

Utredning av den initiala smittspridningen av SARS-CoV-2 till och inom Sverige

Robert Dyrdak och Jan Albert

Underlagsrapport till SOU 2021:89 Sverige under pandemin

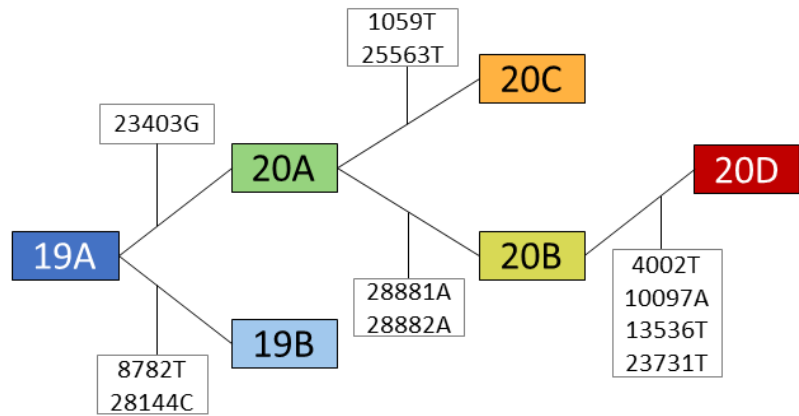
Stockholm 2021

ISBN 978-91-525-0263-1

Errata

På sidan 20 är ”20B” och ”20C” omkastade i Figur 4: Klassifikation av SARS-CoV-2 sekvenser.

Nextstrain	Gisaid	Pangolin
19A	L	B, B.3, B.11
	O	B, B.4, B.6
	V	B, B.40
19B	S	A, A.2, A.2.4
20A	G, GH	B.1
20C	GH	
20B	GR	B.1.1, B.1.1.29
20D		B.1.1.1, C.14



Sammanfattning

I coronakommissionens direktiv ingår att beskriva och analysera olika aspekter av den initiala smittspridningen till och inom Sverige. Denna rapport belyser dessa frågeställningar genom fylogenetiska släktskapsanalyser av SARS-CoV-2 gensekvenser och PCR analyser av tidigare ej testade luftvägsprov från början av 2020. Rapporten grundas på analys av 1 350 svenska sekvenser från prov tagna under perioden 31 januari till 1 juni 2020 samt alla cirka 125 000 internationella sekvenser tillgängliga i databasen GISAID för samma period, vilket inkluderar sekvenser från de andra nordiska länderna.

Inga indikationer på dold smittspridning före vecka 9

För att undersöka om det kan ha förekommit dold smittspridning av SARS-CoV-2 i Sverige innan de första kända fallen diagnostiserades utfördes SARS-CoV-2 PCR på knappt 2 000 sparade luftvägsprov från vecka 6 – 14

år 2020. Dessa prov hade inte tidigare genomgått denna analys utan hade testats för andra luftvägs patogener. Vi fann endast 15 säkra eller misstänkta nya fall. Dessa prov var tagna i vecka 10 eller senare, dvs. vid en tidpunkt då det redan fanns kända fall. Folkhälsomyndigheten (FOHM) har rapporterat liknande resultat vid analys av prov tagna för influensa-övervakning. Skulle dolda fall ändå ha funnits före vecka 9 så verkar de inte påtagligt ha bidragit till den fortsatta smittspridningen.

Många tidiga virusimporter, men få ledde till påtaglig fortsatt smittspridning

FOHMs tidiga rapporter anger att många hundra av de tidigaste fallen hade smittats i utlandet, men det är sannolikt att det verkliga antalet var flera gånger högre. Trots detta så talar våra undersökningar för få av dessa importörer påtagligt har bidragit till den fortsatta smittspridningen i Sverige. Istället kan många fall under första pandemivågen kopplas till smittkedjor som startades av några få tidigt importerade virusvarianter.

Tidig import av virusvarianter från italienska och österrikiska alperna initierade sannolikt en betydande del av den fortsatta smittspridningen i Sverige under den första pandemivågen

Många (67%) av de undersökta virussekvenserna är unikt svenska och mer än hälften kan kopplas till nio svenska eller nordiska smittkedjor. Både reseanamnes och de fylogenetiska analyserna indikerar att import från italienska och österrikiska alperna under vecka 9–10 var mycket viktig för att etablera dessa inhemska smittkedjor och därmed den fortsatta smittspridningen i landet. Ett kluster inom genotyp 20C (20C:G24368T) med troligt ursprung i Österrike förefaller ha varit extra viktigt och utgör knappt 50% av de undersökta sekvenserna.

Tidigt etablerade smittkedjor med ursprung i alpområdet dominerade smittspridningen under våren 2020 även i våra nordiska grannländer

Vi jämförde mönstren för virusets spridning i Sverige med de i Danmark, Finland, Island och Norge. I dessa länder, precis som i Sverige, finns många land-specifika smittkedjor som förefaller ha etablerats genom tidig import av virusvarianter från alpområdet. Detta talar för att skillnaderna i incidensen av SARS-CoV-2-infektioner mellan länderna och stor del beror på hur man lyckats dämpa den fortsatta inhemska smittspridning genom restriktioner. Skillnader i fortsatt virusimport från utlandet kan också ha bidragit till skillnaderna i antalet fall i de nordiska länderna, men bedöms ändå vara en mindre viktig faktor under den första pandemivågen.

Smittspridning i områden med socioekonomiska utmaningar

Vi har inte detaljerad information om var provtagna personer bodde eller arbetade, men provtagande sjukvårdsinrättning ger viss information. Genotyp 20C, som dominerade smittspridningen i Sverige, stod för en extra stor andel av prover tagna på Sankt Görans sjukhus, vars upptagningsområde är nordvästra Stockholm inklusive Järva-området. Detta talar för att viruset som spreds där hade sitt ursprung i österrikiska alperna, snarare än andra delar i världen. Våra undersökningar ger dock ingen exakt information om hur eller varför denna variant kom att dominera bland prover från Sankt Görans sjukhus.

Smittspridning inom äldreården

Vi information om att 161 av de svenska proverna prov var tagna på äldre- och omsorgsboende. Av dessa tillhörde 74% genotyp 20C, jämfört med 53% för det kompletta datasetet. Skillnaden speglar sannolikt att genotyp 20C gradvis fick en alltmer dominerande roll för smittspridningen i landet. Resultaten ger stöd för uppfattningen att den allmänna samhällsspridningen var den huvudsakliga källan till spridning inom äldreården, dvs att det inte var några speciella virusvarianter som spreds där. Dessutom finns exempel på att olika virusvarianter påvisats på ett och samma äldreboende vilket bara förklaras av flera oberoende virusintroduktioner till dessa boenden.

Uppdragsbeskrivning

I regeringens kommittédirektiv till Coronakommissionen (Dir. 2020:74) anges på sidan 6 att kommissionen ska:

- beskriva den initiala smittspridningen till och inom Sverige,
- analysera reseströmmarnas betydelse för smittspridningen i Sverige och i övriga nordiska länder, och
- särskilt beskriva smittspridningen i områden med socioekonomiska utmaningar och smittspridningens konsekvenser för utsatta grupper.

Coronakommissionen har därför uppdragit till Prof. Jan Albert, Karolinska Institutet att inkomma med en rapport som ger ett vetenskapligt underlag till kommissionens återrapportering av punkterna 1 och 2 ovan. I uppdragsbeskrivningen anges att kommissionen förväntar att arbetet fr.a. ska baseras på s.k. fylogenetiska analyser av SARS-CoV-2 gensekvenser och matematisk modellering. Projektgruppen förväntas insamla data för arbetet från Folkhälsomyndigheten och olika öppna källor samt själv generera data om detta bedöms nödvändigt. Resedata och analys av dessa kommer om möjligt göras i samarbete med forskningsprogrammet *Ett forskningsprogram om covid-19 i Sverige: Smittspridning, bekämpning och effekter på individer och samhälle* (under ledning av Torsten Persson, Stockholms universitet).

1 Inga indikationer på spridning av SARS-CoV-2 i Sverige före vecka 9 2020

1.1 Bakgrund – Undersökning av oupptäckt tidig smittspridning

Det finns rapporter om att SARS-CoV-2 retrospektivt har detekterats redan i december 2019 i prov från enstaka patienter i Frankrike [1] och Italien [2], samt i avloppsvatten i Italien [3]. Detta är cirka två månader innan de första fallen rapporterades från dessa länder. I Sverige diagnostiserades det första fallet i Jönköping i vecka 6 hos en person som återvänt från Wuhan-området i Kina. Det finns inga indikationer till att det uppstod sekundärfall från detta första svenska fall (figur 1). Det tog sedan till vecka 9 innan 13 fall rapporterades (6 i Västra Götaland, 4 i Stockholm, 2 i Uppsala och 1 i Jönköping). I svensk media har det dock rapporterats om personer i Dalarna med covid-19-liknande sjukdom redan i december 2019, men vi har ingen information om att dessa misstänkta fall har bekräftats.

Frågan uppkommer om det i Sverige fanns odetekterad spridning av viruset i samhället innan de första fallen diagnostiserades. För att besvara denna fråga analyserade vi 1 979 st sparade luftvägsprov från biobanken på Klinisk Mikrobiologi, Karolinska Universitetslaboratoriet (KUL) med PCR för SARS-CoV-2-RNA. Proven hade inte tidigare undersökts för SARS-CoV-2 utan hade skickats in för analys av andra luftvägs patogener. I urvalet exkluderades barn yngre än fem år samt i möjlig mån patienter med redan påvisat SARS-CoV-2-RNA i annat prov. Urvalet omfattade veckorna 6–14 under år 2020, se figur 2A. Det använde kommersiella PCR-testet riktas mot tre olika virusgener varav minst två ska detekteras för att provet ska anses vara positivt för SARS-CoV-2. I händelse av detektion av endast en gen bedömdes det som ett misstänkt fall av SARS-CoV-2-infektion.

1.2 Resultat – Undersökning av oupptäckt tidig smittspridning

Av 1 979 stycken analyserade luftvägsprov prov upptäcktes 13 stycken tidigare ej kända fall av SARS-CoV-2-infektion samt två misstänkta fall (figur 2A). Provtagningsdatumet för det tidigaste av dessa fall var vecka 10. Två prov var från vecka 11 och resterande tolv prov var tagna veckorna 12–14. Som framgår av figuren hade 4 respektive 140 stycken bekräftade fall rapporterats i Stockholm vecka 9 respektive vecka 10 (Folkhälsomyndighetens statistik [4]). Försök till sekvensering gjordes på samtliga 15 prov och var framgångsrik i 9 fall. Figur 2A visar att sekvenserna i dessa prov tillhörde virustyper som var tidigt etablerade i Stockholm och påvisades också bland tidigare kända fall.

1.3 Diskussion – Undersökning av oupptäckt tidig smittspridning

Analysen av sparade prov från KUL gav inget stöd för någon utbredd förekomst av SARS-CoV-2 i region Stockholm innan vecka 9 eller någon påtaglig dold smittspridning (figur 2A). Denna slutsats stöds också av data från Folkhälsomyndigheten, där prover för influensaövervakning (sentinelprov) analyserades för SARS-CoV-2 utan någon detektion av SARS-CoV-2 i prover från vecka 1–9 (figur 2B). Enstaka fall påvisades fr.o.m. vecka 10, precis som för proverna från KUL.

Resultatet är i linje med den översikt av rapporteringen av tidiga fall som globalt visade en ökning av antalet fall från och med vecka 9, speciellt i Europa [5], samt att större smittspridning inträffade i regioner med skollov vecka 9 [6]. I en liknande undersökning av sparade luftvägsprov i Nottingham, Storbritannien, detekterades tre tidigare oupptäckta fall, varav det tidigaste i ett prov taget den 21 februari (vecka 8) från en patient utan resehistorik. Sekvensen från detta fall klustrar med, samt föregår med en vecka, det första tidigare kända fallet

i Nottingham [7]. Analys av prov från sentinelprovtagning för influensa i Bayern, Tyskland, påvisade inte något fall före det första officiella fallet 27 januari 2020 [8]. Inte heller i USA påvisades dold cirkulation för slutet av februari [9].

2 Molekylärepidemiologisk undersökning av smittspridning till och inom Sverige

2.1 Bakgrund - Molekylärepidemiologi för SARS-CoV-2

För att analysera smittspridningen kan man utifrån virusets arvsmassa studera hur virus från enskilda personer är besläktade med varandra. SARS-CoV-2 arvsmassa består av cirka 30 000 byggstenar (nukleotider) RNA vilka kan kartläggas med sekvensering. När virus förökar sig under en infektion uppstår slumpmässigt mutationer i dess arvsmassa och det är utifrån förekomsten av sådana mutationer som släktskapet mellan enskilda virus kan analyseras med molekylärepidemiologiska metoder. Resultaten från molekylärepidemiologiska undersökningar kan åskådliggöras med så kallade fylogenetiska släktskapsträd och andra metoder.

Liksom andra RNA-virus har SARS-CoV-2 en snabb mutationstakt (evolution) vilket underlättar molekylärepidemiologiska undersökningar. Men trots detta finns det utmaningar med att undersöka den tidiga spridningen av SARS-CoV-2 eftersom samtliga virus har ett gemensamt ursprung som ligger väldigt nära tillbaka i tiden, endast några få månader. Alla mutationer har ju uppstått sedan den tidigaste gemensamma anfadern för SARS-CoV-2 existerade vilket har beräknats till november 2019. Det korta tidspannet gör att arvsmassan för virus från olika länder kan vara identiska eller skilja sig åt med en eller ett fåtal mutationer. Det har rapporterats att tillräcklig genetisk diversitet för fylogenetiska analyser uppnåddes i början av februari 2020 [10]. Reseanamnes är därför ett viktigt komplement till sekvensdata eftersom det kan indikera ett möjligt ursprung för import även om det inte utförts tillräckligt omfattande sekvensering av virus i själva smittlandet. Det ska dock påpekas att reseanamnes kan dock också ha felkällor, t.ex. om en person smittas i Sverige kort efter en utlandsresa och därför felaktigt bedöms vara smittad utomlands.

En annan utmaning med molekylärepidemiologiska undersökningar av spridning av SARS-CoV-2 är att tillgängliga sekvenser inte är helt representativa för smittspridningen i olika länder och över tid. Flertalet fall av SARS-CoV-2-infektion har aldrig diagnostiserats och bara en andel av dessa patienters prov har sekvenserats. I Sverige blev uppskattningsvis en tiondel av alla infektioner som inträffade fram till juni diagnostiserade, och av dessa bekräftade fall finns virussekvens för omkring 3%. Sekvenserna representerar alltså bara en liten andel av alla faktiska infektioner (figur 3). Fylogenetisk analys är beroende att det finns representativa sekvenser. Vilka prov som har sekvenserats är inte heller slumpvis valt och olika länder har sekvenserat olika mycket, t.ex. har det i Europa utförts mycket sekvensering i Danmark och Storbritannien, speciellt under pandemins första fas. Sannolikheten att hitta en speciell virusvariant i ett land är större om det har utförts mycket sekvensering i landet. Det betyder att om en virusvariant har påvisats i t.ex. Danmark eller Storbritannien så kan det spegla skillnader i tillgång på sekvensdata snarare än reella skillnader i förekomsten av virusvarianten. Detta är viktigt att beakta ovanstående utmaningar när resultaten från våra och andras molekylärepidemiologiska undersökningar tolkas.

2.2 Bakgrund - System för typning och klassificering av SARS-CoV-2

Internationellt så används ett antal olika och delvis överlappande klassificeringssystem där virus grupperas i genotyper, de viktigaste systemen är Nextstrain, Pangolin, GISAID och WHO. Enligt Nextclade tilldelas en beteckning (år-bokstav, t.ex. 19A) när en genotyp vid något tillfälle har en frekvens på mer än 20% av globala sekvenser. Pangolin är ett system där nya beteckningar (upp till fyra nivåer av tecken separerade med punkter, t.ex. B.1.1.1 eller A.2) tilldelas när man utifrån vissa fylogenetiska kriterier kan urskilja en genotyp [11]. WHO-beteckningen avser att förenkla kommunikation med allmänheten och betecknar de viktigare genotyperna med en grekisk bokstav (exv. Nextstrain 20I = Pangolin B.1.1.7 = WHO alfa). GISAID-klassificeringen utgår i sin beteckning (bokstäver, t.ex. G, GH) från valda markör-mutationer som karakteriserar en genotyp. Figur 4 visar hur de i rapporten vanligaste varianterna benämns enligt dessa system. WHO-klassificeringen infördes under våren 2021 och är inte applicerbar på sekvenserna som undersökts i denna utredning.

I denna rapport har vi valt att framförallt använda Nextstrain-klassificeringen. Enligt detta system klassificeras merparten av sekvenserna från våren 2020 i de sex genotyperna 19A–B och 20A–D.

2.3 Översikt av undersökta virussekvenser och associerade metadata

När detta uppdrag gavs fanns det knappt 700 svenska sekvenser tillgängliga i den publika internationella databasen GISAID som representerade prov tagna under perioden 31 januari till 1 juni 2020. Den absoluta merparten

av dessa prov hade sekvenserats av FOHM som i sin tur laddat upp dessa data till GISAID. Av dessa sekvenser var 220 från Stockholm. Men hälften av proven från Stockholm var från vecka 10 medan prov från efterföljande veckor endast representerades av endast enstaka sekvenser per vecka. Vi sekvenserade ytterligare prov eftersom vi bedömde att det tillgängliga datasetet var för litet för våra planerade undersökningar. Vår målsättning var att ha tillgång till minst 100 sekvenser per provtagningsvecka under första delen av pandemin. Speciellt var prov från Stockholm underrepresenterade eftersom FOHM hade haft problem att få ut prover från Stockholms Medicinska Biobank. Sekvenseringen av prov från Stockholm kompletterades dels utifrån en lista från Folkhälsomyndigheten på planerade sekvenseringar, samt utifrån jämförelse med data från rutindiagnostiken vid Klinik mikrobiologi, KUL som visade en underrepresentation av äldre patienter. Folkhälsomyndigheten urval baserades på ett antal kluster med epidemiologiska samband, närbesläktade genotyper kan då förväntas. Urvalet på Klinisk Mikrobiologi utgick inte från några kopplingar mellan proven utan valdes utifrån i) tillräcklig virusnivå, ii) provtagningsvecka, iii) patientens ålder, iv) prover med reseanamnes och prover från patienter på äldreboenden.

Analysen i denna rapport omfattar 1 350 svenska virussekvenser. Associerade metadata fanns för en andel av proverna och omfattade information om provtagningsdatum, provtagande sjukhus/enhet, smittland, ålder och kön. Information om smittland har hämtats via utdrag från SMInet där informationen lämnats av behandlande läkare och/eller smittskyddet. Merparten av sekvenserna (62%) kommer från patienter i region Stockholm och därefter Västra Götaland (6,1%), se tabell 1. Andelen sekvenserade fall av bekräftade fall av covid-19 till och med vecka 23 är för Sverige totalt sett 3,0%, på regionsnivå har Halland den högst andel sekvenser av bekräftade fall (7,5%), följt av Stockholm (5,9%) och Värmland (4,7%), se tabell 1. Nästan hälften (45%) av sekvenserna kom från patienter 65 år eller äldre och 25% från patienter 80 år och äldre (tabell 2). En något högre andel av sekvenserna var från män (51%) än kvinnor.

Som jämförelsematerial använde vi samtliga tillgängliga sekvenser i GISAID med provtagningsdatum t.o.m. 31 maj 2020 (n=125 306 sekvenser) som fanns tillgängliga under juli 2021. För den fylogenetiska analysen subsamplades datasetet till att inkludera samtliga sekvenser från Sverige, Danmark, Finland, Island och Norge med en efterföljande subsampling på cirka 1 500 sekvenser vardera av europeiska och utomeuropeiska sekvenser grupperade efter land och månad med prioritet på närbesläktade sekvenser. Det rörde sig om 1 763 sekvenser från Danmark, 757 från Finland, 377 från Norge och 1 447 från Island vilket motsvarar 15%, 11% 4% respektive 80% av de diagnostiserade fallen under tidsperioden i respektive land. Motsvarande siffra för Sverige t.o.m 1 juni 2020 är 3% (1 350 sekvenser och 39 309 bekräftade fall). För övriga länder var sekvenser från Storbritannien (n=36 502) och USA (n=32 848) klart dominerande.

2.4 Översiktligt resultat av typning av svenska SARS-CoV-2 virus.

De 1 350 svenska sekvenserna fördelades i sex genotyper: 19A–B och 20A–D, se figurerna 5 och 6. Redan dessa översiktliga figurer visar att fem av dessa genotyper (19A och 20A–D) påvisades i Sverige redan i mars 2020 dvs tidigt under första vågen av pandemin. Figurerna visar också att huvuddelen av de svenska sekvenserna faller inom genotyperna 20B (29%) och 20C (53%). Trädet och figuren indikerar att smittspridning av dessa två genotyper fortsatte under våren, medan smittspridningen inom genotyperna 19A, 20A och 20D avtog eller helt avstannade. Genotypen 19B påvisades endast hos tre personer i Skåne vilket indikerar mycket begränsad spridning i Sverige av denna genotyp. Nedan undersöker och diskuterar vi de möjliga orsakerna till dessa skillnader och varifrån de olika genotyperna av virus sannolikt importerades.

Smittspridning under andra halvåret 2020 och 2021 berörs egentligen inte i denna rapport men det kan ändå nämnas att inga av de fem(-sex) genotyper som spreds under våren 2020 visade påtaglig fortsatt smittspridning efter sommaren då istället genotyp 20E (B.1.177) med ursprung i Spanien kom att dominera i Sverige och resten av Europa. Genotyp 20E ersattes under början av 2021 av alfa-varianten (20I eller B.1.1.7) som i sin tur ersattes av delta-varianten (21A eller B.1.617.2) under sommaren 2021. Att dessa upprepade skiften av dominerade genotyp skett i Sverige och stora delar av världen visar att nya virusvarianter snabbt kan spridas mellan länder trots resebegränsningar och alla andra åtgärder för att förhindra inflöde och spridning av nya virusvarianter. När de har etablerats i ett nytt land så kommer de framförallt spridas lokalt på samma sätt som skedde med genotyperna 19A–B och 20A–D under våren 2020.

2.5 När och hur många gånger importerades viruset till Sverige?

Som ovan nämnts finns det inga indikationer på närvaro av av SARS-CoV-2 i Sverige före vecka 9 då 13 fall diagnostiserades, förutom importfallet från Kina. Det är troligt att de importer av virus som ledde till dessa tidiga fall skedde samma vecka eller någon vecka innan. För fem av dessa fall finns sekvens. Alla fem tillhör genotyp 19A, ett fall uppges ha smittats i Iran, ett i Sverige och för tre saknas uppgift om smittland. Det är värt att redan här nämna att våra analyser indikerar att import av virus av genotyp 19A från Iran eller andra länder inte påtagligt bidrog till den fortsatta smittspridningen i Sverige under den första vågen (figur 7). Som framgår

av figuren dominerar genotyp 19A i tidiga sekvenser från Iran. Denna genotyp visade begränsad smittspridning i Sverige (se figurerna 5 och 6).

Genom jämförelse med internationella sekvenser kan man identifiera att det skedde introduktion till Sverige av minst 12 genetiskt tydligt skilda virusvarianter till Sverige (se figur 5 och 8). Vissa av dessa introduktioner förefaller ha givit ett begränsat antal sekundärfall, dvs mindre utbrott. Andra introduktioner, fr.a. de inom de genetiska genotyperna 20A, 20B och 20C förefaller ha gett upphov till fortsatt smittspridning i större skala.

Det är mycket troligt att det skedde många fler nästintill simultana introduktioner av virus tillhörande genotyperna 20A, 20B och 20C, men det är mycket svårt att mer exakt uppskatta antalet introduktioner från de fylogenetiska undersökningarna. Som ovan nämnts beror detta på att det var få särskiljande mutationer mellan virus från olika infekterade personer i denna tidiga fas av pandemin samt att tillgängliga sekvenser från Sverige och andra länder inte nödvändigtvis är representativa för alla virus som var i omlopp. En något mer rättvisande uppskattning av antalet importer fås via reseanamnes för identifierade fall. Enbart i vårt material omfattande 1 350 sekvenserade fall finns uppgift om sannolik smitta utomlands för 138 fall (tabell 3). Det är sannolikt att det verkliga antalet fall av utlandssmitta, dvs importer, är många gånger högre. Detta illustreras av en tidig veckorapport från FOHM där man anger utlandssmitta för 88% av drygt 200 fall vecka 10 och 63% av drygt 800 fall vecka 11 [12]. Trots detta, och som vi visar nedan, så kan merparten av fallen under den första pandemivågen i Sverige härledas till fortsatt smittspridning i Sverige av ett mycket mer begränsat antal importerade virus. Det tydligaste exemplet finns inom genotypen 20C där mer än 600 svenska fall i vårt dataset härhör från en virusvariant som sannolikt redan i slutet av februari eller början av mars 2020 importerades till Sverige eller uppstod här (se nedan). I linje med detta indikerar en rapport att det förekommer mer än 1 000 oberoende introduktioner till Storbritannien, främst via resenärer från Spanien, Frankrike och Italien, men bara vissa varianter blev mer generellt spridda [13]. Som vi diskuterar nedan när vi jämför spridningen i Sverige med den i de övriga nordiska länderna så är det inte helt överraskande att ett fåtal tidiga virusimporter dominerar den fortsatta smittspridningen i Sverige och andra länder. Om tidiga gränsstängningar vore helt täta så skulle alla fortsatta fall vara resultat av inhemsk smittspridning.

Det är dock viktigt att återigen påminna om att antalet undersökta sekvenser är mycket litet i relation till det totala antalet fall i Sverige som troligen uppgick hundratusentals personer under den första pandemivågen enligt FOHMs studier av seroprevalens som indikerade att omkring 5% av den svenska befolkningen hade genomgått infektionen under det första halvåret 2020 [14] (figur 3). Vi bedömer dock att vårt sekvensdataset är tillräckligt stort och representativt för att ha spegla de viktigare aspekterna av introduktioner och fortsatt spridning av SARS-CoV-2 i Sverige. Ytterligare en svårighet är att med fylogenetiska metoder uppskatta antalet virusintroduktioner beror på att virus med identisk sekvens kan ha importerats flera gånger vilket är svårt att kartlägga från virussekvenser.

2.6 Från vilka länder skedde viktigare import

En viktig fråga när det gäller den tidiga spridningen av SARS-CoV-2 till och inom Sverige är varifrån virus importerades. Detta gäller speciellt genotyper och varianter av viruset som senare under första vågen utgjorde viktiga källor till fortsatt smittspridning. Det finns flera sätt att undersöka denna fråga. De metoder som vi framförallt använt är: 1) Vilket land som rapporterat sekvensen; 2) Svenska sekvenser med uppgift om var personen smittats (smittland); 3) Kluster med unikt svenska sekvenser. I alla tre fall så kopplas informationen till resultaten från de fylogenetiska analyserna.

I figur 9 visas uppgift om smittland för svenska sekvenser fördelat över de 6 genotyperna 19A–B och 20A–D. Som ovan nämnts så påvisades fem av sex genotyper (19A och 20A–D) i svenska prover tagna redan de första veckorna i mars. Genotyp 19B skiljer sig eftersom den endast har påvisats i tre prov som dessutom togs senare våren (april). Informationen om patienternas smittland finns för 138 av de 1 350 sekvenserade svenska proven (tabell 3). Denna information ger indikation, men inte tvärsäkert svar, på varifrån de virus som sedan spreds i Sverige importerades. Av de 138 sekvenserna med uppgift om utlandssmitta så anges Italien och Österrike som smittland för 66 respektive 38 personer (66%). Dominansen för dessa två länder är ännu större för prov från vecka 10–11. Det finns förstås en selektionsbias i detta eftersom resande från norra Italien, och viss mån Österrike, speciellt uppmanades söka kontakt med vården vid symptom.

Som framgår av figur 9 så finns Österrike angivet smittland för många av tidigaste sekvenserna tillhörande genotyperna 19A och 20C. För 17 av 20 patienterna infekterade med genotyp 20C i Österrike så var proverna tagna under dagarna 9 – 12 mars, dvs tidigt i vecka 11. Med hänsyn till inkubationstid, mm så bör infektionerna ha skett vecka 9 eller vecka 10. Flera internationella forskningsrapporter indikerar att det skedde superspridning av genotypen 20C i Tyrolen och framförallt skidorten Ischgl i slutet av februari och att merparten av den fortsatta spridningen av 20C i Europa och resten av världen går att härleda tillbaka till detta [15, 16, 17].

Förutom de 20 dokumenterade fallen med genotyp 20C med smittland Österrike i vårt svenska material finns liknande rapporter från bland annat Island [16] och Danmark [15]. Grundvarianten av genotypen 20C (se tabell 5 för definierande mutationer) är den vanligaste varianten (n=4 241) i hela databasen av cirka 125 000 internationella sekvenser och totalt tillhör cirka 25 000 sekvenser genotyp 20C. Redan i prover från vecka 10

och vecka 11 återfinns grundvarianten av genotyp 20C i 846 sekvenser från 27 länder, förutom Sverige. Bland svenska sekvenser återfinns grundvarianten hos 15 patienter provtagna 8–31 mars i 7 regioner (Halland, Kalmar, Västra Götaland, Stockholm, Örebro, Sörmland). För tolv av dessa, provtagna 9–16 mars, finns uppgift om smitta i Österrike. Det är värt att notera att varianten initialt inte alls finns påvisad i prov från Skåne som hade sportlov vecka 8. Sammantaget så finns det starka indikationer på att import av virus från österrikiska alperna runt månadsskiftet februari-mars 2020 utgjorde grunden till viktig del av smittspridningen av genotyp 20C, som

dominerade under våren 2020 i Sverige, liksom i Danmark och flera andra europeiska länder. Se även stycke 2.7 för analys av viktiga smittkluster nedan.

Italien finns angivet som smittland för 66 patienter varav 63 provtogs vecka 10 och vecka 11. Merparten av dessa sekvenser finns inom genotyp 20A och 20B, men enstaka tillhör genotyp 19A och 20D (figur 9). Som vi visar nedan så finns det ett par svenska smittkluster inom genotyperna 20A och 20B som med stor sannolikhet kan härledas till import från norra Italien vid månadsskiftet mars–april och som sedan har stått för en betydande del av den fortsatta smittspridningen i Sverige.

Sammantaget kan sägas att mycket talar för att tidig import från norra Italien och Österrike, förmodligen skidorter där, var mycket betydelsefull för att etablera smittspridning i Sverige under våren 2020. Vi ser inga tydliga indikationer på import från andra länder, städer och regioner har varit lika viktig. För att ytterligare belysa betydelsen av import och fortsatt spridning av virusvarianter från österrikiska och italienska alperna vore det värdefullt att analysera reseströmmar från Sverige till olika delar av Europa kopplat till uppskattningar av incidensen av SARS-CoV-2-infektion i rese målen. Arbete med sådana analyser pågår i samarbete med Torsten Perssons forskargrupp på Stockholms Universitet.

2.7 Detaljerad analys av nio smittkluster

Redan tidigt fanns det i Sverige mutationer som inte har påvisats i något annat land globalt (tabell 4). Det finns också flera exempel på kluster med virus som är unikt eller övervägande svenska. I ett fall rör det sig om ett nordiskt, snarare än ett svenskt, kluster. Bakgrunden till dessa kluster är att varianterna med de unika mutationerna antingen har uppstått i Sverige eller att de har importerats och sedan spritts i Sverige, men inte alls eller i mycket liten grad spritts (eller i alla fall sekvenserats) i ursprungslandet och andra länder, se figur 10. Nedan analyserar vi i mer detalj några av viktigaste av dessa kluster. De illustrerar på olika sätt hur SARS-CoV-2 troligen importerades till Sverige och sedan spreds generellt i landet eller mer lokalt. Tillsammans omfattar dessa kluster 788 av de 1 350 (58%) undersökta svenska sekvenserna. Se figurerna S1–S9.

Vi påminner om att antalet analyserade svenska sekvenser är lågt i förhållande till det totala antalet fall under studieperioden (1 350 sekvenser av uppskattningsvis upp mot 500 000 fall; 0,3%) (figur 3). Om vi hade haft tillgång till sekvens från större andel av det totala antalet fall är det sannolikt att de största klustren hade blivit ännu större och att vi funnit fler unikt svenska kluster av mindre storlek. Det vill säga det hade förstärkt, snarare än försvagat, mönstret vi beskriver där en stor andel svenska fall kan kopplas till tidigt etablerad inhemsk spridning utgående från en mycket liten andel av alla importerade varianter.

2.7.1 Kluster 20C:G24368T med 636 svenska sekvenser

Det största svenska, eller snarare nordiska, klustret finns i genotyp 20C och karaktäriseras av mutationen G24368T som leder till aminosyra-substitutionen D936Y i spike-genen. Det finns 636 svenska sekvenser av genotyp 20C med denna mutation och 1 017 utländska. Bland dessa sekvenser finns det 183 svenska och 227 internationella sekvenser (varav 86 brittiska och 82 finska) med enbart denna mutation relativt grundformen av 20C, men de flesta har 1 – 10 ytterligare mutationer. I det svenska sekvensmaterialet tillhör 636 av 1 350 sekvenser (47%) detta kluster. Dessa prov har tagits i 18 av landets regioner under perioden 8 mars till 1 juni vilket visar på stor spridningen vad gäller antal, geografi och tid.

Förutom de 636 svenska sekvenserna tillhörande denna variant så finns det alltså enbart ytterligare 1 017 sekvenser med denna varianten i det internationella referensdatasetet av cirka 125 000 sekvenser (0,8%). Detta visar att varianten är klart överrepresenterad i Sverige och förefaller utgöra en viktig del av smittspridningen i landet under våren 2020. Utöver Sverige, finns varianten i 415 av 757 (55%) av sekvenser från Finland, 60 av 377 (16%) av sekvenser från Norge, 34 av 1 763 (2%) av sekvenser från Danmark och 423 av 36 502 (1%) av sekvenser från Storbritannien. Resterande 90 sekvenser utgör enstaka sekvenser från 22 olika länder. Med reservation för skevheter i sekvensurvalet så talar detta för att G24368T-varianten spreds i stor omfattning i Sverige och Finland, måttlig omfattning i Norge och i låg omfattning i Danmark och Storbritannien (figur 11). Den dominerande och över tid ökande förekomsten av 20C:G24368T-varianten i Sverige och Finland har rapporterats tidigare [18, 19, 20, 21]. Det finns en rapport som indikerar att mutationen D936Y skulle ha genomgått positiv selektion i Sverige [20], vilket föreslås kunna bero på att den skulle kunna påverka struktur och funktion hos spike-proteinet [21]. Men en annan artikel lyfter fram att epidemiologiska faktorer, såsom superspridningshändelser, är en alternativ förklaring [19]. Det faktum att varianten påvisats i Danmark och Storbritannien utan att komma dominera där talar emot att den skulle ha en ökad smittsamhet jämfört med andra varianter av genotyp 20C.

I och med 20C:G24368T-variantens stora betydelse för smittspridningen i Sverige (och Finland) under våren 2020 så är det intressant att försöka klarlägga dess ursprung och fortsatta spridningdynamik. Det tidigaste provet i världen med denna sekvensvariant är taget i Sörmland 2020-03-08 (vecka 10) med uppgiven smittort Eskilstuna/Stockholm, men redan vecka 11 (9–15 mars) återfinns varianten i ytterligare 18 prov från Stockholm och 1 från Örebro, intressant nog finns i SMInet uppgift om att alla bedöms vara smittade i Sverige. Givet bara en liten andel av alla fall har sekvenserats så bör sekvenserade fall med inhemsk smitta rent statistiskt ligga åtminstone två–tre steg in i en smittkedja. Eftersom inkubationstiden (eller mer korrekt det seriella intervallet) är minst 3–5 dagar så innebär detta att 20V:G24368T-varianten bör ha importerats (eller uppstått i Sverige) cirka två veckor tidigare, dvs omkring vecka 8 – vecka 9.

Den första anamnesuppgiften om att en svensk patient infekterad med varianten skulle ha smittats utomlands är en patient provtagen i vecka 12 (Tanzania). I vecka 13 finns ytterligare 5 patienter (en per Albanien, Indonesien, Israel, samt två i Schweiz) och i vecka 14 en patient (Spanien). Vi bedömer att det är osannolikt att dessa anamnesuppgifter stämmer eftersom 20C:G24368T-varianten inte alls har rapporterats från Albanien, Indonesien, Israel, eller Schweiz, och bara en gång från Spanien. Eftersom det pågick omfattande spridning av varianten i Sverige så förefaller mer troligt att dessa patienter har smittats i Sverige i nära anslutning till en utlandsresa. I det internationella dataset saknas uppgift om smittland för 20C:G24368T sekvenser med undantag för 4 sekvenser från Island där Storbritannien är uppgivet som smittland. Men dessa prover togs 19 mars – 1 april vilket gör att anamnesuppgiften inte säger så mycket om variantens ursprung eftersom den då redan var spridd till flera länder. Via smittland har vi alltså ingen klar indikation på varifrån eller hur många gånger 20C:G24368T-varianten importerades till Sverige.

Under vecka 11 rapporterades 20C:G24368T-varianten totalt 13 gånger från andra länder i världen; 6 fall i Storbritannien (med 4 första fall 9 mars), 4 fall från Danmark (med första fall 9 mars), 1 fall från vardera Norge, Saudiarabien, och USA (med första fall 14 mars). Således står Sverige för 20 av de 32 tidigaste fallen av infektion med varianten, dvs under vecka 10 och vecka 11. Under den resterande delen av mars 2020 finns varianten i 241 sekvenser från Sverige och 157 sekvenser från resten av världen. I Finland, som senare hade många fall infekterade med varianten, provtogs det första fallet 23 mars och återfinns endast 23 gånger under mars. Det kan nämnas att ett pre-print som studerar introduktion och spridning av SARS-CoV-2 i Finland, inklusive 20C:G24368T-varianten, anger att det är svårt att tillförlitligt bestämma variantens ursprung [19].

Figur 10 visar tre möjliga scenarion för hur 20C:G24368T-varianten kan ha nått Sverige: A) Varianten uppstod i Österrike och exporterades därifrån till Sverige och andra länder; B) Varianten uppstod i Österrike, men nådde Sverige via ett annat land; C) Varianten uppstod i Sverige och spreds därifrån till andra länder. Vi bedömer att det troligaste scenariot är att 20C:G24368T-varianten uppstod i österrikiska alperna eftersom denna region utgjorde ett nav för superspridning av genotyp 20C vid månadsskiftet februari–mars 2020 [17]. Det som också talar för detta scenario är att 20C:G24368T-varianten nästintill samtidigt påvisades i Sverige, Storbritannien, Danmark och Norge, varifrån sannolikt många skidturister befann sig i Ischgl och andra orter i Tyrolen. Det som talar emot scenariot är att varianten inte påvisats i prover från Österrike och att Österrike inte finns angivet för smittland för några av de tidiga 20C:G24368T-fallen. Antalet sekvenser av genotypen 20C från Österrike från veckorna 10–12 är dock lågt (1, 51 och 33 sekvenser i respektive vecka), vilket kan jämföras med drygt 300 20C-sekvenser från Danmark i veckorna 10–11. Ett lågt antal sekvenserade prov i Österrike kan således vara en orsak till att varianten 20C:G24368T inte har påvisats där. I detta sammanhang kan noteras att de österrikiska sekvenserna av genotyp 20C från vecka 11 uppvisar genetisk diversitet med sekvenser som skiljer sig med en, två respektive tre mutationer från grundvarianten av 20C. Denna diversitet indikerar dels att 20C-spridningen vid denna tid pågått ett par veckor i Österrike och dels att det troligen fanns ytterligare varianter som cirkulerade där men inte återfinns bland de få tillgängliga österrikiska sekvenserna. Motsvarande situation finns för varianterna 20A:G22583A och 20A:A12790G som med stor säkerhet importerades till Sverige från Italien, men ändå saknas bland sekvenser från Italien (se nedan). Scenario B (smitta till Sverige via ett annat land än Österrike) bedömer vi mindre sannolikt eftersom det förutsätter att varianten redan hunnit etablera sig i detta tredje land och spridas till Sverige till slutet av vecka 10. På samma sätt bedömer vi scenario C (varianten uppstod i Sverige) mindre sannolikt eftersom den påvisades i Danmark och England nästan samtidigt som i Sverige. Vi har övervägt att matematiskt modellera sannolikheten för dessa alternativa scenarion, men har avstått eftersom flera viktiga parametervärden är väldigt osäkra.

Vår bedömning om ett troligt österrikiskt ursprung för 20C:G24368T-varianten skiljer sig från rapporter från Folkhälsomyndigheten [18] och Danmark [15] där man föreslår att varianten importerats från andra länder i Europa respektive Storbritannien, vilket vi finner mindre sannolikt, men inte uteslutet, givet ovanstående resonemang. Om varianten uppstod i Österrike så förefaller troligt att det skedde, eller i alla fall fick spridning, bland resenärer från Skandinavien och Storbritannien, såsom en bar, ett hotell eller en transitbuss. Spridningen till Finland har sannolikt till betydande del skett från Sverige, givet att fallen i Sverige föregick de första finska med cirka två veckor, men det kan också tänkas att de tidigaste fallen infekterades med varianten saknas i det finländska sekvensdatasetet. Ett liknande scenario är möjligt för Norge.

2.7.2 Kluster 20A:C14195A med 39 svenska sekvenser

Det näst största klustret återfinns i genotyp 20A och kännetecknas av mutationen C14195A. Klustret innehåller 39 svenska sekvenser och enbart 5 sekvenser från andra länder (Costa Rica 1, Danmark 1, Ryssland 2, Storbritannien 1). Globalt var de tre första proverna med denna genotyp tagna i Sverige (Västra Götaland 11 och 16 mars, samt Halland 17 mars). Resterande svenska prov var tagna i nio olika regioner (Västra Götaland, Halland, Stockholm, Västerbotten, Jämtland Härjedalen, Norrbotten, Värmland, Kalmar, Skåne) under perioden 20 mars – 25 maj. Varianten spreds alltså till stora delar av landet under större delen av våren 2020. För det svenska provet taget 16 mars finns uppgift om att smitta skett i Storbritannien, medan uppgift om smittland saknas eller uppgitt ha skett i Sverige för övriga prover. Även uppgiften om att ett svenskt fall skulle kunna smittats i Storbritannien ska tas på allvar så är det anmärkningsvärt att genotypen endast setts i 1 av mer än 36 000 sekvenser från Storbritannien tagna under vintern – våren 2020. I ljuset av detta förefaller troligt att mutationen som karaktäriserar varianten uppstått i Sverige eller mycket tidigt importerats från ett annat land än Storbritannien. Oavsett vilket, så kan misstankarna riktas mot norra Italien eftersom vi vet att genotyp 20A spreds tidigt där och att flera av de tidiga fallen i Sverige med genotyp 20A uppges ha smittats i Italien. I sammanhanget ska det noteras att det i vårt dataset om 125 000 internationella sekvenser enbart finns 1 409 sekvenser från Italien för hela perioden januari till maj 2020. Redan när det första provet med varianten 20A:C14195A togs i Västra Götaland den 11 mars så hade Italien rapporterat mer än 10 000 sjukdomsfall och det verkliga antalet fall var säkerligen mycket högre.

2.7.3 Kluster 20A:G22583A med 23 svenska sekvenser

Inom genotyp 20A finns även ett kluster som kännetecknas av mutationen G22583A. Klustret har likheter med det föregående i det att det innehåller 23 svenska sekvenser och enbart 5 sekvenser från andra länder (alla från Storbritannien). Tjugo av de 23 svenska prover tagna 3–9 mars, medan det första från Storbritannien var taget 12 mars. Femton svenska prover var tagna i Stockholm 3–8 mars, för dessa prov finns ingen uppgift om smittland. Sex prover var tagna i Västernorrland den 9 mars och alla dessa uppges vara smittade i Italien. Två senare prov var tagna i Västra Götaland 27 mars respektive Uppsala 12 april vilket indikerar att varianten fortsatte att cirkulera i Sverige efter introduktionen i början av mars. Anamnesuppgifterna för fallen i Västernorrland indikerar att varianten uppstod i Italien och sedan importerades till både Västernorrland och Stockholm. De senare fallen i Storbritannien skulle också kunna representera importer från Italien eller möjligen från Sverige.

2.7.4 Kluster 20B:C26447T+C28887T med 25 svenska sekvenser

Nästa kluster finns inom genotyp 20B och kännetecknas av mutationerna G22583A och C28887T. Klustret innehåller 25 svenska sekvenser provtagna 31 mars – 1 juni över större delen av landet (Stockholm, Västernorrland, Uppsala, Jämtland Härjedalen, Jönköping, Värmland, Skåne, Kalmar, Västmanland). Det finns enbart ett utländskt prov med denna variant; i ett dansk prov taget 18 maj som troligen representerar import från Sverige. För ungefär hälften av proverna finns uppgift om att smitta ska ha skett i Sverige medan uppgift saknas för resterande prov. Den fylogenetiska analysen indikerar att varianten uppstått i två steg. Den första mutation skulle i så fall vara C26447T (mot bakgrunden C241T,C3037T,C14408T,A23403G,G28881A,G28882A,G28883) som finns påvisad i 5 av de 125 000 internationella sekvenserna och dessutom i ett svenskt prov taget i Jämtland- Härjedalen 23 mars från en patient som uppges vara smittad i Sverige. Om uppgiften om smittland för detta fall stämmer så betyder detta att C26447T-varianten varit i åtminstone begränsad cirkulation i landet. Varifrån C26447T-varianten har importerats till Sverige är svårt att bedöma i och med att det första och enda svenska fallet diagnostiserades 23 mars då SARS-CoV-2 var spritt över hela Europa och världen. Däremot är det troligt att tilläggsmutationen C28887T har uppstått i Sverige. Därefter har sedan varianten med de två kännetecknande mutationerna C26447T+C28887T spritts ganska brett i Sverige under våren och försommaren. Från Sverige har den exporterats till Danmark vid minst ett tillfälle.

2.7.5 Kluster 20A:A12790G med 19 svenska sekvenser

Detta kluster inom genotyp 20A är ett exempel på en variant som tidigt importerades till Värmland från Italien och spreds lokalt under en begränsad tid (4–12 mars 2020) utan påvisad fortsatt spridning lokalt eller andra delar av landet. Nio av de infekterade personerna uppges vara smittade i Italien, åtta i Sverige och för två saknas uppgift. Inga av de 125 000 internationella referenssekvenserna uppvisar den kännetecknande mutationen A12790G inom genotyp 20A. Det är överraskande att nio svenska personer infekterade med varianten uppges vara smittade i Italien och samtidigt det inte finns någon internationell sekvens med denna variant. Det pekar mot att varianten inte hade bred spridning i Italien. Man kan spekulera i möjligheten att de personer från Värmland som uppges vara smittade i Italien tillhörde samma resesällskap och att en person i sällskapet förde in infektionen i gruppen. Vi har dock inte tillgång till några detaljerade anamnesuppgifter som skulle kunna bekräfta eller förkasta en sådan hypotes. En ytterligare kommentar är att detta kluster visar att ett lokalt utbrott

(som detta i Värmland) kan stoppas så att bredare spridning i landet förhindras. Vi vet inte om detta skett genom aktiva smittskyddsåtgärder eller om det delvis är en slump att detta lokala utbrott snabbt begränsades.

2.7.6 Kluster 20B:C24865T med 16 svenska sekvenser

Detta kluster innehåller 14 svenska sekvenser. Fjorton sekvenser har erhållits från prov tagna i Stockholm 5 april – 28 maj, medan ett prov är från Västerbotten 15 maj och ett från Uppsala 20 maj. Ingen person uppges vara smittad utomlands, men sex i Sverige. Bland de 125 000 internationella sekvenserna så finns C24865T-mutationen i 10 st genotyp-20B-sekvenser, men ingen av dessa matchar den fullständiga genotypen för de svenska sekvenserna, vilket vi tolkar som att de inte tillhör det svenska klustret. Det går inte att säkert bedöma om varianten uppstått i Sverige eller importerats, men våra analyser visar att varianten spritts i Stockholm under april–maj och även påvisats i Västerbotten och Uppsala.

2.7.7 Kluster 20B:G1421A med 13 svenska sekvenser

Detta kluster innehåller 13 svenska sekvenser. Elva sekvenser har erhållits från prov tagna i Östergötland under perioden 12 mars – 5 maj, medan ett prov är från Kalmar 15 april och ett från Stockholm 30 april. Det finns inte uppgift om utlandssmitta för något av de svenska fallen, men 8 uppges vara smittade i Sverige och 5 saknar uppgift. Bland de 125 000 internationella sekvenserna så är G1421A-varianten endast påvisad 3 gånger (Storbritannien 30 mars och 14 april samt Bosnien-Herzegovina 25 mars). Sammantaget så bedömer vi att detta är en svensk variant som fr.a. spritts i måttlig omfattning lokalt i Östergötland under våren. Det går inte att säkert bedöma om varianten importerats till Östergötland eller om den uppstått där. Det går inte heller att bedöma om G1421A-mutationen i de tre utländska fallen representerar spridning från Sverige eller om mutationen oberoende har uppstått i dessa tre fall.

2.7.8 Kluster 20B:C9857T med 10 svenska sekvenser

Detta kluster innehåller 10 svenska sekvenser som är helt identiska. Alla 10 prover är tagna i Stockholm under dagarna 2–6 mars. Uppgift om smittland saknas för alla fallen. Bland de 125 000 internationella sekvenserna så finns tre prov från Österrike som har identisk sekvens med de svenska sekvenserna. Dessa prov är tagna 6, 13, och 28 mars. Det är troligt men inte säkerställt att varianten importerats till Sverige från Österrike. En alternativ förklaring skulle kunna vara att importerats till både Sverige och Österrike från Italien, som var en hub för tidig spridning av genotypen 20B. Det finns ytterligare 12 av de internationella genotyp 20B-sekvenserna som har C9857T-mutationen men som sekvensmässigt skiljer sig lite från det svensk/österrikiska klustret. Dessa sekvenser är från Storbritannien, Brasilien, Frankrike och Australien. Det är tänkbart att dessa infektioner också skulle kunna ha sitt ursprung i Österrike eller Italien.

2.7.9 Kluster 20B:C5654T+C24034T med 7 svenska sekvenser

Det sista kluster som vi valt att presentera innehåller 7 svenska sekvenser och en belgisk. De svenska proverna är tagna i Stockholm under dagarna 7–10 mars. Det belgiska provet var taget 6 mars. Det finns inga andra av de 125 000 referenssekvenserna som har denna genotyp. För ett av proven finns uppgift om smitta i Italien, medan uppgift om smittland saknas för övriga fall. Det är troligt att varianten importerats en eller flera gånger till Sverige från Italien. Det är intressant att denna variant inte förefaller att ha fortsatt att spridas i Sverige eller någon annan del av världen under den första pandemivågen.

2.8 Jämförelse av virusimport och spridning i de nordiska länderna.

I och med att vi har alla tillgängliga sekvenser från de nordiska länderna (Danmark, Finland, Island, Norge, Sverige) med i analyserna så kan vi översiktligt jämföra de molekylärepidemiologiska mönstren mellan länderna med reservation för olikheter i urval och andel sekvenserade prov av alla diagnostiserade fall där det finns betydande skillnader. Under den studerade tidsperioden t.o.m. 1 juni 2020 finns för Danmark sekvens för 1 763 av 11 633 rapporterade fall (15%), Finland 757 av 6 826 fall (11%), Norge 377 av 8 411 fall (4%), Island 1 447 av 1 806 fall (80%), samt för Sverige 1 350 av 39 309 fall (3%).

Figur 12 visar att det finns sekvenser från prover tagna i slutet av februari och början av mars i fyra av länderna, medan antalet är lite lägre för Finland. Precis som för Sverige finns rapporter om tidiga importer av virus från Italien och Österrike till Danmark och Island [15, 16]. Trots skillnaderna i total sjukdomsbörda i de fem länder så finns det många likheter i de fylogenetiska mönstren. Det gäller framförallt att en förhållandevis stor andel av sekvenserna återfinns i land-specifika kluster som etablerades i början av mars och sedan fortsatte att spridas i respektive land under resten av våren. Mönstret är allra tydligast för Danmark och Island där det finns flera stora land-specifika kluster inom genotyperna 20A och 20C respektive 19A, 20A och 20B (figurer 13, 14). För Finland är mönstret något annorlunda eftersom många finländska sekvenser, som ovan nämnts,

förefaller ha sitt ursprung i det stora svenska G24368T-klustret inom genotyp 20C, men det finns också ett par kluster som förefaller domineras av finska sekvenser (figur 15, 16A). Även i Norge är bilden med land-specifika kluster inte lika tydlig som i de övriga nordiska länderna (figur 16). Av de danska sekvenserna återfinns 41% inom tre större kluster (20C:C1302T, 20C:G26951A, 20A:C7834T) och bland de isländska sekvenserna finns ett dominerande kluster som utgör 21% av sekvenserna (20A:A10323G) och ett större på 9% (19A:C5142T). Bland de norska sekvenserna ses ett större kluster (20A:G8371T, 11% av norska sekvenser) medan det största finska klustret utgör 5% av finska sekvenser (20A:C6401T).

De spridningsmönster som indikeras av de fylogenetiska analyserna är egentligen inte överraskande. Vid månadsskiftet februari-mars var den omfattande smittspridningen i norra Italien, inklusive skidorter, och de österrikiska skidorterna ännu inte fullt uppmärksammat. Detta ledde till ett mer eller mindre okontrollerat inflöde av smittade personer vars antal troligen var relativt väl korrelerat till resemönster. De olika nordiska länderna införde sedan smittskyddsåtgärder såsom karantän för återvändande resenärer, stängda landsgränser, och nedstängning av samhället i olika grad och takt. Även om sådana åtgärder inte är 100% effektiva så är det förväntat att den fortsatta smittspridningen domineras av virusvarianter som redan etablerat lokal spridning, vilket är vad de land-specifika klustren indikerar. Omfattningen av denna fortsatta smittspridning kommer mycket att bero på hur effektiva åtgärderna mot fortsatt lokal smittspridning är. Med öppna gränser förväntas ett visst tillskott av nya importerade varianter, men även i Sverige, som hade minst strikta åtgärder, så förefaller dessa stå för mindre än 50% av fallen. Detta gäller i alla fall under våren 2020 då varianterna inom genotyp 20A–C sannolikt hade ungefär samma smittsamhet. I linje med detta resonemang finns en forskningsrapport som indikerar att det i europeiska länder under första halvåret 2020 finns en omvänd korrelation mellan tidiga åtgärder mot smittspridning och mortalitet [22]. Det ska dock påpekas att import av virusvarianter har stor betydelse om de är påtagligt mer smittsamma vilket med all önskvärd tydlighet illustreras av alfa- och delta-varianterna som snabbt kom att dominera i såväl Sverige som andra europeiska länder trots att de inte uppstod i lokalt utan importerades och ersatte tidigare varianter.

2.9 Smittspridning inom Sverige

Som ovanstående analys av de utvalda klustren visar så finns exempel på smittspridning som var begränsad i tid och rum, dvs lokala utbrott såsom kluster 20B:G1421A i Östergötland och kluster 20A:A12790G i Värmland. Men de flesta av de större klustren har fick spridning i stora delar av landet. Det gäller fr.a. det stora klustret 20C:G24368T som också spreds till och i Finland. Dessa mönster är inte överraskande i och med att det inte funnits strikta restriktioner avseende resande inom Sverige. Trots detta skiljde sig omfattningen av smittspridning under första pandemivågen påtagligt mellan olika svenska regioner, t.ex. mindre i Skåne än i Stockholm. Orsaken till dessa skillnader är dock svåra att utreda via de molekylärepidemiologiska undersökningarna.

2.10 Analys av smittspridning i region Stockholm

2.10.1 Skillnader i fördelning av virusgenotyper mellan Stockholmssjukhusen

För prov tagna i region Stockholm har vi information om på vilken vårdinrättning provet var taget på och vi kunde därför analysera eventuella skillnader vad gäller sekvenstyper mellan vårdinrättningar. För prov tagna på akutsjukhusen noteras att prov från Sankt Görans sjukhus nästan uteslutande utgörs av virus tillhörande genotyp 20C. 20C utgör också majoriteten i prov från Danderyd sjukhus samt från Karolinska Universitetssjukhuset Solna (figur 17). En stor andel av dessa sekvenser återfinns inom det stora klustret 20C:G24368T (135 av 147 på Sankt Görans sjukhus, 52 av 84 på Danderyds sjukhus, samt 29 av 42 på Karolinska Universitetssjukhuset Solna). Dessa fynd indikerar att det föreligger vissa regionala skillnader i smittspridningen inom region Stockholm så att genotyp 20C dominerat i upptagningsområdet för Sankt Görans sjukhus (dvs nordvästra Stockholm), medan andra genotyper fått begränsad spridning i detta område. Karolinska Universitetssjukhuset Huddinge dock utgjorde initialt regional vårdavdelning för covid-19, dvs kopplingen till upptagningsområde blir inte lika tydlig då patienter från alla upptagningsområden blev hänvisade dit. Men faktum kvarstår att under resten av våren så var genotyp 20C extra dominerande bland prover från Sankt Görans sjukhus. Som ovan nämnts så utgjorde österrikiska alperna ett nav för tidig spridning av denna genotyp även om ursprunget för den dominerande 20C:G24368T-varianten inte är helt klarlagd.

Fördelat över tid påvisades genotyp 20C i region Stockholm första gången vecka 11. De alla första fallen i vecka 9 utgjordes av 19A och 20A, medan 20B dominerade i vecka 10 (figur 18). Detta visar att det inte verkar ha varit de alla tidigaste introduktionerna i regionen som kom att dominera den fortsatta smittspridningen under första pandemivågen. Att det skulle ha pågått någon påtaglig smittspridning av virus i genotyp 20C före vecka 10–11 som undgått upptäckt stöds inte av analyserna av tidigare ej undersökta luftvägsprov (se stycke 1.2).

2.10.2 Sekvenser från tidigare okända fall

Som ovan redovisats utförde vi SARS-CoV-2 PCR på knappt 2 000 luftvägsprov som tidigare inte genomgått denna diagnostik. Totalt upptäcktes enbart 13 stycken nya fall varav 9 kunde sekvenseras (fem genotyp 20C, två 20A, ett 20B och ett 19A) (figur 2). Det tidigaste fallet var provtaget i vecka 10 och detta virus tillhör genotyp 20A. En patient med virus i genotyp 20C provtogs vecka 11 då det redan fanns 47 kända patienter infekterade med virus ur genotyp 20C. Detta indikerar att det inte förekom någon påtaglig dold/oupptäckt smittspridning av genotyp 20C i samhället innan de första fallen diagnostiserades. För de sekvenserade nyupptäckta fallen har vi ingen information om att de skulle ha smittats utomlands och de bedöms därför troligen ha smittats i Sverige. Att 5 av 9 typbara virus hos de nyupptäckta patienterna tillhör genotyp 20C och tre av dessa detekterades vecka 12 eller senare styrker att det huvudsakligen var denna variant som cirkulerade i samhället, snarare än att 20C dominerar i materialet på grund av riktad testning. Det ska också nämnas att alla fem 20C-fallen tillhörde det stora 20C:G24368T-klustret vilket ytterligare bekräftar dess dominerande roll i den svenska smittspridning i och med att dessa nyupptäckta fall sannolikt provtogs på grund av luftvägssymptom utan specifik misstanke om SARS-CoV-2-infektion eller reseanamnes.

2.10.3 Unikt svenska varianter

Det finns 151 sekvenser från prov tagna i vecka 10. Av dessa har 54 sekvenser (36%) en kombination av mutationer som förekommer unikt i Sverige, dvs saknar identiska motsvarigheter i det internationella datasetet om 125 000 sekvenser. Andelen unikt svenska sekvenser ökade under våren och uppgick vecka 13 till 57% och vecka 14 till 73%. Detta ger ytterligare stöd för att smittspridningen redan i tidigt skede dominerades av inhemsk smitta snarare än kontinuerlig import från utlandet. Viss försiktighet krävs dock när dessa fynd tolkas eftersom både det svenska datasetet och det internationella jämförelsematerialet bara omfattar en liten del av alla SARS-CoV-fall under studieperioden. Dessutom är antalet särskiljande mutationer få och mindre delar av virusets arvsmassa kan sakna sekvens (det internationella kvalitetskravet anger att minst 95% av viruset arvsmassa ska vara bestämd). Vi bedömer dock att jämförelsen av svenska och internationella sekvenser är tillräckligt robust för att ge en bra bild av andelen unikt svenska sekvenser.

2.11 Åldersgrupper

Information om ålder fanns för 1 338 av 1 350 patienter med sekvenserat virus (tabell 2). De flesta sekvenserade virus kom från patienter i åldrarna 25-64 år (20A: 51%, 20B: 52% samt 20C: 40% av påvisade virus). Virus i genotypen 20C var vanligare hos äldre patienter. Hos personer >65 år utgjorde genotyp 20C 67% av alla sekvenser, medan 20C endast återfanns hos 14% av personer <25 år. På motsvarande sätt utgjorde unga vuxna (15-24 år) 8,7% och 9,8% av patienterna med virus i genotyperna 20A respektive 20B, men endast 1,3% av patienterna med virus i genotyp 20C, medan patienter 65 år och äldre utgjorde 34% vardera av patienterna med virus i genotyp 20A respektive 20B men 57% av patienterna med virus i grupp 20C. Det ska dock påpekas att det skedde ett skifte vad gäller dominerade genotyp (20C ökade vartefter) samtidigt som indikationerna för PCR-testning ändrades under den första vågen. Initialt testades alla misstänkta fall och kontakter oavsett ålder. Senare testades fall som krävde sjukhusvård, dvs företrädesvis äldre. Detta kan ha lett till en viss bias i åldersfördelning vad gäller virusgenotyper eftersom andelen provtagning och därmed sekvensering av yngre personer minskade under våren.

2.12 Varianter på äldreboenden

I och med överrepresentationen av dödsfall på äldreboenden (här avser vi mer precist särskilt boende samt vård- och omsorgsboende) har vi separat analyserat molekylärepidemiologin för sekvenser från personer i denna grupp. Analyserna har bara kunnat göras för Stockholm där vi hade tillgång till information om vilken vårdinrättning som sänd in provet. Totalt har vi tillgång till 161 sekvenser från personer provtagna på 88 vård- och omsorgsboenden, äldreboenden samt servicehus. Från 42 boenden fanns minst två sekvenser. För de flesta av dessa 42 boenden så var sekvenserna närbesläktade vilket är förenligt med smitta inom boendet. Men på sex boenden tillhörde sekvenserna olika genotyper (figur 19), vilket betyder att minst två oberoende virusintroduktioner hade inträffat. Det ska dock påpekas att förekomst av två eller flera identiska sekvenser på ett boende inte formellt utesluter oberoende importörer även om lokal smittspridning får anses troligare.

Av de 161 sekvenserna från personer på äldreboenden så var 84 sekvenser unika, dvs återfanns varken bland internationella eller övriga svenska sekvenser. Bland prov från äldre- och omsorgsboende tillhörde 119 av 161 (74%) genotyp 20C, vilket kan jämföras med det totala datasetet där 714 av 1350 (53%) av sekvenserna tillhörde denna genotyp. Andelen av 20C sekvenserna var också högre på äldre- och omsorgsboendena än generellt i åldersgrupperna 65-79 år och ≥80 år där genotypen 20C utgjorde 66% respektive 67% av sekvenserna (tabell 2). Inom genotypen 20C hade 94 av sekvenserna från äldre- och omsorgsboenden mutationen G24368T. Denna fördelning är inte helt överraskande eftersom sjukdomsfallen inom äldreboenden tidsmässigt släpade efter den mer allmänna samhällsspridningen och genotyp 20C gradvis fick en alltmer dominerande roll.

Resultaten ger dock stöd för uppfattningen att den allmänna samhällsspridningen var den huvudsakliga källan till spridning inom äldreården, dvs att det inte var några speciella virusvarianten som spreds inom äldreården.

2.13 Smittsamhet

Under pandemin har det flera gånger visats att virusvarianter med mutationer som ger ökad smittsamhet snabbt har kommit dominera smittspridningen genom att ersätta tidigare varianter. De tydligaste exemplen är den brittiska alfa-varianten (Nextstrain-genotyp 20I/Pangolin B.1.1.7) som under början av 2021 helt ersatte tidigare varianter (fr.a. varianten 20E/B.1.177). Alfa-varianten ersattes i sin tur av delta-varianten (21A/B.1.617.2) under sommaren 2021.

Den första mutationen som visades öka smittsamheten var dock spike-mutationen D614G [23]. Denna mutation är det som framför allt skiljer genotyperna 19A–B från genotyperna 20A–D; de senare har mutationen medan de förra inte har den. Det är därför inte överraskande att genotyperna 19A–B fick begränsad smittspridning i Sverige (liksom i många andra delar av världen) under den senare delen av första pandemivågen. Genotyperna 19–B trängdes helt enkelt undan av de mer smittsamma genotyperna 20A–D, precis på samma sätt som alfa- och delta-varianterna senare gjorde. Det finns dock inga data som tyder på att genotyp 20C skulle vara mer smittsam än 20A, 20B och 20D. Dominansen av 20C beror således troligen på andra faktorer och då sannolikt framförallt på att varianten 'av en slump' hamnade i rätt miljö/population vid rätt tillfälle och därför fick en spridningsfördel (en variant av det som kallas *founder effect* på engelska).

3 Tillkännagivanden

Vi vill tacka alla som bidragit till rapporten, i synnerhet vill vi tacka Richard A. Neher (Basels Universitet) och Emma B. Hodcroft (Berns Universitet) för vägledning i utförandet de fylogenetiska analyserna och för kritisk granskning av utkast på rapporten; Malin Grabbe (KUL) för uppslaget till undersökningen av sparade luftvägsprov; Tanja Normark och Isak Sylvén (KUL, SciLifeLab) för utformningen av den bioinformatisk bearbetning av sekvensdata; samt all personal utvecklat, lett och utfört arbetet med diagnostiska PCR-analyser och sekvensering, i synnerhet Niklas Svedberg, Natalija Gerasimcik, Lynda Eneh, Martin Ekman och Karina Hentrich på KUL samt Sandra Broddesson på KI. Till sist vill vi också rikta vårt tack till alla i det vetenskapliga samfundet globalt som möjliggjort våra analyser genom att dela sina sekvensdata på plattformen GISAID.

Referenser

- [1] A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, C. Alloui, E. Carbonnelle, J. Zahar, S. Brichtler, and Y. Cohen, "SARS-CoV-2 was already spreading in France in late December 2019," *International journal of antimicrobial agents*, vol. 55, no. 6, p. 106006, 2020.
- [2] A. Amendola, S. Bianchi, M. Gori, D. Colzani, M. Canuti, E. Borghi, M. C. Raviglione, G. V. Zuccotti, and E. Tanzi, "Evidence of SARS-CoV-2 RNA in an oropharyngeal swab specimen, Milan, Italy, early December 2019," *Emerging Infectious Diseases*, vol. 27, no. 2, p. 648, 2021.
- [3] G. La Rosa, P. Mancini, G. B. Ferraro, C. Veneri, M. Iaconelli, L. Bonadonna, L. Lucentini, and E. Suffredini, "SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring," *Science of the total environment*, vol. 750, p. 141711, 2021.
- [4] Folkhälsomyndigheten, "Bekräftade fall i Sverige – daglig uppdatering," 2021.
- [5] F. S. Dawood, P. Ricks, G. J. Njie, M. Daugherty, W. Davis, J. A. Fuller, A. Winstead, M. McCarron, L. C. Scott, D. Chen, *et al.*, "Observations of the global epidemiology of covid-19 from the prepandemic period using web-based surveillance: a cross-sectional analysis," *The Lancet Infectious Diseases*, vol. 20, no. 11, pp. 1255–1262, 2020.
- [6] J. Björk, K. Mattisson, and A. Ahlbom, "Impact of winter holiday and government responses on mortality in Europe during the first wave of the COVID-19 pandemic," *European journal of public health*, vol. 31, no. 2, pp. 272–277, 2021.
- [7] J. G. Chappell, T. Tsoleridis, G. Clark, L. Berry, N. Holmes, C. Moore, M. Carlile, F. Sang, B. J. Debebe, V. Wright, *et al.*, "Retrospective screening of routine respiratory samples revealed undetected community transmission and missed intervention opportunities for sars-cov-2 in the united kingdom," *Journal of General Virology*, vol. 102, no. 6, p. 001595, 2021.

- [8] U. Eberle, S. Heinzinger, R. Konrad, C. Wimmer, B. Liebl, K. Katz, N. Ackermann, and A. Sing, “Virological COVID-19 surveillance in Bavaria, Germany suggests no SARS-CoV-2 spread prior to the first German case in January 2020,” *Infection*, pp. 1–4, 2021.
- [9] C. COVID, R. Team, M. A. Jorden, S. L. Rudman, E. Villarino, S. Hoferka, M. T. Patel, K. Bemis, C. R. Simmons, M. Jespersen, *et al.*, “Evidence for limited early spread of COVID-19 within the United States, January–February 2020,” *Morbidity and Mortality Weekly Report*, vol. 69, no. 22, p. 680, 2020.
- [10] S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, and G. Baele, “Temporal signal and the phylodynamic threshold of sars-cov-2,” *Virus evolution*, vol. 6, no. 2, p. veaa061, 2020.
- [11] A. Rambaut, E. C. Holmes, A. O’Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus, “A dynamic nomenclature proposal for sars-cov-2 lineages to assist genomic epidemiology,” *Nature microbiology*, vol. 5, no. 11, pp. 1403–1407, 2020.
- [12] Folkhälsomyndigheten, “Veckorapport om covid-19, vecka 11,” 2020.
- [13] L. du Plessis, J. T. McCrone, A. E. Zarebski, V. Hill, C. Ruis, B. Gutierrez, J. Raghwani, J. Ashworth, R. Colquhoun, T. R. Connor, *et al.*, “Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK,” *Science*, vol. 371, no. 6530, pp. 708–712, 2021.
- [14] Folkhälsomyndigheten, “Påvisning av antikroppar mot sars-cov-2 hos blodgivare,” 2021.
- [15] A. Bluhm, M. Christandl, F. Gesmundo, F. Ravn Klausen, L. Mančinska, V. Steffan, D. Stilck Franca, and A. H. Werner, “Sars-cov-2 transmission routes from genetic data: A danish case study,” *PloS one*, vol. 15, no. 10, p. e0241405, 2020.
- [16] D. F. Gudbjartsson, A. Helgason, H. Jonsson, O. T. Magnusson, P. Melsted, G. L. Norddahl, J. Saeundsdottir, A. Sigurdsson, P. Sulem, A. B. Agustsdottir, *et al.*, “Spread of SARS-CoV-2 in the Icelandic population,” *New England Journal of Medicine*, vol. 382, no. 24, pp. 2302–2315, 2020.
- [17] A. Popa, J.-W. Genger, M. D. Nicholson, T. Penz, D. Schmid, S. W. Aberle, B. Agerer, A. Lercher, L. Endler, H. Colaco, *et al.*, “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2,” *Science translational medicine*, vol. 12, no. 573, 2020.
- [18] Folkhälsomyndigheten, “Helgenomsekvensering av svenska sars-cov-2 som orsakar covid-19, del 2,” 2020.
- [19] P. T. Nguyen, R. Kant, F. Van den Broeck, M. T. Suvanto, H. Alburkat, J. Virtanen, E. Ahvenainen, R. Castren, S. L. Hong, G. Baele, *et al.*, “The phylodynamics of SARS-CoV-2 during 2020 in Finland—Disappearance and re-emergence of introduced strains,” *Research square (preprint)*, 2021.
- [20] J. Ling, R. A. Hickman, J. Li, X. Lu, J. F. Lindahl, Å. Lundkvist, and J. D. Järhult, “Spatio-temporal mutational profile appearances of Swedish SARS-CoV-2 during the early pandemic,” *Viruses*, vol. 12, no. 9, p. 1026, 2020.
- [21] R. Oliva, A. R. Shaikh, A. Petta, A. Vangone, and L. Cavallo, “D936Y and other mutations in the fusion core of the sars-cov-2 spike protein heptad repeat 1: Frequency, geographical distribution, and structural effect,” *Molecules*, vol. 26, no. 9, p. 2622, 2021.
- [22] J. A. Fuller, A. Hakim, K. R. Victory, K. Date, M. Lynch, B. Dahl, O. Henao, C. COVID, and R. Team, “Mitigation policies and COVID-19-associated mortality—37 European countries, January 23–June 30, 2020,” *Morbidity and Mortality Weekly Report*, vol. 70, no. 2, p. 58, 2021.
- [23] B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, *et al.*, “Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus,” *Cell*, vol. 182, no. 4, pp. 812–827, 2020.
- [24] H. Li and R. Durbin, “Fast and accurate short read alignment with Burrows–Wheeler transform,” *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.
- [25] N. D. Grubaugh, K. Gangavarapu, J. Quick, N. L. Matteson, J. G. De Jesus, B. J. Main, A. L. Tan, L. M. Paul, D. E. Brackney, S. Grewal, N. Gurfield, K. K. A. Van Rompay, S. Isern, S. F. Michael, L. L. Coffey, N. J. Loman, and K. G. Andersen, “An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using primaseq and ivar,” *Genome Biology*, vol. 20, no. 1, p. 8, 2019.

- [26] L.-T. Nguyen, H. A. Schmidt, A. Von Haeseler, and B. Q. Minh, "Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies," *Molecular biology and evolution*, vol. 32, no. 1, pp. 268–274, 2015.
- [27] P. Sagulenko, V. Puller, and R. A. Neher, "Treetime: Maximum-likelihood phylodynamic analysis," *Virus evolution*, vol. 4, no. 1, p. vex042, 2018.

4 Appendix

4.1 Internationella SARS-CoV-2 sekvenser

Dataset av samtliga globala sekvenser med tillhörande metadata med provtagningsdatum fram till 31 maj 2020 tillgängliga i GISAID sammanställdes under juli 2021

4.2 Svenska sekvenser

Från GISAID fanns 725 svenska sekvenser tillgängliga, huvudsakligen sekvenserade av Folkhälsomyndigheten (n=714). För utredningen generades ytterligare 625 nya sekvenser. Prov på Karolinska Universitetslaboratoriet valdes ut enligt följande kriterier för sekvenseringen: i) enligt urval och uppdrag från Folkhälsomyndigheten, ii) enligt eget urval utifrån provtagningsdatum, provtagare, ålder och tillräcklig virusnivå, iii) enligt eget urval utifrån SMI-net-data om smitta utomlands, iv) på tidigare ej diagnostiserade personer som upptäckts genom SARS-CoV-2 PCR på sparade luftvägsprov på KUL, samt v) inom annat projekt om återinfektion.

4.3 Metadata

Sekvenserna kompletterades med metadata. Metadata för de svenska sekvenserna erhöles från: i) statistikmodulen för laboratoriets informationssystem (ålder, kön, provtagare, provtagningsdatum) ii) utdrag från SMI-net (smittland, smittort, anledning till provtagning).

4.4 SARS-CoV-2 helgenomsekvensering

Helgenomsekvensering av SARS-CoV-2 från sparade luftvägsprov utfördes den metod som är i drift inom rutindiagnostik på KUL som bygger på ett samarbete mellan Klinisk Mikrobiologi, KUL och Clinical Genomics, SciLifeLab.

Luftvägsproverna späddes 1:2 i MagNA Pure 96 External Lysis Buffer (kat. nr. 06374913001, Roche Diagnostics AB) för inaktivering av virus och sedan extraherades RNA på MagNA Pure 96 (MP96) (Roche Diagnostics) med MagNA Pure 96 DNA and Viral NA SV Kit (kat. nr. 06543588001, Roche Diagnostics AB). cDNA tillverkades med Illumina COVIDSeq Test Kit (Document # 1000000126053 v02) i enlighet med tillverkarens instruktion: 8,5 µL RNA-eluat blandades med 8,5 µL Elution Prime Fragment 3HC Mix HT följt av annealing vid 63°C - 3 minuter och 4°C - indefinitely. Därefter tillsattes 0,8 µL Reverse Transcriptase HT och 7,2 µL First Strand Mix HT till det annealade RNA:t. Protokoll för cDNA-syntes: 25°C - 5 min, 50°C - 10 min, 80°C - 5 min, samt 4°C - indefinitely.

Erhållet cDNA användes för amplifiering av hela SARS-CoV-2 genomet med överlappande amplicon enligt ARTIC-systemet. Två reaktioner per prov sattes upp med följande innehåll: 12,5 µL Illumina PCR Mix HT, 3,9 µL nukleasfritt vatten, 3,6 µL COVIDSeq primer pool 1 HT respektive COVIDSeq primer pool 2 HT, samt 5 µL cDNA. Protokoll för amplifiering: 98°C - 3 min, 35 cykler av (98°C - 15 s och 63°C - 5 min), samt 4°C indefinitely. Efter amplifiering poolades reaktionerna till en slutvolym om 30 µL.

PCR amplifierat DNA användes för att tillverka sekvenseringsbibliotek med tagmenteringsmetodik. Till varje reaktionsmix tillsattes 10 µL Tagmentation Buffer 1 HT, 3,3 µL Enrichment BLT HT samt 16,7 µL nukleasfritt vatten. Proverna inkuberades i 55°C - 5 min följt av 10°C - indefinitely. Efter tagmentering renades ampliconen genom att tillsätta 10 µL Stop Tagment Buffer 2 HT följt av inkubation i rumstemperatur i 5 min och magnetisk separation. Magnetkulorna tvättades två gånger med Tagmentation Wash Buffer HT. Amplifiering av tagmenterade amplicon skedde genom att tillsätta 20 µL Enhanced PCR Mix HT, 10 µL Index Adapters och 20 µL nukleasfritt vatten till varje prov och inkubera i 72°C - 3 min, 98°C - 3 min, 7 cykler av (98°C - 20 s, 60°C - 30 s, 72°C - 1 min), 72°C - 3 min samt 4°C indefinitely. Amplifierade bibliotek poolades och 0.9X volym Illumina Tune Beads tillsattes. Efter 5 min inkubation i rumstemperatur separerades reaktionen på magnet och magnetkulorna tvättades två gånger med 80% etanol. Biblioteken resuspenderades i 55 µL Buffer HT och supernatanten togs tillvara. Biblioteken kvantifierades med Qubit Fluorometer och Qubit dsDNA HS Assay kit enligt tillverkarens instruktioner.

Sekvensering skedde på NovaSeq 6000 Sequencing System, där sekvensdata med ett måldjup på 1,000,000 sekvenspar (2 × 150 bp) generades. Konvertering och demultiplexning av rådata utfördes med bcl2fastq v2.20.

4.5 Bioinformatisk behandling och analys av sekvensdata

Rådata från sekvenseringen (s.k. reads) analyserades med hjälp av Sveriges nationella pipeline för SARS-CoV-2 analyser, [gms-artic](#). Denna analys utgår ifrån det internationellt rekommenderade analysflöde ([ncov2019-artic-nf](#)) som framtagits av ARTIC Network och Public Health Wales för SARS-CoV-2 sekvensdata. I analysen används Trim Galore för att trimma sekvenserna som sedan mappas mot SARS-CoV-2 referensgenomet Wuhan- Hu-1 (accessionsnummer MN908947.3) med bwa [24]. Ivar [25] används för trimning av primersekvenser samt

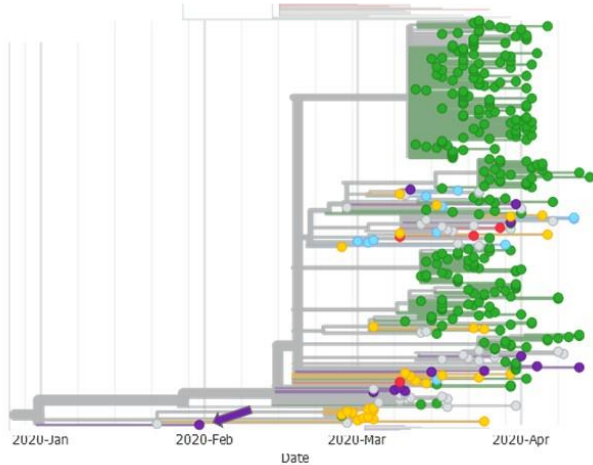
generering av konsensussekvens för varje prov. Analysen tilldelar automatiskt varje sekvens en så kallad pangolin- sekvenstyp. I enlighet med kvalitetskrav från GISAID och FOHM krävdes för godkänt sekvensresultat att >95% av genomet behövde ha ett sekvensdjup på minst 10x. Enbart godkända sekvenser gick vidare till fortsatt analys. Analys av godkända sekvenser utfördes med Nextstrains bioinformatiska pipeline augur för visualisering med programmet auspice. Karakterisering av förekomst av mutationer av sekvenserna utfördes med Nextstrains öppet tillgängliga program Nextclade. Sekvensanalys av mutationsprofiler och gruppering i genotyperna [19A-B, 20A-D] utfördes för de svenska och internationella sekvenserna med nextclade med Wuhan-Hu-1 som referens. Sekvenserna grupperades enligt genotyper [19A-B, 20A-D] utifrån definierade mutationer (tabell 5). Fylogenetisk trädanalys utfördes med Nextstrains pipeline augur: alignment av sekvenserna utfördes med nextalign med Wuhan-Hu-1 som referens, IQ-tree [26] användes för att härleda ett fylogenetiskt träd (substitutionsmodell GTR) följt av härledning av ett maximum-likelihood träd med TreeTime [27] rotat med Wuhan-Hu-1, med parametern clock rate satt till 0,0008 mutationer per site per år med standarddeviationen 0,0004. Sekvenser med ett avstånd som avvek mer än 4 interkvartiler från clock rate (enligt regression root-to-tip vs tid) exkluderades.

4.6 SARS-CoV-2 PCR analys av tidigare ej undersökta prover

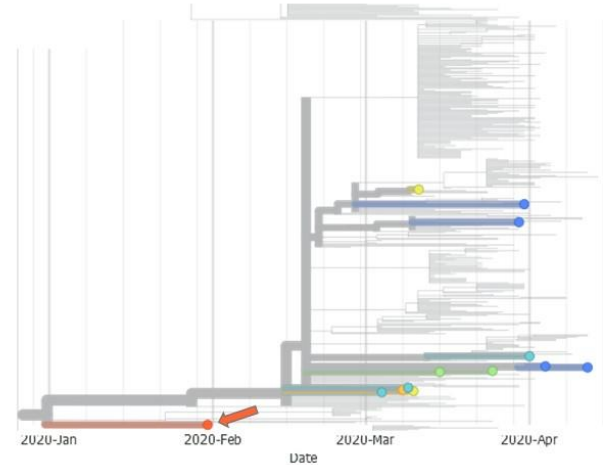
Luftvägsprover, som ej tidigare testats med avseende på SARS-CoV-2, screenades för eventuell förekomst av SARS-CoV-2-RNA. Pooler om max 10 st prover i 25 µL späddes 1:10 i 250 µL MagNA Pure 96 External Lysis Buffer (kat. nr. 06374913001, Roche Diagnostics AB) för inaktivering av eventuella virus. För automatiserad extraktion och PCR-setup användes HAMILTON STARlet KingFisher Presto® med MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit (Kat.nr. A42352, ThermoFisher) och TaqPath™ COVID 19 CE IVD RT PCR Kit (Kat.nr. A48067, ThermoFisher). Quantstudio 5 (ThermoFisher/Applied Biosystems) användes för qPCR. Analysen baserades på fluorofor-konjugerade hydrolyseringsprober för detektion av de SARS-CoV-2-specifika generna ORF1ab (FAM), N (VIC) och S (ABY). En internkontroll i form av enkelsträngad MS2-bakteriofag (JUN) ingick också i analysen. SARS-CoV-2-RNA bedömdes påvisat om en pool visade Ct-värde <40 i minst en av de SARS-CoV-2-specifika målgenerna. Vid påvisat SARS-CoV-2-RNA analyserades de ingående proverna enskilt med spädning 1:10 (50 µL prov + 450 µL ELB). Totalt analyserades 200 pooler och 189 prov från 19 pooler analyserades enskilt. De enskilt analyserade proven resulterade i 18 stycken verifierade fall, 3 misstänkta fall, samt 161 negativa. Material för enskild om-analys saknades för 7 prov. Av de 18 verifierade fallen var 12 stycken tidigare ej kända patienter (en patient hade detektion i två prov). Av de 3 misstänkta fallen hade 1 patient SARS-CoV-2-RNA påvisat i annat prov. Sekvensering utfördes på prover med ct-värde upp till 35.

4.7 Figurer och tabeller

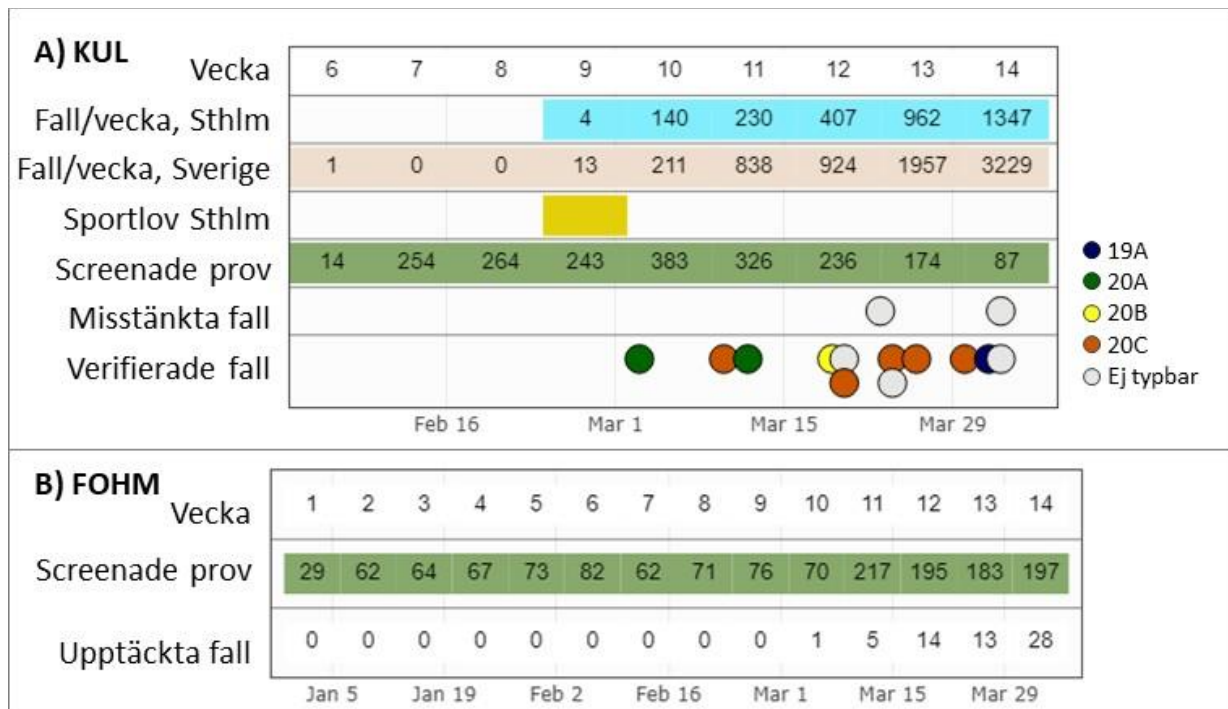
A



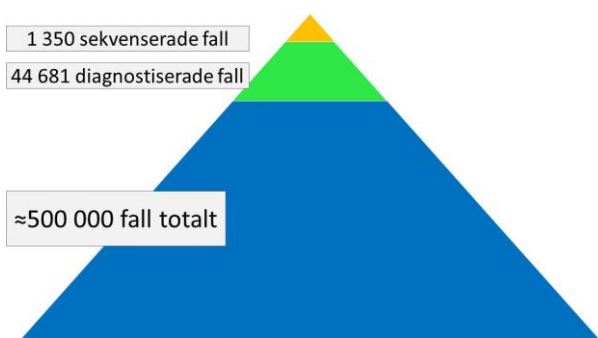
B



Figur 1: **Detaljbild inom genotyp 19A.** Förstoring av gren inom genotyp 19A i vilken 13 sekvenser från Sverige återfinns färgat efter land för detektion (A), samt sekvenserna från Sverige färgade efter smittland (B). Det finns inga svenska sekvenser som är närbesläktade med det tidigaste svenska fallet av covid-19 smittad i Kina i Wuhan-området (utmärkt med pil i båda panelerna), dvs inga belegg för att just den introduktionen skulle orsakat fortsatt smittspridning i Sverige.

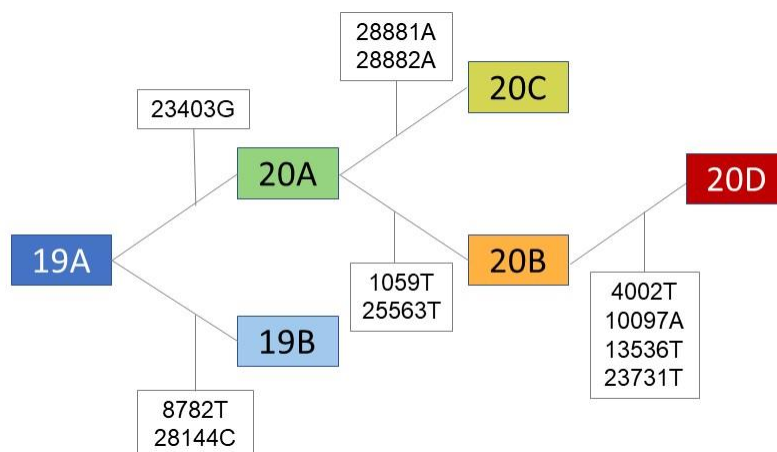


Figur 2: **Undersökning av eventuell oupptäckt tidig smittspridning av SARS-CoV i Stockholm och Sverige.** Antalet analyserade prov och upptäckta fall av SARS-CoV-2-infektion i relation till antalet kända fall. A) Sparade luftvägsprov från klinisk rutindiagnostik analyserade för SARS-CoV-2 på Klinisk mikrobiologi, Karolinska Universitetssjukhuset (KUL). B) Luftvägsprov från Folkhälsomyndighetens (FOHMs) sentinelövervakningen för influensa analyserade för SARS-CoV-2 (källa: FOHMs *Veckorapport om covid-19, vecka 16*, samt personlig kommunikation Mia Brytting, FOHM).



Figur 3: **Sekvenser relaterat till antal fall.** Pyramiden visar den ungefärliga relationer mellan antalet svenska sekvenser under tidsperioden för utredningen, antalet bekräftade fall. Det senare baserat på FOHMs studier av seroprevalens som indikerade att omkring 5% av den svenska befolkningen hade genomgått infektionen under det första halvåret 2020 [14]

Nextstrain	Gisaid	Pangolin
19A	L	B, B.3, B.11
	O	B, B.4, B.6
	V	B, B.40
19B	S	A, A.2, A.2.4
20A	G, GH	B.1
20C	GH	
20B	GR	B.1.1, B.1.1.29
20D		B.1.1.1, C.14



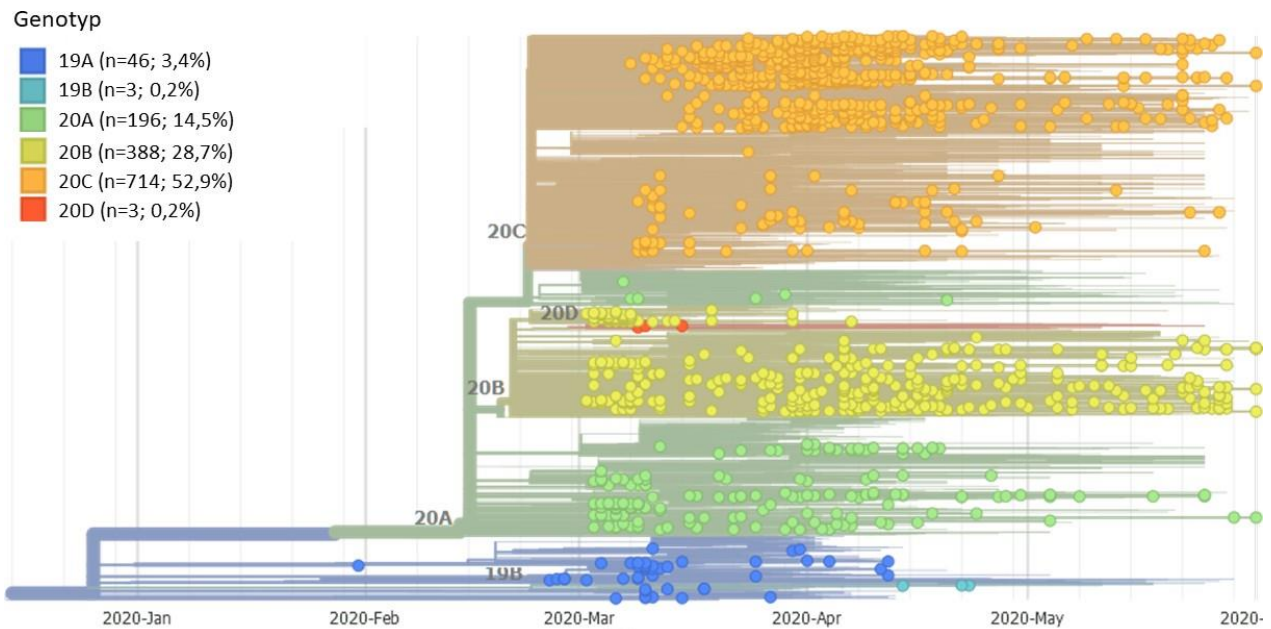
Figur 4: **Klassifikation av SARS-CoV-2 sekvenser.** Tabellen visar de tre vanligaste systemen för genetisk klassificering av SARS-CoV-2 och de i rapporten vanligaste genotyperna med de ungefärliga motsvarigheterna mellan systemet. WHO-klassificering (alfa, beta, osv.) avser enbart varianter som förekommit under 2021 och är inte applicerbar på genotyperna i tabellen eller denna rapport. Till höger visas hur de sex Nextstrain-genotyperna i rapporten relaterar till varandra, de vita boxarna anger de definierande mutationerna som skiljer genotyperna.

Region	Fall (antal)	Sekvenser (antal)	Andel av svenska sekvenser (%)	Andel av fall sekvenserade (%)
Blekinge	297	0	0,0	0,0
Dalarna	1319	40	3,0	3,0
Gotland	92	0	0,0	0,0
Gävleborg	1526	7	0,5	0,5
Halland	916	69	5,1	7,5
Jämtland	813	11	0,8	1,4
Härjedalen				
Jönköping	1773	10	0,7	0,6
Kalmar	408	14	1,0	3,4
Kronoberg	932	0	0,0	0,0
Norrbottn	504	17	1,3	3,4
Skåne	1994	20	1,5	1,0
Stockholm	14247	842	62,4	5,9
Sörmland	1756	20	1,5	1,1
Uppsala	2243	56	4,1	2,5
Värmland	643	30	2,2	4,7
Västerbotten	535	18	1,3	3,4
Västernorrland	880	26	1,9	3,0
Västmanland	1437	17	1,3	1,2
Västra Götaland	8133	83	6,1	1,0
Örebro	1975	21	1,6	1,1
Östergötland	2259	45	3,3	2,0
Uppgift saknas	0	4	0,3	
Total	44682	1350	100	3,0

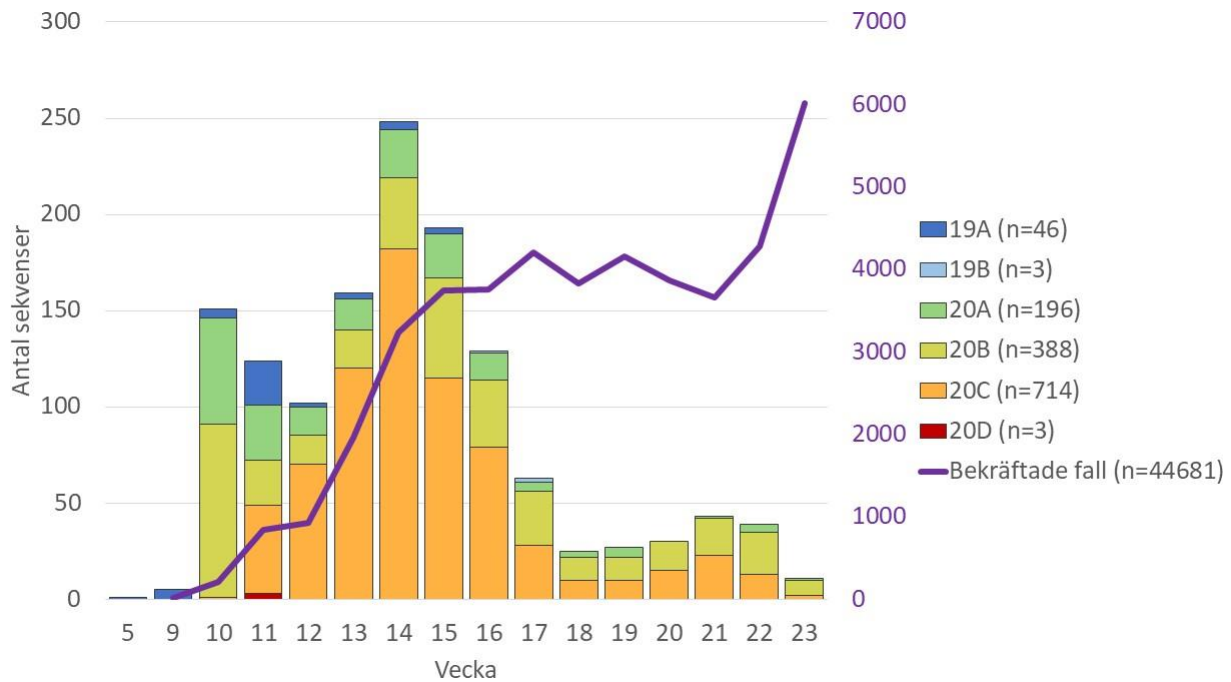
Tabell 1: **Andel sekvenserade fall per region.** Antal diagnostiserade SARS-CoV-2 fall per region i Sverige till och med vecka 23 2020 (källa: Folkhälsomyndighetens hemsida) samt antal sekvenser inkluderade i analysen.

Ålder (år)	19A	19B	20A	20B	20C	20D	Total	Andel (%)
0 till 4				3	3		6	0,4
5 till 14	1	1	11	11	3		27	2,0
15 till 24	10		17	38	9		74	5,5
25 till 64	27	2	99	202	285	3	618	45,8
65 till 79	2		28	65	184		279	20,7
≥80	5		39	66	224		334	24,7
Ingen uppgift	1		2	3	6		12	0,9
Total	46	3	196	388	714	3	1350	100

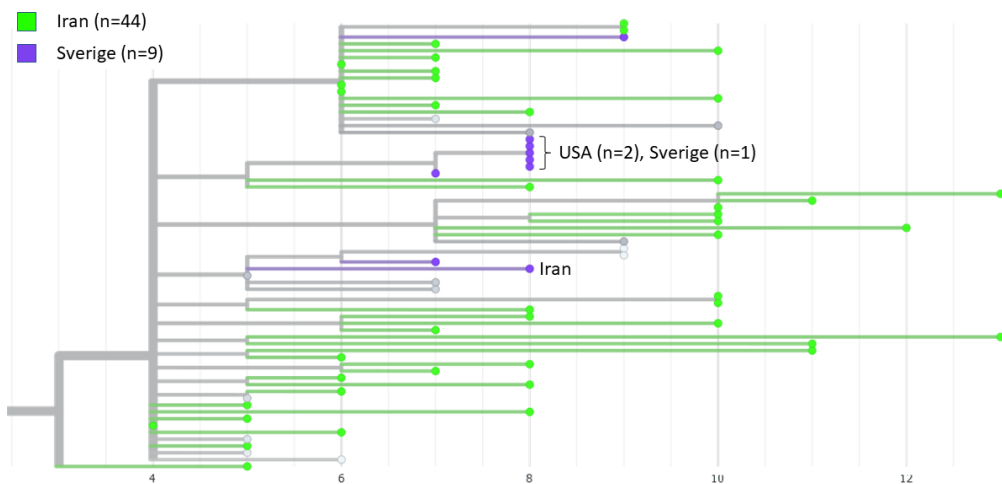
Tabell 2: ålderskategori per genotyp för de svenska SARS-CoV-2 patienter vars prov sekvenserats i det svenska sekvensdatasetet.



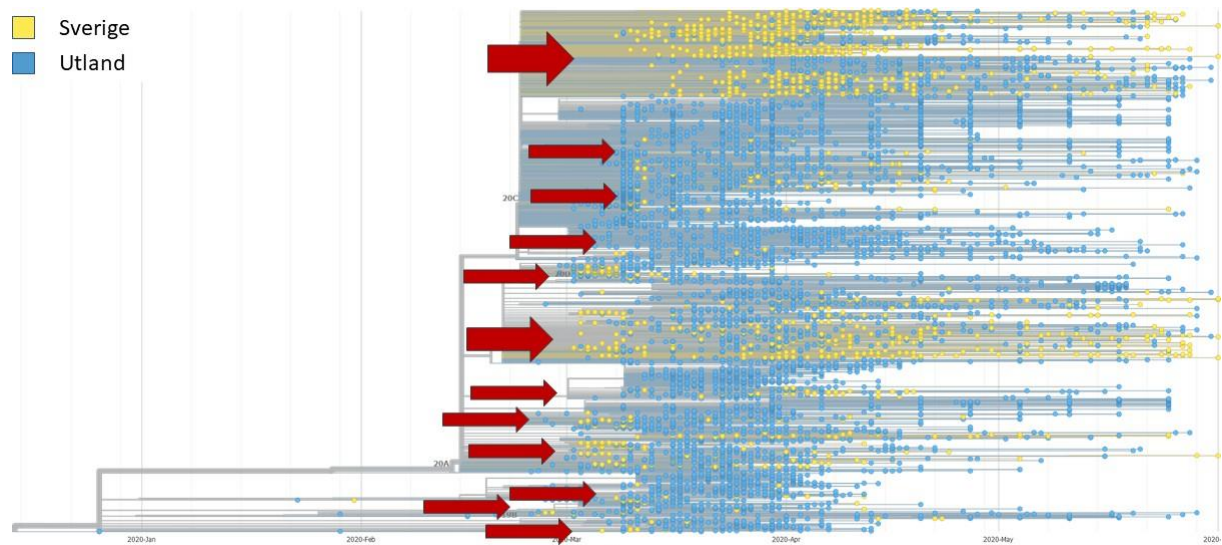
Figur 5: **Svenska sekvenser.** Fylogenetiskt släktskapsträd som visar släktskap mellan svenska SARS-CoV-2 sekvenser. De sex genotyperna är indikerade med olika färger. X-axeln visar provtagningsdatum.



Figur 6: **SARS-CoV-2 genotyper i Sverige per vecka 2020.** Fördelning av SARS-CoV-2 genotyper per provtagningsvecka för sekvenser från svenska patienter provtagna fram till och med 1 juni 2020 (måndag v 23) (staplar färgade per genotyp) samt antal diagnostiserade fall under dessa veckor (linje). Totalt antal sekvenser = 1 350.



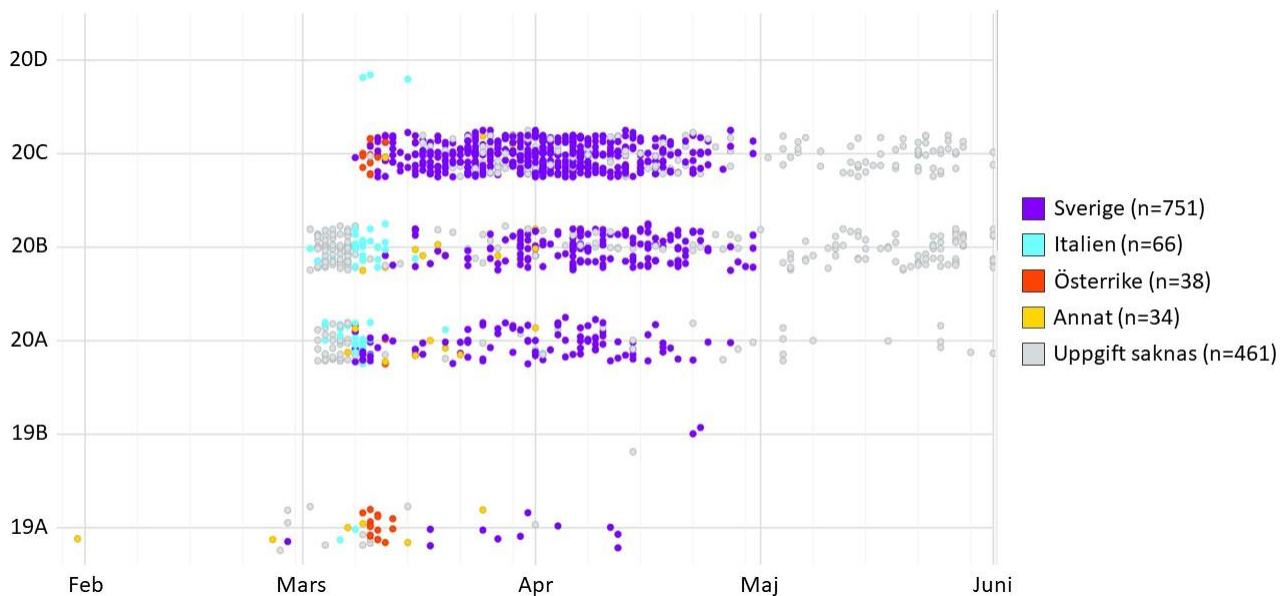
Figur 7: **Sekvenser Iran.** Detaljbild av fylogenetiska trädet inom genotyp 19A med färgmarkering av svenska och iranska respektive övriga sekvenser. Texten i figuren anger smittland för de fyra sekvenser där uppgift som det finns i SMI-net. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



Figur 8: **SARS-CoV-2 introduktioner till Sverige.** Fylogenetiskt träd som visar sekvenser från virus i Sverige i relation till utländska sekvenser. Pilarna ger en indikation om det minimala antalet introduktioner av SARS-CoV-2 till Sverige. Det samma antalet introduktioner är dock med all säkerhet många gånger högre (se text). X-axeln anger provtagningsdatum.

Smittland / Vecka	5	...	9	10	11	12	13	14	15-23	Total
Italien				30	33	2		1	0	66
Österrike					35	2	1		0	38
Frankrike					3	2	1		0	6
Spanien					1	1		2	0	4
Schweiz					1		2		0	3
Tyskland				1		1		1	0	3
USA				2	1				0	3
Storbritannien						2			0	2
Albanien							1		0	1
Danmark					1				0	1
Finland					1				0	1
Indonesien							1		0	1
Iran			1						0	1
Israel							1		0	1
Kina	1								0	1
Lettland							1		0	1
Nederländerna				1					0	1
Portugal						1			0	1
Tanzania						1			0	1
Thailand								1	0	1
Östeuropa					1				0	1
Subtotal utland	1		1	34	77	12	8	5	0	138
Sverige			1	5	37	75	113	191	329	751
Ingen uppgift			3	112	10	15	38	52	231	461
Total	1		5	151	124	102	159	248	560	1350

Tabell 3: **Smittland.** Smittland enligt uppgifter i SMInet för svenska sekvenser uppdelat per provtagningsvecka.



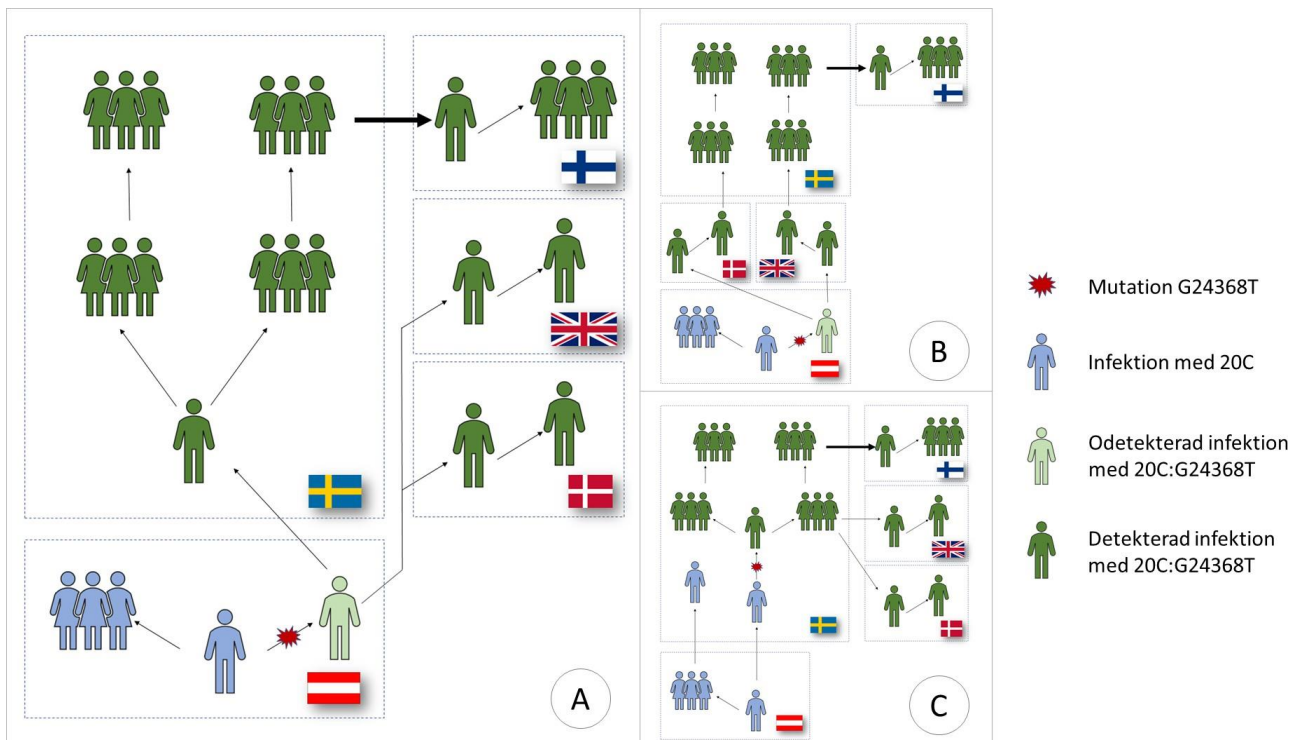
Figur 9: **Smittland svenska sekvenser.** Svenska sekvenser uppdelade per genotyp och färgade efter smittland enligt uppgifter i SMInet. X-axeln anger provtagningsdatum. Se tabell 3 för specificering av länder i kategorin 'Annat'.

Sekvenser/Vecka	5	...	9	10	11	12	13	14	15-23	Total	Antal varianter
Unik (0)	1		5	54	45	39	90	180	484	898	748
Singel (1)				14	13	4	4	4	13	52	22
Fåtal (2-10)				24	9	3	7	2	6	51	27
Flertal (11-100)				14	17	3	1		3	38	18
Hundratals (101-1000)				25	22	46	51	48	31	223	11
Tusentals (> 1000)				20	18	7	6	14	23	88	3
Total	1		5	151	124	102	159	248	560	1350	829

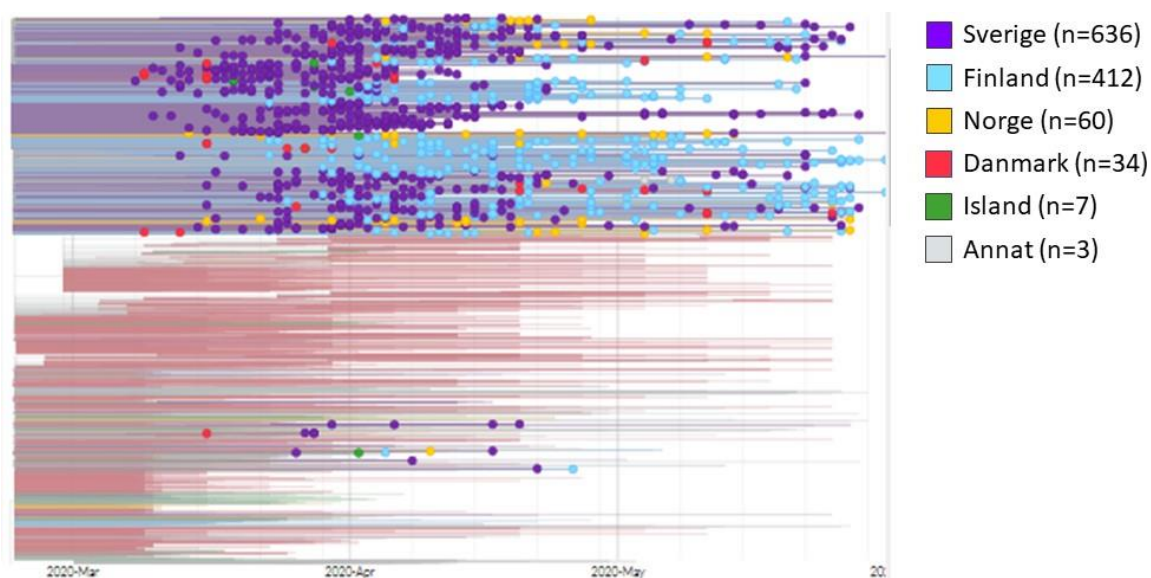
Tabell 4: **Förekomst av sekvenser från Sverige i det globala datasetet.** Tabellen visar hur många identiska sekvenser som finns i det internationella datasetet om cirka 125 000 sekvenser för varje svensk sekvens. Kategori 'Unik' betyder att det inte finns någon identisk sekvens i det internationella datasetet, 'Singel' att det finns en identisk utländsk sekvens, 'Fåtal' att det finns två till tio identiska sekvenser, osv. Kolumnen 'Antal varianter' visar att i t.ex. kategorin 'Singel' utgörs de 52 sekvenserna av 22 olika mutationsprofiler, de 88 sekvenserna i 'Tusentals' görs av 3 mutationsprofiler, osv. Resultaten är uppdelade per provtagningsvecka för de svenska sekvenserna (vecka 5-23).

Genotyp	Definierande mutationer
19A	C8782-, C14408-
19B	C8782T, T28144C
20A	C8782-, C14408T, A23403G
20B	C8782-, C14408T, A23403G, G28881A, G28882A
20C	C1059T, C8782-, C14408T, G23403G, G25563T
20D	C4002T, G10097A, C13536T, C23731T

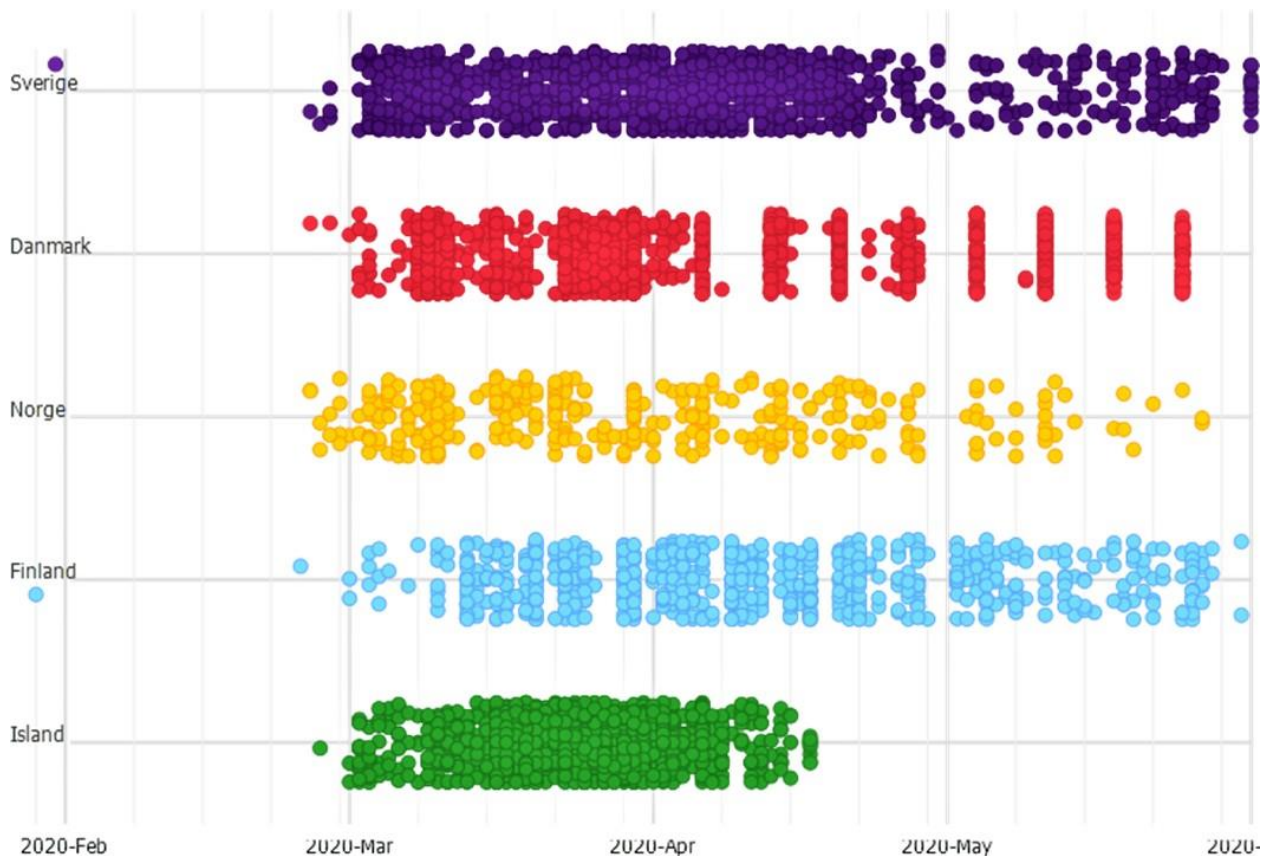
Tabell 5: Definierande mutationer för de sex nextstrain genotyperna 19A-B och 20A-D. Bokstaven före positionsnumret avser ursprungsvarianten, bokstaven efter positionsnumret avser mutationen, dvs 'C8782T' avser mutation från C till T på position 8782. '-' indikerar här avsaknad av mutation, dvs 'C8782-' avser C på position 8782.



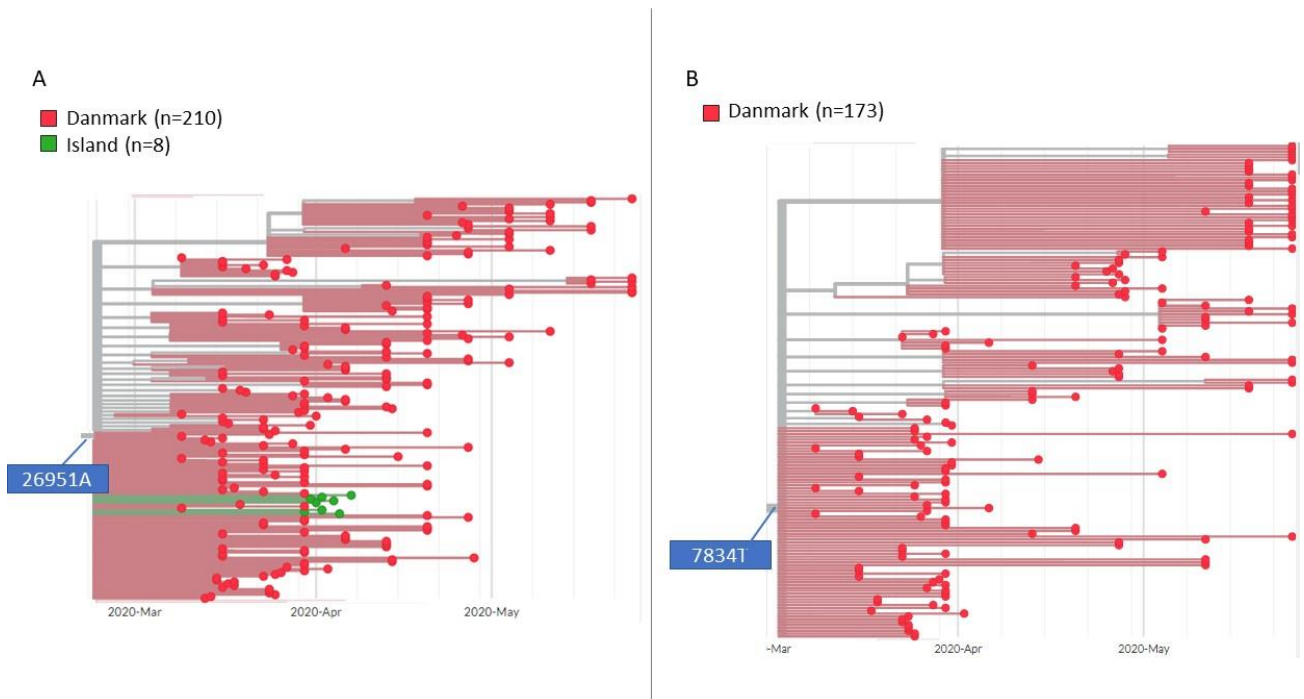
Figur 10: **Alternativa scenarion för spridningen av 20C:G24368T-varianten.** A) Mutationen G24368T uppstår i alpmrådet men detekteras inte där. Varianten 20C:G24368T sprids därifrån till Sverige och andra länder, bland annat Danmark och Storbritannien. Från Sverige sprids den senare vidare till Finland. B) Mutationen G24368T uppstår i alpmrådet men detekteras inte där. Varianten 20C:G24368T sprids till andra länder, bland annat Danmark och Storbritannien, varifrån den vidare sprids till Sverige. Från Sverige sprids den senare vidare till Finland. C) Genotypen 20C importeras till Sverige från alpmrådet. Mutationen G24368T uppstår i Sverige varifrån den sprids vidare till andra länder, bland annat Danmark, Storbritannien och Finland.



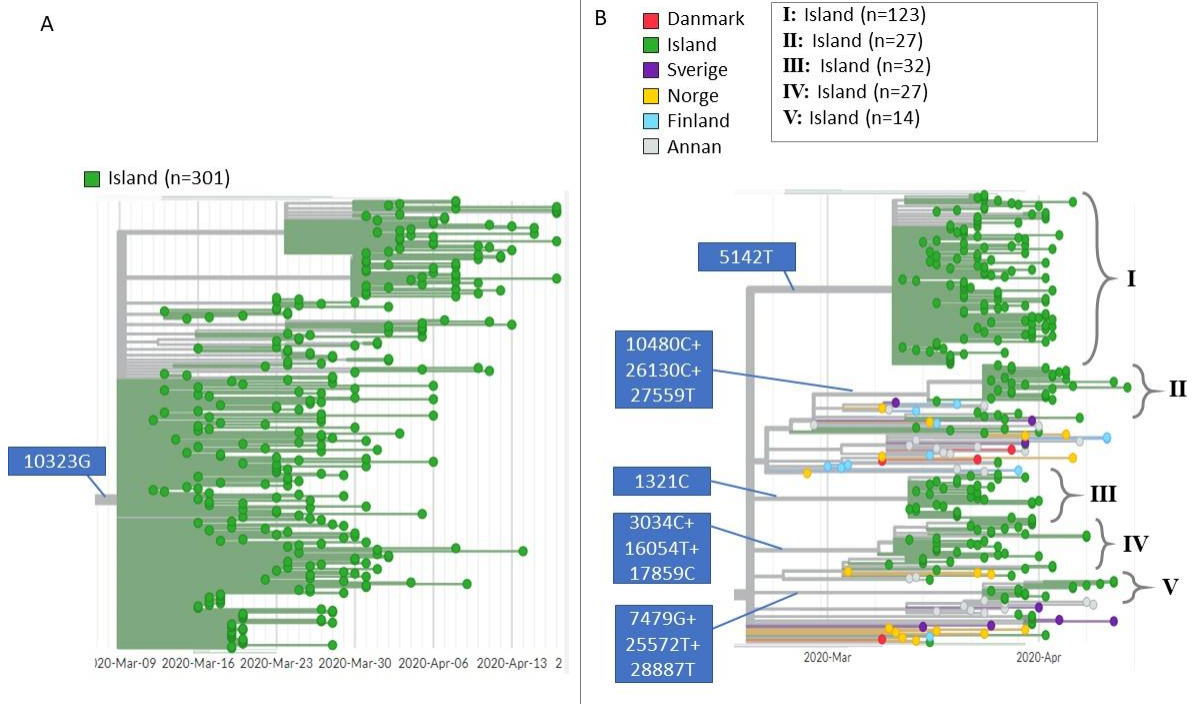
Figur 11: **20C:G24368T i Norden.** Sekvenser inom genotyp 20C med mutationen G24368T färgade per nordiskt land. Huvudsakligen har varianten påvisats i Sverige och Finland. X-axeln anger provtagningsdatum.



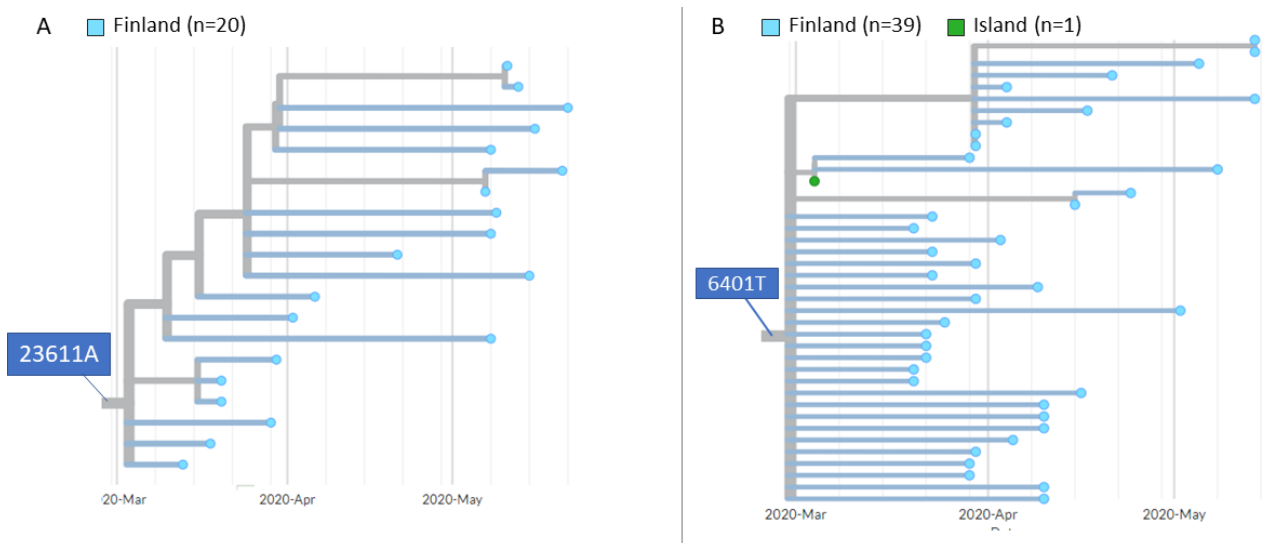
Figur 12: **Sekvenser från de nordiska länderna.** Figuren visar fördelningen över tid för sekvenser från de nordiska länderna. X-axlen anger provtagningsdatum. Sverige (n=1350), Danmark (n=1762), Norge (n=377), Finland (n=757), Island (1447).



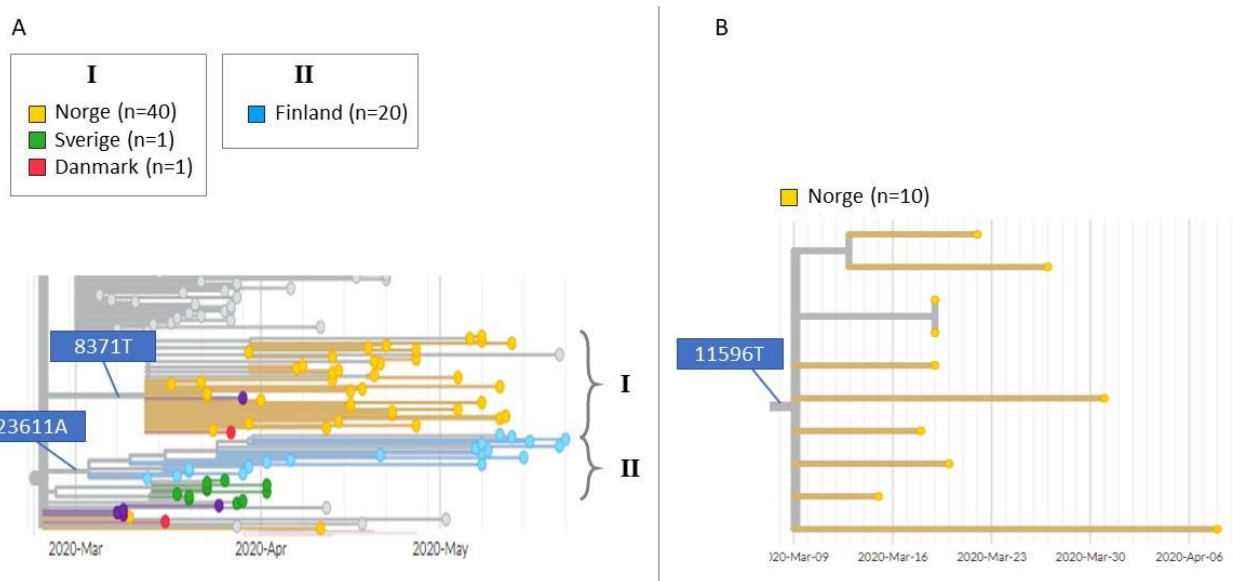
Figur 13: **Exempel danska kluster.** Figuren visar exempel på två kluster med huvudsakligen danska sekvenser i genotyp 20C (A) respektive 20A (B). X-axlen anger provtagningsdatum.



Figur 14: **Exempel isländska kluster.** Figuren visar exempel på kluster med huvudsakligen isländska sekvenser i genotyp 20A (A) respektive 19A (B). X-axeln anger provtagningsdatum.



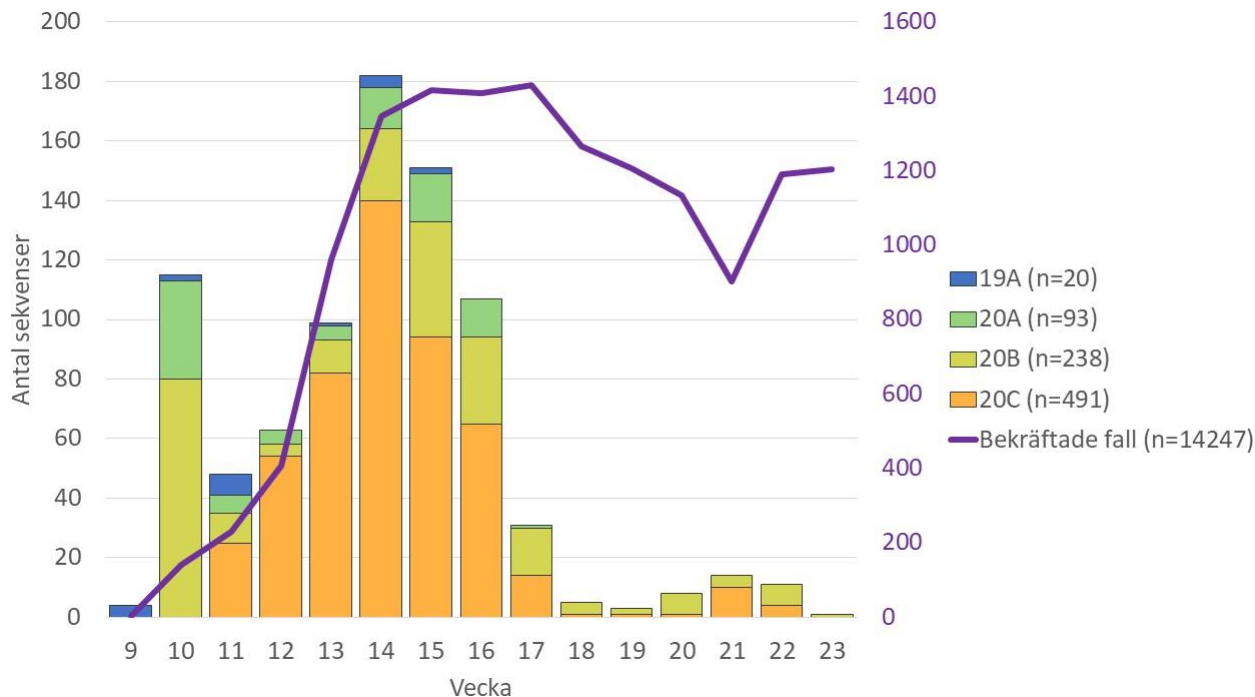
Figur 15: **Exempel finska kluster.** Figuren visar exempel på kluster av finska sekvenser inom genotyp 20A. X-axeln anger provtagningsdatum.



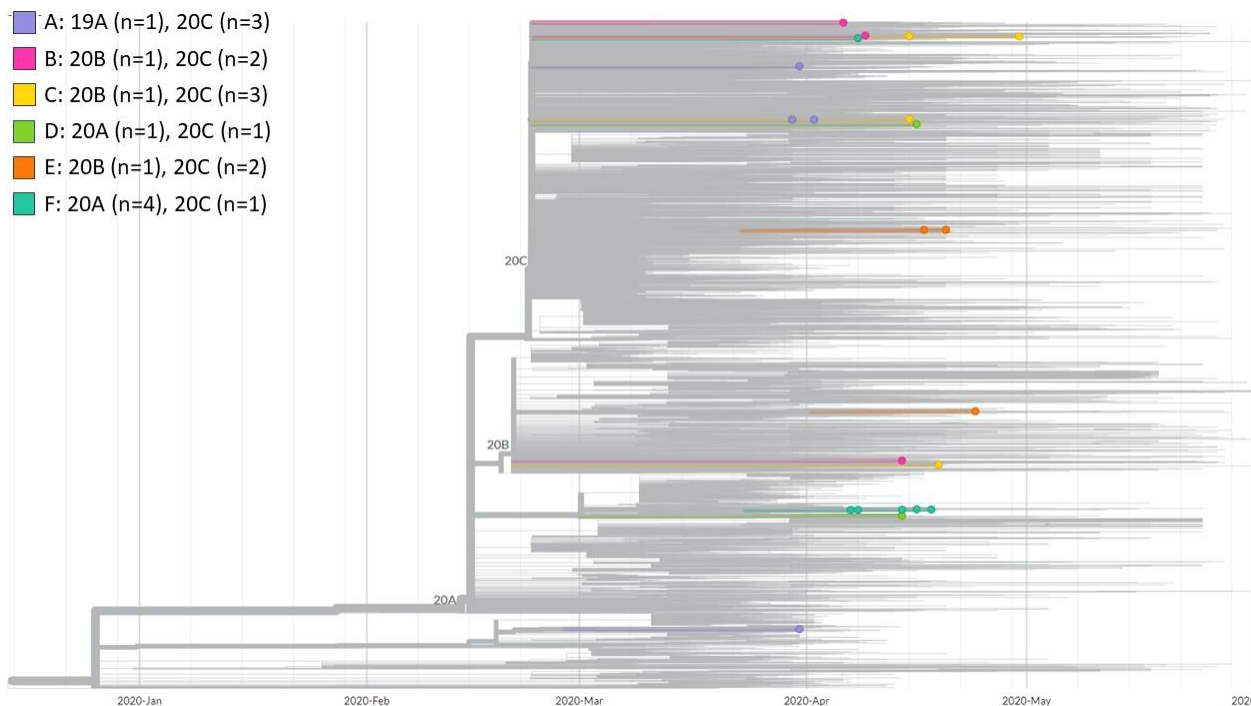
Figur 16: **Exempel norska och ett finskt kluster.** Figuren visar exempel på två kluster med huvudsakligen norska sekvenser (A.I, B) samt ett finskt kluster i genotyp 20A (A.II). X-axeln anger provtagningsdatum.



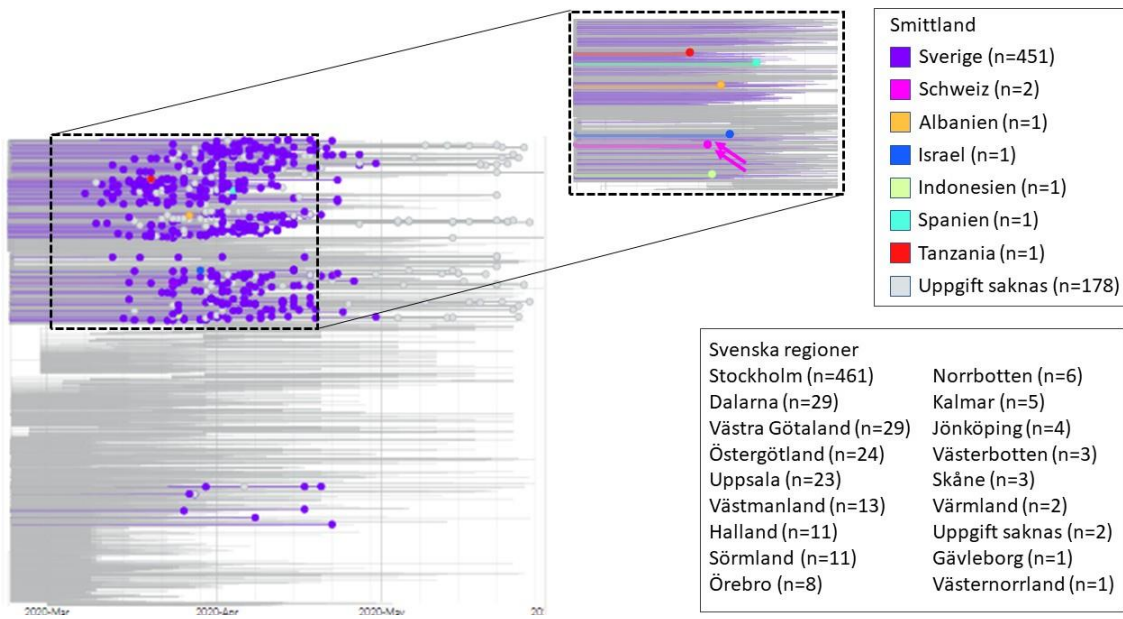
Figur 17: **Virusgenotyp per akutsjukhus.** Grafiken visar 441 sekvenser av prov skickade från Stockholms akutsjukhus färgade enligt A) smittland enligt uppgift i SMInet och B) genotyp. Karolinska Universitetssjukhuset Huddinge, n= 168. Södertälje sjukhus, n= 51. Danderyds sjukhus, n= 53. Karolinska Universitetssjukhuset Solna, n= 28. Sankt Görans sjukhus, n= 122. Södersjukhuset, n= 19.



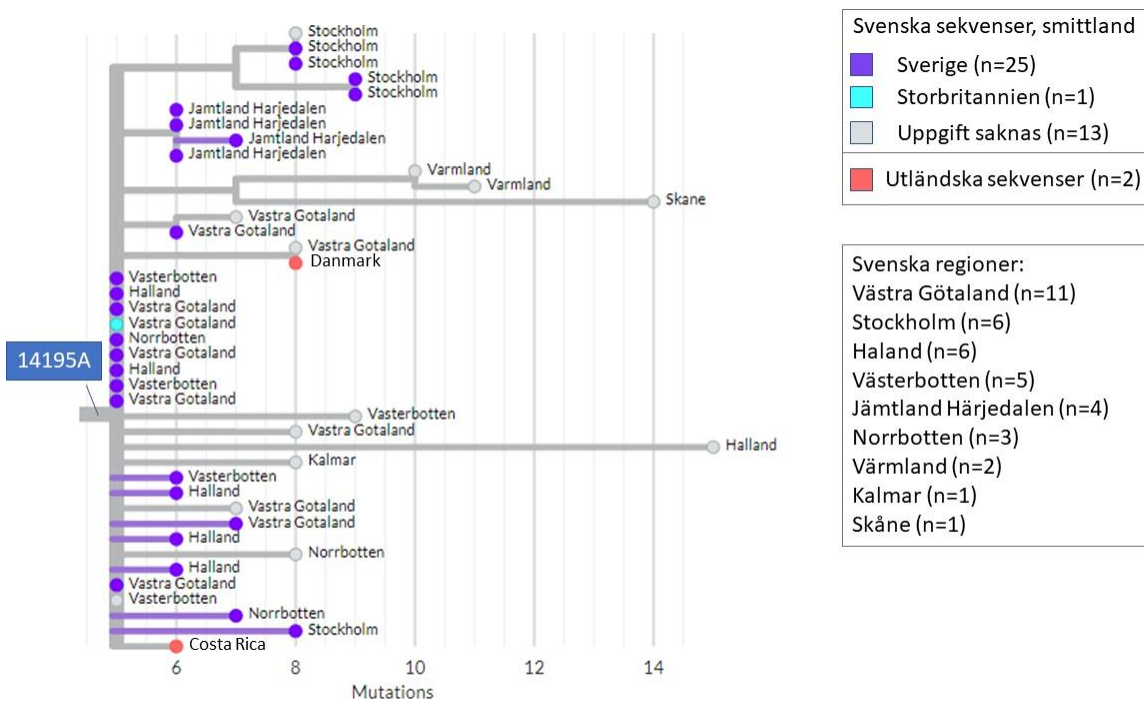
Figur 18: **Virusgenotyper i region Stockholm per vecka 2020.** Fördelning av virusgenotyper i sekvenserade prov från region Stockholm per vecka för provtagning fram till och med 1 juni 2020 (måndag v 23) (staplar färgade per genotyp) samt antalet bekräftade fall i regionen (linje). Totalt antal sekvenser = 842.



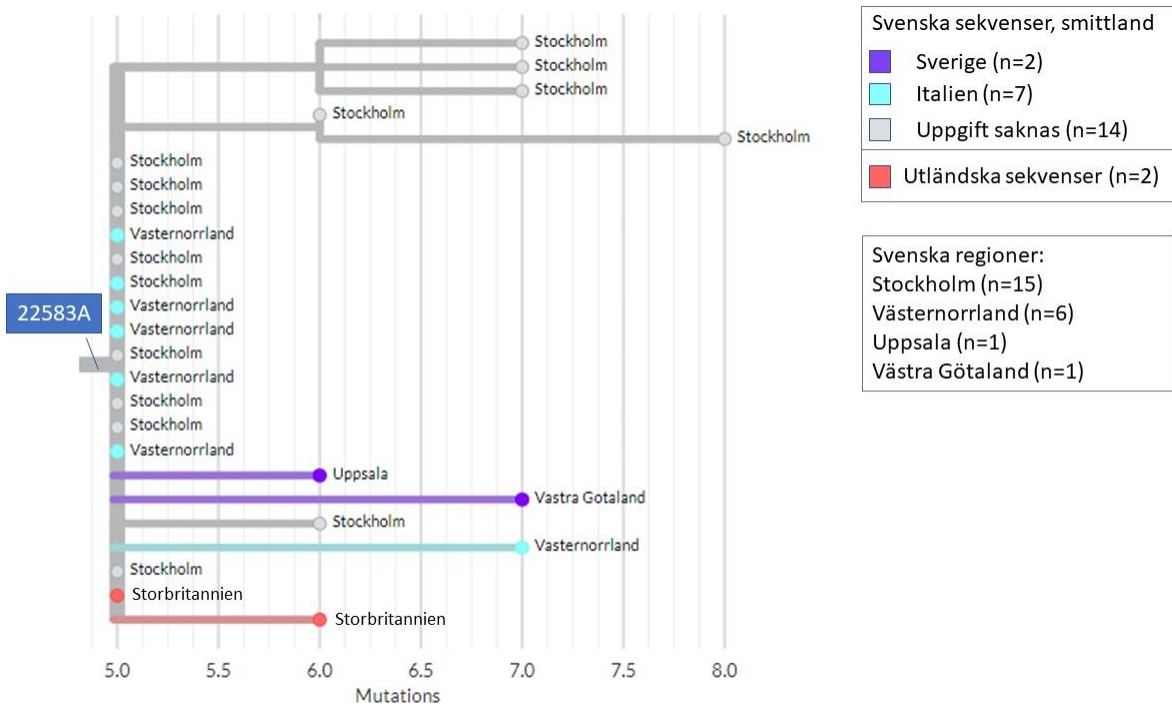
Figur 19: **Multipla virusintroduktioner på boenden.** Trädet visar sex boenden i region Stockholm där sekvenser med olika genotyper påvisats. X-axeln anger provtagningsdatum.



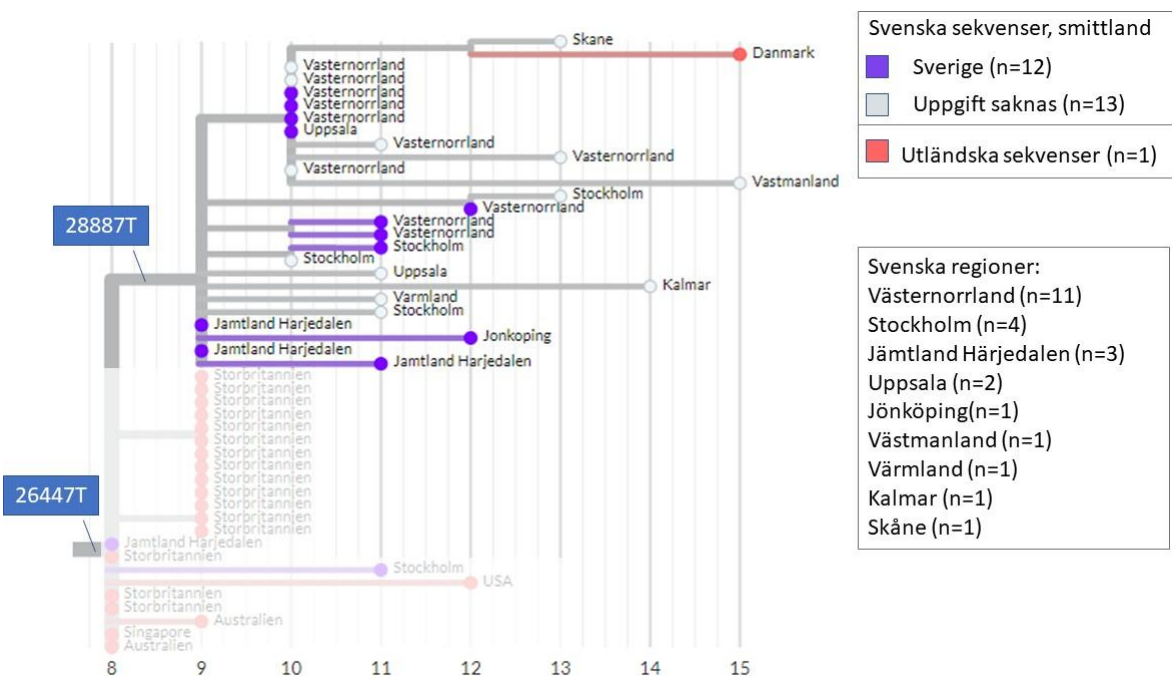
Figur S 1: **20C:G24368T**. Detaljbild av kluster med 636 svenska sekvenser definierat av mutationen G24368T. I utsnittet visas sekvenser med utländskt smittland (pilarna indikerar att två sekvenser med smittland Schweiz överlappar). X-axeln anger provtagningsdatum.



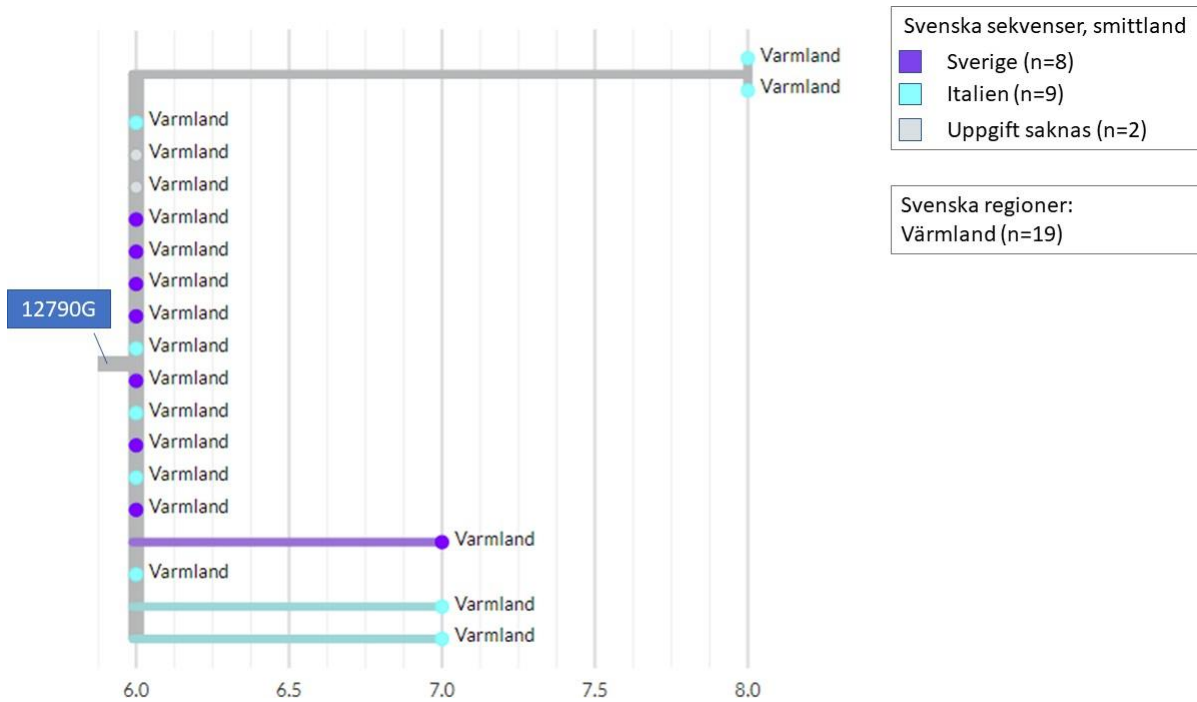
Figur S 2: **20A:C14195A**. Detaljbild av kluster med 36 svenska sekvenser definierat av mutationen C14195A. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



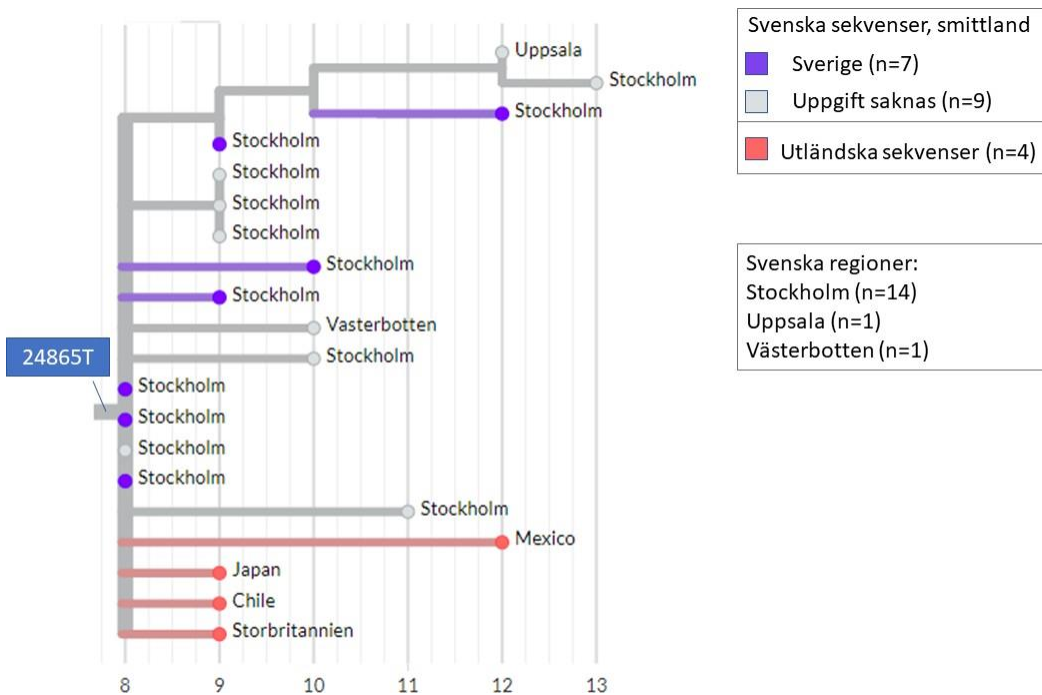
Figur S 3: **20A:G22583A**. Detaljbild av kluster med 23 svenska sekvenser definierat av mutationen G22583A. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



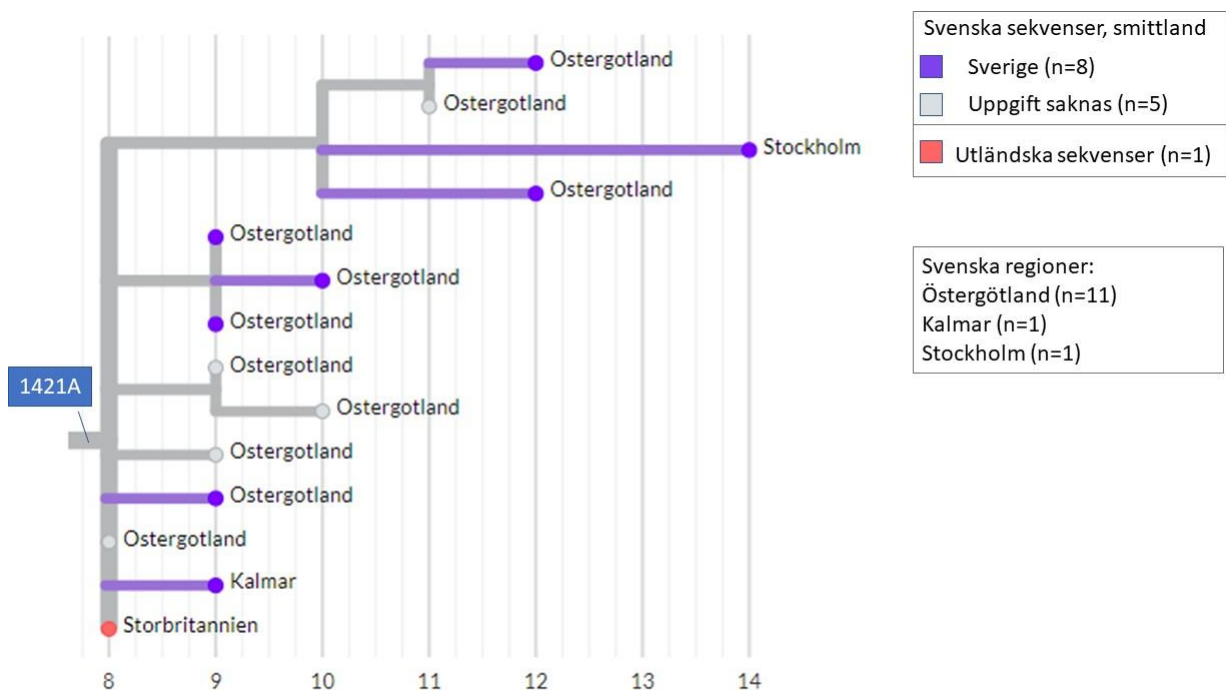
Figur S 4: **20B:C28887T**. Detaljbild av kluster definierat av mutationen C28887T innehållande 25 svenska sekvenser. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



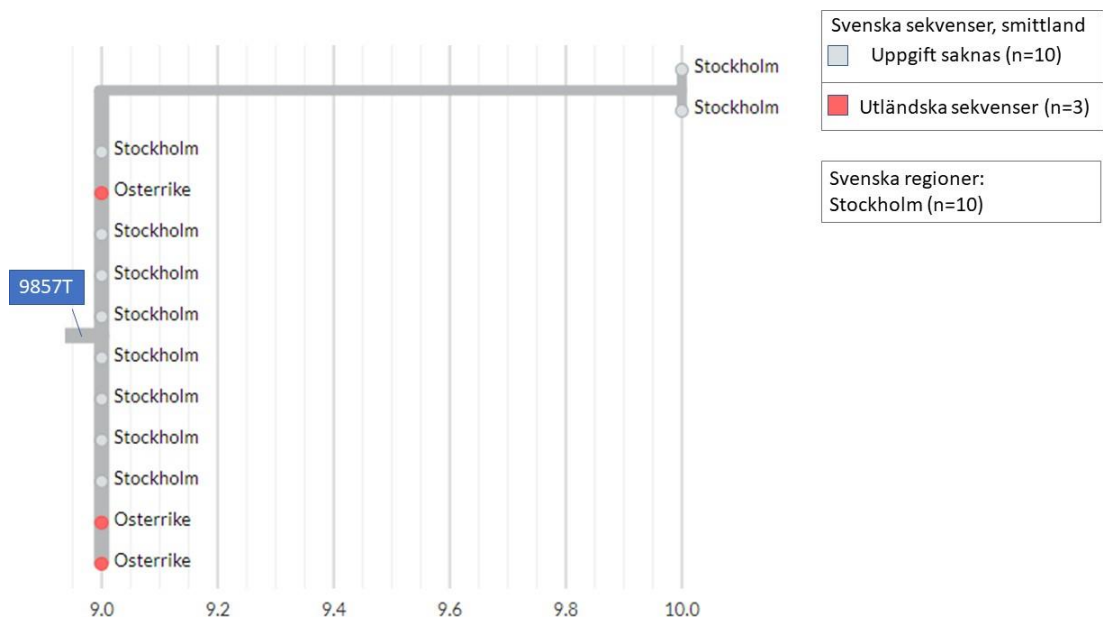
Figur S 5: **20A:A12790G**. Detaljbild av kluster med 19 svenska sekvenser definierat av mutationen A12790G. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



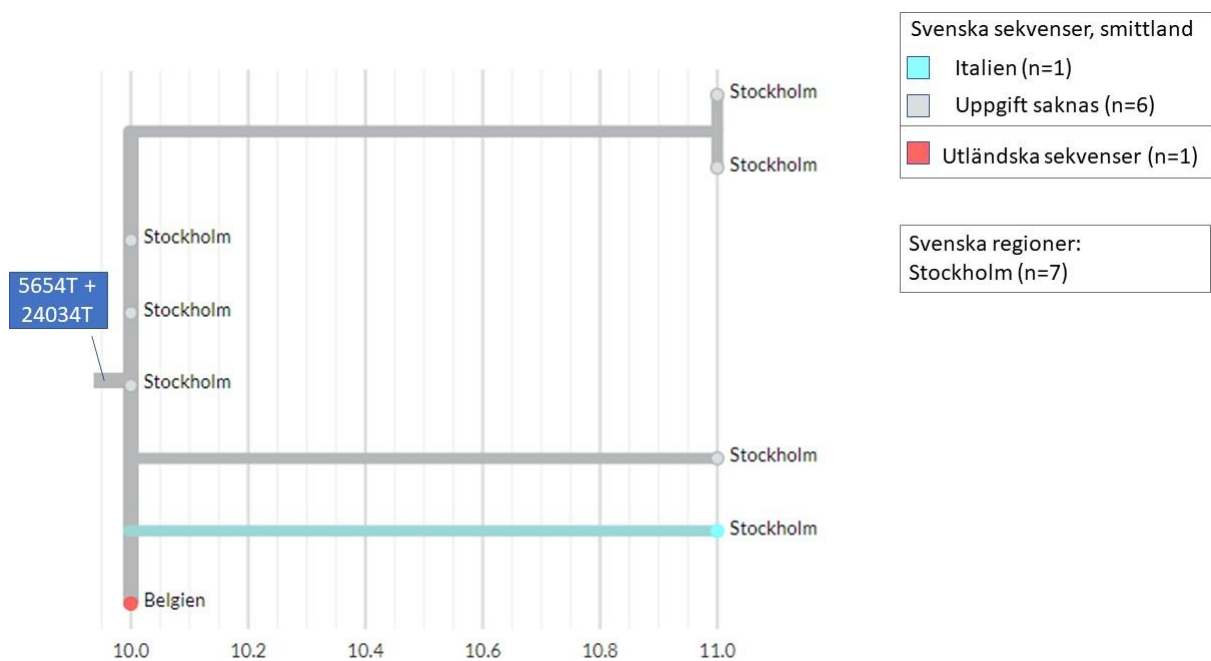
Figur S 6: **20B:C24865T**. Detaljbild av kluster med 14 svenska sekvenser definierat av mutationen C24865T. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



Figur S 7: **20B:G1421A**. Detaljbild av kluster i 20B med 13 svenska sekvenser definierat av mutationen G1421A. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



Figur S 8: **20B:C9857T**. Detaljbild av kluster med 10 svenska sekvenser definierat av mutationen C9857T. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.



Figur S 9: **20B:C5654T,C24034T**. Detaljbild av kluster med 7 svenska sekvenser definerat av mutationerna C5654T och C24034T. X-axeln anger antal mutationers skillnad från referenssekvensen Wuhan-Hu-1.