

# Spåra hepatitis C

## Lärarhandledning

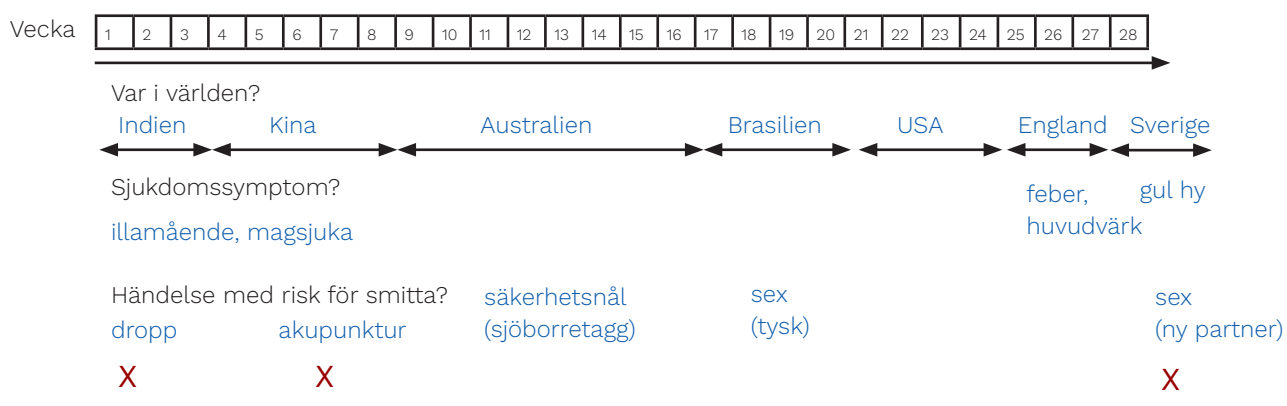
Den här bioinformatikövningen kan användas i undervisning om genteknik eller i avsnitt om kropp och hälsa om smittsamma sjukdomar. Övningen ger ett exempel på hur man kan använda databaser för systematiska undersökningar.

Elevinstruktionen inleds med bakgrundsinformation om patientens resa, fakta om hepatitis C-infektion och vanliga smittvägar. Sedan presenteras en virussekvens och en handledning till hur man undersöker denna i NCBI's virusdatabas. Uppgifterna beskrivs kort här nedan tillsammans med kommentarer.

Övningen har prövats av lärare på gymnasienivå. Den kan förmodligen även fungera för högstadieelever om man ger lite mer stöd och tätare genomgångar av svar på deluppgifterna (se avsnittet "Didaktiska kommentarer").

### Uppgift 1 - Troligaste smittvägen?

Bilden nedan visar ett förslag på hur informationen som ges från samtalet mellan patienten och läkaren kan sammanställas som stöd för att resonera sig fram till sannolika smittotillfällen. Tänkbara hypoteser om smittotillfället kan tas fram efter att man uteslutit några. Sex med den nya partnern kan uteslutas då det gått för kort tid sedan det jämfört med sjukdomens inkubationstid på en till tre månader (se rött kryss i bilden nedan). Även besöket på sjukhus i Indien med dropp och akupunktur i Kina kan förmodligen uteslutas då det var mer än tre månader innan symptomen, till och med för febern i England. Däremot är säkerhetsnålarna i Australien och den sexuella aktiviteten i Brasilien båda misstänkta smittotillfällen då de ligger inom tidshorizonten för den kända inkubationstiden.



## Uppgift 2 - Virusets ursprung?

Genom praktiskt arbete i en virusdatabas med bioinformatiska verktyget BLAST kan man få fram information om patientens hepatit C-virus liknar något som finns lagrat i databasen. Namnet på viruset i databasen skrivs som *Hepacivirus hominis* men även som *Hepacivirus C*.

### Kommentarer:

I de olika sökstegen öppnas nya webbflikar. Gå tillbaka en flik för att ändra i sökningen.

Tabellen med träffar har många kolumner. Man får gå längst ned på sidan för att kunna scrolla långt till höger för att se kolumnen "Geo Location". Smidigast är att klicka på *Select columns* överst till höger om tabellen och ta bort alla kolumner utom *Accession*, *Coverage*, *Identity*, *Score*, *Geo Location* (och klicka på *Apply*).

**Steg 4:** Det finns tre sekvenser med 99% överensstämmelse med patientens HepC-virus-sekvens (alla med *Geo Location* Australien).

**Steg 7:** BLAST-resultatfönstret visar skillnader från den sekvens man matat in (*user\_sequence*) som röda streck. Man kan börja med att på en översiktlig nivå titta på dessa streck för att se vilka sekvenser som har minst antal röda streck och därmed är mest lika patientens virus. Accessionsnummer som det står att man ska notera är beteckningarna som står i början av varje sekvens (exempelvis MN147316.1).

**Steg 8:** Förstora för att se skillnaderna på nukleotidnivå. Manuell genomgång av sekvenserna ger en uppfattning om likheter och skillnader. Ett "Y" motsvarar C eller T och "R" antingen G/A (osäkerhet i data). Skillnader mellan sekvenserna visas ej som röda streck för skillnader för Y/R men syns när man stegar igenom sekvensen.

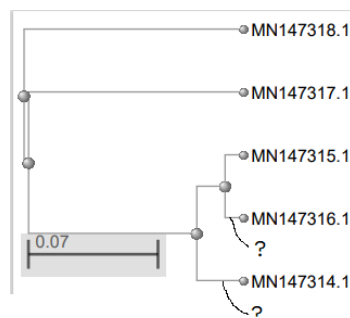
**Steg 9:** Här ska man titta på poängvärden för "Score". Träff med högst värde är den som är mest lik den sekvens vi matat in (patientens virussekvens). Det finns två sekvenser som får samma höga värde (beteckning med 16.1 respektive 14.1 på slutet).

**Steg 10:** Ett fylogenetiskt träd ger inte mycket ny information för smittspårningen, men illustrerar att även virus evolverar. Man kan resonera om att det egna viruset som patienten nu bär på potentiellt, om det sprids vidare, skulle bilda en egen gren i det evolutionära trädets. Men om smittan stoppas så blir det slutet på en gren.

Det fylogenetiska trädets man får fram (se bild nedan) är lite förvånande eftersom inte samma två sekvenser som har högst Score hamnar som närmast släkt. Anledningen är att det finns osäkerheter i sekvenseringen (se tidigare kommentar om Y och R) och att algoritmerna bakom det fylogenetiska trädets och beräkningen av Score hanterar dessa på olika sätt.

I bilden finns inritat två möjliga sätt att visa hur patientens sekvens ("??") skulle kunna bilda en gren tillsammans med antingen den träff som slutar på 16.1 eller 14.1 som är de två som har högst Score-värden.

Den inlagda skalan (0.07) visar bara att längden på grenarna i trädets motsvarar genetiska skillnader (ju längre gren, desto större skillnader).



**Steg 11:** När man klickar på en sökträff visas en ruta med kort information om när och hur viruset samlats in (exempelvis blodprov och tidpunkt då datat registrerats). Det varierar mellan olika sökträffar vilka detaljer som finns

Om man klickar på accessions-numret (se röd pil i bilden intill) som är en kombination av bokstäver och siffror, exempelvis MN123456) kommer man till databasens "hjärta": sidan för en träff i databasen (se bild nedan).

Här finns ofta referenser till vetenskapliga artiklar. Det går inte att få fram namn på personer (integritetskydd). Röda ringar i bilden nedan visar på *Title* på publikationen. Den säger något om vad det är för slags studie som gjorts där sekvensen undersökts och sedan lagts in i databasen.

*Source* talar om vilken organism som sekvensen kommer ifrån och i detta fall med ett virus, även ifrån vilken värdorganism provet tagits (*Homo sapiens*).

*Gene* visar namnet på den gen som undersökts (POLY).

*CDS* betyder *Coding Sequence* och därefter följer sekvensen som kodar för något.

28)	RefSeq Genome (1)
Species	Nucleotide Details
Hepacivirus C	LC435023 Hepatitis C virus subtype 6e N1 for polyprotein, complete cds (Ya
Hepacivirus C	
Hepacivirus C	<b>Attributes</b> <b>Nuc Completeness:</b> partial <b>Length:</b> 9341 <b>Mol Type:</b> RNA <b>Geo Location:</b> Cambodia <b>Collection Date:</b> 2012-08-22
Hepacivirus C	

GenBank

**Hepacivirus C isolate Subject-007663-Week96 polyprotein, RNA-dependent polymerase NS5B region, (POLY) gene, partial cds**

GenBank: MN147314.1  
[FASTA](#) [Graphics](#)

Go to: ☺

LOCUS MN147314 1773 bp RNA linear VRL 10-NOV-2019  
 DEFINITION Hepacivirus C isolate Subject-007663-Week96 polyprotein, RNA-dependent RNA polymerase NS5B region, (POLY) gene, partial cds.  
 ACCESSION MN147314  
 VERSION MN147314.1  
 KEYWORDS .  
 SOURCE Hepacivirus C  
 ORGANISM Hepacivirus C  
 Viruses; Riboviria; Orthornavirae; Kitrinoviricota; Flusviricetes; Amarillovirales; Flaviviridae; Hepacivirus.  
 REFERENCE 1 (bases 1 to 1773)  
 AUTHORS Ragonnet-Cronin,M., Hostager,R., Hedskog,C., Osinusi,A., Svarovskaia,E. and Wertheim,J.O.  
 TITLE HIV co-infection is associated with increased transmission risk in patients with chronic hepatitis C virus  
 JOURNAL J. Viral Hepat. 26 (11), 1351-1354 (2019)  
 PUBMED 31194901  
 REFERENCE 2 (bases 1 to 1773)  
 AUTHORS Ragonnet-Cronin,M., Hostager,R., Hedskog,C., Osinusi,A., Svarovskaia,E. and Wertheim,J.O.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUL-2019) Department of Medicine, University of California, San Diego, San Diego, CA, USA  
 COMMENT ##Assembly-Data-START##  
 Assembly Method :: Vicuna v. 1.3  
 Sequencing Technology :: Sanger dideoxy sequencing; Illumina  
 ##Assembly-Data-END##  
 FEATURES  
 Location/Qualifiers  
 source 1..1773  
 /organism="Hepacivirus C"  
 /mol\_type="genomic RNA"  
 /isolate="Subject-007663-Week96"  
 /isolation\_source="plasma"  
 /host="Homo sapiens"  
 /db\_xref="taxon:11103"  
 /country="Australia"  
 /collection\_date="07-Jan-2015"  
 /note="genotype: 3a"  
 <1..>1773  
 gene /gene="POLY"  
 <1..>1773  
 CDS /gene="POLY"  
 /codon\_start=1  
 /product="polyprotein"

## Uppgift 3 - Värdering av resultat

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

- I. Om patientens virussekvens är mycket lik en sekvens i databasen visar det att patienten blivit smittad av den person som lämnat prov till databasen.
- II. Om patientens virussekvens är mycket lik en sekvens i databasen visar det att patienten blivit smittad av en person som kommer från samma geografiska område som den som lämnat prov till databasen.
- III. Om flera virussekvenser från samma geografiska område liknar patientens virussekvens stärker påstående II.
- IV. Om det var möjligt att få fram nukleotidsekvensen för det hepatit C-virus som infekterat den person som de facto har smittat patienten skulle man få 100 procent överensstämmelse mellan sekvenserna.

### Kommentarer:

- I. Här kan man motivera ett nej med t ex sannolikhet och när virusdata lagrats. Alla som har hepatit C får inte sina virus lagrade i denna databas.
- II. Här kanske man kan motivera ett ja med att risken för att bli smittad av virus av typen som person X bär på kanske är högre nära området, men beteendet som påverkar spridningen av viruset är viktig. Geografiskt område säger en del, men inte allt då vi har global rörlighet idag.
- III. Flera virusekvenser från Australien liknar patientens virus. Det tyder på ett gemensamt ursprung för den gruppen virus. Och om det är många i samma geografiska område som bär på samma typ av virus så tyder det nog på smitta inom regionen, skulle kunna stärka II.
- IV. Här kan man argumentera för nej med hänvisning till ständiga mutationer och pågående evolution. Kanske kommer frågor om virusets mutationshastighet upp. Hepatit C-virus har bland de högsta mutationshastigheterna bland RNA-virus (se tips på artikel nedan),  $10^{-4}$ /bp/replikation (motsvarar 0.01% sannolikhet att en viss nukleotid mutaterar när viruset replikerar). Detta är så högt att det under en och samma infektion i en person hinner bildas en population av virus där det kan finnas en viss variation mellan enskilda virus. När man skattat mutationshastighet hos människa till  $10^{-8}$ /bp/generation har man jämfört arvs massa mellan föräldrar och barn. Då har det skett flera replikationer på vägen (bildning av könsceller, celldelningar i växande barnet) så sannolikheten är ännu lägre för varje replikation.

### Tips!

Läs artikeln "Upptäckten av hepatit C-viruset" i Bi-lagan nr 3 2020.

Echeverría, N., Moratorio, G., Cristina, J., & Moreno, P. (2015). *Hepatitis C virus genetic variability and evolution*. World journal of hepatology, 7(6), 831–845. <https://doi.org/10.4254/wjh.v7.i6.831>

Studera Sveriges statistik över hepatit C-fall via Folkhälsomyndighetens webbplats. Sök på hepatit C (<https://www.folkhalsomyndigheten.se/>)

## Uppgift 4 - Diskutera vilken smittväg som är mest trolig

Vilka smittvägar är tänkbara? Är någon mer sannolik än andra? Motivera med stöd av de slutsatser du kan dra utifrån analysen av virussekvensen med sjukdomens biologi och patientens resa.

### Kommentarer:

Finns det ett rätt svar på uppgiften? Nej, flera osäkerheter gör att det är svårt att säkert avgöra smittotillfället. Men det går att utesluta några saker. Öppenheten i svaret ska stimulera till diskussion och övning i att formulera argument.

Virusinformationen från databasen pekar på likheter med hepatit C-virus som registrerats hos personer som undersökts i Australien. Men personer med andra nationaliteter kan ju mycket väl också ha virus med liknande sekvenser om vi tänkar på vilken global rörlighet som råder bland människor idag. Exempelvis kan den person från Tyskland som nämns (sexuell kontakt i Brasilien) mycket väl vara smittbärare av ett virus med "australiensiskt ursprung". Hen kan ju ha blivit smittade av någon som haft virus med samma ursprung som det som matchar i virusdatabasen. Sjöborre-händelsen med säkerhetsnålen involverar en situation där stickverktyg (säkerhetsnålen) delar mellan människor. Eftersom det är en smittväg som är välkänd för viruset och det dessutom har skett inom en tidsperiod som kan stämma med den symptombild som patienten uppvisar (feber inom 1-3 månader efter händelsen)

Det som är lagrat i databasen kommer ofta från forskningsstudier där patienter som haft Hepatit C ingått. Det är alltså inte en komplett databas på något sätt över alla smittade som man känner till. Exempelvis finns bara ett fåtal sekvenser som överensstämmer till 99% med den sekvens som eleverna testar. Men skulle man provta fler människor skulle antalet möjliga sekvenser bli fler.

Inkubationstiden för sjukdomen är oftast 1-3 månader. Det skulle kunna användas för att utesluta att smittan skedde under sjukhusvistelsen i Indien, eftersom att då borde patienten ha märkt av symptom av feber och gulnande hud med mera tidigare under sin resa. Med samma argument kan det också uteslutas att det är den nya förälskelsen efter hemkomst som är den som smittat (det har gått för kort tid).

## Uppgift 5 - Smittspårning - planera samtal

- Vilka personer är viktiga att kontakta för att hindra fortsatt smitta av hepatit C?
- Vilken information behöver tas upp i ett samtal som är en del av smittspårning?
- Hur kan man bidra till ett bra samtal?

### Kommentarer:

Tanken med den här uppgiften är att komma in på betydelsen av att våga kommunicera för att bryta en smittkedja. Betydelse av att vara ärlig och berätta om fakta man fått fram om viruset. Kanske ta upp betydelsen av att använda kondom för att skydda sig mot den här typen av sexuellt överförbara sjukdomar. Smittväg via sprutdelning och/eller undermålig hygien vid omhändertagande av sår och vid sjukvård kan ju vara bra att ta upp också.

Frågor om skam och skuld - hur hanterar man det? Kanske omöjligt att komma fram till rätt svar på det, men tanken är att genom att lyfta frågorna öka elevernas handlingskompetens - veta att samtal behövs och att samtalen kan påverka hur en smitta lever vidare eller inte.

## Uppgift 6 - Utvärdera övningen

- I övningen har du använt en virusdatabas och verkyget BLAST. Har övningen påverkat dina kunskaper i och om bioinformatik? Berätta!
- Har arbetet väckt några frågor?

### Kommentarer:

Bioinformatik är stort inom biologin, men märks ännu lite i skolämnet biologi. Trots att initiativ tagits till skolövningar i bioinformatik är det ännu förhållandevis få elever som möter bioinformatik i biologiundervisningen internationellt. Det finns relativt få didaktiska studier på elevers lärande i bioinformatik. Elevers hinder i övningar kan ofta bero på bristande kunskaper i biologi snarare än själva bioinformatik-tekniken. Kontakta gärna oss på Bioresurs med återkoppling kring hur denna övning fungerat för dig: [info@bioresurs.uu.se](mailto:info@bioresurs.uu.se).

### Gör ett eget case

Ett förslag på fördjupning är att låta eleverna konstruera egna kluriga uppgifter. Enklast är kanske att använda samma arbetsgång som i den befintliga uppgiften om hepatit C, det vill säga att man utgår från ett fynd av en virussekvens.

Den separata instruktionen som heter "Gör ett eget case" visar hur man gör för att skapa den kända virussekvensen. Instruktionen är skriven för att leta just efter hepatit C-sekvenser, men det går att på liknande sätt hitta sekvenser för många andra virus i samma databas.

Det går också att öppna upp för att göra helt andra typer av "case" för andra organismer och andra sammanhang om man använder en databas för många fler arter, exempelvis NCBI's databas *Nucleotide*.

### Didaktiska kommentarer

Forskning om bioinformatikundervisning pekar på att bioinformatik bör integreras i undervisningen genom intressanta kontexter snarare än att man bara provar på verktygen. Lärare som prövat övningen med gymnasieklasser har rekommenderat att Uppgift 1 görs klart någon lektion innan arbetet med BLAST. Uppgift 1 ger, förutom sammanhanget, träning i att bearbeta och tolka information samt att formulera hypoteser. Eventuellt kan man låta eleverna läsa artikeln "Upptäckten av hepatit C-viruset" i Bi-lagan nr 3 2020 innan uppgift 2.

Enligt forskning om bioinformatikundervisning kan det vara bra att genomföra någon slags "pappersövning" som ger en grundläggande förståelse för principen bakom en BLAST-sökning innan man går in i den digitala miljön. Sekvensinpassning (alignment) kan exempelvis demonstreras genom att visa ett exempel, eller att eleverna får olika utklippta sekvenser som ska jämföras med en analog databas (ett papper med några givna sekvenser).

BLAST och Score-värden kan illustreras med att alternativa kostnadsscheman (algoritmer) ger olika resultat (se exempel nedan). I BLAST prövas inmatad sekvens mot de i databasen med en viss algoritm och ger träffar med så höga poäng som möjligt. Antagandet är att högt score motsvarar homologí (evolutionärt släktskap).

Sekvens 1: ATCGGCCAGTTTAT  
          || | | || |  
Sekvens 2: ATACGGAGTTTATT

Sekvens 1: AT-CGGCCAGTTTAT  
          || ||| | |||||  
Sekvens 2: ATACGG--AGTTTAT

Kostnadsschema I (inte tillåtet öppna):  
matchning ger +1, mismatch ger -1.  
Resultat: -2 poäng.

II: öppning -3 och -1 för förlängning,  
matchning +1 och mismatch -1.  
Resultat: 5 poäng