11. BioVis
11. BioVis

1. Introduction
2. Sequence-Alignments
3. RNA-Secondary Structure
4. Phylogenetic Trees
5. Multiscale Synteny Browser
11.1 Introduction

- Genome – Chromosome – DNA – Gene

- DNA:
  Double helix with nucleotides from sugar, phosphate, and bases
  (Adenine, Cytosine, Guanine, Thymine)

[Image of DNA structure and nucleotide bases]

[Links to genome resources]
11.1 Introduction

- Genome: only 5% of genes, consisting of exons and introns

<table>
<thead>
<tr>
<th>Organism</th>
<th>Base pair</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baking yeast</td>
<td>12 Mil.</td>
<td>6000</td>
</tr>
<tr>
<td>Fruit fly</td>
<td>137 Mil.</td>
<td>13,601</td>
</tr>
<tr>
<td>Human</td>
<td>3 Bil.</td>
<td>~30,000</td>
</tr>
</tbody>
</table>

- RNA: Uracil instead of thymine, single strands
11.1 Introduction

- Proteins are built up from amino acids
- 20 amino acids: encoded by nucleotide-triplets (codons)
  - E.g.: \texttt{GCU}, \texttt{GCC}, \texttt{GCA}, \texttt{GCG} → Alanine (mRNA)

Primary Structure | Secondary Structure | Tertiary Structure

```
ADMVIKAPAGA
KVTKAPVAFSH
KGHASMDC
```

- Protein synthesis
  - Transcription: DNA → mRNA
  - Translation: mRNA → amino acid → protein
11.1.1 Bioinformatics

- Support biologists in processing, analysis, and interpretation of large amount of data

- Approaches:
  - Biological databases
  - Methods for the comparison and function-prediction of sequences
  - Function-prediction of proteins
  - Discovery of new connections
  - Simulation of biological processes
11.1.2 Visualizations

- Visualizations of sequences and related topics
  - Sequence alignments
  - RNA-secondary structure
  - Phylogenetic trees

- Proteins
  - Structure prediction
  - Structure comparison
  - Protein dynamics

[http://bc.imb.sinica.edu.tw/online_tool.php]
11.1.2 Visualizations

**Protein structure:**
- A main research area of molecular biology
- This research area is older than sequence analysis (beginning of the 70s)

**Why is the structure so important?**
- Folded protein has an irregular shape
  - Function of the protein depends on its shape
  - Bonding capacity with other molecules or proteins (interaction)
- The gap between the number of known sequences and structures is very large (approx. 100:1)
11.2 Sequence-Alignments

Comparison of two (or more) sequences

- Given sequences: \text{GACGGATTAG} \n\text{GATCGGAATAG}

Possible alignments

- Goal: the best possible alignment
  - Global alignment: whole sequences
  - Local alignment: substrings
  - Semi-global alignment: prefixes or suffixes
  - [Multiple Sequence Alignment (MSA)]

DB-Server
- BLAST
- FASTA
- ...

- \text{GACGGATTAG}\n\text{GATCGGAATAG}

- \text{GA-} \text{CGGATTAG}\n\text{GATCGGAATAG}
11.2 Sequence-Alignments

Similarity of two sequences

- Given alignment

```
GA-CGGATTTAG
GATCGGAATAG
```

- Easy assessment
  - **Match:** 1 e.g. G with G
  - **Mismatch:** -1 e.g. T with A
  - **Space:** -2 e.g. – with T

- Total value: $9 \cdot 1 + 1 \cdot (-1) + 1 \cdot (-2) = 6$
- **Similarity:** value of the best alignment
11.2.1 Visualizations

Textual

Color coding

[http://www.uib.no/aasland/chromo/chromo-aln.html]
11.2.1 Visualizations

- JALVIEW (www.jalview.org)
11.2.1 Visualizations

- **JALVIEW 2** (www.jalview.org)

11.2.1 Visualizations

- **MEME suite**

- **Tasks** [http://meme.sdsc.edu/meme/intro.html]
  - Discover motifs using MEME (Multiple EM for Motif Elicitation) or GLAM2 (Gapped Local Alignment for Motifs) on groups of related DNA or protein sequences
  - Search sequence databases using motifs
  - Compare a motif to all motifs in a database of motifs
  - Associate motifs with Gene Ontology terms via their putative target genes

- **Visualization**
  - 2D-representation
  - Interaction only through clicking
11.2.1 Visualizations
11.2.1 Visualizations
11.2.1 Visualizations

- **SEQUENCE LOGO**


  [http://www-lecb.ncifcrf.gov/~toms/sequencelogo.html]
11.2.1 Visualizations

- PATTVISION
  - 3D-representation
  - Patterns preserve the respective colors
  - Blending Planes
  - Interaction by zooming
11.2.1 Visualizations
11.2.1 Visualizations

- ARC DIAGRAMS (for sequence structures)
  

Theme of a DNA-subsequence of a yeast

Also suitable for musical themes: *Für Elise*
11.3 RNA-Secondary Structure

- Single-stranded RNA folds itself at suitable places forming a double-rope
- Secondary structure is associated with the function of the RNA

CGACAGGAGUAGGGUUUUUGUCCUCCUCGCG

- Guanine - Cytosine
- Uracil - Adenine
- Guanine - Uracil
### 11.3 RNA-Secondary Structure

- **Structure forms**
  - E: Single strand
  - M: Multiple loop
  - I: Internal loop
  - H: Hairpin loop
  - B: Bulge loop
  - Helix
11.3 RNA-Secondary Structure

- **Visualization variants**
  - Bracket-form
  - Matrix-form
  - Circle graph
  - Structured graph

- **Criteria that should be fulfilled by the visualization**
  - Readable representation of the structure
  - No overlapping of structure elements
  - Consistent representation of similar structures
  - High aesthetics
  - Minimal user-interaction (?)
11.3.1 Bracket Form

- Typical representation of a bracket-form:

```
GGGGGAAAGGGAAAAACCCAAAGGGAAAAACCCAAAACCCCC
((((((...(((((.....)))...(((....)))....))))....))))
```

- For every pair, one opening and one closing bracket
- For every un-bounded base, a point
11.3.2 Matrix form

- A point is plotted at position \((i, j)\), if the bases \(i\) and \(j\) are paired
- Size of the point: probability of this pairing
11.3.2 Matrix form

Hairpin loop

Multiple loop

Inner loop

red: Further pairings
green: No base pairing
11.3.3 Circle Graph

- Bases placed around a circle
- Connections between two bases are the pairings
11.3.3 Circle Graph

- Hairpin loop
- Bulge loop
- Multiple loop
- Inner loop
11.3.4 Structured Graph

- Undirected Graph
  - Nodes: Bases
  - Edges: Connections between bases

- Representation of a hydrogen bond
  - Points: Guanine - Cytosine
  - Lines

- Data
  - Number of the bases

- Start symbol
11.3.4 Structured Graph
11.4 Phylogenetic Trees

- **Problem**
  - Explanation of the evolution history of today’s species and their relations

- **Example:**

![Phylogenetic Tree Diagram]

- Siamang
- Gibbon
- Orangutan
- Gorilla
- Human
- Chimpanzee

*Today’s species*

*Hypothetical ancestors*
11.4 Phylogenetic Trees

- **Objects**
  - Species, populations, DNA, proteins

- **Tree**

  ![Phylogenetic Trees Diagram]

  - There are two ways to determine a root:
    - Outgroup method: with non-related objects
    - Hypothesis of a molecular clock: root has the same distance to all objects
11.4.1 Visualizations

- PhyloDraw
  [Choi et al., "PhyloDraw: a phylogenetic tree drawing system, Bioinformatics 2000"]
11.4.1 Visualizations

- PhyloDraw

[http://pearl.cs.pusan.ac.kr/phylodraw/#test]
11.4.1 Visualizations

- Evolutionary Trees
  [Nina Amenta, Jeff Klingner, Case Study: Visualizing Sets of Evolutionary Trees, InfoVis 2002.]
  - 2 views
  - Points = trees
  - MDS-Layout
  - Consensus-tree
  - Interaction: Selection of individual points or pointsets (multiple selection possible)
11.4.1 Visualizations

- **Evolutionary Trees**
  [Nina Amenta, Jeff Klingner, Case Study: Visualizing Sets of Evolutionary Trees, InfoVis 2002.]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]
11.5 Multiscale Synteny Browser

- [M. Meyer et al., MizBee: A Multiscale Synteny Browser, InfoVis 2009]