WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES,
AND DNA-PROFILING IN PARTICULAR

Thirteenth Session
Brasilia, November 22 to 24, 2011

ADDENDUM

BIONUMERICS: A UNIVERSAL PLATFORM FOR DATABASING AND ANALYSIS
OF BIOLOGICAL DATA

Document prepared by an expert from the Netherlands
BioNumerics
A UNIVERSAL PLATFORM FOR DATABASING AND ANALYSIS OF BIOLOGICAL DATA

Content and Aim

- Overview of the capacity and applications of BioNumerics
- Current use of BioNumerics at Naktuinbouw
- Future use of BioNumerics at Naktuinbouw
Introduction - Basic features

- Relational database containing taxonomic, typing, and genomic data of biological entities
- Possibility to store many different data sets ("experiments") for each organism or sample studied
- Extensive data import, and export functions
- Advanced statistical analysis, comparison, and identification functions
- Possibility to combine the information from different methods in one single analysis
- Exchange of data over the Internet
Relational database set-up

All information is stored using databases like SQL/Oracle.
Working interface is BioNumerics.

BioNumerics Interface

Database panel, Experiment types panel, Experiment presence panel, Experiment files panel, Comparisons panel, Other panels.
What you can store

**Information fields**
- Up to 100 fields (each up to 80 characters)
- Link with external databases

**Attachments**
- Bitmap images
- HTML and hyperlinks
- Word documents
- Excel spreadsheets
- PDF files
- Text documents

**Fingerprints**
- 1-D electrophoresis gels scanned as bitmaps (RFLP, PFGE, Ribotyping, RAPD, DGGE & TGGE, etc.)
- Sequencer chromatogram files (AFLP, VNTR, HDA, etc.)
- Spectrophotometric files
- MALDI & SELDI profiles
- All other kinds of densitometric profiles

**Character data**
- Phenotypic test panels (API, Biolog, Vitec, etc.)
- Antibiotic resistance profiles
- Fatty acid and quinolone profiles
- Hybridization blots such as Spoligo typing
- Biochemical & morphological features
- Microarray & Genechip data
- Etc.

**Sequence data**
- Sequence trace (chromatogram) files
- Formatted sequences from public databases (EMBL, GenBank)
- Aligned sequences such as from RDP
- Amino acid sequences

**2-D gels**

Trendcurve and kinetic reading data
What you can do

Querying
- Using fields
- Using experiments
- Using ranges of values
- Using any combinations

Cluster analysis
- UPGMA, Nearest Neighbor, Furthest Neighbor, Ward...
- Minimum Spanning Trees
- Consensus trees
- Calculation of degeneracy, error estimation on trees

Identification
- Library construction
- Statistical confidence
- Neural Networks
- Decision Networks
**What you can do**

**Dimensioning, ordination**
- Principal Components
- Multi-Dimensional Scaling
- Self-Organizing Maps

**Phylogeny**
- Pairwise & multiple sequence alignment
- Neighbor Joining
- Parsimony
- Maximum likelihood

**Statistics**
- MANOVA
- Discriminant analysis
- K-means partitioning, Jackknife,...
- Numerous statistical tests and charts

**Current use Naktuinbouw**
1. Storage of sample information and experiment data
2. Query
3. Import of raw fingerprint data into BN
4. Analysis of the DNA fingerprints
5. Clustering/Ordination/ Verification
6. Identification Libraries
Current use Naktuinbouw

Example AFLP

- Import gel
- Define lanes
- Link lanes to Entries
- Use internal controle band for normalisation

Normalization

Before normalization

After normalization
**Comparison**

Open comparison window
To compare several different experiments from different samples

**Scoring of markers**
Overview

Composite data sets

Fingerprint 1

Fingerprint 2

Character

Sequence

ACCTGTAAT
GCTGGTACA
ACTGGTACT

Combined picture?
Option 1: averaging similarity matrices

Fingerprint 1

Fingerprint 2

Weighted average

\[ S_c = \frac{w_1 S_1 + w_2 S_2}{w_1 + w_2} \]

Average similarity matrix

Option 2: creation of combined character tables

Character set 1

Character set 2

Combined character table

BMT/13/31 Add. page 12
Conversion to character tables

Fingerprint

Band matching table

(Align) sequence

Sequence table

Categorical data (e.g. A, T, C, G or gap)

Identification Libraries

Fingerprints (AFLP, SSR)

Characters (scores)

Composite data set

Identification library

Variety A 0 0 0 1 0 1 1 0 1 0 1 0 0
Variety B 0 1 0 1 1 0 1 0 1 1 1 0 1 0
Variety C 1 1 0 1 1 1 0 1 1 1 0 1 0
Variety D 1 1 1 1 0 1 1 0 1 1 0 1 1 1

...
Identification Libraries

Best hit for unknown Sample with samples in library

What you can do

- Database sharing
  - BioNumerics Server
- External databases
- Printing of reports and analyses
- External databases
- BioNumerics Database
  - BioNumerics Platform
  - Other BioNumerics shares
- Genomics
  - Transcriptomics
  - Proteomics
  - Public databases
- What you can do
- BMT/13/31 Add. page 14
Database Sharing

Outside LAN
- BN workstation
- Data exchange via XML files

Inside Local Area Network (LAN)
- Server with BN ODBC connection
- BN workstation

Data exchange
- Citrix, RDP
- Over internet

Future use Naktuinbouw

1. Import raw NGS data into BN

2. Combine different types of data
   - Sequence and fingerprint
   - Morfological measurements and fingerprints/sequences
   - Management of reference collections

3. Collaborations with other labs
   - Common database accessible from different labs
### BioNumerics Plugins

- Multilocus VNTR Analysis (MLVA)
- MIRU VNTR
- Multilocus Sequence Typing (MLST)
- Polymorphic VNTR sequence typing
- Spa typing
- Diversilab
- Geographical mapping
- HIV drug resistance
- MLPA
- Band scoring
- HDA
- SNP genotyping

### Bionumerics modules

- Fingerprints, Characters, Tree & Network Inference.
- Sequences, Characters, Tree & Network Inference.
- Fingerprints, Tree & Network Inference, Dimensioning.
- Database Sharing.
- Sequences, Characters, Identification.
- Fingerprints, Characters.
- Fingerprints, Characters, Dimensioning, Database Sharing.
- Characters.

### Modules and Plugins

#### Application Modules

<table>
<thead>
<tr>
<th>Module</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fingerprint Types Module</td>
<td>€5500</td>
</tr>
<tr>
<td>Character Types Module</td>
<td>€1900</td>
</tr>
<tr>
<td>Sequence Types Module</td>
<td>€2600</td>
</tr>
<tr>
<td>Matrix Types Module</td>
<td>€1000</td>
</tr>
<tr>
<td>Trend Data Module</td>
<td>€2300</td>
</tr>
<tr>
<td>2-D gel Types Module</td>
<td>€6300</td>
</tr>
</tbody>
</table>

#### Analysis Modules

<table>
<thead>
<tr>
<th>Module</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree and Network Inference Module</td>
<td>€3000</td>
</tr>
<tr>
<td>Sequence Molecular Analysis Module</td>
<td>€2300</td>
</tr>
<tr>
<td>Chromosome Analysis Module</td>
<td>€3200</td>
</tr>
<tr>
<td>Identification Module</td>
<td>€2400</td>
</tr>
<tr>
<td>Dimensioning Techniques Module</td>
<td>€2500</td>
</tr>
<tr>
<td>Database Sharing Tools Module</td>
<td>€2700</td>
</tr>
<tr>
<td>Versioning &amp; Audit Trails Module</td>
<td>€5000</td>
</tr>
</tbody>
</table>
Conclusions

BioNumerics is an all in one system:

Database (collection of information and query)

Input of all kinds of biological data
several different experiment types

Statistics of the data on combinations of
experiment types

Database sharing tools, compatible to many
different database systems, any number of users
on different locations

[End of document]