



Folkhälsomyndigheten

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

Version 1, 2020-06-11



Denna titel kan laddas ner från: [www.folkhalsomyndigheten.se/publicerat-material/](http://www.folkhalsomyndigheten.se/publicerat-material/). En del av våra titlar går även att beställa som ett tryckt exemplar från Folkhälsomyndighetens publikationsservice, [publikationsservice@folkhalsomyndigheten.se](mailto: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: 20089

## 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 viruset som föranleder viktiga strategiska åtgärder i hanteringen av pandemin.

Denna rapport är den första versionen av analys av helgenomsekvenseringsdata från svenska fall av covid-19. Rapporten kommer löpande uppdateras allt eftersom nya sekvenser och analyser tillkommer.

Folkhälsomyndigheten

Karin Tegmark Wisell

Avdelningschef, avdelningen för mikrobiologi

# Innehåll

Om publikationen .....	3
Ordlista .....	5
Sammanfattning .....	7
Avgränsning .....	9
Syftet med helgenomsekvenseringen .....	10
Bakgrund.....	11
Resultat.....	12
Översikt av genetisk grupptillhörighet hos de svenska sekvenserna .....	12
Kombinerad analys av genetiska grupper .....	13
Analys av resehistorik i relation till genetisk grupp .....	14
Fördelning av genetiska grupper internationellt .....	15
Sammanfattande diskussion .....	19
Pågående och framtida analyser .....	20
Bilaga 1. ....	21
Nextstrains klassifikationssystem.....	21
PANGOLINs klassifikationssystem .....	21
Metodsammanfattning .....	26
Källor .....	28
Författarnas tack till GISAID (Engelska) .....	29

# Ordlista

Helgenomsekvensering	Karakteriseringen av hela arvsmassan för en organism.
PANGOLIN	Phylogenetic Assignment of Named Global Outbreak LINEages är ett mjukvaruverktyg för att klassifiera sekvenser från SARS-CoV-2 mot de internationellt tillgängliga sekvenserna och därmed bestämma släktskap
GIASID	Global Initiative on Sharing All Influenza Data, en internationell databas med möjlighet att snabbt dela sekvenseringsdata mellan länder och forskargrupper
Nextstrain	Ett mjukvarusystem bestående av två delar, en analytisk del som jämför arvsmassa från bakterier och virus, kallad Augur, och en del som visualiserar sagda data, kallad Auspice
Augur	Ett mjukvaruflöde, paketerat inom Nextstrain, för jämförelse av sekvenser av arvsmassa
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2 som betyder ungefär "allvarligt akut respiratoriskt syndrom-coronavirus 2"
Covid-19	Coronavirus disease, alltså coronavirus sjukdom, och året, 2019, då den diagnoserades första gången.
Fylogenetik	Evolutionsbiologisk terminologi för jämförande studier av släktskap mellan organismer. Här använt för att beskriva den process med vilken man jämför SARS-CoV-2 virus och särskiljer dem från varandra i genetiska grupper.
Provtagningsindikationen	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
Genetisk grupp	De grupper som bildas av jämförande analyser mellan olika virus arvsmassa
Fylogenetiskt träd	Ett hierarkiskt system i form av ett dendrogram som beskriver genetiskt släktskap mellan organismer, baserat på analys av virus arvsmassa.

## Sentinelövervakning

I detta sammanhang åsyftas sentinelprovtagning där patienter provtas på vårdcentraler så man kan fastställa hur många personer med influensasymtom som faktiskt har influensa eller covid-19. Sentinel är det engelska ordet för spejare.

# Sammanfattning

Helgenomsekvensering av SARS-CoV-2 har etablerats vid Folkhälsomyndigheten och kan utföras genom samarbete med landets kliniska mikrobiologiska laboratorier. Detta är den första genetiska karakteriseringen av SARS-CoV-2 i Sverige. Den fokuserar på de genetiska grupper som identifierats i diagnostiken under den tidiga fasen av pandemins utbredning i Sverige och omfattar prov från fall diagnosticerade under perioden 31 januari när landets första fall av covid-19 diagnosticerades, till och med den 15 mars då samhällssmitta av covid-19 hade konstaterats i Sverige. Utöver prov från fall inom diagnostiken ingår även prov från fall identifierade genom sentinelövervakningssystemet, vars provtagning i rapporten, sträcker sig från första veckan i mars till mitten av maj.

Utifrån analysen av virus arvs massa kan den smitta vi ser i landet så här långt, i huvudsak grupperas två eller tre genetiska grupper beroende på system. De genetiska grupper som identifieras i Sverige är samma som förekommer i den omfattande smittspridning som utvecklades under februari i Europa. En av de genetiska grupperna är även förekommande i den smittspridning som konstaterats i New York i slutet av februari och början av mars. Majoriteten av de genetiska grupper som finns rapporterade från de tidiga smittkedjorna, främst från Asien till Europa, kan endast påvisas med enstaka fall i det analyserade svenska materialet vilket indikerar att smitta från dessa genetiska grupper hittills inte fått någon omfattande spridning i Sverige och potentiellt brutits helt.

Ett stort antal fall av covid-19 finns konstaterade efter resa till Italien och två av de huvudsakliga genetiska grupperna som konstateras i sentinelövervakningen under mars tillhör också denna sekvensgrupp. Sekvenser från fall efter resa till Österrike utgör merparten av fallen i den tredje stora genetiska gruppen och även denna genetiska grupp återfinns i sentinelövervakningen under mars månad. Alla dessa tre genetiska grupper finns representerade i flera andra europeiska länder vid tiden för när dessa fall identifieras i Sverige och är länder till vilka svenskar reser frekvent. Merparten av fallen i det tidiga svenska materialet, är i enlighet med den då aktuella provtagningsindikationen, provtagna efter resa Italien och Österrike och utgör därför fram till mitten av mars en stor del av de svenska fallen i dessa genetiska grupper. Det finns också bland de svenska fallen med virus tillhörande dessa tre huvudsakliga genetiska grupper uppgift om vistelse i flera andra europeiska länder samt USA.

Internationell helgenomsekvensdata över utbrottet är fortfarande ofullständig, då alla länder inte haft kapacitet eller möjlighet att prioritera helgenomsekvensering för att följa utbrottet, detta gör underlaget mycket begränsat och saknar representativitet från vissa länder i världen. Provtagningsindikation för SARS-CoV-2 samt diagnostisk förmåga att identifiera tidiga introduktioner var även den ofullständig i många länder, med påföljden att delar av det initiala utbrottet sannolikt inte identifierades.

De tidiga svenska fallen som diagnosticerade och därmed också underlaget med virus som finns tillgängliga för helgenomsekvensering och denna analys är styrd av den initiala provtagningsindikationen. Detta gör att de sekvenser som finns tillgängliga inom diagnostiken motsvarar den omfattning misstänkta fall blivit provtagna. Representativiteten för fall importerade från andra länder än de som innefattades i provtagningsindikationen är därför sannolikt begränsad och utgörs endast av sekvenser från svårt sjuka fall som kommit till hälso- och sjukvårdens kännedom. Genom sentinelövervakningen identifieras fall med influensaliknande symtom oavsett provtagningsindikationen för covid-19 varför dessa data har en högre representativitet av den faktiska förekomsten av olika genetiska typer i landet. Mer data rörande nu cirkulerande genetiska grupper kommer fortsatt att erhållas genom det etablerade nationella mikrobiella övervakningsprogrammet för SARS-CoV-2 som kommer att generera ytterligare helgenomsekvenser löpande och publiceras i den internationella databasen GISAID, med påföljande fortsatt analys av den fylogenetiska förändringen över tid i utbrottet.



## Avgränsning

Analysen av helgenomsekvenser omfattar svenska prov med SARS-CoV-2 identifierade under perioden 31 januari – 15 mars. Ytterligare sekvenser kommer att tillföras analysen allt eftersom dessa sekvenseras och därmed möjliggöra fortsatt analys. Ett nationellt övervakningsprogram är etablerat och sekvensering kommer ske löpande inom ramen för detta, samt inom ramen för influensa och covid-19 sentinelövervakningssystemet.

## Syftet med helgenomsekvenseringen

Genom att noga analysera prov med SARS-CoV-2 från svenska fall kan förståelse för hur covid-19 introducerades i Sverige liksom hur smittspridningen i Sverige sett ut öka. Analyser från helgenomsekvenseringen kan också påvisa förändringar i virus över som kan vara av relevans för att anpassa de diagnostiska verktyg som finns etablerade, upptäcka förändrad förmåga hos immunförsvaret att känna igen viruset vid en eventuell förnyad exposition liksom mottagligheten för olika läkemedel och på sikt effekten av ett vaccin.

# Bakgrund

Covid-19 pandemin, som orsakas av viruset SARS-CoV-2, har i dagsläget över 40,000 diagnostiserade sjukdomsfall i Sverige och över 6,5 miljoner smittade i världen. Folkhälsomyndigheten beslutade tidigt att utveckla en sekvenseringsbaserad metodik för att följa virusets utveckling i Sverige.

Initialt provtagningsindikation inkluderande resande från Kina, Iran, Sydkorea och Italien med feber, hosta eller andningspåverkan. Efter rapporter om samhällssmitta i de österrikiska alperna tillkom även Österrike till provtagningsindikationen den 9 mars. Detta då internationell data visade på att en ansenlig del SARS-CoV-2 positiva prover kunde smittspåras till resande i Österrike (Gudbjartsson, D. F. et al. 2020). Analysen i rapporten är baserad på helgenomssekvenser av SARS-CoV-2 från de tidigaste fallen som diagnosticerades i landet samt från det nationella övervakningsprogrammet som inkluderar sentinelövervakning influensa och covid-19, vilket ger en god indikation av spridningen i Sverige under perioden januari – maj utan riktad provtagning baserat på resehistorik. Dessa sekvenser jämförs med internationell publicerade sekvenser genom databasen GISAID (Shu, Y. & McCauley, J. 2017).

En del i analysen av helgenomssekvenserade data är att identifiera nära släktskap mellan virus påvisade i Sverige och internationella sådana, som indikation på att en i tiden nära gemensam smittkälla finns. En rad internationella system för klassifikation av fylogener (släktskapssamband) finns publicerade som de svenska sekvenserna kan jämföras mot för att förstå smittspridningen mellan olika länder.

För närvarande används i huvudsak tre sätt att klassificera virusets släktskap för att härleda virusets släktskap och relation till andra virus, GISAID, Nextstrain och PANGOLIN (Shu, Y. & McCauley, J. 2017, Hadfield, J. et al. 2018, Rambaut, A. et al., 2020). Metodiken finns närmare beskriven i bilaga 1. I rapporten förekommer hänvisningar till samtliga system, men slutsatser är baserade på en kombination av GISAID och PANGOLINs system.

För denna genetiska karakteriseringsrapport har ett urval gjorts där prov från fall identifierade baserat på den mellan januari 2020 och 15 mars 2020 aktuella provtagningsindikationen, samt sentinelprover fram till och med 14 maj. Tabell 1 beskriver fördelningen av prover i analysen. Rapporten syftar till att karaktärisera de olika genetiska gruppernas spridning i Sverige under de första månaderna av utbrottet 2020. Detta ger en överblick över det initiala utbrottet samt en indikation kring utbrottet över tid samt efterföljande introduktioner av SARS-CoV-2.

**Tabell 1.** Antal sekvenser inkluderade i rapporten.

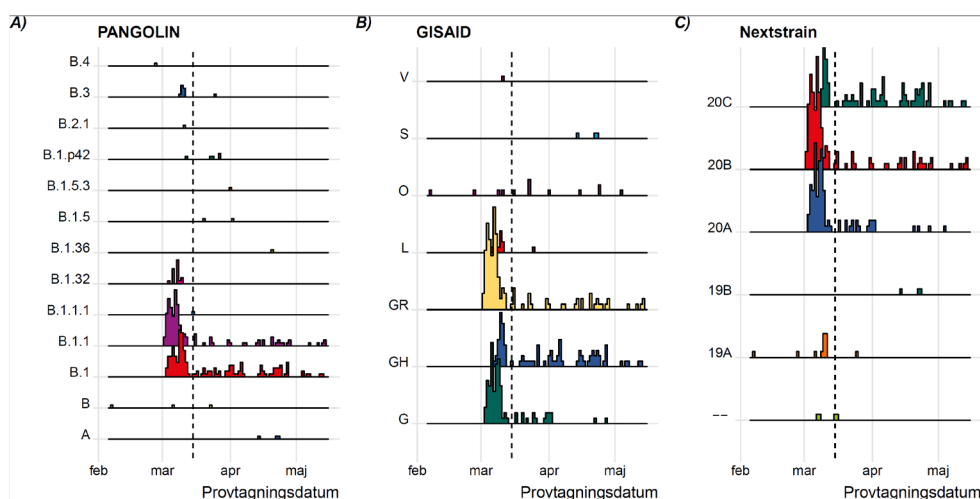
Prover	Antal	Tidsperiod
Tidiga prover	219	2020-02-07 – 2020-03-15
Sena prover	125	2020-03-16 – 2020-05-14
Totalt	344	

# Resultat

## Översikt av genetisk grupptillhörighet hos de svenska sekvenserna

De 344 prover som hittills har sekvenserats vid Folkhälsomyndigheten tillhör framför allt två eller tre genetiska grupper, beroende på vilket klassificeringssystem som används. En sammanfattning av hur sekvenserna fördelas över de genetiska grupperna visas i figur 1. En mer detaljerad översikt där de genetiska grupperna även visas separat för de tidiga proverna och sentinelprover finns i bilaga 1 (bifogad figur 1- 3, samt bifogad tabell 2-4)

Figur 1. Fördelning av de svenska sekvenserna mellan de olika genetiska grupperna över tid, enligt de tre klassifikationssystemen PANGOLIN (A), GISAID (B) och Nextstrain (C). Den streckade vertikala linjen markerar 15 mars, Y-axeln visar genetisk grupp och X-axeln visar provtagningsdatum.



I figur 1 ser man en initial introduktion av flera olika genetiska grupper. Det är dock bara ett fåtal grupper som dominerar bland de prov som sekvenserats i Sverige, i PANGOLINs system är dessa B.1 och B.1.1, se figur 1 samt tabell 2. Under en kortare period finns även en mestadels svensk grupp, B.1.32, representerad i landet, i enlighet med PANGOLINs klassifikationssystem. B.1.32 representerar ett avgränsat utbrott av en genetisk grupp, ett så kallat kluster, där importsmittan kan härledas till resa men den genetiska gruppen inte identifierats i något annat land. Sedan i slutet av mars har B.1.32 inte påvisats i den provtagning som ligger till grund för rapporten. En begränsning är att antalet prov som är sekvenserade efter mars är förhållandevis låg, vilket möjliggör att en låg förekomst av B.1.32 inte synliggörs i det relativt begränsade underlag av den totala smittspridningen som de sekvenserade proven utgör. Internationellt har flera tusen sekvenser publicerats under tidsperioden och B.1.32 kvarstår som en unikt svensk grupp. Även om grupp B.1.32 är unikt svensk i den tillgängliga data som analysen utgår från, härstammar gruppen troligen från importsmitta, vilket har stöd i smittspårningen under provtagningen.

Grupper med härstamning från grupp A är ovanliga i detta material, men finns representerade i sentinelproverna (2,8%; figur 1A). I princip alla fall av A går att knyta till importsmitta från Asien och i dagsläget syns ingen tydlig indikation på cirkulation av dessa grupper i Sverige.

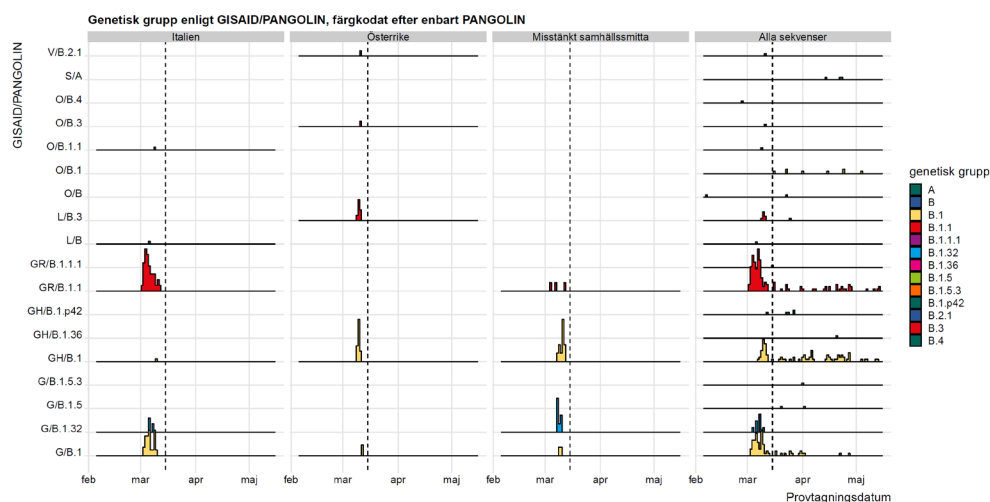
GISAID klassifikation indikerar att tre genetiska grupper etablerade sig i det initiala utbrottet, G, GH och GR. Det indikerar även att grupp L fanns bland den initiala spridningen, men då sentinelproverna senare bara fångar upp något enstaka fall med virus tillhörande denna genetiska grupp, är den troligen lågfrekvent i utbrottet. För sentinelproverna påvisas grupp O under hela perioden och är spridd i hela landet. Grupp O är en samlingsgrupp och representerar flera olika mindre genetiska grupper. I fallet med Nextstrain grupper ser vi ett liknande mönster som med GISAID klassifikationen, tre grupper etablerar sig under utbrottets initiala fas.

Samtliga tre metoder stödjer en dominans av två eller tre genetiska grupper i Sverige. I samtliga system ser vi även ett antal lågfrekventa genetiska grupper, som indikerar att även om de stora grupperna haft en tongivande roll i utbrottet finns introduktioner och mindre utbrott av andra grupper i landet. De huvudsakliga genetiska grupperna har etablerat sig i Sverige och bidrar löpande till nya infektioner med bibehållen genetisk grupp.

## Kombinerad analys av genetiska grupper

För att få en högre upplösning av de genetiska grupper som cirkulerar i Sverige har en kombinerad analys utifrån klassifikation i GISAID och PANGOLIN genomförts (Se figur 2). Detta då spridningen internationellt av framförallt två genetiska grupper gör att härledning baserat på enbart ett system saknar upplösning nog för att dra slutsatser kring smittans ursprung.

**Figur 2.** Kombinerad analys av genetiska grupper utifrån klassifikation i GISAID/PANGOLIN, fördelat på provtagning riktad mot resenärer från Italien, Österrike samt övriga.



Tredje panelen visar misstänkt samhällsmitta och den sista panelen visar det totala utbrottet över tid. Den streckade vertikala linjen visar på brytpunkten 15 mars, Y-axeln utgör de genetiska grupperna och X-axeln datum för provtagningsstillfälle.

Prover från Italien och Österrike bröts även ut ur analysen för att tydligare särskilja den riktade provtagningen mot sporadiska fall som detekterades under perioden samt fallen i sentinelövervakningen.

Den kombinerade analysen av GISAID/PANGOLIN genetiska grupper visar på en separat introduktion av dels två genetiska grupper från Italien och två från Österrike, se figur 2 och tabell 2. Det visar även på en bred variation bland de identifierade genetiska grupperna i materialet. Genom smittspårning kan även data kring samhällsspridning kopplas till det initiala utbrottet. I Figur 2 är möjlig samhällssmitta plottad i panel tre.

Ett flertal av de i analysen inkluderade proven (Figur 4) saknar dock smittland. Då flertalet av de genetiska grupper som importerats eller kartlagts saknar en direkt koppling till resande och var vida spridda i Europa vid tidpunkten är det i dagsläget inte möjligt att dra en slutsats kring smittland. Det finns dock ett antal genetiska grupper i analysen, som inte går att härleda direkt till de dominerande genetiska grupperna i Europa, en möjlig förklaring är att ett antal mindre utbrott även har orsakat spridning, men att de flesta av dessa smittkedjor har brutits.

**Tabell 2.** Antal sekvenser sekvenserade vid folkhälsomyndigheten med rapporterat smittland. För två sekvenser finns ett angivet smittland även vid provtagning efter 15 mars, även dessa är nu inkluderade och det indikeras i provtagningsintervallet vilka de är.

Resland	Antal	Tidigaste	Senaste	Genetisk grupptillhörighet
Kina	1	"2020-02-07"		O/B
Iran	1	"2020-02-27"		O/B.4
Italien	129	"2020-03-02"	"2020-04-02"	G/B.1; G/B.1.32; G/B.1.5; GH/B.1; GR/B.1.1; GR/B.1.1.1; L/B; O/B.1.1
Nederländerna	1	"2020-03-06"		GR/B.1.1
Portugal	1	"2020-03-20"		G/B.1
USA	1	"2020-03-08"		GH/B.1
Österrike	24	"2020-03-09"	"2020-03-12"	G/B.1; GH/B.1; L/B.3; O/B.3; V/B.2.1
Frankrike	1	"2020-03-11"		GH/B.1
Spanien	1	"2020-03-12"		GH/B.1
Tyskland	1	"2020-03-07"		G/B.1

## Analys av resehistorik i relation till genetisk grupp

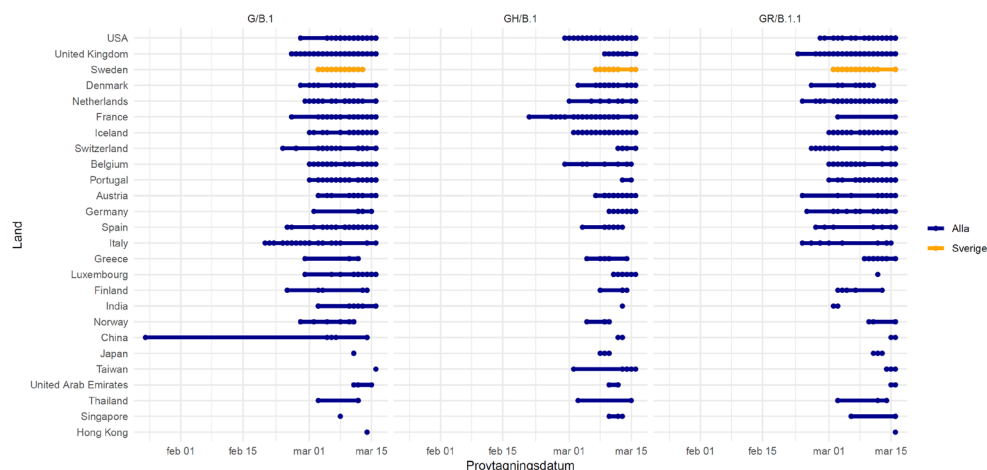
En stor andel tidiga sekvenser i urvalet härstammar från Italien och Österrike, se tabell 2. Detta är inte förvånande då ett fokus i den initiala provtagningen var baserad på provtagningsindikation och riktad mot Italien, Österrike, Iran, Kina och Sydkorea. De få fall som identifierats från Iran, Sydkorea och Kina tyder på att enbart ett fåtal importer skedde från dessa länder, då en mycket grundlig provtagning genomfördes av de hemvändande resande från de länderna i den initiala smittspridningen, sporadiska fall kan dock vara länkade till dessa länder då

underlaget av kartläggning är svagt i vissa länder och hemvändande medborgare anlände under en längre tid, se figur 2. Importsmitta från Italien och Österrike utgör en majoritet av proverna i den tidiga provtagningen, då en riktade provtagning gjordes i enlighet med provtagningsindikationen. Trots en god provtagning av resande från dessa länder finns det ytterligare genetiska grupper bland de övriga proverna (de med okänd smittväg).

## Fördelning av genetiska grupper internationellt

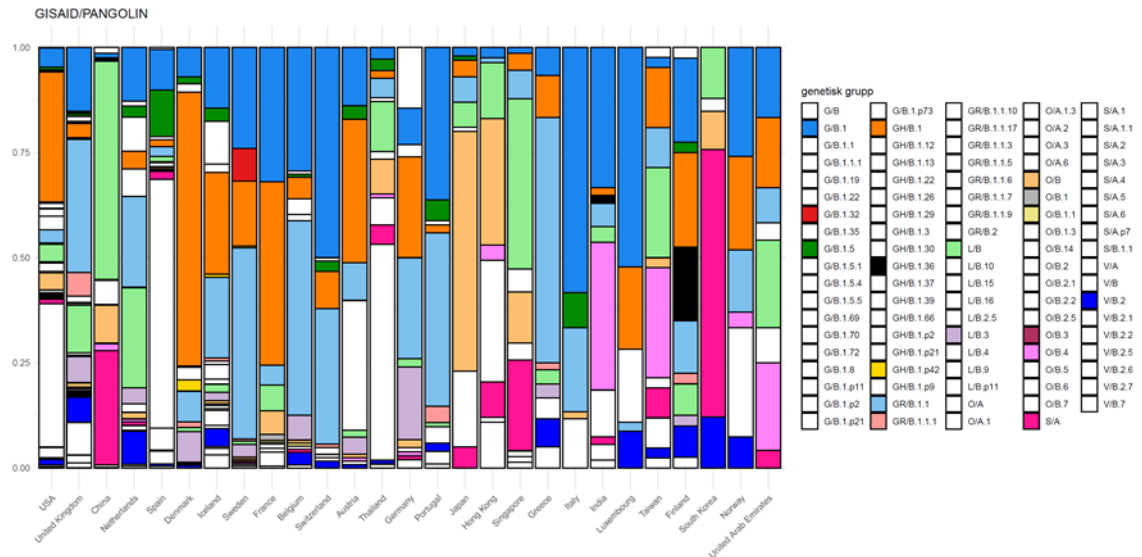
Genom genetisk kartläggning internationellt går det att få en uppfattning om spridningen av vanliga grupper i länder inom Europa och USA, vanliga resedestinationer med en hög grad av resande för Svenska resenärer. Figur 3 visualiserar de huvudsakliga genetiska grupperna som etablerades i Sverige i förhållande till Europa, USA och Asien under perioden fram till 16 mars. Givet data i den här analysen, såväl som i internationella analyser, är data från den tidiga introduktionen utanför Kina ofullständig för de flesta länder, med ett fåtal undantag. Det innebär att en klar fylogenetisk härledning till introduktion är svår att utföra. Vad vi vet från analysen är dock att de i Sverige dominerande grupperna även fanns i övriga Europa under tidsperioden för introduktion till Sverige. Värt att notera är att G/B.1, GH/B.1 och GR/B.1.1 är vanligt förekommande i USA och Europa, men frånvarande i Sydkorea och Kina, vilket ger ytterligare stöd till att smittkedjor direkt från Asien kraftigt begränsades i Europa, se figur 4.

**Figur 3.** Sveriges vanligaste genetiska grupper kronologiskt i ett internationellt sammanhang, baserat på ett utdrag ur dataset i GISAID. En prick representerar ett provtagningsdatum, separerade på land och genetisk grupp med 16 mars som sista datum.



Italien speglar mycket av utbrottet i Europa, med en majoritet av sekvenserna i grupp G/B.1. Även om klara smittkedjor finns kopplat till resande till Italien, är spridningen av den predominant högst förekommande genetiska gruppen i Italien så stor i Europa att det är högst troligt att den importerats från ytterligare länder.

**Figur 4.** Fördelning av genetiska grupper internationellt under tidsperioden till och med 16 mars. Urvalet är de länder i Europa, Asien samt USA som har deponerat minst 20 sekvenser i GISAID databasen, under tidsperioden. Enbart de genetiska grupper som identifierats i Sverige färgkodande.



Även om materialet i rapporten indikerar en förväntad stor import genom resande till kända smittland, visar den också på att de genetiska grupper som importerades fanns vitt spridda i Europa och USA vid den tidpunkten. Detta i kombination med en större än förväntad samhällssmitta kunde konstateras i mitten av mars, indikerar att smitta importerades från ytterligare länder, troligen under en tidigare smittspridning i Europa än vad som vid tidpunkten var beskrivet, även om rapporter senare har tillkommit om tidig spridning (Deslandes, A. *et al.* 2020). Analysen visar även på att ett flertal andra genetiska grupper importerades, skilda från det stora utbrottet i Italien och Österrike, troligen från andra länder, detta då de typer av viruset som karakteriseras i sentinelprover från perioden mars till maj har genetiska grupper som inte är beskrivna i Österrike eller Italien, se tabell 2.

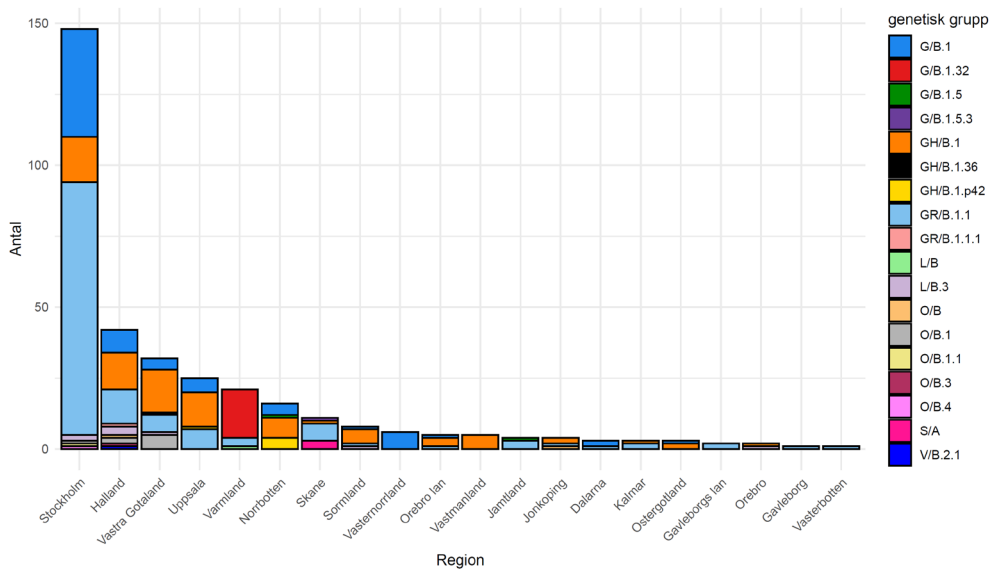
Baserat på den variation av genetiska grupper som finns i det svenska materialet, uppgifter som samhällssmitta hos svenska fall i mitten av mars samt dagens kunskap om förekomst av virus tillhörande samma genetiska grupper som återfinns i Sverige, i flera europeiska länder till vilka svenskar reser frekvent visar på att introduktion av SRAS-CoV-2 skedde från ett flertal olika länder. Detta är kanske tydligast med de fall av genetiska grupper som inte finns beskrivna alls i data från Italien och Österrike men dyker upp tidigt i den svenska provtagningen. Uppföljande analyser kommer klargöra om någon av dessa ännu så länge mindre frekvent förekommande genetiska typer etablerade sig i Sverige. Grupperingen av proverna visar ett mönster där proverna från Italien är tydligt formerad i två grupper, G/B.1 och GR/B.1.1. Prover från Österrike är tydligt grupperade i GH/B1 och L/B.3. Prover vars smittland inte har en tydlig anknytning till de i provtagningsindikationen beskrivna länderna är spridda över ett flertal genetiska grupper. Samtliga sekvenser från de övriga Europeiska länderna, samt USA, under



tidsperioden tillhör även de, de i Sverige vanligt förekommande genetiska grupperna, GH/B.1, GR/B.1.1 samt G/B.1.

Utifrån det faktum att tre grupper dominerar i det tidiga svenska materialet finns en jämförelse över tid för dessa genetiska gruppers förekomst i länder med högt resande till och från Sverige, se Figur 3.

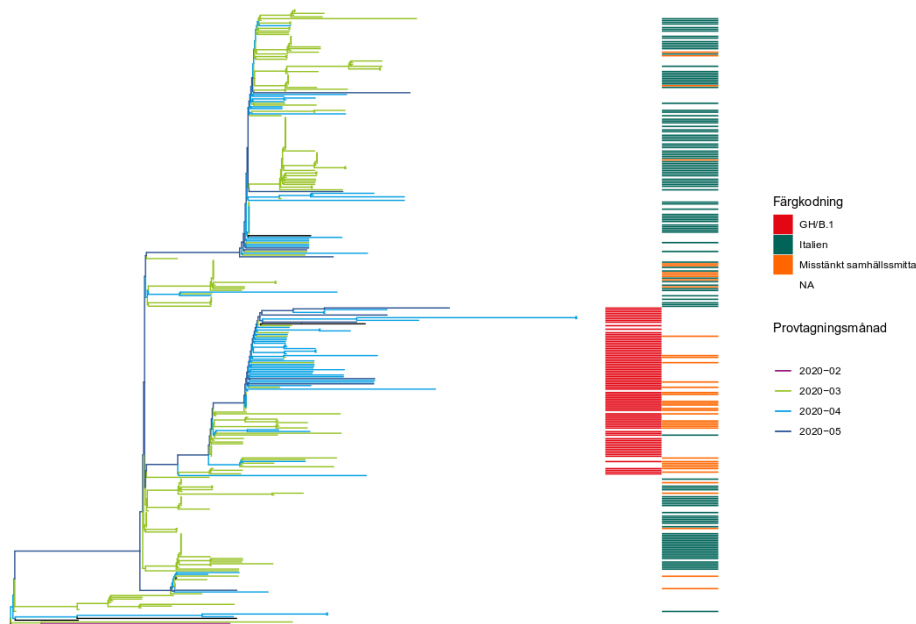
**Figur 5.** Fördelning av genetiska grupper baserat på angiven region



Regionalt är provtagningen överrepresenterad från Stockholmsregionen då denna region hade ett stort antal resande under slutet av februari/början av mars som uppfyllde provtagningsindikationen, se figur 5. Fördelningen av olika genetiska grupper skiljer sig något över landet, men de övergripande mönstren är snarlika. G/B.1.32 är enbart lokaliserad till region Värmland, och har i fallrapporteringen angivet smittland Italien eller samhällssmitta i Sverige. G/B1.32 har dock aldrig fångats upp i tidigare prover från Italien och internationellt finns den ännu så länge bara beskriven bland de svenska proven.

Figur 6 visualiserar det svenska utbrottet som ett fylogenetiskt träd. Som kan ses för de prov som påvisats positiva för SARS-CoV-2 under april (blå färg i figuren), domineras dessa av en genetiska grupp (GH/B.1) som saknar koppling till Italien (endast ett svenskt fall i denna genetiska grupp har smittland Italien angivet). De sekvenser som har Sverige som angivet smittland är överrepresenterade i denna grupp jämfört med deras representation i de övriga genetiska grupperna. Detta indikerar att den under april inom sentinelövervakningen dominerande smittan tillhör en genetisk grupp som inte direkt går att härleda till importsmitta från Italien. GH/B.1 är inte vanligt förekommande i Italien men har karakteriserats bland prover från resande till Österrike och andra länder. Detta är dock inte utmärkt i figur 3, men visualiseras i figur 3.

**Figur 6.** Det svenska utbrottet som ett fylogenetiskt träd. Sekvenser genetisk grupp GH/B.1 är utmärkta med röda sträck i högerkanten.



Trädets grenar är kodade för provtagningsdatum, där prover tagna i mars är representerade i grönt och prover tagna i april och maj i olika nyanser av blått. Ytterst i högerkanten är sekvenser från bekräftad importsmitta från Italien markerade i grönt, prov från personer som blivit smittade i Sverige i orange.

Analysen visualiserad i figur 6, visar även på en trend att GH/B.1 är den grupp som aktivt sprids och utvecklas i de prover som inkommer från sentinelprovtagningen. Då GH/B.1 inte var representerad i provtagningen från Italien, men fanns i åtskilliga andra länder i Europa samt Österrike är det troligt att delar av den samhällsmitta som ses nu är en effekt av introduktion från länder utanför provtagningsindikationen.

Baserat på fördelningen av genetiska grupper i Europa, se figur 4 och figur 5, den tidiga samhällsspridningen, se figur 2, och den breda representationen av genetiska grupper, tabell 2 är det troligt att smitta introducerades från ett flertal länder utanför provtagningsindikationen. I underlaget för rapporten är dock Italien och Österrike representerade i hög grad, då det initialt var många positiva fall hos återvändande resenärer. De genetiska grupper som primärt importerats från Italien och Österrike var dock del av det stora Europeiska utbrottet och introduktion från ett flertal länder är troligt. De genetiska grupper som då etablerades tidigt i landet förekommer fortfarande hos fall diagnosticerade i Sverige, även om GH/B.1 dominerar från fall påvisade inom sentinelövervakningen under april.

## Sammanfattande diskussion

Den omfattande provtagningen vid det svenska utbrottets start har möjliggjort för Sverige och Folkhälsomyndigheten att följa utbrottets utveckling. Baserat på data från det initiala utbrottet och från sentinelövervakningen kan konstateras att smittspridningen i Sverige domineras av tre genetiska grupper. Dessa grupper är vanligt förekommande i Europa och USA. Under slutet av februari och början av mars spreds dessa genetiska grupper inom Europa liksom mellan USA och Europa, (Pybus O & Rambaut A, *et al.*, 2020, Gonzalez-Reiche, A. S. *et al.*, 2020).

Efterföljande provtagning visar även att dessa genetiska grupper sprider sig globalt. Därmed är det svårt att dra slutsatser om smittans ursprung med enbart dessa data som underlag. De sekvenser som är tillgängliga för jämförelse är även ojämnt fördelade över Europas länder, då kapaciteten till och möjligheten att prioritera helgenomsekvensering varierar.

Baserat på den spridning av genetiska grupper som fanns i Europa vid månadsskiftet februari -mars och som korrelerar med de genetiska grupper som detekterats i Sverige är det troligt att, även om en majoritet av proverna i undersökningen härrör från import från Italien och Österrike, så var spridningen i Europa en sådan vid tidpunkten att smitta importerades från ett antal andra länder i samma tidsperiod.

Det är troligt att ytterligare importfall av dessa genetiska typer skedde med hemvändande resande även efter att provtagningsindikatorn ändrats till att inkludera en generell provtagning. De genetiska grupper med låg förekomst som kan konstateras från fall inom sentinelövervakningen indikerar fortsatt introduktion av nya varianter av viruset över tid även under mars.

## Pågående och framtida analyser

Folkhälsomyndigheten arbetar nu med att fördjupa analysen samt att inkludera sekvenser från fall diagnosticerade efter den 15 mars för att ytterligare få förståelse för spridningen i landet och förändringar över tid. För att fortsatt inhämta kunskap om vilka typer som cirkulerar i landet kommer sentinelövervakningent fortsätta tillsvidare.

Diagnostiska prov kommer fortsatt också systematiskt samlas in från de kliniska mikrobiologiska laboratorerna systematisk kartläggning av genetiska grupper som cirkulerar i landet analyseras. Tillsammans kommer dessa aktiviteter ge en god bild av utveckling över tid av de nu konstaterade genetiska grupperna och eventuellt tillkommande sådana liksom om förändringar sker i virusets arvs massa som kan påverka befintliga diagnostiska verktyg, upptäcka förändrad förmåga hos immunförsvaret att känna igen viruset vid en eventuell förnyad exposition liksom mottagligheten för olika läkemedel och på sikt effekten av ett vaccin

# Bilaga 1.

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.

**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.

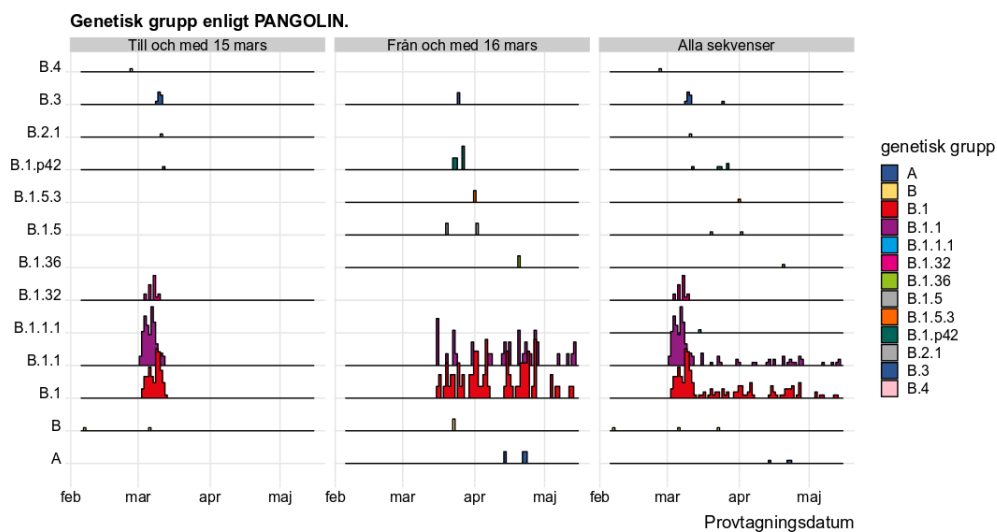
## PANGOLINs klassifikationssystem

- 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 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 C.1).
- c. Alla sekvenser tilldelas ett släkte, så om de inte uppfyller kriteriet för en lägre nivå, tilldelas de en högre

**Bifogad figur 1.** Figuren visar hur antalet sekvenser tillhörande de olika genetiska grupperna, enligt PANGOLIN, förändrats över tid. Staplarna är proportionerliga mot antalet fall.



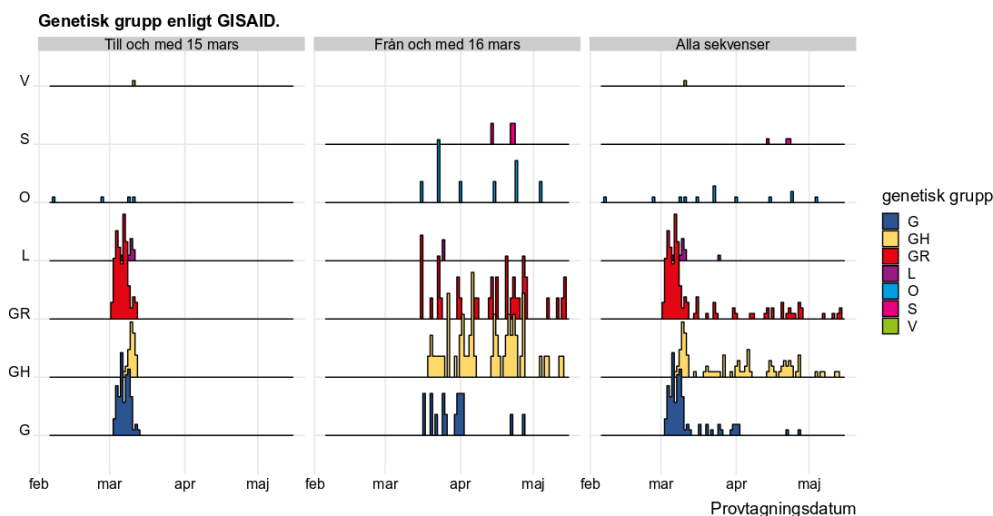
En sammanställning av de olika genetiska grupperna i enlighet med PANGOLINs systemet finns i tabell 2 nedan.

**Bifogad tabell 2.** Fördelningen av genetiska grupper givet klassifikation med PANGOLIN.

Genetisk grupp (huvudgrupp och subgrupp)	Antal
A	3
B	3
B.1	161
B.1.1	139
B.1.1.1	1
B.1.32	17
B.1.36	1
B.1.5	2
B.1.5.3	1
B.1.p42	5
B.2.1	1
B.3	9
B.4	1
Totalt	344

Initialt sker ett antal introduktioner av flera olika genetiska grupper. Det är dock bara ett fåtal grupper som får fäste i landet, i PANGOLINs system är dessa B.1 och B.1.1, se figur 1 samt tabell 2. Under en kortare period finns även en mestadels svensk grupp, B.1.32, representerad i landet, i enlighet med PANGOLINs klassifikationssystem. B.1.32 representerar ett avgränsat utbrott av en genetisk grupp, ett så kallat kluster, där importsmittan kan härledas till resa men den genetiska gruppen inte identifierats i något annat land. Sedan i slutet av mars har B.1.32 inte påvisats i den provtagning som ligger till grund för rapporten. En begränsning är att antalet prov som är sekvenserade efter mars är förhållandevis låg, vilket möjliggör att en låg prevalens av B.1.32 kan missas. Internationellt har flertalet tusen sekvenser publicerats under tidsperioden och B.1.32 kvarstår som en unikt svensk grupp. Även om grupp B.1.32 är unikt svensk i den tillgängliga data som analysen utgår från, härstammar gruppen troligen från importsmitta, vilket har stöd i smittspårningen under provtagningen. Grupper med härstamning från grupp A är ovanliga i detta material, men finns representerade i sentinelproverna (2,8%). I princip alla fall av A går att knyta till importsmitta från Asien och i dagsläget syns ingen tydlig indikation på smitta av dessa grupper.

**Bifogad figur 2.** Figuren visar hur antalet sekvenser tillhörande de olika genetiska grupperna, enligt GISAID, förändrats över tid. Staplarna är proportionerliga mot antalet fall.



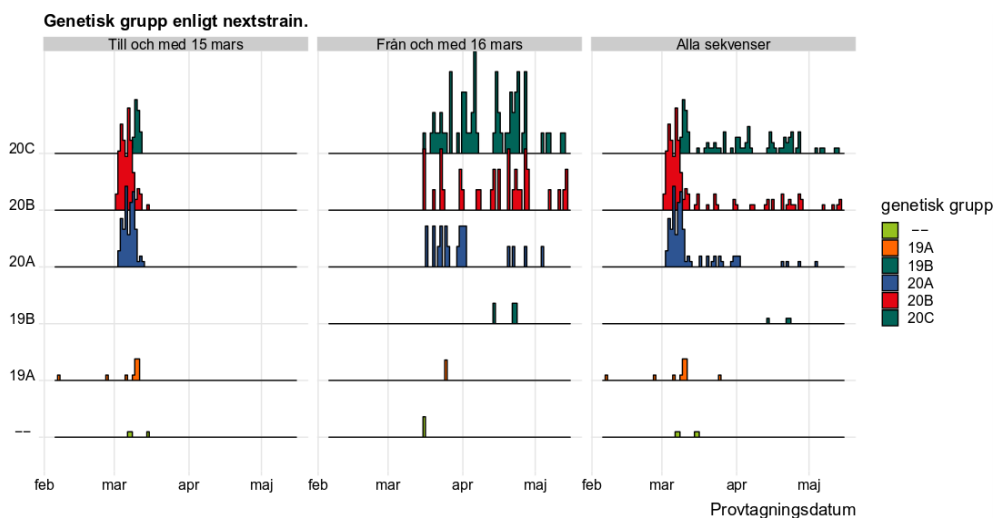
De totala antalet fall, givet GISAID klassifikation, redovisas i tabell 3 nedan.

**Bifogad tabell 3.** Det totala antalet fall givet GISAID klassifikation.

Genetisk grupp	Antal
G	90
GH	89
GR	139
L	9
O	13
S	3
V	1
Totalt	344

GISAID klassifikation indikerar att tre genetiska grupper etablerade sig i det initiala utbrottet, G, GH och GR. Det indikerar även att grupp L fanns bland den initiala spridningen, men då sentinelprovorna senare bara fångar upp något enstaka fall med virus tillhörande denna genetiska grupp, är den troligen lågfrekvent i utbrottet. För sentinelprovorna påvisas grupp O under hela perioden och är spridd i hela landet. Grupp O är en samlingsgrupp och representerar flera olika mindre genetiska grupper.

**Bifogad figur 3.** Figuren visar hur antalet sekvenser tillhörande de olika genetiska grupperna, enligt Nextstrain, förändrats över tid. Staplarna är proportionerliga mot antalet fall.



**Bifogad tabell 4.** Det totala antalet fall givet Nextstrain klassifikation.

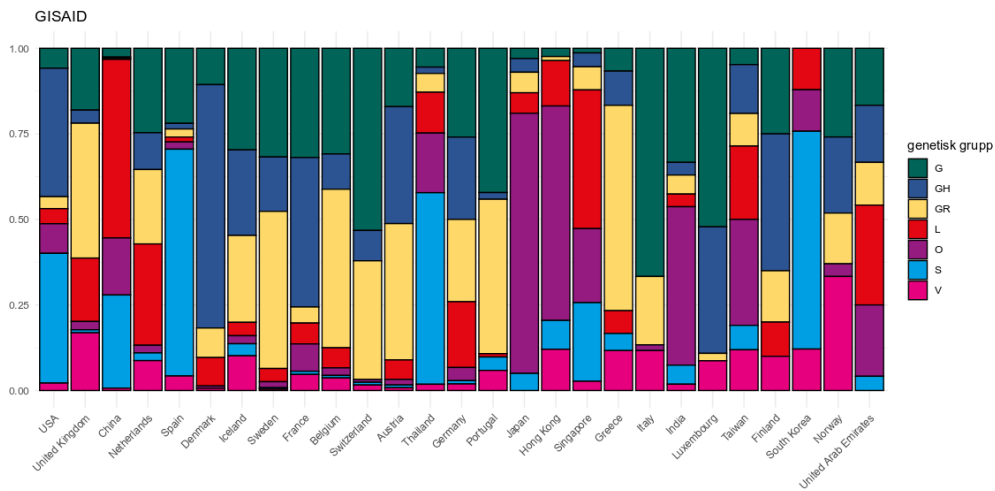
Genetisk grupp	Antal
Saknar genetisk grupp	4
19A	12
19B	3
20A	98
20B	138
20C	89
Totalt	344

I fallet med Nextstrain grupper ser vi ett liknande mönster som med GISAID klassifikationen, tre grupper etablerar sig under utbrottets initiala fas.

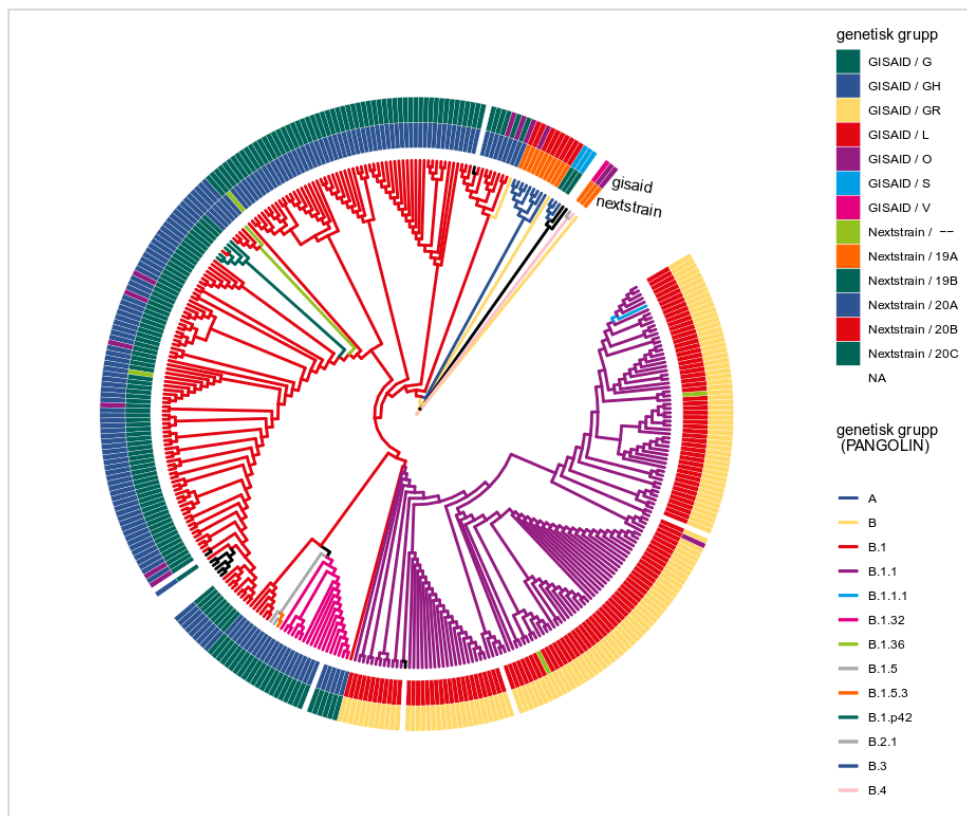
Samtliga tre metoder stödjer en dominans av två eller tre genetiska grupper i Sverige. I samtliga system ser vi även ett antal lågfrekventa genetiska grupper, som indikerar att även om de stora grupperna haft en tongivande roll i utbrottet finns introduktioner och mindre utbrott av andra grupper i landet.



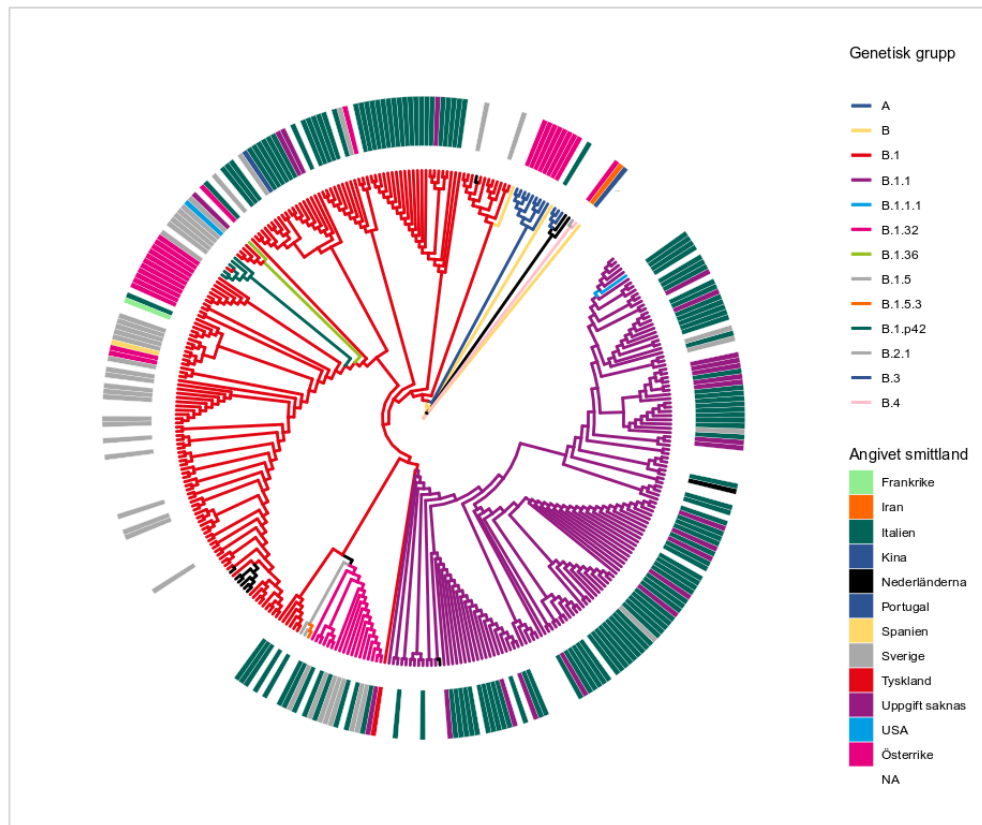
**Bifogad figur 4.** Fördelning av genetiska grupper internationellt, enligt GISAIIDs klassifikationssystem. Urvalet är de länder i Europa, Asien samt USA som har deponerat minst 20 sekvenser i GISAIID databasen.



**Bifogad figur 5.** En jämförelse av de tre olika klassificeringssystemen och hur de överlappar för de svenska sekvenserna. Det fylogenetiska trädet innehållande alla svenska sekvenser deponerade i GISAIID är färgkodat för genetisk grupp enligt PANGOLIN.



**Bifogad figur 6:** Det fylogenetiska trädet är skapat enligt beskrivning i metodsammanfattningen och inkluderar alla svenska sekvenser. Trädet är färgkodat efter genetisk grupp enligt PANGOLIN, för att visualisera vilka genetiska grupper importsmittan tillhör.



## Metodsammanfattning

Ion Torrent targeted Next Generation Sequencing (NGS), även kallad AmpliSeq, användes för helgenomsekvensering SARS-CoV-2 RNA som isolerats från covid-19 positiva patienter. Metoden bygger på en uppsättning av mycket specifika primrar, ”Ion AmpliSeq SARS-CoV-2 Research Panel” (Thermo Fisher Scientific) som i kombination med AmpliSeq reagenser producerar genomiska DNA segmentkopior som sedan kan sekvenseras på ett Ion Torrent-instrument.

Sekvensanalys genomfördes med hjälp av Torrent Mapping Alignment Program (TMAP) och plug-in i Torrent Suite (Torrent Variant Caller och Coverage Analysis) mot referensen Wuhan-Hu-1 NCBI Reference Genome (Accession number: MN908947.3. Wu *et al.* 2020). Endast genomsekvenser med minst 20x täckning på alla amplicon godkändes. Bland dessa valdes de högst frekventa varianter ut ur varje enskilt dataset och en sammanfattande, så kallad consensus, sekvens (inkluderat de förändringar som gavs stöd i sekvensanalysen) producerades. 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 deponerades).

Fylogener genomfördes i enlighet med samtliga existerande metodiker, GISAID, Nextstrain och PANGOLIN. Sekvensjämförelse skedde med MAFFT, träd byggdes med IQ-tree. Med antingen typningsschema (GISAID) eller script för klassifikation (Nextstrain och PANGOLIN). Träd skapades med ett urval av enbart svenska sekvenser, svenska sekvenser med representativa sekvenser för samtliga genetiska typer i PANGOLIN, samt svenska sekvenser och samtliga unika genetiska typer i PANGOLIN. Träd och data visualiserades i R, med hjälp av paketen ape, ggplot2, ggtree, dplyr, ggrepel, reshape2, tidyr, treeio.

## Källor

- Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020.04.17.046086 (2020)  
doi:10.1101/2020.04.17.046086
- Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, (2017)
- Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018)
- Wu, F., Zhao, S., Yu, B. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
- Gudbjartsson, D. F. *et al.* Early Spread of SARS-Cov-2 in the Icelandic Population. *Epidemiology* (2020) doi:10.1101/2020.03.26.20044446
- Deslandes, A. *et al.* SARS-CoV-2 was already spreading in France in late December 2019. *Int. J. Antimicrob. Agents* **55**, 106006 (2020)
- Pybus O & Rambaut A, *et al.* Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages. *Virological*  
<https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507> (2020)
- Gonzalez-Reiche, A. S. *et al.* Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* (2020) doi:10.1126/science.abc1917

## Författarnas tack till GISAID (Engelska)

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's hCoV-19 database, used in the genetic analyses. All submitters of data may be contacted directly via the GISAID website.

[www.gisaid.org](http://www.gisaid.org).

The appendix GISAIDsekvenser.xlsx indicated by a paperclip below is embedded in this file. To access it we recommend you to first download this file and opening it using Adobe Acrobat Reader. Double klick on each paperclip to open the appendix of choice.

GISAIDsekvenser.xlsx

Detta är den första genetiska karakteriseringen av SARS-CoV-2 i Sverige. Den fokuserar på de genetiska grupper som identifierats i diagnostiken under den tidiga fasen av pandemins utbredning i Sverige och omfattar prov från fall diagnosticerade under perioden 31 januari när landets första fall av covid-19 diagnosticerades, till och med den 15 mars då samhällssmitta av covid-19 hade konstaterats i Sverige.

Utifrån analysen av virus arvsmassa kan den smitta vi ser i landet så här långt, i huvudsak grupperas två eller tre genetiska grupper beroende på system. De genetiska grupper som identifieras i Sverige är samma som förekommer i den omfattande smittspridning som utvecklades under februari i Europa. En av de genetiska grupperna är även förekommande i den smittspridning som konstaterats i New York i slutet av februari och början av mars. Majoriteten av de genetiska grupper som finns rapporterade från de tidiga smittkedjorna, främst från Asien till Europa, kan endast påvisas med enstaka fall i det analyserade svenska materialet vilket indikerar att smitta från dessa genetiska grupper hittills inte fått någon omfattande spridning i Sverige och potentiellt brutits helt.

---

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.



Folkhälsomyndigheten

**Solna** Nobels väg 18, 171 82 Solna. **Östersund** Campusvägen 20. Box 505, 831 26 Östersund.

[www.folkhalsomyndigheten.se](http://www.folkhalsomyndigheten.se)