

## Henrik Brändén

Vetenskapsrådet  
103 78 Stockholm  
Sverige

### Undersökning av influensavirus med hjälp av släkträd

#### Introduktion

Influensavirus delas in i olika stammar beroende på vilka varianter viruset bär med avseende två ytproteiner. Dessa ytproteiner kallas hemagluttinin (HA) och neuroaminidas (NA). De olika stammarna bär därför namn som H5N1 eller H4N2. Idag klassificeras virus genom att man analyserar DNA-sekvensen för dessa proteiner. Därigenom kan man inte bara snabbt konstatera vilken stam ett virus tillhör utan också släktskapen mellan olika virus och man kan dessutom få ledtrådar till att förklara spridningen av viruset.

I denna övning ska du få sekvensen för hemagluttinigen som hittats hos en influensasjuk människa eller från en upphittad sjuk fågel. Sedan skall du ta reda på vilken stam viruset kommer från genom att bygga ett släkträd där din sekvens för HA-genen ingår, samt ett antal andra sekvenser från HA-gener från olika stammar av influensavirus som hittats i fåglar och människor runt om i världen.

#### 1. Förberedelser

Till övningen finns tio olika virussekvenser att undersöka, benämnda virus 1 – 10. Ni skall fördela dessa mellan er, så att inte alla arbetar med samma sekvens.

För att hämta sekvenserna:

a. Gå till

<http://forskning.se/nyabiologin/larare/material>, klicka på länken till "Underlag för lärare" och sedan på "Influensasekvenser – underlag övning". I det word-dokument du då får upp, kopierar du den sekvens du skall arbeta med och skapar ett nytt word-dokument, där du klistrar in sekvensen.

b. Om du får problem med radlängden i worddokumentet, gå in på Arkiv>Utskriftsformat och välj liggande papper, samt anpassa radlängden och välj eventuellt en mindre teckenstorlek. Notera på ett papper varifrån sekvensen är tagen.

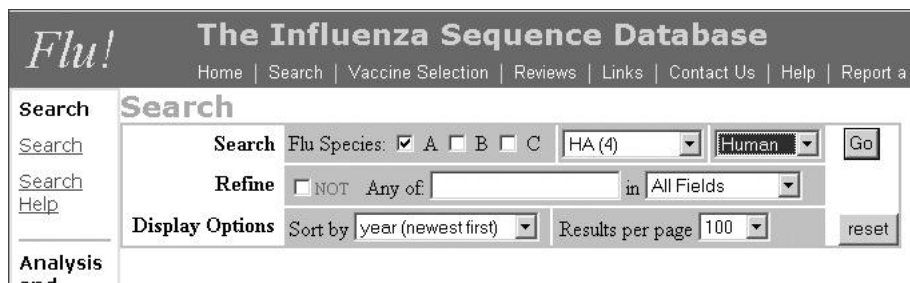
c. Ovanför sekvensen skriver ni en rad där det står ">Min\_okända\_gen" (Med understrukna mellanslag.)

KORRESPONDENS TILL  
Henrik Brändén  
Email: [henrik.branden@vr.se](mailto:henrik.branden@vr.se)

2. Hämta fler sekvenser till släkträdet

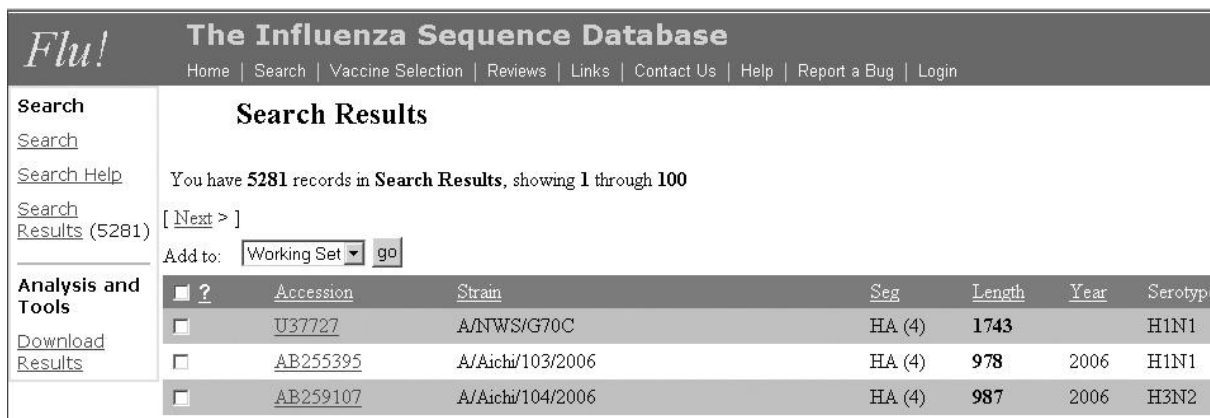
Du skall nu gå till en databas och hämta 8 – 10 ytterligare hemagluttinin-sekvenser från olika stammar av influensaviruset, hittade hos olika arter i olika delar av världen.

- a. Gå till <http://www.flu.lanl.gov>. Klicka på "Search" i listan upptill på sidan.



Figur 1.

- b. På sidan du då ser (Fig 1), gör du följande val: vid "Flu species", klicka "A". I fönstret där man kan välja segment av viruset, välj "HA(4)". I fönstret invid där man kan välja värddjur, välj "Human". Klicka på "Go".
- c. Du får nu upp en lång lista (Fig 2) på olika sekvenser från influensavirus hittade i människa i olika delar av världen. Välj en post där sekvensens "Length" anges till över 1600 baser. Klicka på den understruken koden till vänster på raden.



Figur 2.

- d. Du kommer nu till en sida (Fig 3) med sekvens och (oftast) information om var och när provet är taget. Om sådan information saknas, välj en annan sekvens.

ISD Nucleotide Record	
Accession	ISDN140818
Gi	ISDN140818
Locus	ISDN140818
Strain	A/Indonesia/538H/2006
Definition	
Source	
Collection Country	Indonesia
Length of Sequence	1716
Collection Date (YYYY)	2006
Host Species	Human
Segment	4
Serotype	H5N1
DBComments	ISOLATE_DATE: 05/09/2006 additional information: Bronchial suction from Indonesia Case 538 (19M)
Comments	Direct Web Submission by h5. Virus provided by Drs T. Soendoro & N. Kandun, Ministry of Health, Republic of Indonesia.
Raw Sequence	<pre> ATCTGTCAAAATGGAGAAATAGTGCTTCTTCTTGCAACAGTCAGTCTTGTAAAAAGTGATCAGATTGCATTGG TTACCATGCAAAACAATTCAACAGAGCAGGTTGACACAAATAATGGAAAAGAACGTCACGTGCACACATGCCAAGA CATACTGGAAAAGACACACACACGGGAAGCTCTGCGATCTAGATGGAGTGAAGCCTCTAATTTTAAGAGATTGTAG TGTAGCTGGATGGCTCCTCGGGAACCCAATGTGTGACGAATTCATCAATGTACCGGAATGGTCTTACATAGTGGGA GAAAGGCCAATCCAGCCAATGGCCTCTGTGTACCCAGGGAATTTCAACGACTATGAAGAACTGAAACACCTATTGAG CAGAATAAATCATTTTGA AAAA ACTTCAGATAATCCCAAAAGTTCTTGGTCCGATCATGAAGCCTCATTAGGGGT GAGCTCAGCATGTCATACCTGGGAAGGTCCTCCTTCTTTAGAAATGTGGTATGGCTTATCAAAAAGAACATAC ATACCCAAACAATAAAGAGAAGCTACATAATAATACCAACCAAGAAAGATCTTTTGGTACTGTGGGGGATTCACCATCC TAATGATGCAGCAGAGCAGACAAGGCTATATCAAAAACCAACCACTTATATTTCCGTTGGGACATCAACACTAAA CCAGAGATTGGTACCAAAAATAGTACTAGATCCAAAAGTAAATGGGCAAAAGTGAAGGATGGAATTCCTCTGGAC </pre>

Figur 3.

e. När du valt en sekvens, återvänd till ditt worddokument, och skriv på raden efter den tidigare inklistrade sekvensen en rad med information om den nya sekvensen. Du skall ange virusets värddjur, plats och år det hittats samt virusets stam (serotyp). Börja raden med ">" och använd understrukena mellanslag för att separera ord. Raden kan t ex se ut så här:

>human\_cambodja\_2004\_H5N1.

f. Återvänd till databasens sida med sekvensen och kopiera denna. Klistra in den i ditt worddokument så att den börjar på raden efter den identifikationsrad du just skrivit.

g. Sedan fortsätter du på samma sätt och fyller på ditt worddokument med 4-5 ytterligare DNA-sekvenser från virus funna i människor. Varje sekvens du klistrar in skall föregås av en identifikationsrad, som specificerar var och när viruset är hittat och vilken stam det tillhör. Välj bland de olika sekvenserna i databasen enligt följande principer: (a) Sekvensen ska vara minst 1600 baser lång (annars är endast del av genen sekvenserad). (b) Ta med virus av olika stammar (serotypes). (Observera att H3 är lika med H3N2, och att H1 innebär H1N1). Om du inte hittar lämpliga sekvenser på den sida som visas, scrolla och använd "Next-knappen" i början eller slutet av sidan.

h. När du har 4-5 sekvenser från virus funna i människan går du tillbaka till "Search" (se Fig 1) och ändrar sökningen till "avian". Välj därefter ut 4 -5 olika se-

kvenser från virus i fåglar efter samma principer som tidigare (minst 1600 baser, olika orter, olika stammar). Klistra in dessa en efter en i worddokumentet, föregångna av var sin rad i samma form som tidigare (värdart, plats, år och virusstam).

### 3. Redigera dokumentet

Nu ska du redigera worddokumentet, så det får en form som accepteras av det dataprogram som gör släkträd.

- a. Genom att på Arkiv>Utskriftsformat välja liggande papper, och mixtra med teckenstorlek och radlängd kan du få raderna att rymmas i dokumentet.
- b. Kontrollera att varje sekvens du klistrat in föregås av en rad med information om vilken art sekvensen kommer från och att dessa rader börjar med ett ">", inte innehåller några "tomma" mellanslag utan bara\_tecken\_och\_understrukna\_mellanslag.
- c. Se till att det inte finns några tomma rader mellan eller inne i de olika sekvenserna, samt att det inte finns några mellanslag i början av raderna. Mellanslag inne i sekvenserna är däremot inget problem, inte heller om raderna är olika långa. Ett exempel på hur ditt worddokument kan se ut efter redigeringen visas i rutan nedan.

```
>Hum_Irak_2006_H5N1
GATCAGATTTGCATTGGTTACCATGCAAAACAACCTCGACAGAGCAGGTTGACACAATAATGGAAAAGAACGTAAC
GTTACACACGCCCAAGACATACTGGAAAAGACACACAACGGGAAACTCTGCGATCTAGATGGAGTGAAAGCCTCTA
ATTTAAGAGATTGTAGTGTAGCTGGATGGCTCCTCGGGAACCCAATGTGTGACGAATTCCTCAATGTGCCGGAA
TGGTCTTACATAGTGGAGAAGATCAATCCAGCCAATGACCTCTGTTACCCAGGGAATTTCAACGACTATGAAGAA
CTGAAACACCTATTGAGCAGAATAAACCATTTTGGAGAAAATTCAGATCATACCCAAAAGTTCTTGGTCAGATCAT
GAAGCCTCATCAGGGGTGAGCTCAGCATGTCCATACCAGGGAAGGTCCTCTTTTTAGAAAATGGTGGATGGCTT
ATCAAAAAGGACAATGCATACCCAAACAATAAAGAGAAGTTACAATAATACCAATCAAGAAGATCTTTTGGTACTG
TGGGGGATTCACCATCCAAGTGATGCGGCAGAGCAGACAAGGCTCTATCAAAAACCAACCCTATATTTCCGTT
GGGACATCAACACTAAACCAGAGATTGGTACCAAAAATAGCTACTAGATCCAAGGTAAACGGGCAAAAGTGGAAAG
ATGGAGTTCTTTGGACAATTTAAAACCGAATGATGCAATAAACTTTGAGAGTAATGGAAAATTCATTGTCTCA
GAAAATGCATACAAAATTGTCAAGAAAAGGGGACTCAACAATTATGAAAAGTGAATTGGAATATGGTAACTGCAAC
ACCAAGTGTCAAACTCCAATAGGGGCGATAAATCTAGTATGCCATTCCACAACATCCACCCTCTACCATAGGG
GAATGCCCAAAATATGTGAAATCAAACAGATTAGTCTTGGGACTGGGCTCAGAAATAGCCCTCAAGGAGAGAGA
AGAAGAAAAAGAGAGGATTATTTGGAGCTATAGCAGTTTTATAGAGGGAGGATGGCAGGGAATGGTAGATGGT
TGGTATGGGTACCACCATAGCAATGAGCAGGGGATGGGTAACGTTGTTTCGAGTCTATCAGACAAAAGAATCCACT
GATGGAGTCAACAATAAGGTCAACTCGATCATTGACAAAATGAACACTCAGTTTGGAGGCTGTTGGAAGGGAATTT
AATAACTTAGAAAGGAGAATAGAAAATTTAAACAAGAAGATGGAAGACGGATTCTAGATGTCTGGACTTATAAT
GCTGAACTTCTGGTTCTCATGAAAAATGAGAGAACTCTAGACTTTTATGACTCAAATGTCAAGAACCTTTACGAC
AAGGTCCGACTACAGCTTAGGGATAATGCAAAGGAGCTTGGTAAACGTTGTTTCGAGTCTATCAGATGTGAT
AATGAATGTATGGAAGTGTAAAGAAACGGAACGTATGACTACCCGAGTATTCAGAAGAAGCAAGATTAAGAA
GAGGAAATAAGTGGAGTAAAATTGGAATCAATAGGAACTTACCAAACTAGTCAATTTATTCAACAGTGGCGAGC
TCCCTAGCACTGGCAATCATGGTGGCTGGTCTATCTTTATGGATGTGCTCCAATGGATCGTTACAATGCAAA
>Hum_Kambodja_2005_H5N1
ATGGAGAAAATAGTGCTCTTTTTGCGATAGTCAGTCTTGTTAAAAGTGATCAGATTTGCATTGGTTACCATGCA
AACAACCTCGACAGAGCAGGTTGACACAATAATGGAAAAGAACGTTACTGTTACACATGCCCAAGACATACTGGAA
AAGACCCACAACGGGAAGCTCTGTGATCTAGATGGAGTGAAGCCTCAATTTTGGAGAGATTGTAGTGTAGCTGGA
TGGCTCCTCGGAAACCAATGTGTGACGAATTCATCAATGTGCCGGAATGGTCTTACATAGTGGAGAAGGCCAAT
```

### 4. Låt datorn göra ett släkträd

Du är nu klar att låta en dator jämföra dessa sekvenser med varandra.

- a. Gå till <http://align.genome.jp>. Kopiera och klistra in dokumentet du gjort med sekvenserna i den stora

rutan en bit ner på sidan (Fig 4), se till att du har valt DNA, inte protein i raden strax ovanför rutan, klicka på "Execute multiple alignment" och vänta.

Figur 4.

b. Programmet undersöker nu den parvisa likheten mellan de olika sekvenser du matat in. Efter en stund skapas en sida med namnet "CLUSTALW RESULT". Scrolla till botten av sidan där du ser ett fönster, där du kan välja olika slags släkträd. Välj " N-J Tree with branch length" och klicka på knappen "Exec" till höger.

c. Du får efter en stund en ny sida på skärmen med ett släkträd för dessa gener, avseende hemagluttinin. Genom att peka på trädet med musen och högerklicka, kan du kopiera bilden, klistra in den i ett nytt word-dokument och skriva ut det.

### 5. Analysera trädet

Titta på släkträdet. Beteckningen H1, H3 o s v i början av namnet på virusets serotyp (stam) betecknar vilken variant viruset har av detta hemagluttinin-protein.

#### Fundera över:

- Samlas virus som enligt databasen har samma grundtyp av hemagluttinin på samma huvudgenar i trädet?
- Vilken variant förefaller det virus ha, som du skulle undersöka?
- Kan du urskilja spår av en utveckling även inom de olika virusstammarna?

### Till läraren

*Syfte:* Denna övning illustrerar hur generna för influensavirusets ytprotein hemagglutinin skiljer sig dramatiskt mellan virusstammar och hur denna kunskap utnyttjas när man idag avgör vilken stam ett nyhittat virus kommer ifrån.

*Förkunskaper: Eleverna måste veta*

- vad en gen är
- att mutationer ändrar baser i en gen och att man kan avgöra grad av släktskap genom att studera skillnader i gensekvenser.
- att influensavirus, och i synnerhet generna för deras ytproteiner, förändras mycket snabbt och att detta leder till att förra årets influensa sällan ger en fullständig immunitet mot årets.

*Genväg:*

Om ni har ont om tid, kan eleverna hämta ett dokument där ett antal virussekvenser redan är samlade, och endast föra in sin egen okända sekvens och ytterligare ett par som de hämtar ur databaserna. Ett sådant dokument finns att hämta på

[www.forskning.se/nyabiologin/larare/material](http://www.forskning.se/nyabiologin/larare/material)

>Underlag för lärare

>släkträd influensavirus – underlag lärarkurs.

*Om de "okända" sekvenserna:* De sekvenser eleverna tilldelas är något "muterade" varianter av sekvenser från influensavirus som deponerats i databasen, med korrekt angivelse av var och när de hittats. Anledningen till att sekvenserna är förändrade är bara att undvika att några elever skall råka hitta en exakt kopia av sitt virus i databasen.