

Biologisk information

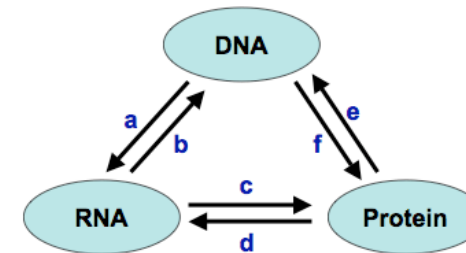
Med fokus på DNA

- Læringsmål
 - Hvad er biologisk information
 - Informations flow
 - Teknikken bag DNA sekventering
 - Fil formater for biologisk data
 - Søgning i GenBank databasen

Gruppearbejde: Flow af biologisk information

- Løses i grupper på 2-3.
- Brug ~15 minutter på opgaven.
- Diskutér i grupperne.
- Opgaven gennemgås i fælleskab.

Quiz: Flow af **information**

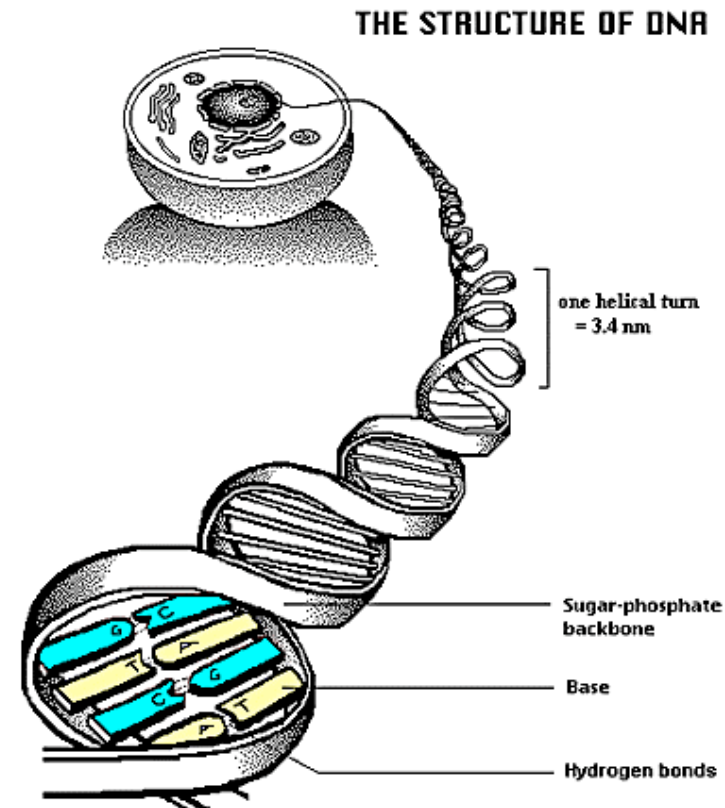


Test af forhåndsviden om flow af biologisk information.

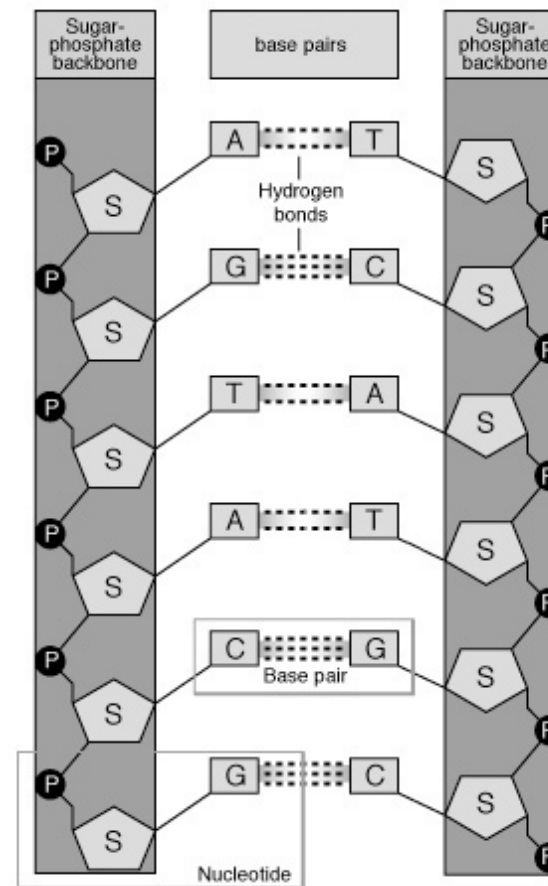
Ovenover ses en skematisk tegning af flow af information mellem tre vigtige typer af biologiske makromolekyler. Det er gruppens opgave at diskutere og beskrive følgende:

- 1) Hvilken rolle spiller hhv. Protein, DNA og RNA i cellen?
Skriv i stikordsform.
- 2) Hvilke af de viste pile (a - f) repræsenterer flow af information i cellen?
Skriv en **kort** beskrivelse af processen for hver pil I mener er relevant.

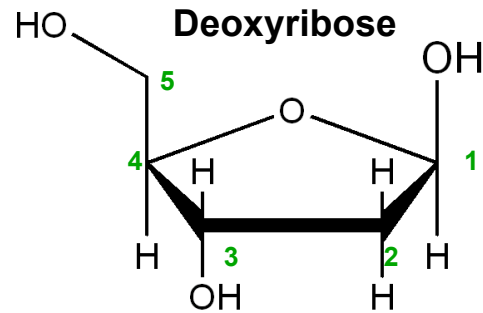
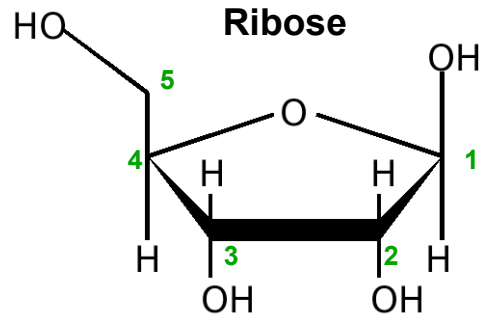
- DNA indeholder cellens opskrift på proteiner og enzymer.
- Ved celledeling får hver dattercelle en komplet kopi af DNA.



- Informationen indeholdt i DNA strengen kan skrives i en fire-bogstav kode: **A**, **T**, **G**, **C**.
- DNA kan sekventeres og resultatet kan gemmes i en fil på en computer.
- ATGGCCCTGTGGAT



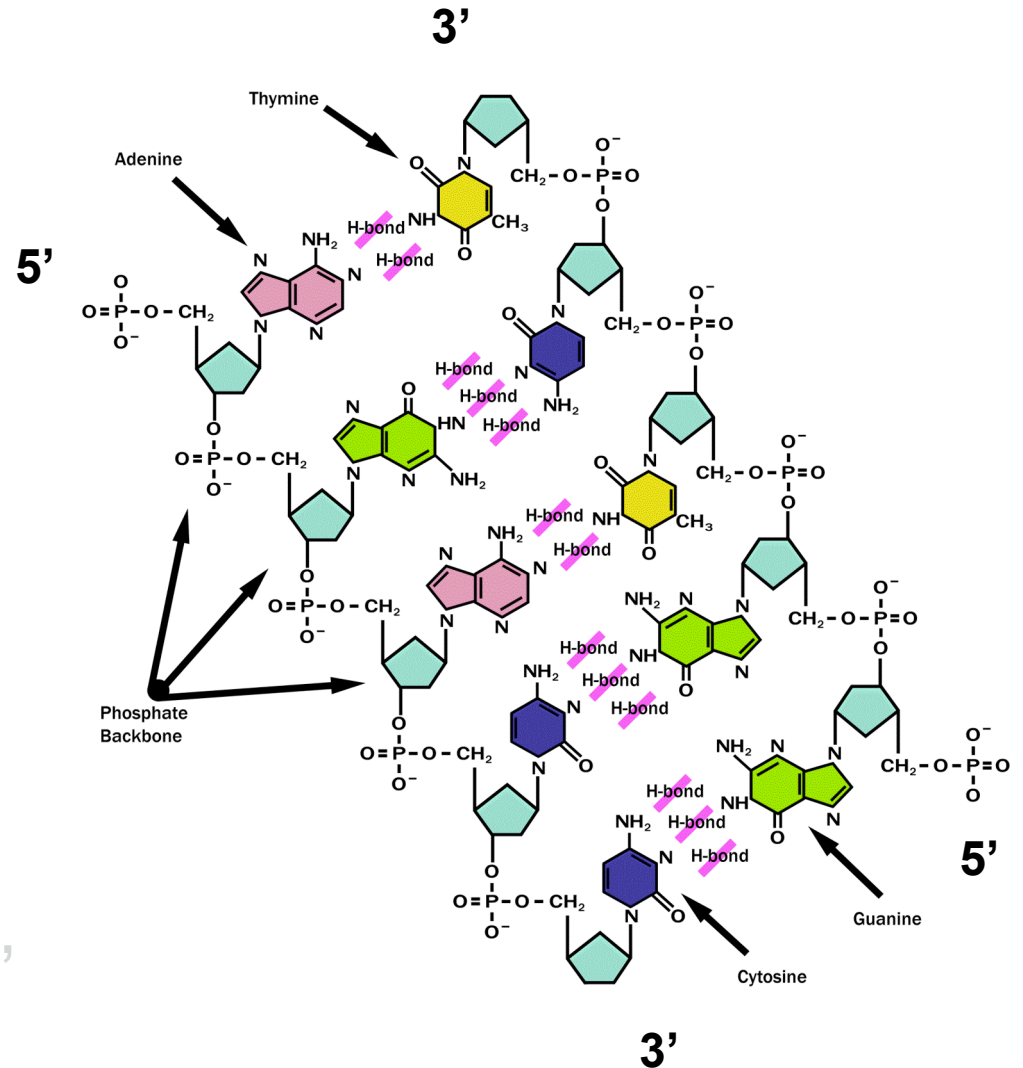
DNA skrives 5' → 3'



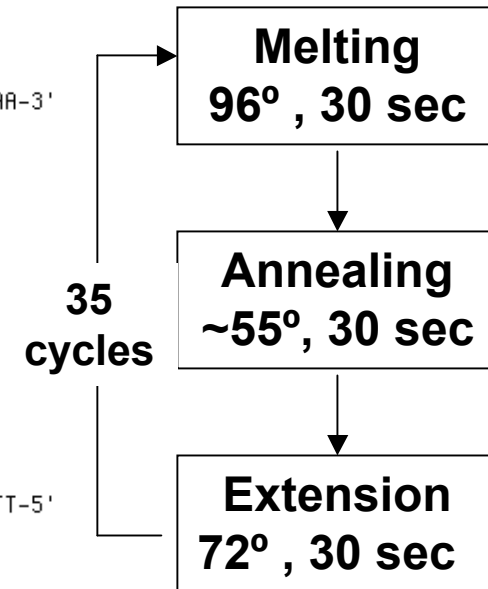
5' AGCC 3'

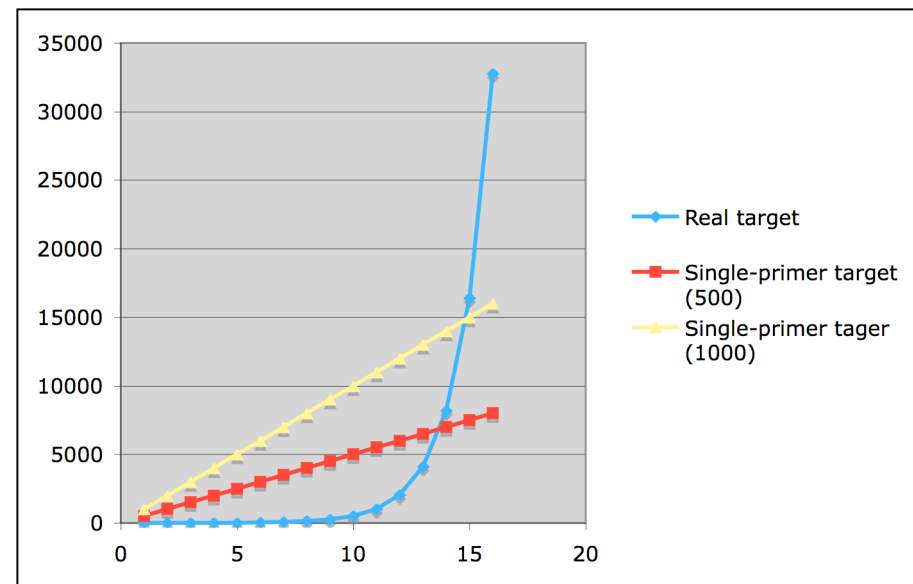
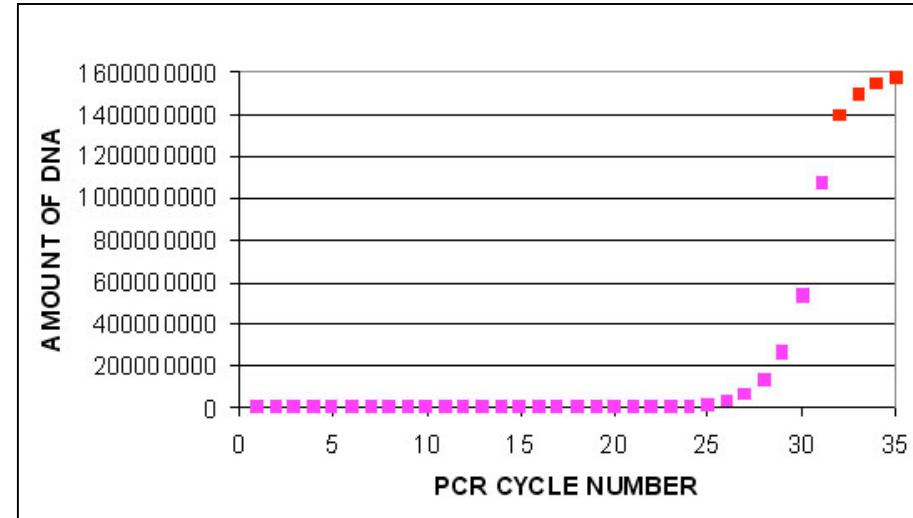
3' TCGG 5'

5' ATGGCCAGGTAA 3'



Cycle 1



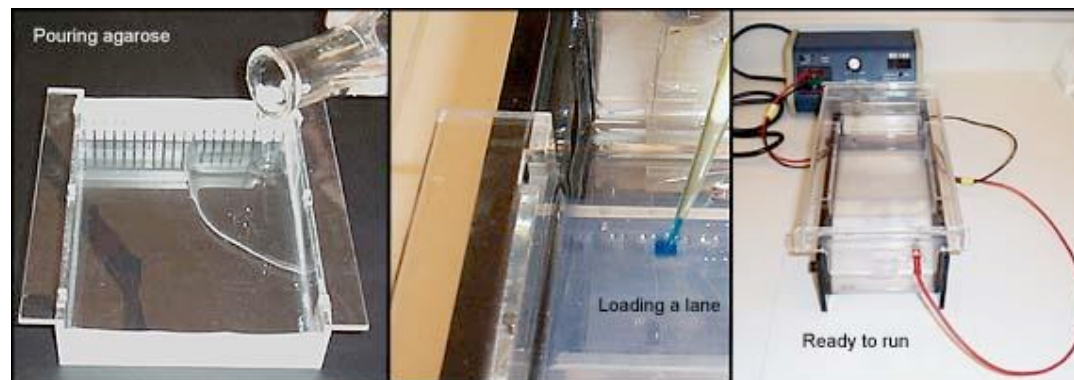
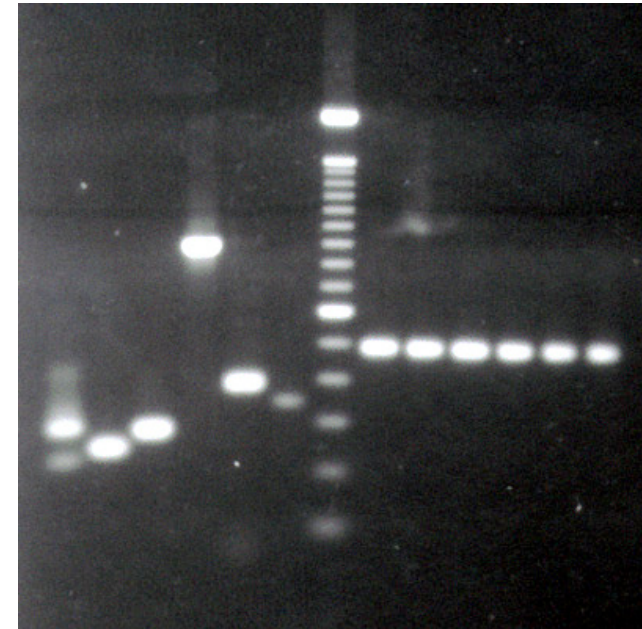


Animation: <http://www.people.virginia.edu/~rjh9u/pcranim.html>

PCR graph: <http://pathmicro.med.sc.edu/pcr/realtime-home.htm>

Gel-elektroforese

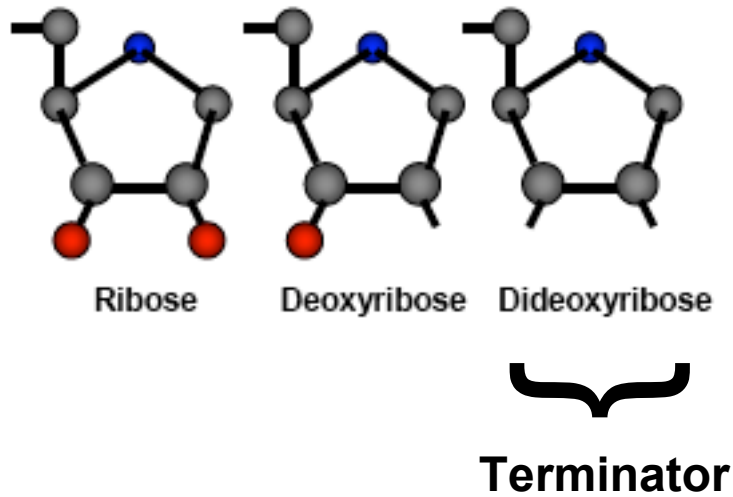
- DNA fragmenterne separeres via gel-elektroforese.
 - Typisk 1% agarose
 - Farves med EtBr (lyser under UV).
 - DNA "ladder" bruges indikation af kendte størrelser.



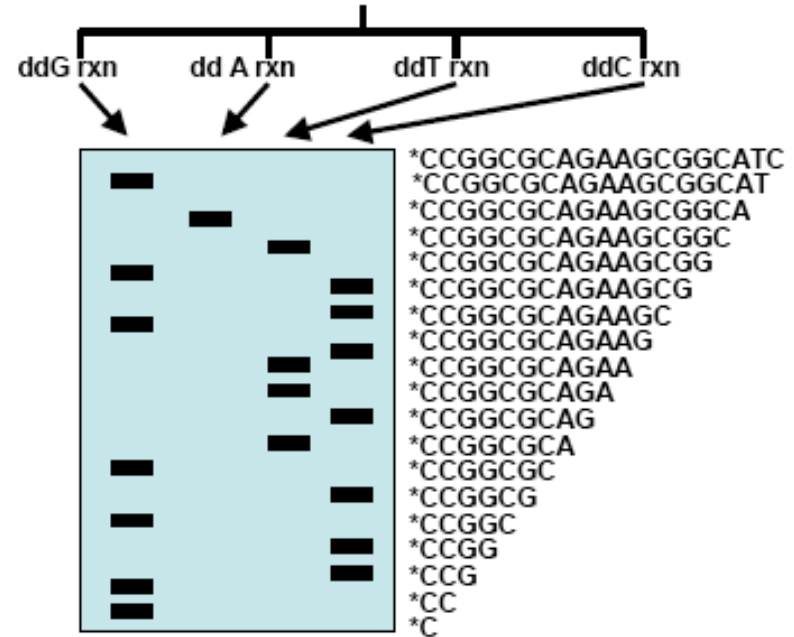
Gel billede: <http://www.pharmaceutical-technology.com/projects/roche/images/roche3.jpg>

PCR opsætning: <http://arbl.cvmb.colostate.edu/hbooks/genetics/biotech/gels/agardna.html>

DNA sekventering



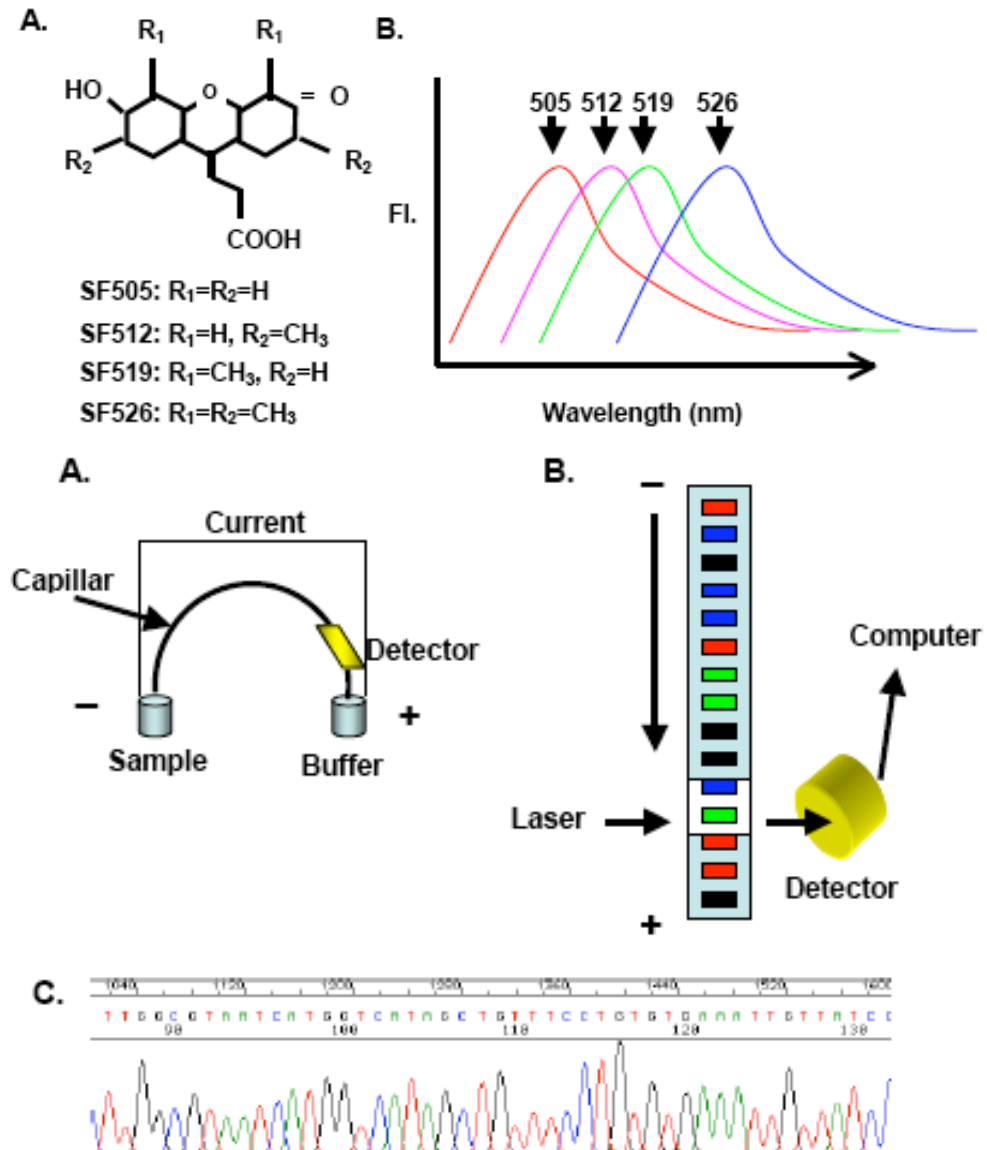
5' pCpCpGpGpCpGpCpApGpApApGpCpGpGpCpApTpCpApGpCpApApA 3'



Røntgen sekventeringsgel

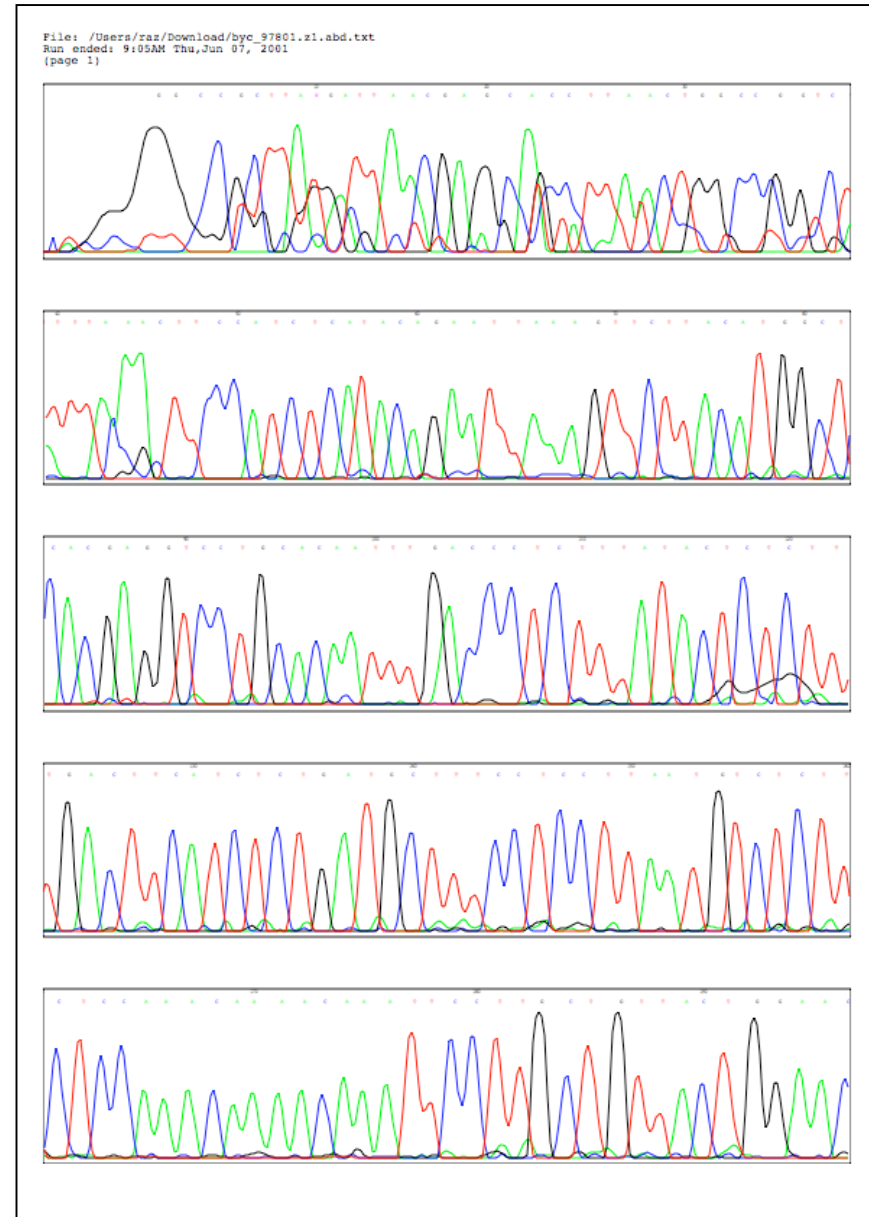
Automatiseret sekventering

- Det store gennembrud for DNA sekventering skete gennem automatisering.
- Fluorescerende mærkning
- Laser-baseret aflæsning
- Kapilær istedet for gel
- Computer baseret analyse
 → samling af sekvensen.



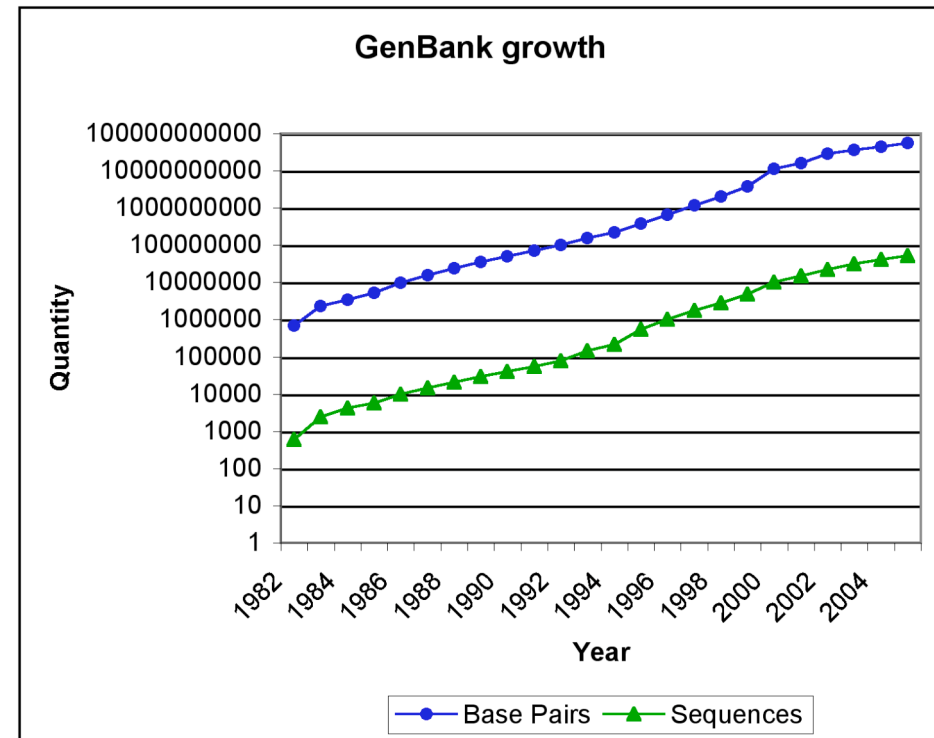
Øvelse: "base-calling"

- Handout: Plot af "trace-file".
- Arbejd sammen i grupper på 2-3.
- Opgave: Find det optimale interval.



- GenBank databasen
- Formater
 - FASTA
 - GenBank

- GenBank er standard databasen for DNA sekvenser.
- Drives af NCBI: *National Center for Biotechnology Information*.
- Har eksisteret siden 1982.
- Databasen er offentlig: Ingen restriktioner på brug af data.



>alpha-D

```
ATGCTGACCGACTCTGACAAGAAGCTGGTCTGCAGGTGTGGGAGAAGGTGATCCGCCAC
CCAGACTGTGGAGCCGAGGCCCTGGAGAGGTGCGGGCTGAGCTTGGGGAAACCATGGGCA
AGGGGGGCGACTGGGTGGGAGCCCTACAGGGCTGCTGGGGGTGTTTCGGCTGGGGGTCAG
CACTGACCATCCCGCTCCCGCAGCTGTTACCACCTACCCCCAGACCAAGACCTACTTCC
CCCCTTCGACTTGCACCATGGCTCCGACCAGGTCCGCAACCACGGCAAGAAGGTGTTGG
CCGCCTTGGGCAACGCTGTCAAGAGCCTGGGCAACCTCAGCCAAGCCCTGTCTGACCTCA
GCGACCTGCATGCCTACAACCTGCGTGTGACCCCTGTCAACTTCAAGGCAGGCGGGGGAC
GGGGGTGAGGGGCCGGGGAGTTGGGGGCCAGGGACCTGGTTGGGGATCCGGGGCCATGCC
GGCGGTACTGAGCCCTGTTTTGCCTTGCAGCTGCTGGCGCAGTCTTCCACGTGGTGCTG
GCCACACACCTGGGCAACGACTACACCCCGGAGGCACATGCTGCCTTCGACAAGTTCCTG
TCGGCTGTGTGCACCGTGTGGCCGAGAAGTACAGATAA
```

>alpha-A

```
ATGGTGTCTGTGCCAACGACAAGAGCAACGTGAAGGCCGTCTTCGGCAAAATCGGCGGC
CAGGCCGGTGAAGTGGGTGGAAGCCCTGGAGAGGTATGTGGTCATCCGTCATTACCCC
ATCTCTTGTCTGTCTGTGACTCCATCCCATCTGCCCCATACTCTCCCCATCCATAACTG
TCCCTGTTCTATGTGGCCCTGGCTCTGTCTCATCTGTCCCCAACTGTCCCTGATTGCCTC
TGTCCCCCAGGTTGTTTCATCACCTACCCCCAGACCAAGACCTACTTCCCCCACTTCGACC
TGTACATGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTGGCGGAGGCACTGGTTG
AGGCTGCCAACCACATCGATGACATCGCTGGTGCCCTCTCCAAGCTGAGCGACCTCCACG
CCAAAAGCTCCGTGTGGACCCCGTCAACTTCAAAGTGAAGCATCTGGGAAGGGGTGACCA
GTCTGGCTCCCCCTCCTGCACACACCTCTGGCTACCCCTCACCTCACCCCTTGCTCACC
ATCTCCTTTTGCCTTTTCAAGCTGCTGGGTCACTGCTTCCCTGGTGGTTCGTGGCCGTCCACTT
CCCCTCTCTCCTGACCCCGGAGGTCCATGCTTCCCTGGACAAGTTCGTGTGTGCCGTGGG
CACCGTCCTTACTGCCAAGTACCGTTAA
```

(Udleveret som handout)

GenBank format

```

LOCUS       CMGL04                1185 bp    DNA     linear   VRT 18-APR-2005
DEFINITION  Cairina moschata (duck) gene for alpha-D globin.
ACCESSION   X01831
VERSION    X01831.1   GI:62724
KEYWORDS   alpha-globin; globin.
SOURCE     Cairina moschata (Muscovy duck)
  ORGANISM  Cairina moschata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
REFERENCE   1  (base 1 to 1185)
AUTHORS     Erdil,C. and Nisasingh,T.
TITLE       The primary structure of the duck alpha D-globin gene: an unusual
            5' splice junction sequence
JOURNAL     SMO 2, 2 (8), 1339-1343 (1983)
PUBMED     10872328
COMMENT     Data kindly reviewed (13-NOV-1985) by J. Nisasingh.
FEATURES             Location/Qualifiers
     source          1..1185
                    /organism="Cairina moschata"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:8855"
     CAAT_signal     20..24
     TATA_signal     69..73
     p1cursor_RNA   101..1114
                    /note="primary transcript"
     exon            101..234
                    /number=1
     CDS             join(143..234,387..591,939..1067)
                    /codon_start=1
                    /product="alpha D-globin"
                    /protein_id="CA23496.2"
                    /db_xref="GI:4455876"
                    /db_xref="Gene:PO2003"
                    /db_xref="InterPro:IPR000971"
                    /db_xref="InterPro:IPR002338"
                    /db_xref="InterPro:IPR002340"
                    /db_xref="InterPro:IPR009050"
                    /db_xref="UniProt/Swiss-Prot:PO2003"
                    /translation="MLTAEKDKLIVYMRVAVGEGEFGSEALQHEPLATPQKTFPP
                    HPLRERQKQVWDRKVAALDAWVSLERLQALRELSRATRELRVDPVDFKLLA
                    DCFQVLAARLHCSTFSDMAADRELSVAVAALDRK"
     repeat_region  227..246
                    /note="direct repeat 1"
     intron          235..386
                    /number=1
     repeat_region  287..309
                    /note="direct repeat 1"
     exon            387..391
                    /number=2
     intron          392..398
                    /number=2
     exon            940..1114
                    /number=3
     polyA_signal    1079..1100
     polyA_signal    1114
ORIGIN
1  ctggtgagcc  taagcccttc  cacccctcca  cgtgataag  ataagccacg  gggggagag
61  cagggtgcta  taagagctgg  gcccgggggg  tgtctcaac  acagaaccc  gtaagtgcc
121  agcctgcaac  gggctggcgg  ccatgtgag  cggcagagac  aagaagctca  togtgaggt
181  gtggaaagac  gtatcgggca  acagagagca  atggaaagc  gaagctctgc  aagatgtgg
241  gctgagccca  gggggcactc  acaggtggg  cagcagggag  caggagccct  gaaggggta
301  tgggtgggga  cccagagcgc  caaggggtg  gggctgagat  gggcaagaca  gaagggcacc
361  aaaaatgact  ggcctctgca  cagagagata  tctctgact  accccagagc  aagaactaac
421  tttcccaact  taagactgca  tccaggtctt  gaacagctcc  gtggcaatg  caagaagatg
481  ggggtgccc  tgggcaatg  cgtgaaagac  ctggacaacc  taagccaggg  cctgtctgag
541  ctcaagcaac  tgaatgctca  caactctggt  gttagccttg  taactcttca  ggaagcggg
601  gactaggttc  ctggctctg  ggggtctcag  gttcaggtt  caaggtctga  ggggtccagg
661  ggtctgagtt  tontgggctc  tggagcttcc  ggggtgtag  ggcagggctc  ctgtgttctt
721  ggttaccagg  gtcttgggg  ctagagccca  gaagcaggg  gctggagttg  aatctggat
781  gtggcccaag  ggttgagtt  ctgcttgaaa  tggagctctg  caaggtgcta  gggccaggtt
841  ggggactctc  gggctcagg  gggactcggg  ggggactgca  gggagactca  gggccatctg
901  tcaggagcag  ggttactcag  ccttggtttg  ccttgagct  gctggcaacg  tggctccagg
961  tggctctggt  ggcacactgt  ggaagagct  acagcccca  gatgagctg  gctttgaca
1021  agttctgtct  ggcctgagct  ggcgtgctgg  ctgaaagata  caagtagacc  actgactgca
1081  cccctgcaac  tcaataaag  acacaattac  caagctctg  tctctgtgtg  tggctggact
1141  gggctcagg  gttccagggt  aggtctggct  tggctccaca  catcc
  
```

Header

Indeholder information ang. Organisme, publikation, Accession ID mm.

FEATURE blok

Indeholder en beskrivelse af forskellige elementer i DNA sekvensen.

CDS: Coding Sequence.
 Indeholder koordinater på den protein kodende del af et gen. Bemærk de tre intervaller.

ORIGIN blok

Indeholder selve DNA sekvensen.

- Stammer fra NCBI's GenBank Database
- Indeholder både DNA sekvens og annotering af "features" - herunder gener.

(Udleveret som handout)

GenBank format - HEADER

LOCUS CMGLOAD 1185 bp DNA linear VRT 18-APR-2005
DEFINITION Cairina moschata (duck) gene for alpha-D globin.
ACCESSION X01831
VERSION X01831.1 GI:62724
KEYWORDS alpha-globin; globin.
SOURCE Cairina moschata (Muscovy duck)
ORGANISM Cairina moschata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
REFERENCE 1 (bases 1 to 1185)
AUTHORS Erbil,C. and Niessing,J.
TITLE The primary structure of the duck alpha D-globin gene: an unusual
5' splice junction sequence
JOURNAL EMBO J. 2 (8), 1339-1343 (1983)
PUBMED 10872328
COMMENT Data kindly reviewed (13-NOV-1985) by J. Niessing.

GenBank format - ORIGIN blok

ORIGIN

```
1 ctgcgtggcc tcagcccctc caccctcca cgctgataag ataaggccag ggcgggagcg
61 cagggtgcta taagagctcg gccccgcggg tgtctccacc acagaaacc gtcagttgcc
121 agcctgccac gccgctgccg ccatgctgac cgccgaggac aagaagctca tcgtgcaggt
181 gtgggagaag gtggctggcc accaggagga attcggaagt gaagctctgc agaggtgtgg
241 gctgggcca gggggcactc acagggtggg cagcaggag caggagccct gcagcgggtg
301 tgggctggga cccagagcgc cacgggtgfc gggctgagat gggcaaagca gcagggcacc
361 aaaactgact ggcctcgctc cggcaggatg ttctcgcct acccccagac caagacctac
421 ttccccact tcgacctgca tcccggctct gaacagggtc gtggccatgg caagaaagtg
481 gcggctgccc tgggcaatgc cgtgaagagc ctggacaacc tcagccaggc cctgtctgag
541 ctcagcaacc tgcatgccta caacctgcgt gttgaccctg tcaacttcaa ggcaagcggg
601 gactagggtc cttgggtctg ggggtctgag ggtgtggggt gcagggctctg ggggtccagg
661 ggtctgagtt tcctggggtc tggcagtcct gggggctgag ggccagggtc ctgtggtctt
721 gggtagcagg gtccctggggg ccagcagcca gacagcaggg gctgggattg catctgggat
781 gtgggcccaga ggctgggatt gtgtttggaa tgggagctgg gcaggggcta gggccagggt
841 gggggactca gggcctcagg gggactcggg gggggactga gggagactca gggccatctg
901 tccggagcag gggactaag ccctggtttg ccttgagct gctggcacag tgcttccagg
961 tgggtgctggc cgcacacctg ggcaaagact acagccccga gatgcatgct gcctttgaca
1021 agttcttgtc cgccgtggct gccgtgctgg ctgaaaagta cagatgagcc actgcctgca
1081 cccttgacc ttcaataaag acaccattac cacagctctg tgtctgtgtg tgctgggact
1141 gggcatcggg ggtcccaggg agggctgggt tgcttccaca catcc
```

//

GenBank format - FEATURE blok

```

FEATURES             Location/Qualifiers
     source            1..1185
                        /organism="Cairina moschata"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:8855"
     CAAT_signal       20..24
     TATA_signal       69..73
     precursor_RNA    101..1114
                        /note="primary transcript"
     exon              101..234
                        /number=1
     CDS               join(143..234,387..591,939..1067)
                        /codon_start=1
                        /product="alpha D-globin"
                        /protein_id="CAA25966.2"
                        /db_xref="GI:4455876"
                        /db_xref="GOA:P02003"
                        /db_xref="InterPro:IPR000971"
                        /db_xref="InterPro:IPR002338"
                        /db_xref="InterPro:IPR002340"
                        /db_xref="InterPro:IPR009050"
                        /db_xref="UniProt/Swiss-Prot:P02003"
                        /translation="MLTAEDKKLIVQVWEKVAGHQEEFGSEALQRMFLAYPQTKTYFP
HF'DLHPGSEQVRGHGKKVAAALGNVKS'LDNLSQALSEL'SNLHAYNLRVDPVNFKLLA
QCFQVVLA'AH'LGKDYSPEMHA'AFDKFLSAVA'AVLAEKYR"
     repeat_region    227..246
                        /note="direct repeat 1"
     intron            235..386
                        /number=1
     repeat_region    289..309
                        /note="direct repeat 1"
     exon              387..591
                        /number=2
     intron            592..939
                        /number=2
     exon              940..1114
                        /number=3
     polyA_signal      1095..1100
     polyA_signal      1114
  
```

Computer-øvelse: GenBank, data og søgning

- Arbejd sammen i grupper af 2-3.
- Der er et link'et til øvelsesvejledningen direkte fra kursusprogrammet.

The screenshot shows the Entrez Nucleotide database interface. At the top, there is a search bar with the text "Search Nucleotide for" and buttons for "Go" and "Clear". Below the search bar are tabs for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The main content area features a yellow box with text: "The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of June 2005, there are over 89 billion bases in GenBank and RefSeq alone." Below this is a section titled "Human Genome" with the text: "Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#)." Further down, there is a section titled "Building the human genome" with text: "The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the [Genome View](#) page and includes an accompanying set of [statistics](#) and [release notes](#)." At the bottom, there is a section titled "Homo sapiens genome view" with the text "build 35 version 1 statistics" and a bar chart showing the number of hits for each chromosome (1-22, X, Y, MT).