



# Tomat och banan – hur är de släkt?

Bildkälla: <https://pixabay.com/>

## Alignment

### Introduktion

I denna övning studeras släktskapet mellan olika växtarter genom att jämföra DNA-sekvenser från olika växtarter (s.k. alignment) och därefter skapa ett släkträd. DNA-sekvenser från olika växtarter kan hämtas från:

- egna sekvenserade DNA-sekvenser. Se uppgiften ”*Tomat och banan - hur är de släkt? Laborationsbeskrivning*”
- sekvenser som hämtats från GenBank, se nedan.
- en fil med färdiga sekvenser från Skolprojekt Linné (bifogas sist i denna övning)

### Bioinformatik

= molekylärbiologi + datavetenskap + matematik

Bioinformatik är en snabbt växande gren inom biologin. Det är kombinationen av nya metoder och kunskaper inom molekylärbiologi, datavetenskap och matematik som gör det möjligt att hantera de enorma datamängder som kommer fram vid analys av protein och DNA. Databaser med sekvenser av nukleotider eller aminosyror från en mängd organismer är fritt tillgängliga. Den största DNA-databasen är *GenBank*, medan *SwissProt* innehåller flest sekvenser från proteiner. Det finns också gratisprogram att hämta på nätet för att arbeta med de stora datamängderna.

Utvecklingen av de levande organismerna kan studeras på molekylnivå och intressanta evolutionära frågor väcks. Inom området systematisk biologi har DNA-data och avancerade matematiska beräkningsmetoder lett till ökad förståelse för släktskap och evolution.

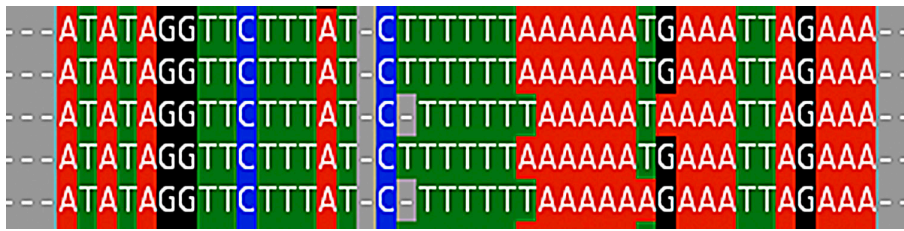
### Alignment

Gratisprogrammet *ClustalOmega* användas för att passa ihop sekvenserna i en matris så att de partier som överensstämmer placeras ovanför varandra. Detta kallas att *aligna*. Sekvenserna kan ha insertioner (en extra DNA-bit) och deletioner (bitar saknas), och om så är fallet ska övriga bitar passas ihop så att rätt delar av sekvensen jämförs med varandra. Detta sker automatiskt i programmet. Se figur på nästa sida med fem korta alignade sekvenser.

Gå in på webbsidan *European Bioinformatics Institute*, [www.ebi.ac.uk](http://www.ebi.ac.uk), välj *Services, DNA & RNA och ClustalOmega*. På den sida som öppnas kopierar du in de sekvenser som ska jämföras i den tomma rutan. Sekvenserna ska läggas in i s.k. *Fasta format*. Det innebär att en rad med först tecknet > och omedelbart efter namnet på växten (inga mellanslag på denna rad) läggs innan själva sekvensen, se figur till vänster. Det får inte vara några siffror i sekvensen, men däremot gör det inget om det är mellanslag mellan nukleotiderna. De förvalda alternativen behålls. Välj *DNA sequences* ovanför rutan med dina sekvenser. Klicka sedan på *Submit*. I det nya fönster som öppnas visas alignment för de valda sekvenserna.

**Fasta format:**

```
>Broccoli  
TCAGAGAAA....
```



Figuren visar DNA-sekvenser från fem olika växtarter där nukleotiderna har markerats med olika färger. Jämför exempelvis sekvens nummer 3 och 5 med övriga sekvenser. Vilka avvikelser finns?

## Släktträd från DNA-sekvenser

Den alignade matrisen analyseras för att få fram ett släktträd genom att klicka på *Phylogenetic Tree*.

Urvalet av arter till släktträdet är mycket betydelsefullt och kan göra att släktträdet inte får det förväntade utseendet och att arter som är närstående inte hamnar intill varandra i släktträdet.

## Släktträd på andra sätt

### Taxonomy

För att ta reda på släktskap mellan organismer kan man använda funktionen *Taxonomy Genbank* (<http://www.ncbi.nlm.nih.gov/guide/taxonomy/>). Välj *Taxonomy Common tree* längre ner på sidan. I det nya fönster som öppnas skrivs valfria arter in genom att ange ett engelskt eller vetenskapligt namn i sökrutan (*Enter name or id*). Ett släktträd byggs efterhand där arterna anges med latinska namn. Trädet byggs utifrån den information som finns i databasen om de valda arterna.

Ett exempel på ett träd som byggts med TaxBrowser finns i uppgiften *Tomat och banan – hur är de släkt? Släktträdsövning*.

### DNA-sekvenser från GenBank

Den sekvens, *trnL* som varit fokus i den praktiska delen i denna serie av övningar kan användas för att söka efter liknande sekvenser i databasen GenBank. Denna sekvens passar för att identifiera skillnader mellan närstående växtarter. En annan sekvens *rbcl* (chloroplast ribulosebiphosphate carboxylase large subunit) kan användas för att analysera växtarter som är mer avlägset släkt. Denna sekvens kommer från kloroplaster och kodar för den större subenheten av enzymet *RuBisCO*. Växter måste ha ett fungerande *RuBisCO* för att kunna fotosyntetisera. Genen som kodar för *RuBisCO* är därför relativt stabil. Generna i kloroplasterna liksom de gener som finns i mitokondrierna ärvs endast från moderplantan till skillnad från generna i cellkärnan, vilka ärvs både från moderplantan och faderplantan.

Gå in på Genbank: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Skriv exempelvis *tomato rbcl* eller *tomato trnL* i sökrutan överst på *GenBanks* webbsida och välj *Nucleotide* i rutan *Search*. Flera träffar visas. Välj en av träffarna som visar den specifika sekvensen (ej hela genomet) och kopiera DNA-sekvensen som finns i slutet av sidan. Använd denna sekvens för att låta GenBank söka efter liknande sekvenser. Gå tillbaka till startsidan för GenBank och välj *BLAST* (t.h.). En ny sida öppnas där *Nucleotide blast* väljs. Kopiera in sekvensen i den tomma rutan. Ta bort siffrorna från sekvensen (mellanslagen behöver inte tas bort) och lägg till en rad överst (>artnamnet) för att få *Fasta format* och klicka på *BLAST* längst ner på sidan. Programmet söker nu efter liknande sekvenser och presenterar dem som en lång lista. Bygg ett släktträd av dessa sekvenser genom att klicka på *Distance tree of results*. Trädet som visas är komplicerat, men ger en uppfattning om vilka möjligheter som finns.

Välj själv de arter du vill undersöka genom att hämta sekvenser från den BLAST-sökning du gjort. Alternativt kan du söka efter varje art specifikt, se överst i föregående stycke. Lättast är nog att spara alla sekvenser i en separat wordfil för att sedan göra alignment och

bygga släkträd, se rubrikerna *Alignment* och *Släkträd från DNA-sekvenser* enligt ovan.

## Referenser

Övningen ingår i Idéhäfte 6 *Efter Linné*, serien *Linnélektioner*. Red. Britt-Marie Lidesten. Nationellt resurscentrum för biologi och bioteknik. 2008

Övningen har utvecklats av Elisabeth Långström, FD, botaniker, informatör, Inst. f. Evolution, genomik och systematisk botanik, Uppsala universitet och Britt-Marie Lidesten (se ovan)

# DNA-sekvenser från några olika växter

## >Grass

GTAAC TTCAAATTCAGAGAAACCCTGGAATTA AAAAAGGG-  
CAATCCTGAGCCAAATCCGTGTTTTGAGAAAACAAAGGGGTTTC-  
TCGAATCGAACTATAATACAAAGGAAAAGGATAGGTGCAGAGAC-  
TCAATGGAAGCTGTTCTAACGAATCGAGTTAATTACGTTGTGTTGT-  
TAGTGGAACTACTTCTAAATTTGAGAAAAGAAGGGCTTTATACATC-  
TAATAAACACGTATAGATACTGACATAGCAAACGATTAATCACA-  
GAATGCATATTATATATAATATGGGTTCTTTATTCTTTTTTAGGTGGAT-  
TAATCGGACGAGGACAAAGAGAGAGTCCCATTCTACATGTCAA-  
TACTGACAACAATGAAATTTCTAGTAAAAGGAAAATCCGTCGACTT-  
TAGAAGTCGTGAGGGTTCAA

## >Broccoli

TCAGAGAAACCCTGGAATTACAATGGGCAATCCTGAGCCAAATC-  
CTGGGTTACGCGAACAAAACAGAGTTTAGAAAGCGGGATAGGT-  
GCAGAGACTCAATGGAAGCTGTTCTACAAATGGAGTTCAATCC-  
CTTGTGTTGAATCAAACGATTCACTTCATAGTCTGATAGATCCTT-  
GGTGGAAC TTATTAATCGGACGAGAATAAAGATAGAGTCCCATTCT-  
TACATGTCAATACTGACAACAATGAAATTTATAGTAAGATGAAAATC-  
CGTTGACTTTTAAAATCGTGAGGGTTCAAG

## >Cress

ATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATC-  
CTGAGCCAAATCCTGGGTTACGCGAACAAACCAGAGTTTAGAAAG-  
CGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAAT-  
GGAGTTCAATCCCTTGTGTTGAATCAAACGATTCACTTCATAGTCT-  
GATAGATCCTTGGTGGAAC TTATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACTGACAACAATGAAATTTATAGTAA-  
GATGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTCAAG

## >Tatsoi

ATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATC-  
CTGAGCCAAATCCTGGGTTACGCGAACAAACCAGAGTTTAGAAAG-  
CGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAAT-  
GGAGTTCAATCCCTTGTGTTGAATCAAACGATTCACTTCATAGTCT-  
GATAGATCCTTGGTGGAAC TTATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACTGACAACAATGAAATTTATAGTAA-  
GATGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTCAAGTC

## >Sprout

ATTCAGAGAAACCCTGGAATTACAATGGGCAATCCTGAGCCAAATC-  
CTGGGTTACGCGAACAAAACAGAGTTTAGAAAGCGGGATAGGT-  
GCAGAGACTCAATGGAAGCTGTTCTAACAAATGGAGTTCAATCC-  
CTTGTGTTGAATCAAACGATTCACTTCATAGTCTGATAGATCCTT-  
GGTGGAAC TTATTAATCGGACGAGAATAAAGATAGAGTCCCATTCT-  
TACATGTCAATACTGACAACAATGAAATTTATAGTAAGATGAAAATC-  
CGTTGACTTTTAAAATCGTGAGGGTTCAAGTCCTCCTATCC

>GreenPepper

TGGAATGAACAAAATGGGCAATCCTGAGCCAAATCCTATTTTCT-  
GAAAACAAATAAAGGTTTCAGAAAAAAAGGATAGGTGCAGAGAC-  
TCAATGGAAGCTATTCTAACAAATGGAGTTAAATGTGTTGGTA-  
GAGGAATCTTTACATCGAACTTCAGAAAGAAAAAGAATGAAGT-  
GAAGGATAAACGTATATACATACGTATTGAATACTATATCAAATGAT-  
TAATGACGACCCGAATCCCTGATCAAATCATTCACTCCATAGTCTGA-  
TAGATCTTTTGAAGAACTGATTAATCGGACGAGAATAAAGATA-  
GAGTCCCGTTCTACATGTCAATACCGGCAACAATGAAATTTATAG-  
TAAGAGGAAAATCCGTCGACTTTTAAAATCGTGAGGGTTC

>SavoyCabbage

ATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATC-  
CTGAGCCAAATCCTGGGTTACGCGAACAAAACAGAGTTTAGAAAG-  
CGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAAT-  
GGAGTTCAATCCCTTGTGTTGAATCAAACGATTCACTTCATAGTCT-  
GATAGATCCTTGGTGGAACTTATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACTGACAACAATGAAATTTATAGTAA-  
GATGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTC AAGT

>Parsley

CGGAATTAATAAAAATGGGCAATCCTGAGCCAAATCCTATTTTC-  
CAAAAACAAACAAAGGCCAGAAAGGTGAAAAAAGGATAGGTGCA-  
GAGACTCAATGGAAGCTGTTCTAACAAATGGAATTGACTGTGTT-  
GCATTGGTAGAGGAATCCTTCCATTGAACTTCCTAAAAGAT-  
GAAGGATATACGTATATACATACGTATACTGAAATAC-  
TCTATCAAATGATTAATGACGACCTTAATCTGTATTTTTCTAT-  
GAAAAAAGGAAAAGTTGTTGTGAATCGATTCCATATTGAAGAAA-  
GAATCGAATATTCATTGATCAAAGCATTACCACACAGTCTGA-  
TAGTTCTTTTGCAGAACTAATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACCGGCAACAATGAAATTTATAG-  
TAAGAGGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTC A

>MangeTout

CTGGAATTCAAATGGGCAATCCTGAGCCAAATCCTTCTTTCTGAAAA-  
CAAATAAAAGTTTCAGAAAGTGAAAATCAAAAAAGGATAGGTGCA-  
GAGACTCAATGGAAGCTGTTCTAACAAATGGAGTTGACAACATT-  
CAATTGATTAATGAAGATTTCTAACTTCTATTTGTA ACTATTTT-  
GATTCTATCACAATGGAAACATTAGAATCAATTACA ACTGGAA-  
GAAAAAATGACTGAATATTCATTGCTCAAATCAGTCACTCCACCATA-  
ATCTGATGGATCTTTTGAATAACTGATTAATCAGACGAGAATAAAGA-  
TAGAGTCCCATTCTACATGTCAATACCGACATCAATGAAATTTT TAG-  
TAAGAGGAAAATCCGTCGACTTTAGAAATCGTGAGGGTTC A

>Rocket

AACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATC-  
CTGAGCCAAATCCTGGGTTACGCAAACAAACCAGAGTTTAGAAAG-  
CGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAAT-  
GGAGTTCAATCCCTTGTGTTGATCAACTGATTCACTTCATAGTCT-  
GATAGATCCTTGGTGGAACTTATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACTGACAACAATGAAATTTATAGTAA-  
GATGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTCAAG

>Corriander

TCAGAGAAACCCCGGAATAATAAAAATGGGCAATCCTGAGC-  
CAAATCCTATTTTCCAAAAACAAACAAAGGCCAGAAGGT-  
GAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAA-  
CAAATGGAATTGACTGTGTTGCATTAGTAGAGGAATCCTTCCACT-  
GAAACTTCGAAAAGATGAAGGATATACGTATATACATACGTA-  
TACGTACTGAAATACTCTATCAAATGATTAATGACGACCTTAATCTG-  
TATTTTTCTATGAAAAAGGGAAAAAATTGTCGTGAATCGATTCCATA-  
TTGAAGAAAAAATCGAATATTCATTGATCAAAGCATTACCACA-  
CAGTCTGATAGTTCTTTTGCAGAACTAATTAATCGGACGAGAATAAA-  
GATAGAGTCCCATTCTACATGTCAATACCGGCAACAATGAAATTTA-  
TAGTAAGAGGAAAATCCGTTGACTTTAAAATCGTGAGGGTTCA

>SpringOnion

TCAGAGAAACCCCTGGAATAAAAATGGGCAATCCTGAGC-  
CAAATCTTTATTTTTGAAAAACAAGGGTTTAAAAAAGAGAA-  
TAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTC-  
TAACGAATGAAGTTGACTACGTTGCGTTGGTAACTCAAATTCTT-  
CTATAACAAAAAATGATTAATCGGACGAGAATAAAGAGAGAGTC-  
CCATTCTACATGTCAATAGCGACAACAATGAAATTTATAGTAAGAG-  
GAAAATCCGTCGACTTTAGAAAATCGTGAGGGTTCA

>Mizuna

ATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGC-  
CAAATCCTGGGTTACGCGAACAACAGAGTTTAGAAAGCGG-  
GATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAATGGAGTT-  
CAATCCCTTGTGTTGAATCAAACGATTCACTTCATAGTCTGATAGATC-  
CTTGGTGGAACTTATT

>GreenCabbage

ATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATC-  
CTGAGCCAAATCCTGGGTTACGCGAACAACAGAGTTTAGAAAG-  
CGGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAAT-  
GGAGTTCAATCCCTTGTGTTGAATCAAACGATTCACTTCATAGTCT-  
GATAGATCCTTGGTGGAACTTATTAATCGGACGAGAATAAAGATA-  
GAGTCCCATTCTACATGTCAATACTGACAACAATGAAATTTATAGTAA-  
GATGAAAATCCGTTGACTTTTAAAATCGTGAGGGTTCAAG