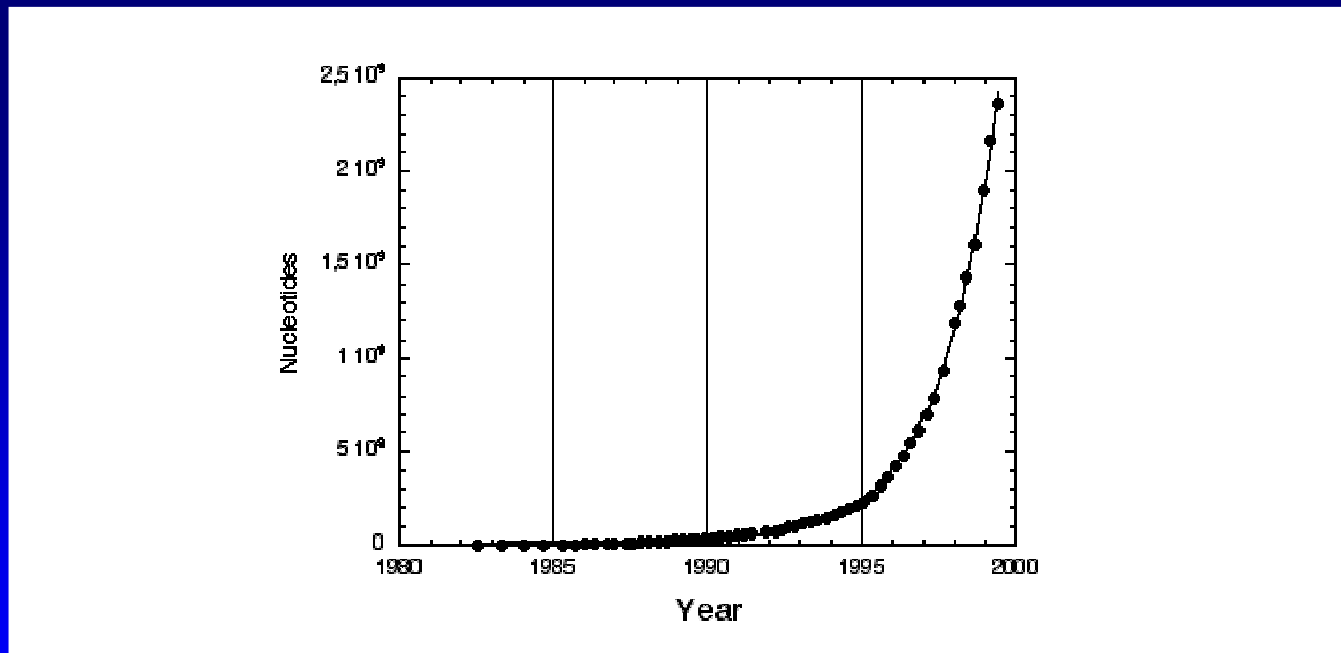


Bioinformatika je nová disciplína na rozhraní počítačových věd, informačních technologií a biologie. Bioinformatika zahrnuje studium biologických dat a jejich praktické uchovávání, vyhledávání a modelování.



Významná data sekvenace DNA

tRNA - (1964) - 75 bases (old, slow, complicated method)

First complete DNA genome: X174 DNA (1977) - 5386 bases

human mitochondrial DNA (1981) - 16,569 bases

tobacco chloroplast DNA (1986) - 155,844 bases

First complete bacterial genome (*H. Influenzae*)(1995) - 1.9×10^6 bases

Yeast genome (eukaryote at $\sim 1.5 \times 10^7$) completed in 1996

Several archaeobacteria

E. coli -- 4×10^6 bases [1997 & 1998]

Several pathogenic bacterial genomes sequenced

Helicobacter pylori (ulcers)

Treponema pallidum (Syphilis)

Borrelia burgdorferi (Lyme disease)

Chlamydia trachomatis (trachoma - blindness)

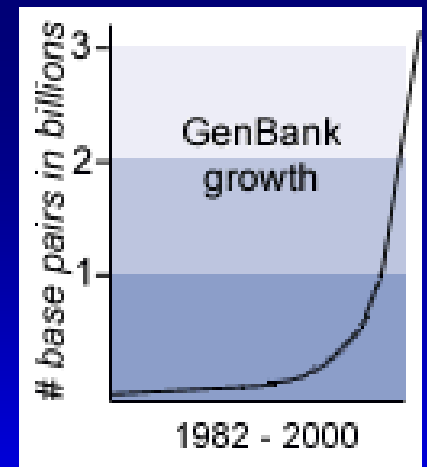
Rickettsia prowazekii (epidemic typhus)

Mycobacterium tuberculosis (tuberculosis)

Nematode *C. elegans* ($\sim 4 \times 10^8$) - December 1998

Drosophila (fruit fly) (2000)

Human genome (rough draft completed 5/00) - 3×10^9 base



Základní zdroje a aplikace bioinformatiky

Výpočetní základy

algoritmy
grafika
zpracování signálu
architektura hardwaru
informační teorie
správa databází
statistika
umělá inteligence
zpracování obrazu
robotika
softwarové inženýrství

Zdroje dat

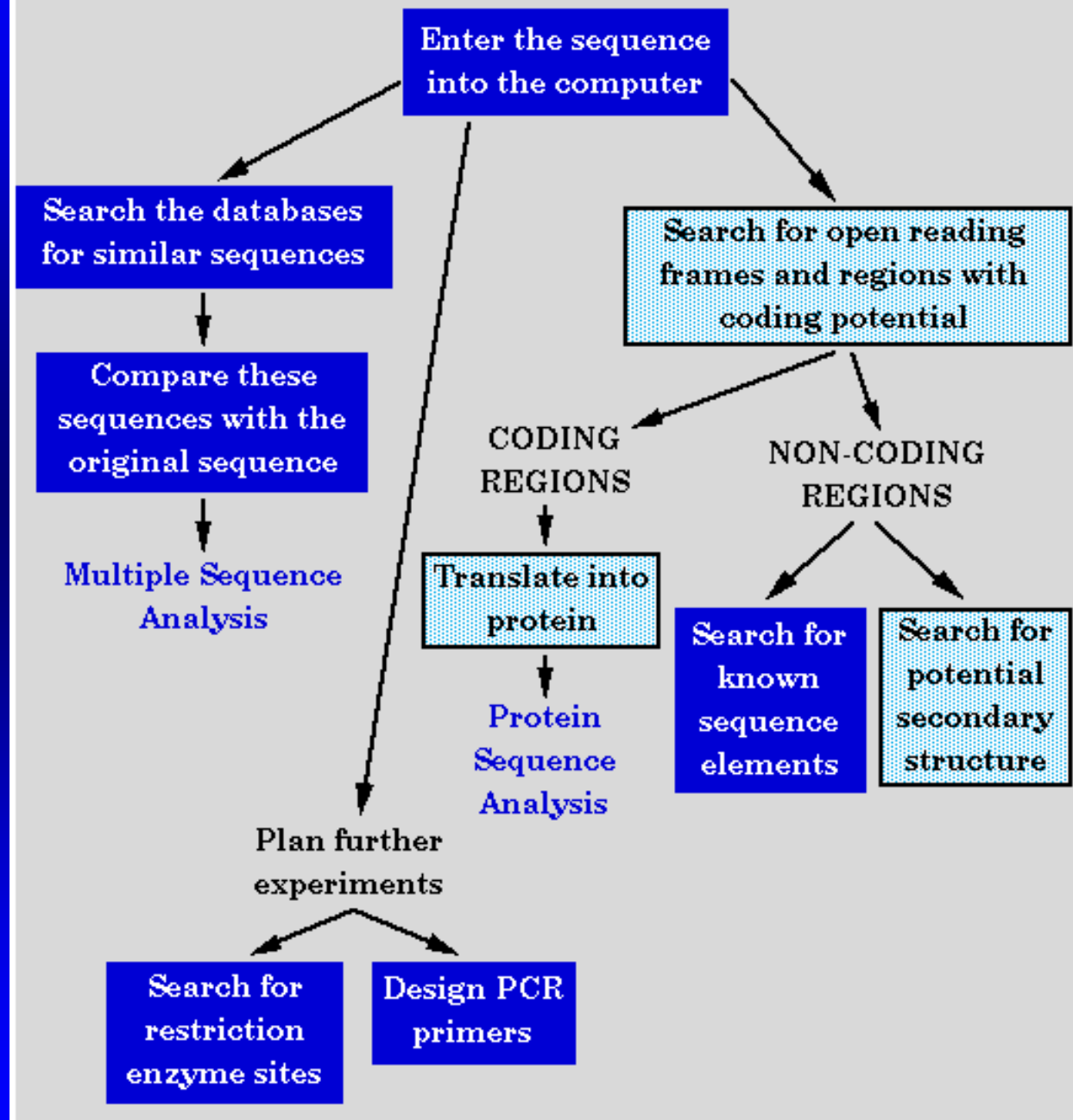
**Obecně dostupné
databáze**

**Zpracování
laboratorních dat**

Aplikace bioinformatiky

získávání dat
nástroje pro přístup k databázím
mapování a srovnávání genomů
seřazení sekvencí
identifikace genů
funkční identifikace proteinů
molekulární evoluce
molekulární modelování
predikce struktur
srovnávání struktur
vývoj léčiv na základě struktur

Nucleotide Sequence Analysis



Instituce zabývající se správou dat a vývojem nástrojů pro jejich analýzu a poskytováním informací

- Evropský institut pro bioinformatiku (EBI)
 - Hinxton, Velká Británie, genomová databáze EMBL
 - <http://www.ebi.ac.uk>
- Národní centrum pro biotechnologické informace (NCBI)
 - USA, genomová databáze GenBank, literární databáze MEDLINE, OMIM - Online Mendelian Inheritance in Man
 - <http://www.ncbi.nlm.nih.gov>
- Národní genetický institut (NIG)
 - Mishima, Japonsko, genomová databáze DDBJ
 - <http://www.cib.nig.ac.jp>

ExPASy Molecular Biology server

– <http://www.expasy.ch>

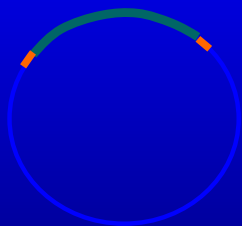
Např. PROSITE - Database of protein families and
domains

Pdb Viewer - <http://www.expasy.org/spdbv/>

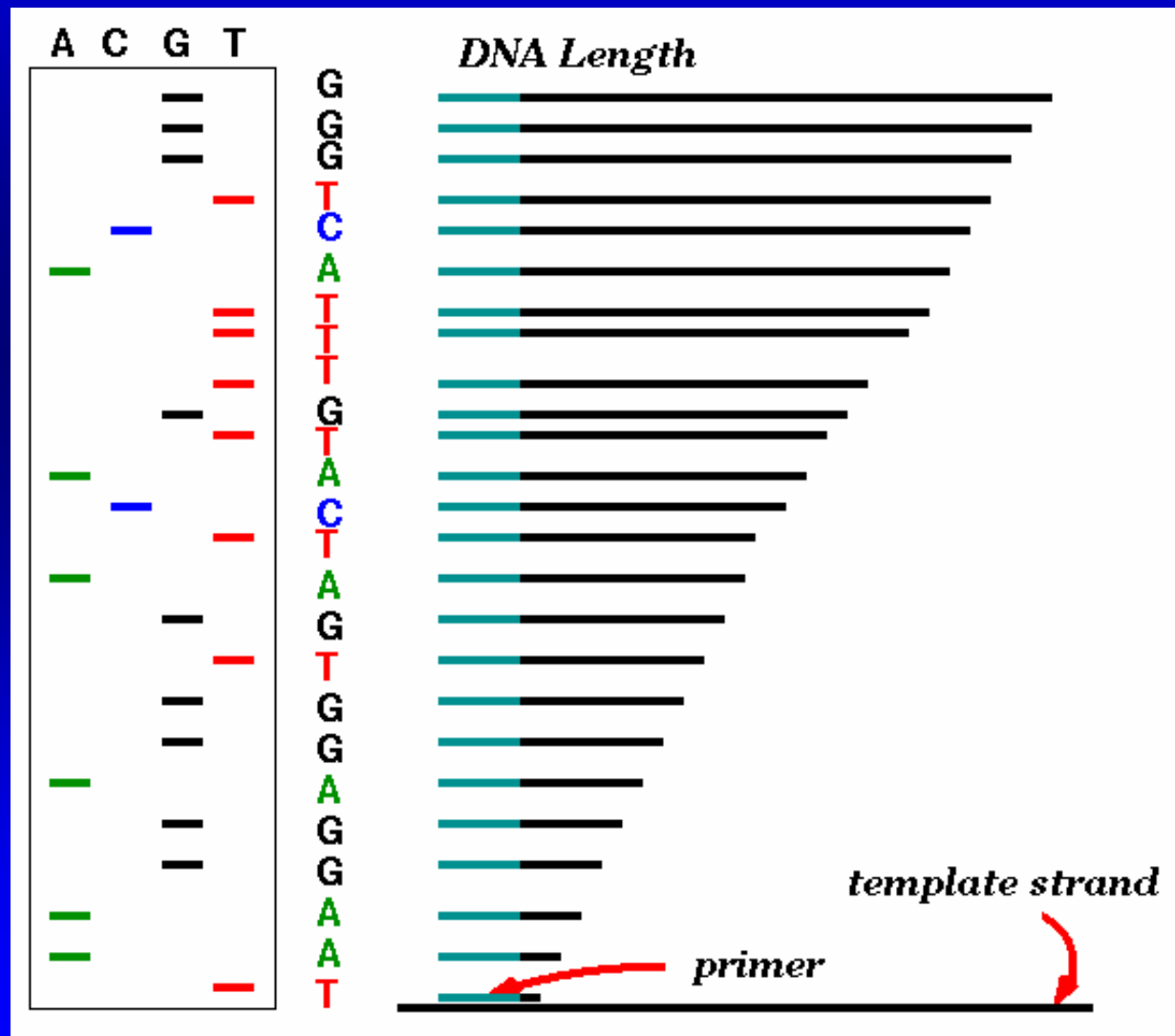
Vyhledávání podobností sekvencí

- BLAST <http://www.ncbi.nlm.nih.gov/BLAST>
- FASTA <http://www.ebi.ac.uk/fasta3>

Sanger Method: Generating Read



1. Start at primer
(restriction site)
2. Grow DNA chain
3. Include ddNTPs
4. Stops reaction at all possible points
5. Separate products by length, using gel electrophoresis



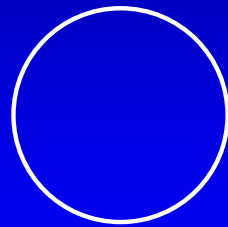
Traditional DNA Sequencing

DNA

Shake

DNA fragments

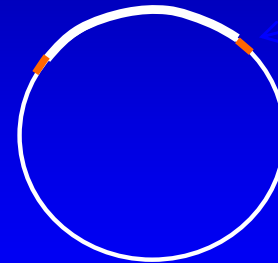
Vector
Circular genome
(bacterium, plasmid)



+



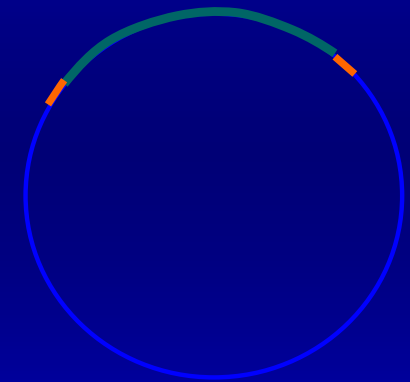
=



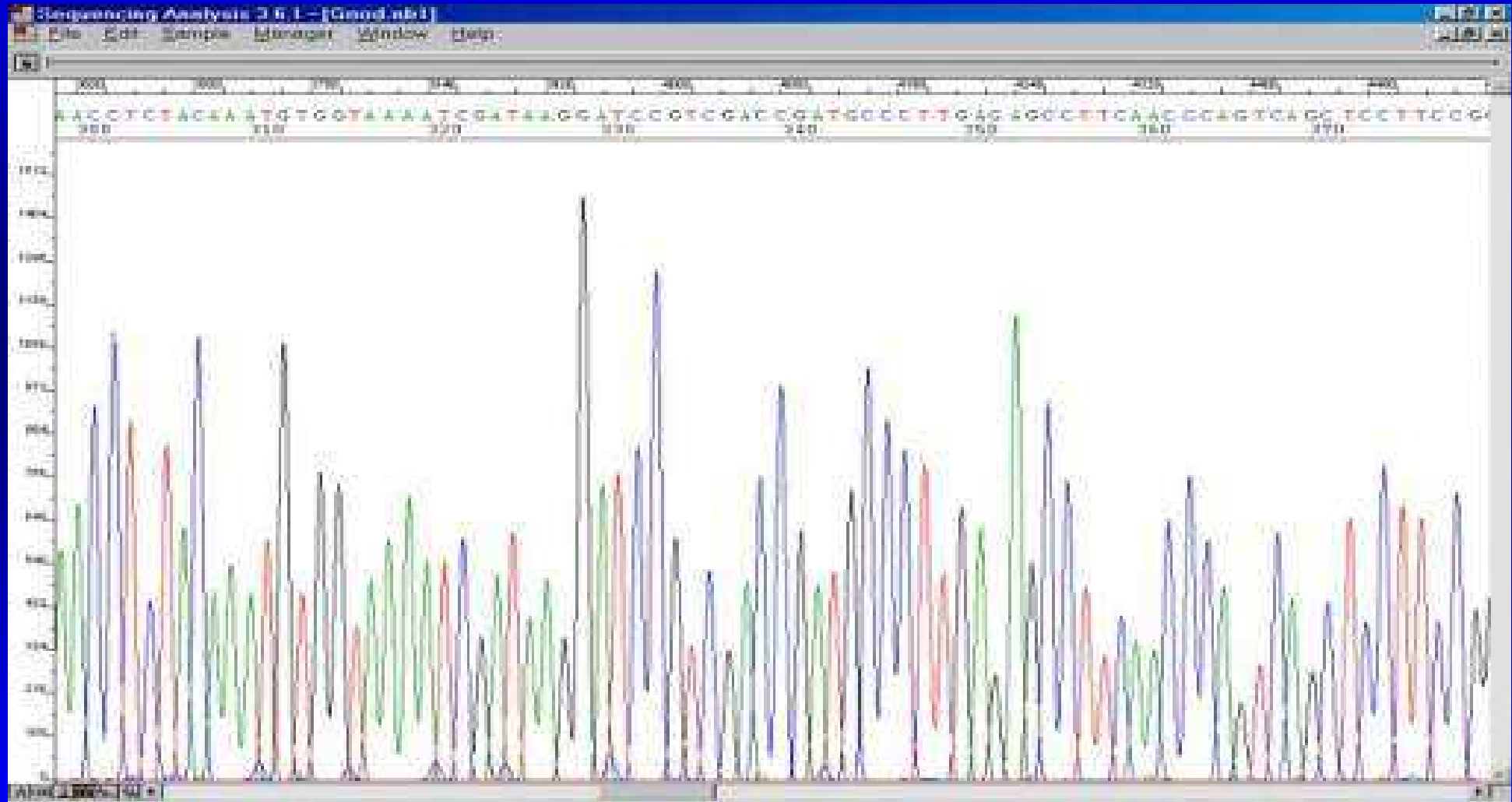
Known
location
(restriction
site)

Different Types of Vectors

<u>VECTOR</u>	<u>Size of insert (bp)</u>
Plasmid	2,000 - 10,000
Cosmid	40,000
BAC (Bacterial Artificial Chromosome)	70,000 - 300,000
YAC (Yeast Artificial Chromosome)	> 300,000 Not used much recently



Electrophoresis Diagrams



Reading an Electropherogram

- Filtering
- Smoothing
- Correction for length compressions
- A method for calling the nucleotides – **PHRED**

Finding Overlapping Reads

Create local multiple alignments from the overlapping reads

The diagram illustrates overlapping DNA reads. Each read is represented by a cyan text string and a red horizontal line below it. The reads are:
1. TAGATTACACAGATTACTGA (topmost, full length)
2. TAGATTACACAGATTACTGA (shifted right by 1 base)
3. TAG TTACACAGATTATTGA (shifted right by 2 bases, with a space before TTACACAGATTATTGA)
4. TAGATTACACAGATTACTGA (shifted right by 3 bases)
5. TAGATTACACAGATTACTGA (shifted right by 4 bases)
6. TAGATTACACAGATTACTGA (shifted right by 5 bases)
7. TAGATTACACAGATTACTGA (shifted right by 6 bases)
8. TAG TTACACAGATTATTGA (shifted right by 7 bases, with a space before TTACACAGATTATTGA)
9. TAGATTACACAGATTACTGA (bottommost, full length)
The red lines are horizontal bars of varying lengths and positions, corresponding to the cyan text above them, showing how they overlap.

Sequencing by Hybridization (SBH): History

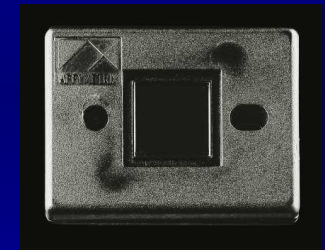
- **1988:** SBH suggested as an alternative sequencing method. Nobody believed it will ever work

First microarray prototype (1989)



- **1991:** Light directed polymer synthesis developed by Steve Fodor and colleagues.

First commercial DNA microarray prototype w/16,000 features (1994)



- **1994:** Affymetrix develops first 64-kb DNA microarray

500,000 features per chip (2002)



Komparativní genomové sekvenování (NimbleGen)

29-mery překrývající se o 22 bp



Každý nukleotid je detegován nejméně 8 oligonuk.

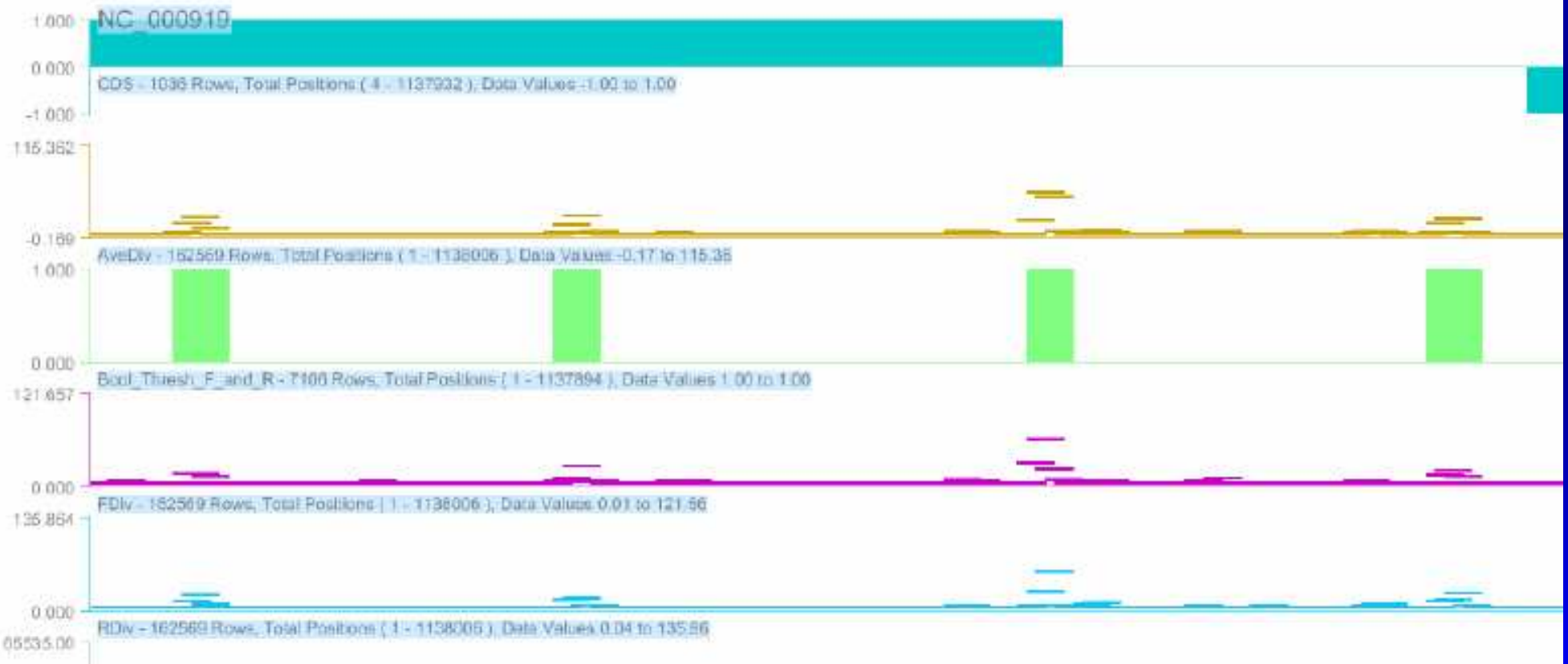
↓
Výsledky jsou použity pro přípravu sekvenačního čipu

DNA



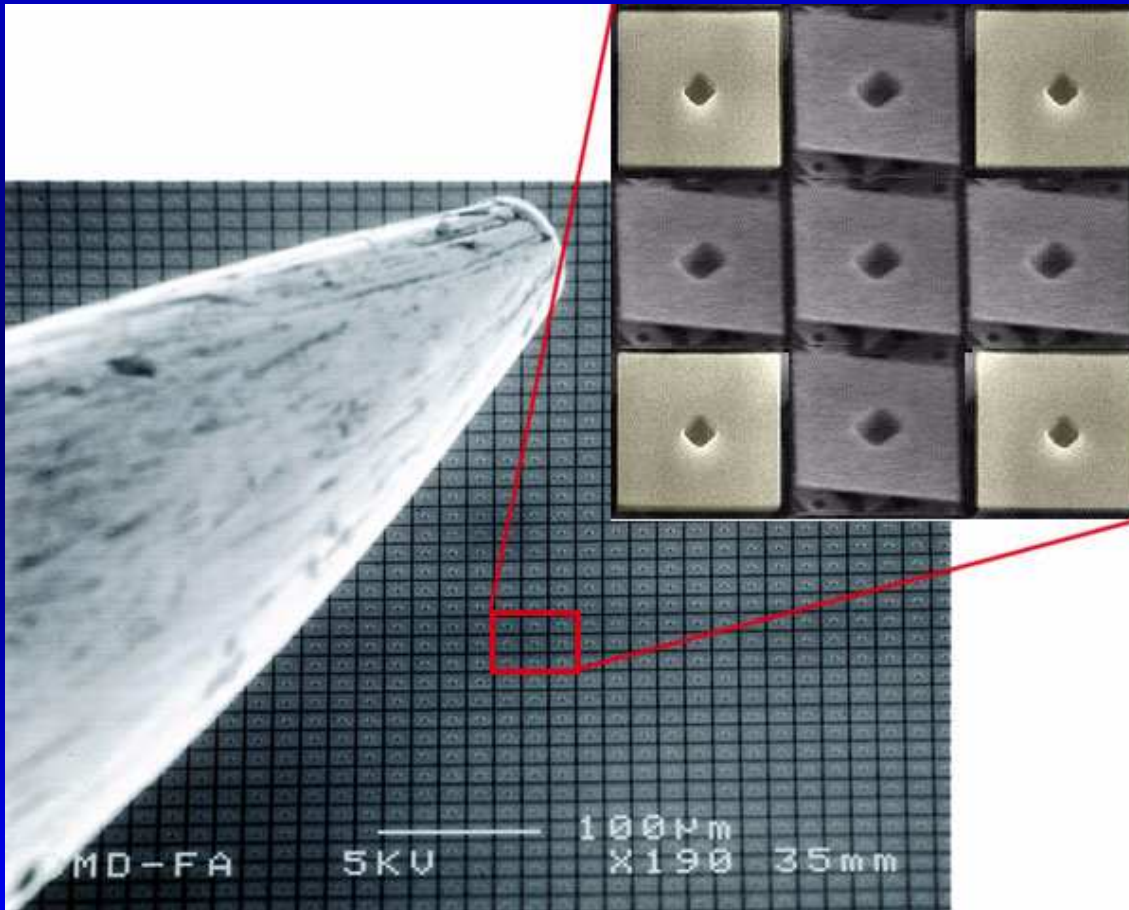
Sekvenování DNA pomocí
oligonukleotidů rozpoznávajících
SNP

A82220 A82280 A82340 A82400 A82460 A82520 A82580 A82640 A82700 A82760 A82820 A82880 A82940 A83000 A83060 A83120 A83180 A83240 A83300



in situ oligonucleotide synthesis

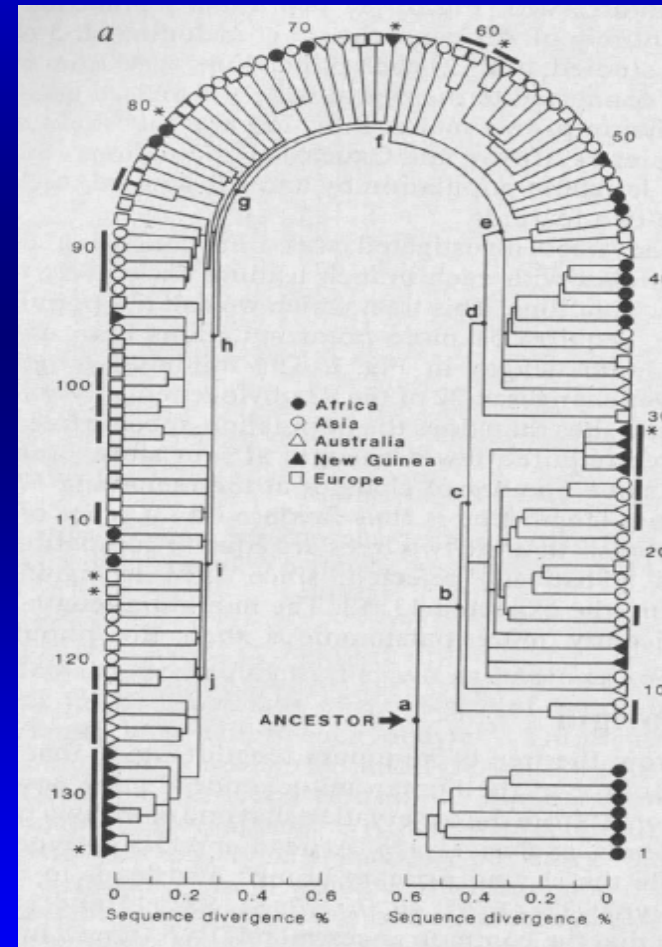
NimbleGen Systems Inc.



- Digital Micromirror Device (DMD)
- Up to 386.000 features per chip

Evolutionary Tree of Humans (mtDNA)

The evolutionary tree separates one group of Africans from a group containing all five populations.



Human Migration Out of Africa

