bli A.1 och sedan A.1.1). c) Denna process kan bara upprepas för tre nivåer (A.1.1.1), sedan uppstår ett nytt huvudsläkte (Så, A.1.1.1.1 blir alltså B).
- c. Alla sekvenser tilldelas ett släkte, så om de inte uppfyller kriteriet för en lägre nivå, tilldelas de en högre.

Källor

1. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* [Internet]. 2017 Mar 30;22(13). Available from: <http://dx.doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
2. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018 Dec 1;34(23):4121–3.
3. Rambaut A, Holmes EC, Hill V, O'Toole Á, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology [Internet]. *bioRxiv*. 2020 [cited 2020 Jun 4]. p. 2020.04.17.046086. Available from: <https://www.biorxiv.org/content/10.1101/2020.04.17.046086v1>
4. Bluhm A, Christandl M, Gesmundo F, Klausen FR, Mančinska L, Steffan V, et al. SARS-CoV-2 Transmission Chains from Genetic Data: A Danish Case Study [Internet]. *bioRxiv*. 2020 [cited 2020 Jul 2]. p. 2020.05.29.123612. Available from: <https://www.biorxiv.org/content/10.1101/2020.05.29.123612v1>
5. [Pybus O, Rambaut A, L. du P, Zarebski AE, M. K, J. R, et al. Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages \[Internet\]. <https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507>. University of Oxford, University of Edinburgh; 2020 \[cited 2020 Jun 11\]. Available from: <https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507>](https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507)
6. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*. 2020 Mar 11;27(3):325–8.
7. [R Core Team. R: A Language and Environment for Statistical Computing \[Internet\]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: <https://www.R-project.org/>](https://www.R-project.org/)
8. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse [Internet]. Vol. 4, *Journal of Open Source Software*. 2019. p. 1686. Available from: <http://dx.doi.org/10.21105/joss.01686>
9. Yu G, Smith D, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data [Internet]. Vol. 8, *Methods in Ecology and Evolution*. 2017. p. 28–36. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12628/abstract>
10. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, et al. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data [Internet]. Vol. 37, *Molecular Biology and Evolution*. 2020. p. 599–603. Available from: <http://dx.doi.org/10.1093/molbev/msz240>

Denna rapport redogör för Folkhälsomyndighetens fortlöpande arbete att helgenomkaraktisera de olika genetiska grupperna av SARS-CoV-2 som cirkulerar i samhället. Rapporten syftar dessutom till att förstärka kunskapsläget kring viruset och dess möjliga utveckling, samt att bidra med möjligheten att upptäcka nya introduktioner av viruset.

En fördjupad analys av SARS-CoV-2 har gjorts i det pågående utbrottet i Sverige. Resultaten stärker slutsatserna från delrapport 1, det vill säga att tre genetiska grupper etablerade sig i Sverige vid utbrottets start, varav främst två fortsätter att spridas. Dessa genetiska grupper har skillnader jämfört med de varianter av genetiska grupper som importerades med resande från Italien och Österrike, vilket stärker hypotesen att ett flertal separata introduktioner skedde in i landet. Slutligen kan man fortsatt se spridning av främst två av de genetiska grupperna baserat på resultat från sekvensering av prover från sentinelövervakningen.

Folkhälsomyndigheten är en nationell kunskapsmyndighet som arbetar för en bättre folkhälsa. Det gör myndigheten genom att utveckla och stödja samhällets arbete med att främja hälsa, förebygga ohälsa och skydda mot hälsohot. Vår vision är en folkhälsa som stärker samhällets utveckling.