E

TWC/29/30
ORIGINAL: English
DATE: June 1, 2011

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Ninth Session
Geneva, June 7 to 10, 2011

BIONUMERICS SOFTWARE FOR DATABASING AND DATA ANALYSIS

Prepared by an expert from the Netherlands

BioNumerics

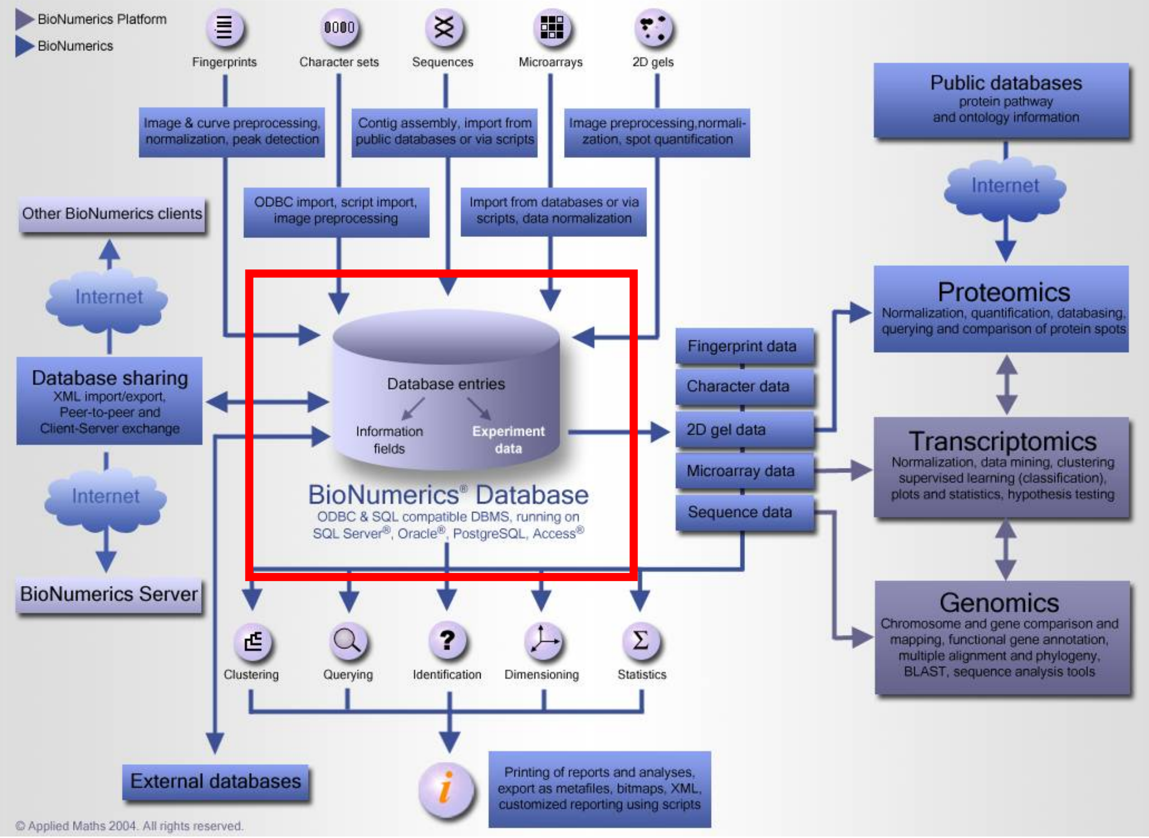
A UNIVERSAL PLATFORM FOR DATABASING AND ANALYSIS OF BIOLOGICAL DATA

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

APPLIED MATHS

BN Scheme



BioNumerics Platform
BioNumerics

Fingerprints
Character sets
Sequences
Microarrays
2D gels

Image & curve preprocessing, normalization, peak detection
Contig assembly, import from public databases or via scripts
Image preprocessing,normalization, spot quantification
ODBC import, script import, image preprocessing
Import from databases or via scripts, data normalization

Public databases
protein pathway and ontology information

Internet

Other BioNumerics clients

Internet

Database sharing
XML import/export, Peer-to-peer and Client-Server exchange

Internet

BioNumerics Server

Database entries
Information fields
Experiment data

BioNumerics® Database
ODBC & SQL compatible DBMS, running on SQL Server®, Oracle®, PostgreSQL, Access®

Fingerprint data
Character data
2D gel data
Microarray data
Sequence data

Proteomics
Normalization, quantification, databasing, querying and comparison of protein spots

Transcriptomics
Normalization, data mining, clustering supervised learning (classification), plots and statistics, hypothesis testing

Genomics
Chromosome and gene comparison and mapping, functional gene annotation, multiple alignment and phylogeny, BLAST, sequence analysis tools

Clustering
Querying
Identification
Dimensioning
Statistics

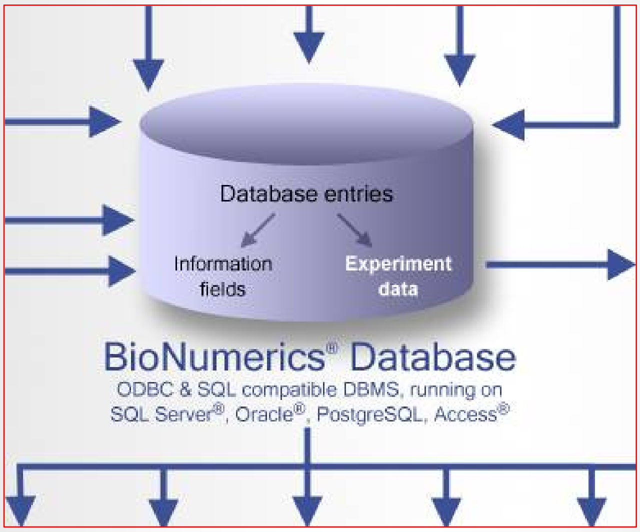
External databases

Printing of reports and analyses, export as metafiles, bitmaps, XML, customized reporting using scripts

APPLIED MATHS

What you can do

input



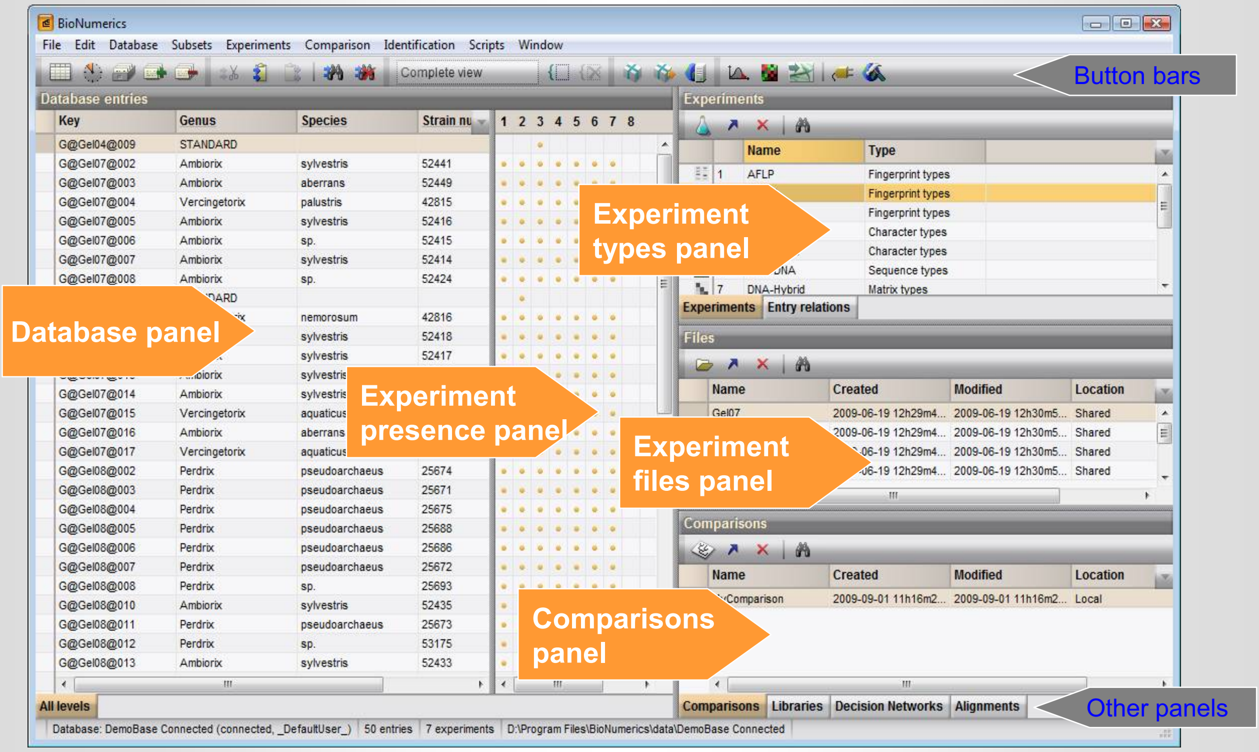
Database entries
Information fields
Experiment data

BioNumerics® Database
ODBC & SQL compatible DBMS, running on SQL Server®, Oracle®, PostgreSQL, Access®

output

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

What you can do



Key	Genus	Species	Strain nu
G@Gel04@009	STANDARD		
G@Gel07@002	Ambiorix	sylvestris	52441
G@Gel07@003	Ambiorix	aberrans	52449
G@Gel07@004	Vercingetorix	palustris	42815
G@Gel07@005	Ambiorix	sylvestris	52416
G@Gel07@006	Ambiorix	sp.	52415
G@Gel07@007	Ambiorix	sylvestris	52414
G@Gel07@008	Ambiorix	sp.	52424
	...DARD		
	...ix	nemorosum	42816
		sylvestris	52418
		sylvestris	52417
	...biorix	sylvestris	
G@Gel07@014	Ambiorix	sylvestris	
G@Gel07@015	Vercingetorix	aquaticus	
G@Gel07@016	Ambiorix	aberrans	
G@Gel07@017	Vercingetorix	aquaticus	
G@Gel08@002	Perdrix	pseudoarchaeus	25674
G@Gel08@003	Perdrix	pseudoarchaeus	25671
G@Gel08@004	Perdrix	pseudoarchaeus	25675
G@Gel08@005	Perdrix	pseudoarchaeus	25688
G@Gel08@006	Perdrix	pseudoarchaeus	25686
G@Gel08@007	Perdrix	pseudoarchaeus	25672
G@Gel08@008	Perdrix	sp.	25693
G@Gel08@010	Ambiorix	sylvestris	52435
G@Gel08@011	Perdrix	pseudoarchaeus	25673
G@Gel08@012	Perdrix	sp.	53175
G@Gel08@013	Ambiorix	sylvestris	52433

	Name	Type
1	AFLP	Fingerprint types
		Fingerprint types
		Fingerprint types
		Character types
		Character types
	...DNA	Sequence types
7	DNA-Hybrid	Matrix types

Name	Created	Modified	Location
Gel07	2009-06-19 12h29m4...	2009-06-19 12h30m5...	Shared
	2009-06-19 12h29m4...	2009-06-19 12h30m5...	Shared
	...06-19 12h29m4...	2009-06-19 12h30m5...	Shared
	...06-19 12h29m4...	2009-06-19 12h30m5...	Shared

Name	Created	Modified	Location
...Comparison	2009-09-01 11h16m2...	2009-09-01 11h16m2...	Local

Database: DemoBase Connected (connected, _DefaultUser_) | 50 entries | 7 experiments | D:\Program Files\BioNumerics\data\DemoBase Connected

Button bars

Experiment types panel

Database panel

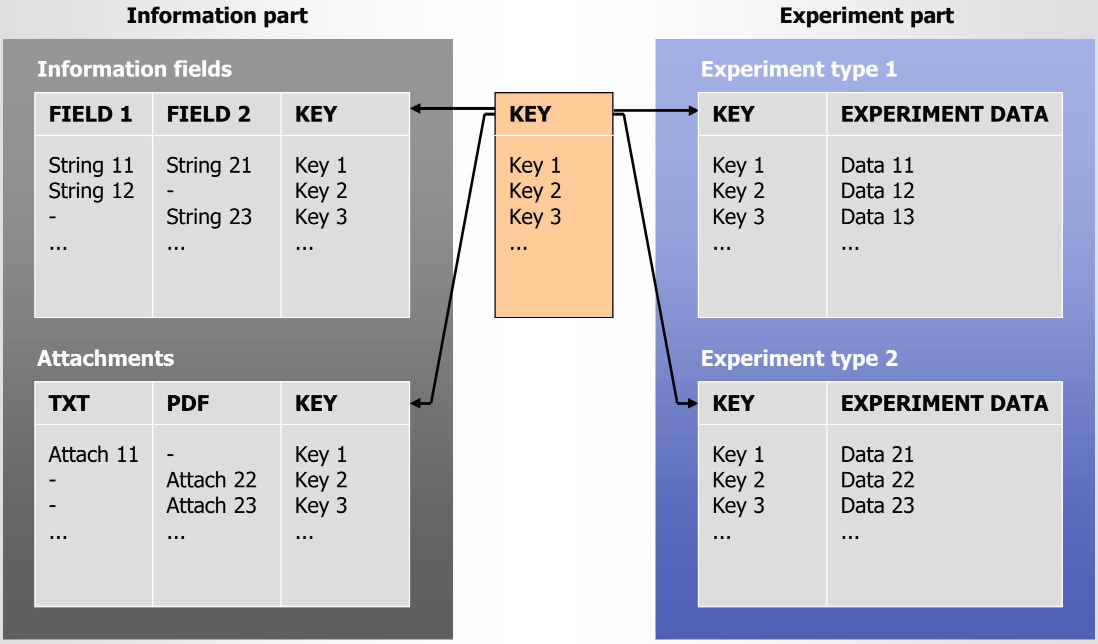
Experiment presence panel

Experiment files panel

Comparisons panel

Other panels

Database concept



Information part

Information fields

FIELD 1	FIELD 2	KEY
String 11	String 21	Key 1
String 12	-	Key 2
-	String 23	Key 3
…	…	…

Attachments

TXT	PDF	KEY
Attach 11	-	Key 1
-	Attach 22	Key 2
-	Attach 23	Key 3
…	…	…

KEY
Key 1
Key 2
Key 3
…

Experiment part

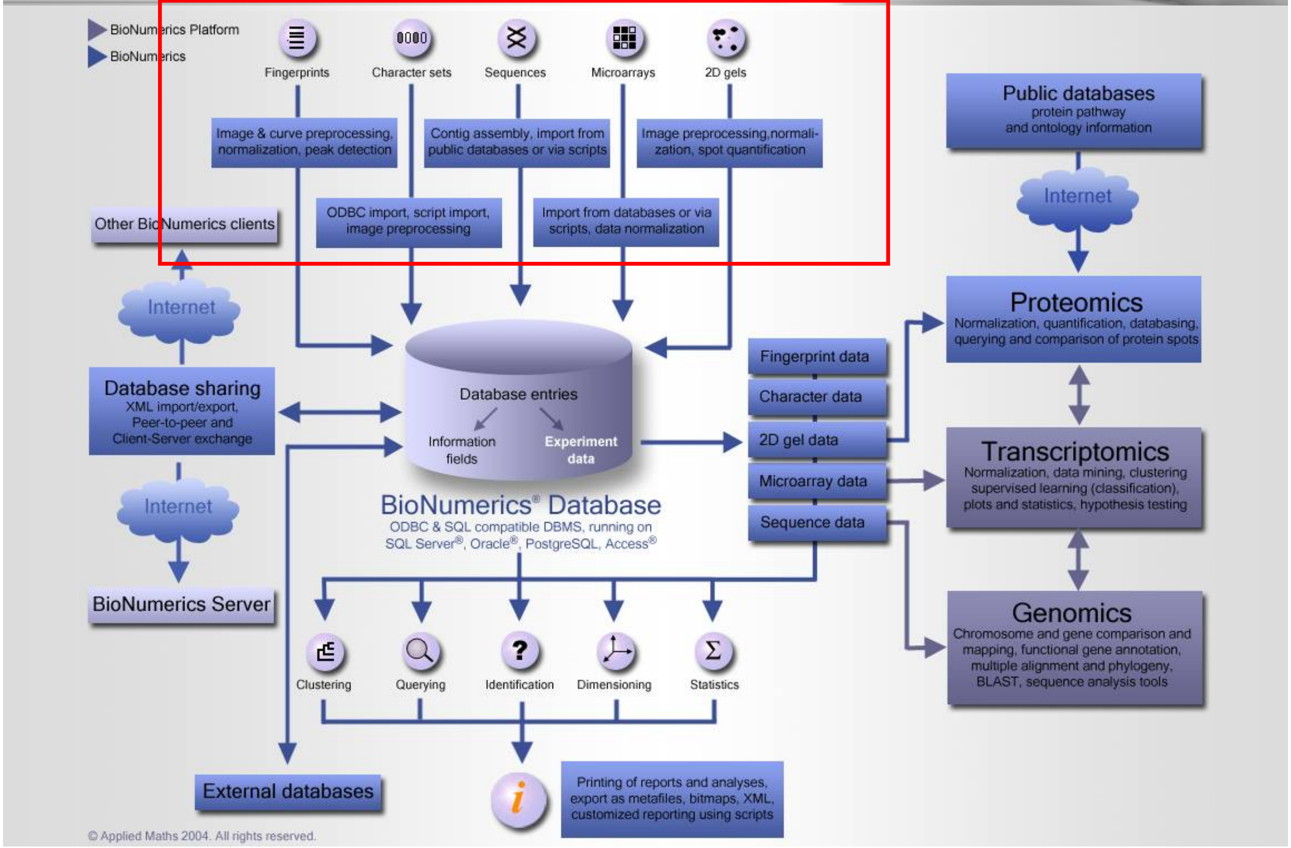
Experiment type 1

KEY	EXPERIMENT DATA
Key 1	Data 11
Key 2	Data 12
Key 3	Data 13
…	…

Experiment type 2

KEY	EXPERIMENT DATA
Key 1	Data 21
Key 2	Data 22
Key 3	Data 23
…	…

What you can do



- BioNumerics Platform
- BioNumerics

Fingerprints

Character sets

Sequences

Microarrays

2D gels

Image & curve preprocessing, normalization, peak detection

Contig assembly, import from public databases or via scripts

Image preprocessing,normalization, spot quantification

ODBC import, script import, image preprocessing

Import from databases or via scripts, data normalization

Other BioNumerics clients

Internet

Database sharing
XML import/export,
Peer-to-peer and
Client-Server exchange

Internet

BioNumerics Server

Database entries

Information fields

Experiment data

BioNumerics® Database
ODBC & SQL compatible DBMS, running on
SQL Server®, Oracle®, PostgreSQL, Access®

Fingerprint data

Character data

2D gel data

Microarray data

Sequence data

Public databases
protein pathway
and ontology information

Internet

Proteomics
Normalization, quantification, databasing, querying and comparison of protein spots

Transcriptomics
Normalization, data mining, clustering supervised learning (classification), plots and statistics, hypothesis testing

Genomics
Chromosome and gene comparison and mapping, functional gene annotation, multiple alignment and phylogeny, BLAST, sequence analysis tools

Clustering

Querying

Identification

Dimensioning

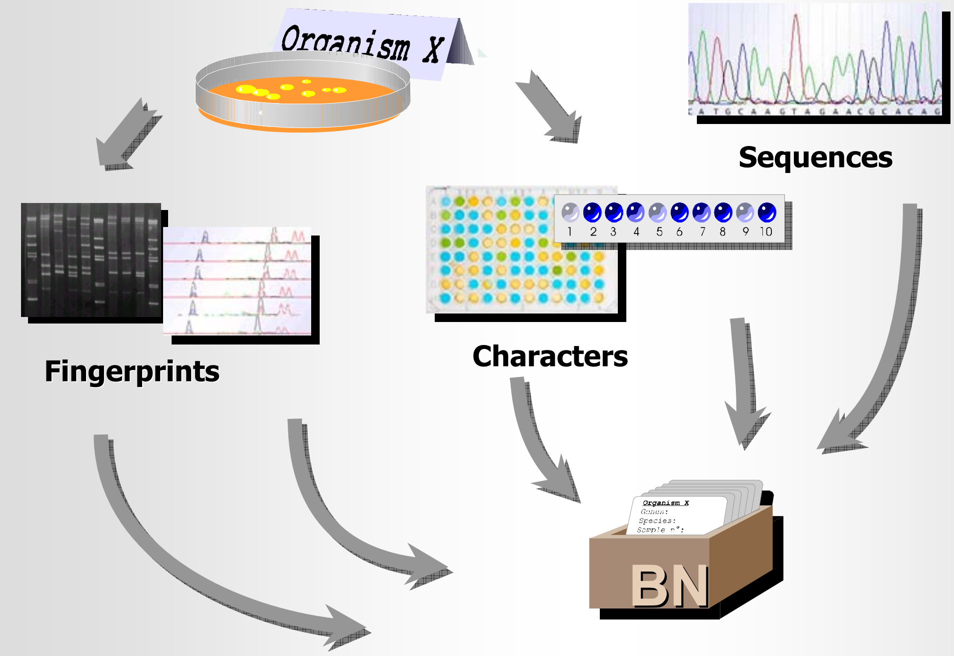
Statistics

External databases

Printing of reports and analyses, export as metafiles, bitmaps, XML, customized reporting using scripts

What you can store

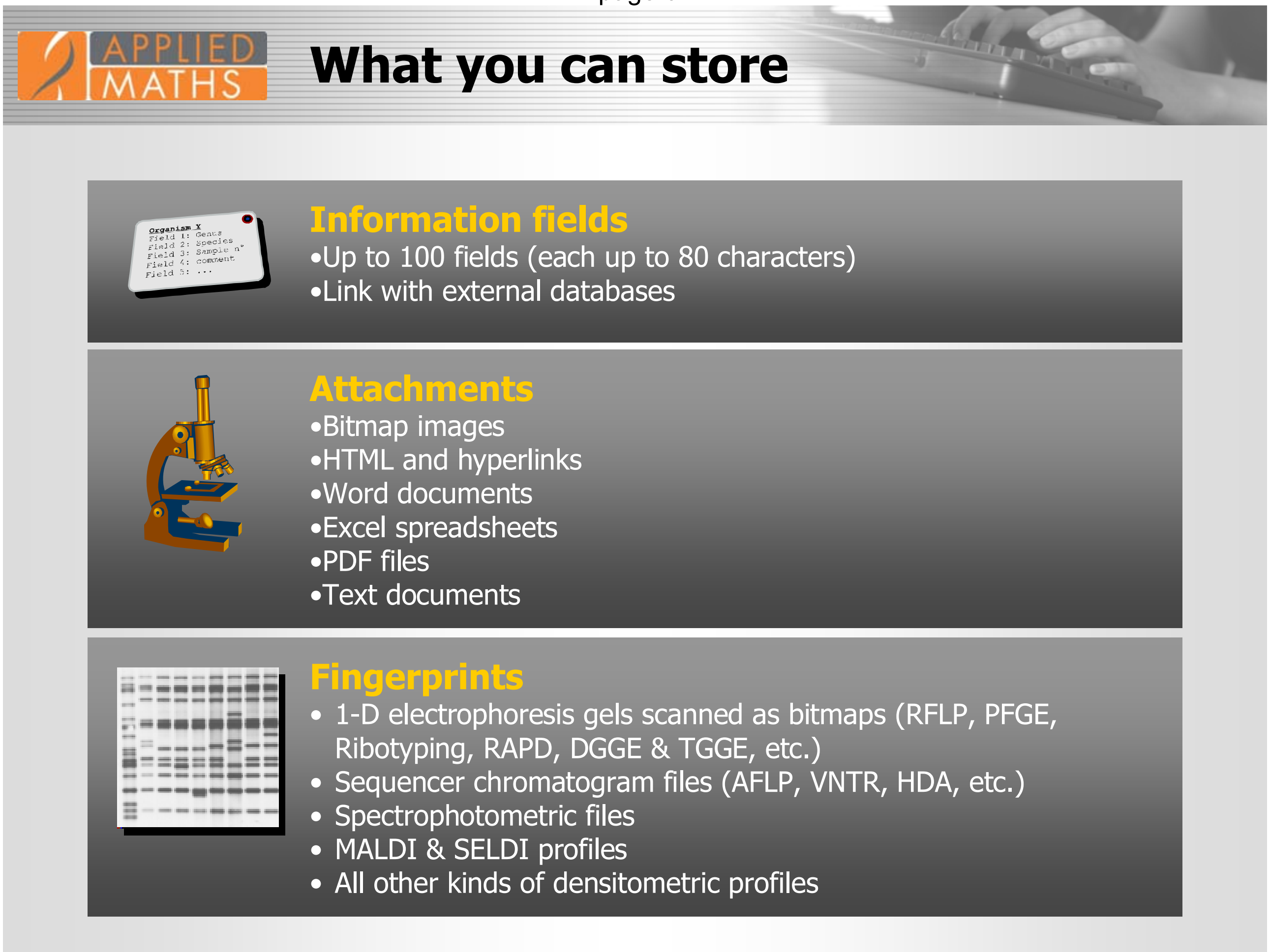
- BioNumerics can store virtually all types of biological data that are used to compare organisms or organical material:



Sequences

Fingerprints

Characters



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

What you can store

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.

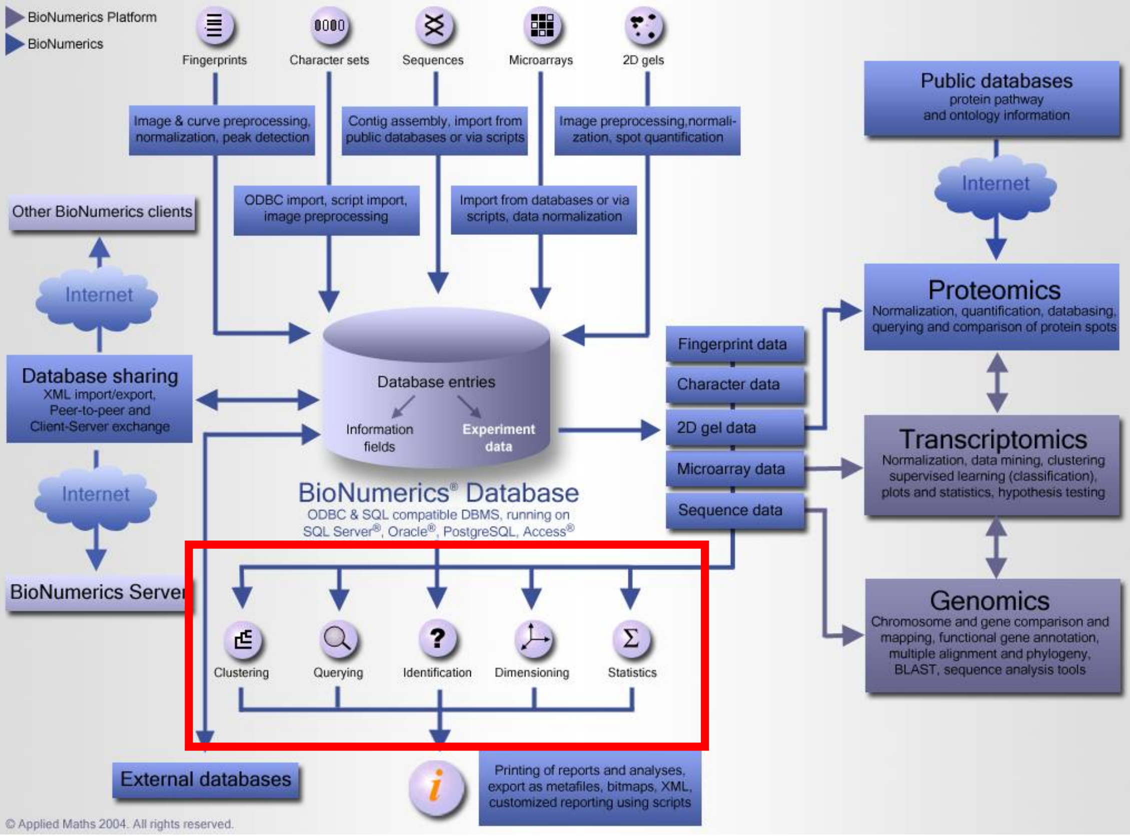
Sequence data

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

2-D gels

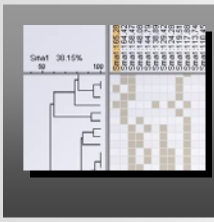
Trendcurve and kinetic reading data

What you can do



What you can do

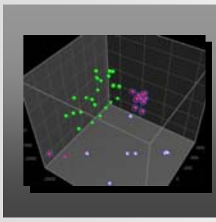
Querying
- Using fields
- Using experiments
- Using ranges of values
- Using any combinations



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

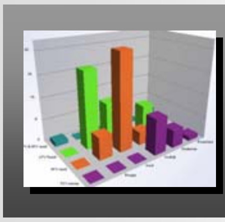


Dimensioning, ordination

- Principal Components
- Multi-Dimensional Scaling
- Self-Organizing Maps

Phylogeny

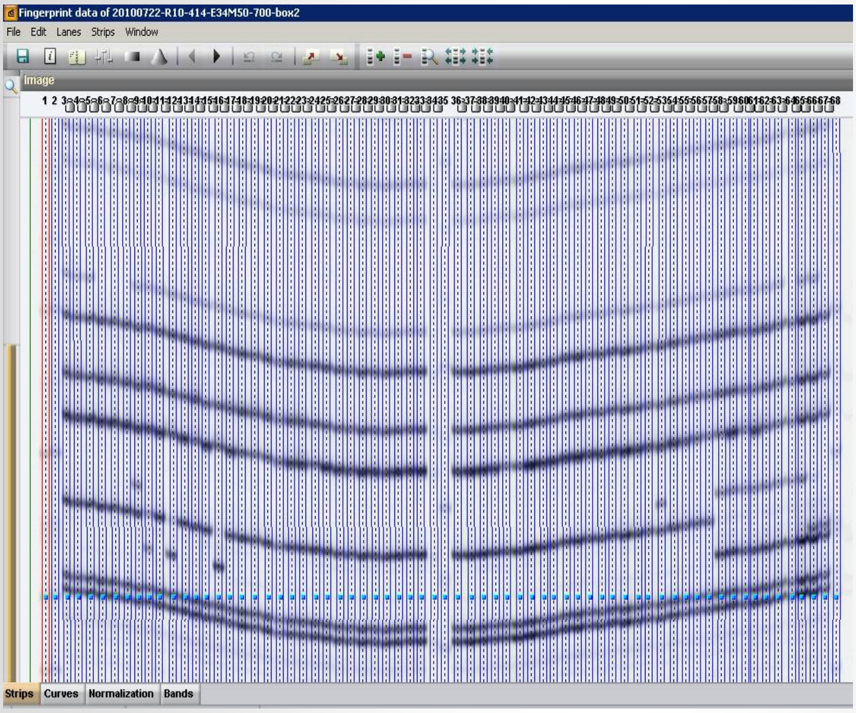
- Pairwise & multiple sequence aligment
- Neighbor Joining
- Parsimony
- Maximum likelihood



Statistics

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

Example AFLP



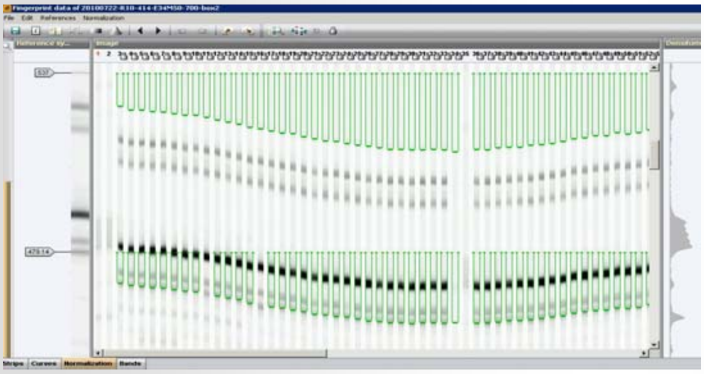
Import gel

Define lanes

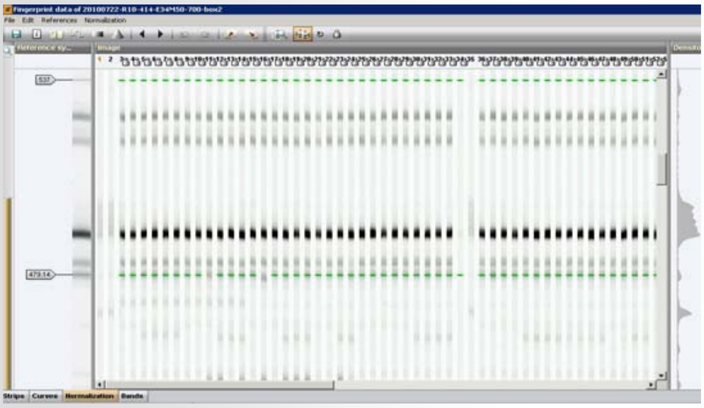
Link lanes to Entries

Use internal controle band for normalisation

Example AFLP

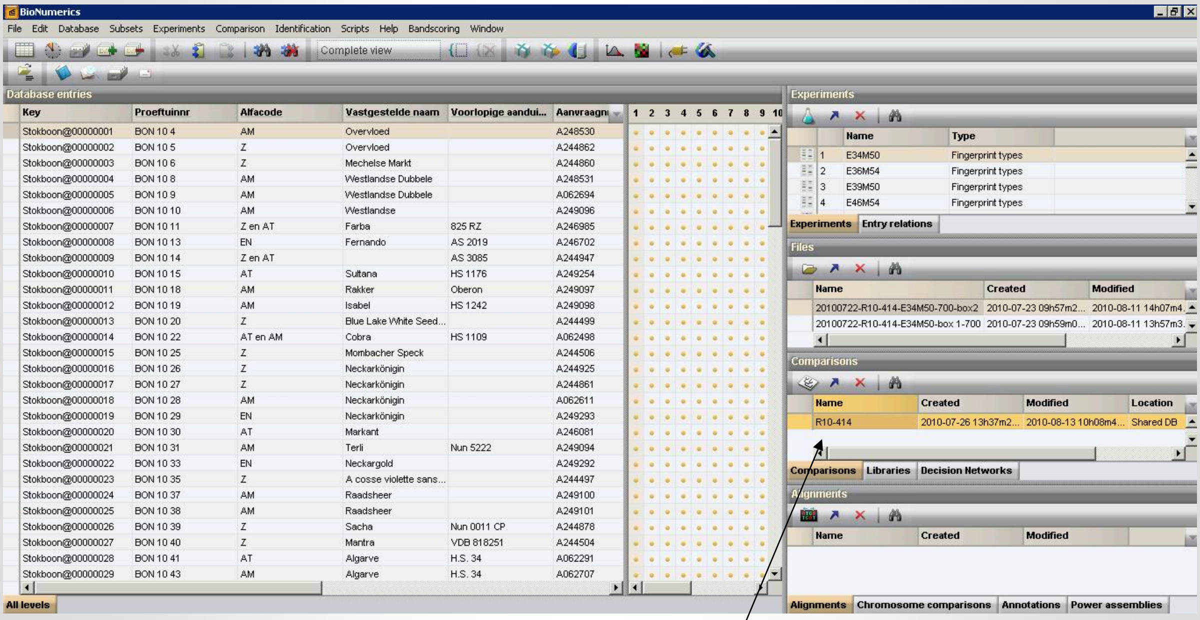


Before normalization



After normalization

Example AFLP



Key	Proeftuinnr	Alfacode	Vastgestelde naam	Voorlopige aandui...	Aanvraagn
Stokboon@00000001	BON 10 4	AM	Overvloed		A248530
Stokboon@00000002	BON 10 5	Z	Overvloed		A244862
Stokboon@00000003	BON 10 6	Z	Mechelse Markt		A244860
Stokboon@00000004	BON 10 8	AM	Westlandse Dubbele		A248531
Stokboon@00000005	BON 10 9	AM	Westlandse Dubbele		A062694
Stokboon@00000006	BON 10 10	AM	Westlandse		A249096
Stokboon@00000007	BON 10 11	Z en AT	Farba	825 RZ	A246985
Stokboon@00000008	BON 10 13	EN	Fernando	AS 2019	A246702
Stokboon@00000009	BON 10 14	Z en AT		AS 3085	A244947
Stokboon@00000010	BON 10 15	AT	Sultana	HS 1176	A249254
Stokboon@00000011	BON 10 18	AM	Rakker	Oberon	A249097
Stokboon@00000012	BON 10 19	AM	Isabel	HS 1242	A249098
Stokboon@00000013	BON 10 20	Z	Blue Lake White Seed...		A244499
Stokboon@00000014	BON 10 22	AT en AM	Cobra	HS 1109	A062498
Stokboon@00000015	BON 10 25	Z	Mombacher Speck		A244506
Stokboon@00000016	BON 10 26	Z	Neckarkönigin		A244925
Stokboon@00000017	BON 10 27	Z	Neckarkönigin		A244861
Stokboon@00000018	BON 10 28	AM	Neckarkönigin		A062611
Stokboon@00000019	BON 10 29	EN	Neckarkönigin		A249293
Stokboon@00000020	BON 10 30	AT	Markant		A246081
Stokboon@00000021	BON 10 31	AM	Terli	Nun 5222	A249094
Stokboon@00000022	BON 10 33	EN	Neckargold		A249292
Stokboon@00000023	BON 10 35	Z	A cosse violette sans...		A244497
Stokboon@00000024	BON 10 37	AM	Raadsheer		A249100
Stokboon@00000025	BON 10 38	AM	Raadsheer		A249101
Stokboon@00000026	BON 10 39	Z	Sacha	Nun 0011 CP	A244878
Stokboon@00000027	BON 10 40	Z	Mantra	VDB 818251	A244504
Stokboon@00000028	BON 10 41	AT	Algarve	H.S. 34	A062291
Stokboon@00000029	BON 10 43	AM	Algarve	H.S. 34	A062707

	Name	Type
1	E34M50	Fingerprint types
2	E36M54	Fingerprint types
3	E39M50	Fingerprint types
4	E46M54	Fingerprint types

Name	Created	Modified
20100722-R10-414-E34M50-700-box2	2010-07-23 09h57m2...	2010-08-11 14h07m4.
20100722-R10-414-E34M50-box 1-700	2010-07-23 09h59m0...	2010-08-11 13h57m3.

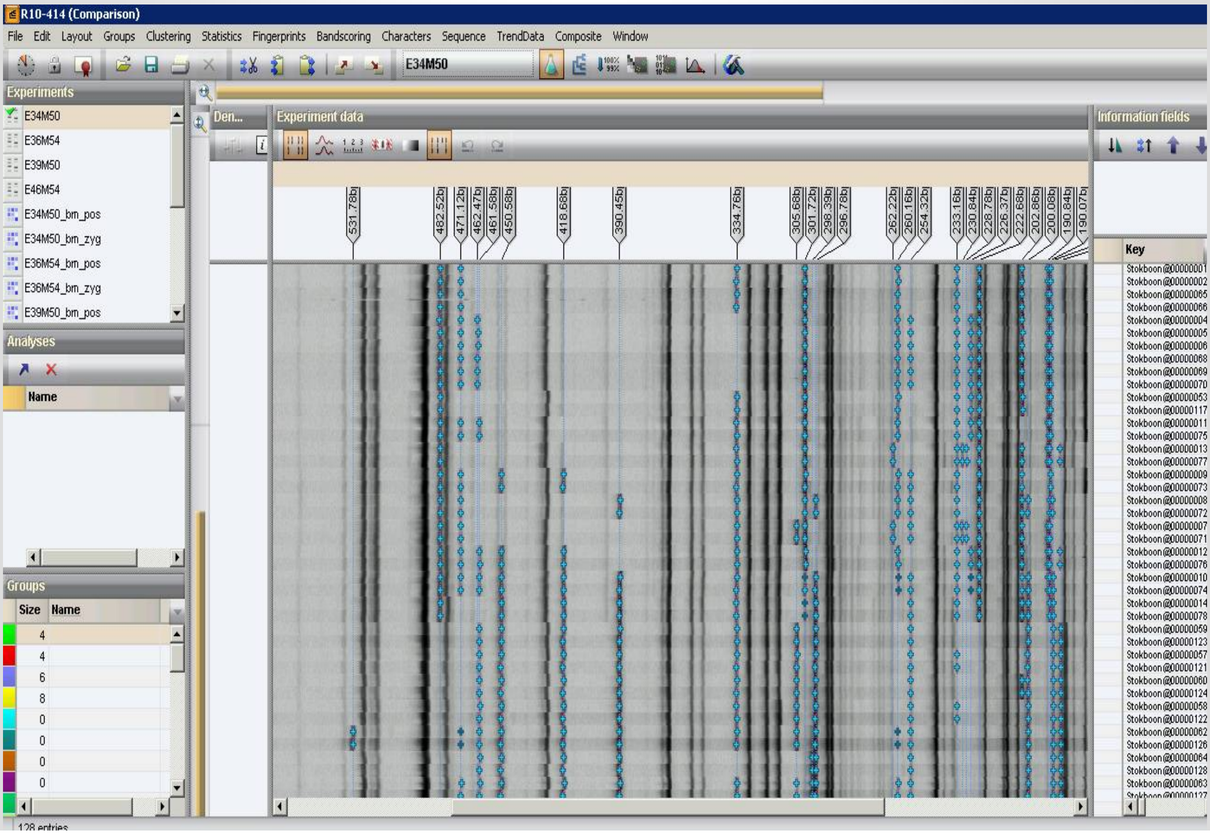
Name	Created	Modified	Location
R10-414	2010-07-26 13h37m2...	2010-08-13 10h08m4...	Shared DB

Name	Created	Modified

Open comparison window
To compare several different experiments from different samples

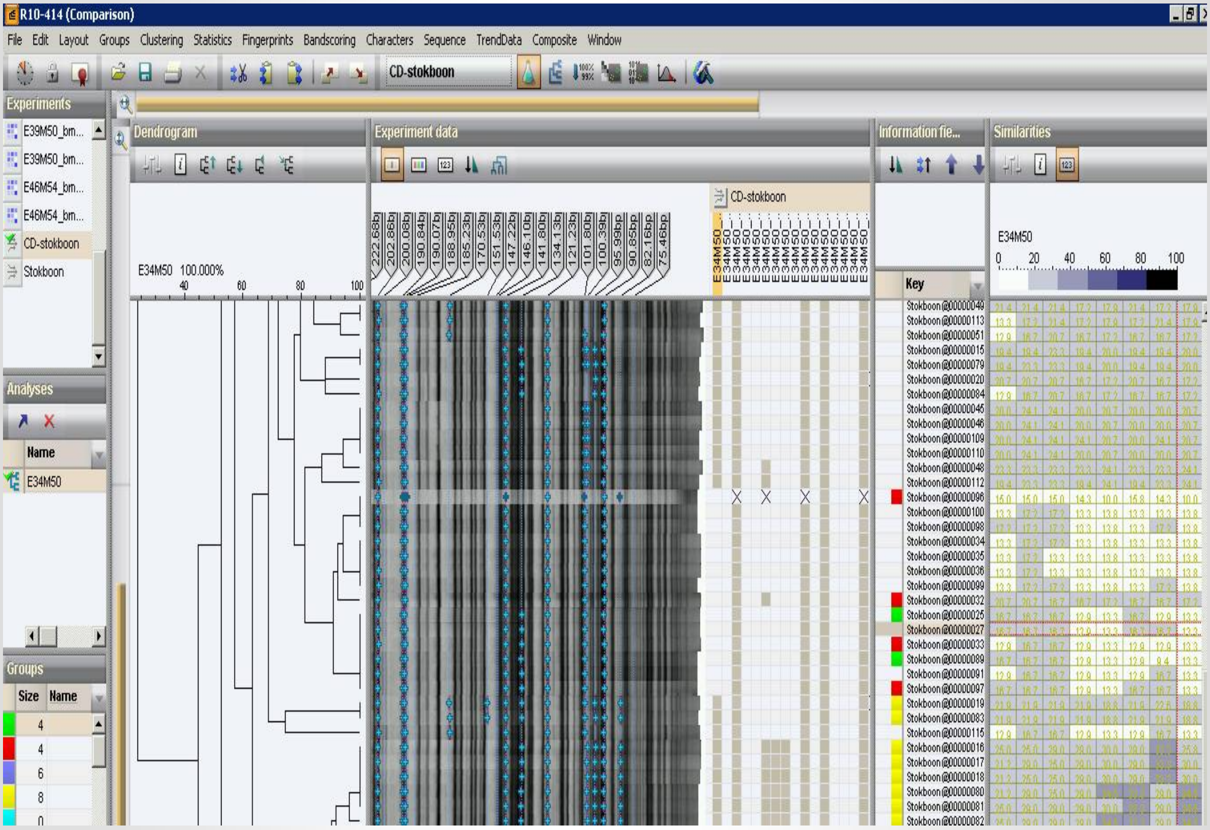
APPLIED MATHS

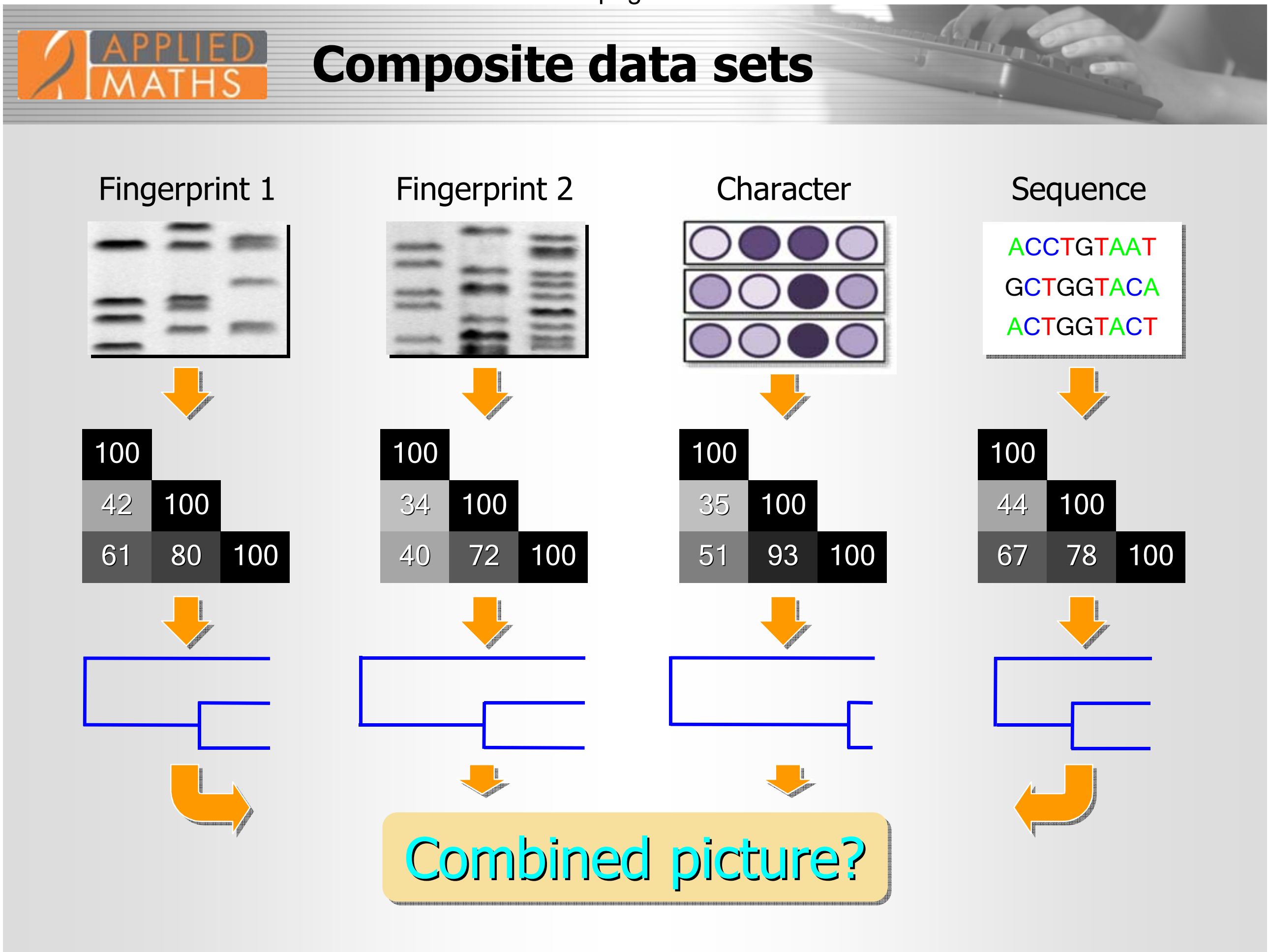
Example AFLP



APPLIED MATHS

Example AFLP





Composite data sets

Fingerprint 1		
100		
42	100	
61	80	100

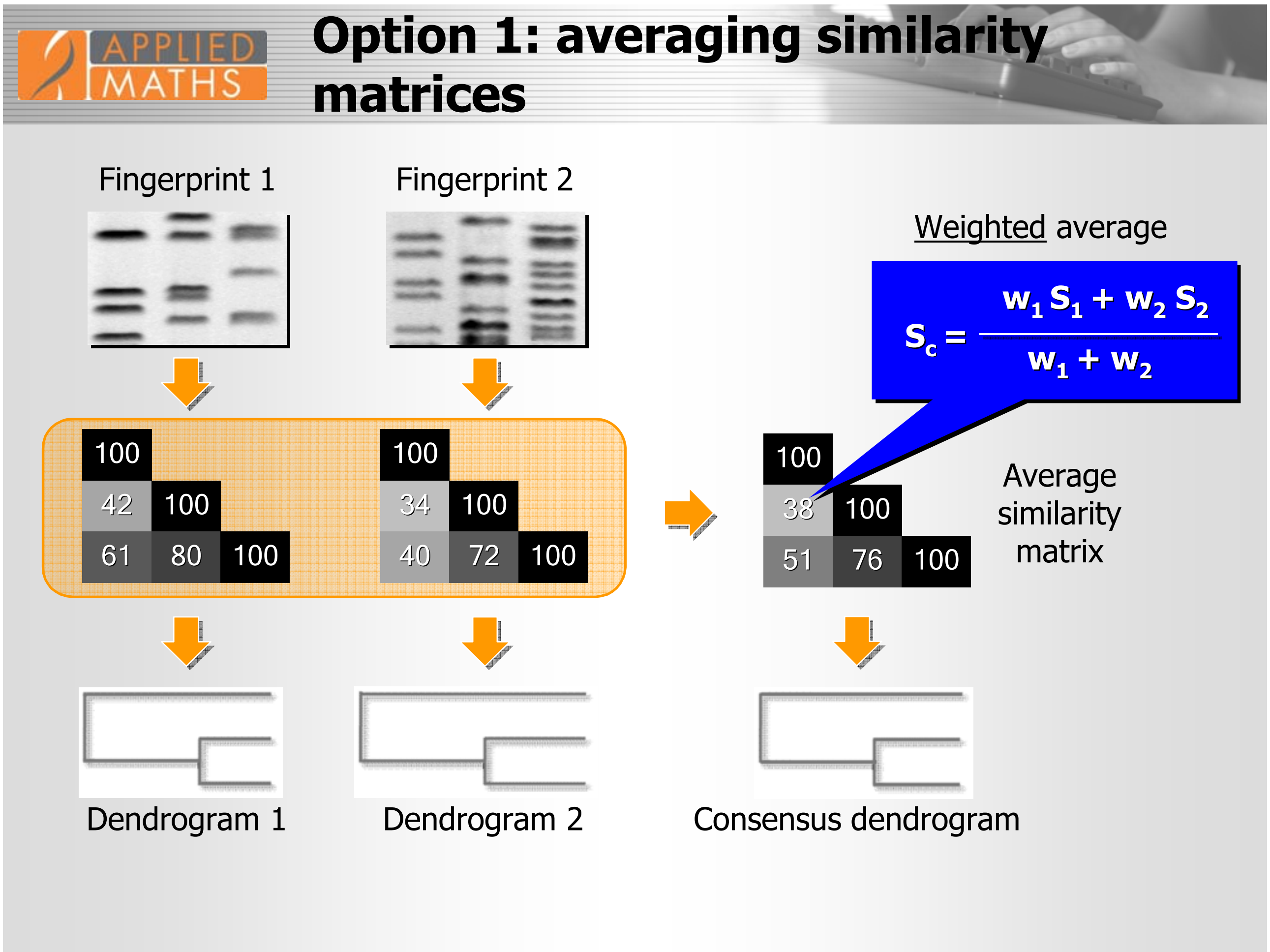
Fingerprint 2		
100		
34	100	
40	72	100

Character		
100		
35	100	
51	93	100

Sequence		
100		
44	100	
67	78	100

ACCTGTAAT
GCTGGTACA
ACTGGTACT

Combined picture?



Option 1: averaging similarity matrices

Fingerprint 1		
100		
42	100	
61	80	100

Fingerprint 2		
100		
34	100	
40	72	100

Weighted average

$$S_c = \frac{w_1 S_1 + w_2 S_2}{w_1 + w_2}$$

100		
38	100	
51	76	100

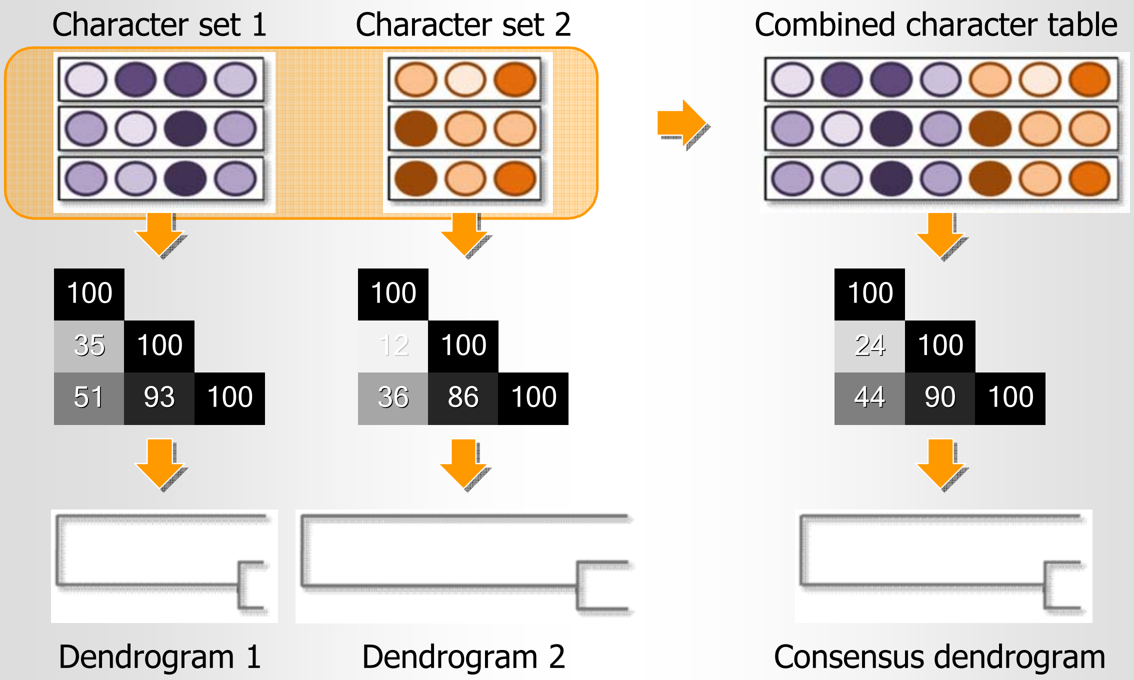
Average similarity matrix

Dendrogram 1

Dendrogram 2

Consensus dendrogram

Option 2: creation of combined character tables



Character set 1

100		
35	100	
51	93	100

Dendrogram 1

Character set 2

100		
12	100	
36	86	100

Dendrogram 2

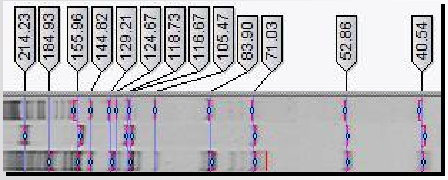
Combined character table

100		
24	100	
44	90	100

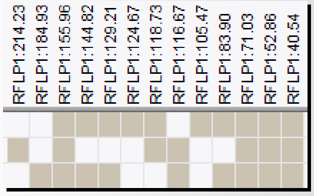
Consensus dendrogram

Conversion to character tables

Fingerprint



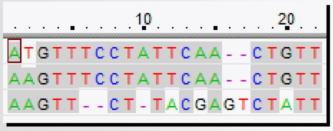
Band matching table



RFLP1:214.23	RFLP1:184.93	RFLP1:155.96	RFLP1:144.82	RFLP1:129.21	RFLP1:124.67	RFLP1:118.73	RFLP1:116.67	RFLP1:105.47	RFLP1:83.90	RFLP1:71.03	RFLP1:52.86	RFLP1:40.54
		▨	▨	▨	▨	▨		▨	▨	▨	▨	▨
▨		▨				▨	▨			▨	▨	▨
	▨	▨	▨	▨			▨		▨	▨	▨	▨

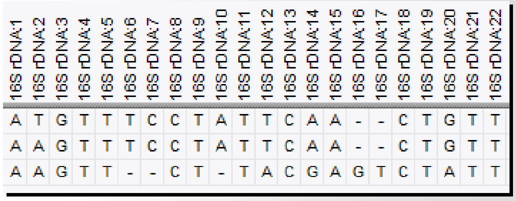
Binary (presence/absence) or quantitative data

(Aligned) sequence



```
         10        20
ATGTTTCCTATTCAA--CTGTT
AAGTTTCCTATTCAA--CTGTT
AAGTT--CT-TACGAGTCTATT
```

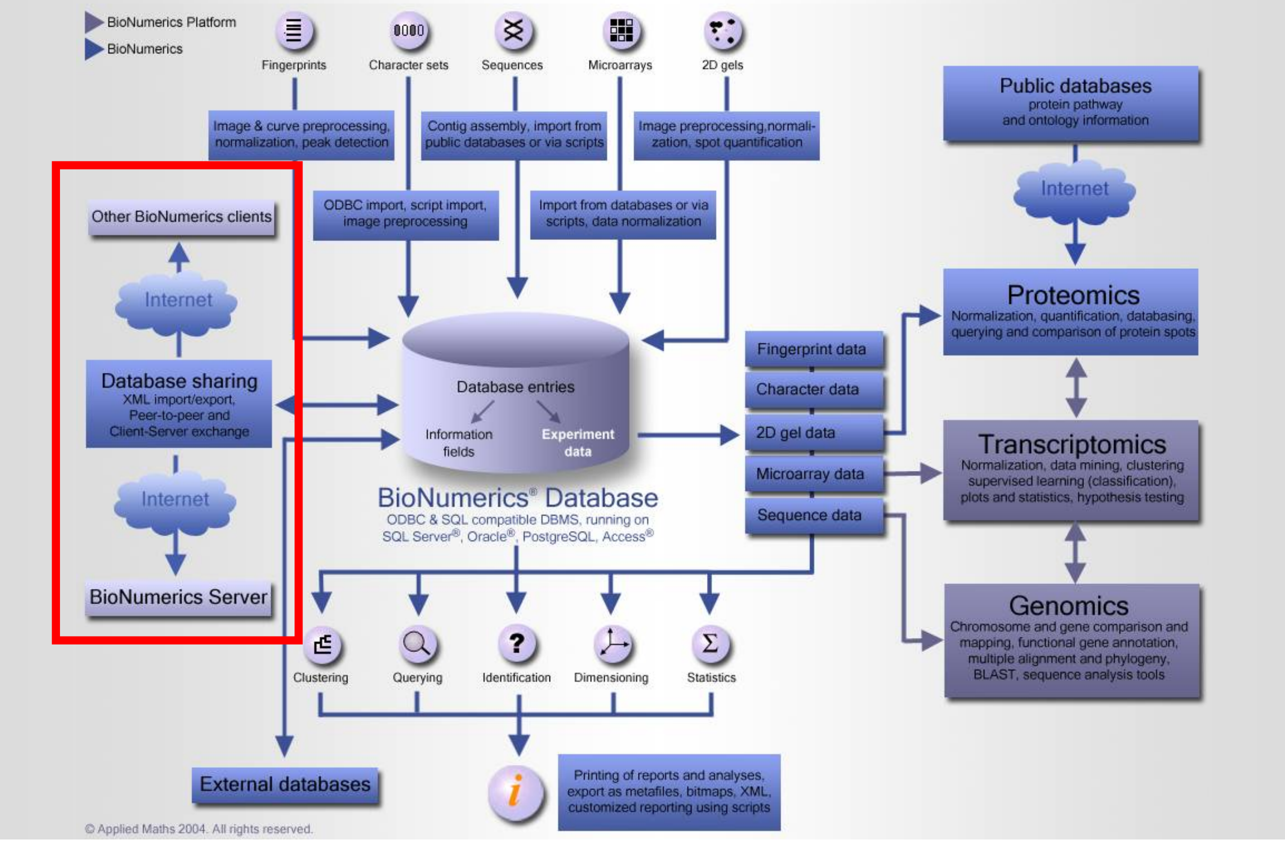
Sequence table



16S rDNA:1	16S rDNA:2	16S rDNA:3	16S rDNA:4	16S rDNA:5	16S rDNA:6	16S rDNA:7	16S rDNA:8	16S rDNA:9	16S rDNA:10	16S rDNA:11	16S rDNA:12	16S rDNA:13	16S rDNA:14	16S rDNA:15	16S rDNA:16	16S rDNA:17	16S rDNA:18	16S rDNA:19	16S rDNA:20	16S rDNA:21	16S rDNA:22
A	T	G	T	T	T	C	C	T	A	T	T	C	A	A	-	-	C	T	G	T	T
A	A	G	T	T	T	C	C	T	A	T	T	C	A	A	-	-	C	T	G	T	T
A	A	G	T	T	-	-	C	T	-	T	A	C	G	A	G	T	C	T	A	T	T

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

What you can do



BioNumerics Platform

BioNumerics

Fingerprints

Character sets

Sequences

Microarrays

2D gels

Image & curve preprocessing, normalization, peak detection

Contig assembly, import from public databases or via scripts

Image preprocessing,normalization, spot quantification

ODBC import, script import, image preprocessing

Import from databases or via scripts, data normalization

Other BioNumerics clients

Internet

Database sharing
XML import/export,
Peer-to-peer and
Client-Server exchange

Internet

BioNumerics Server

Database entries

Information fields

Experiment data

BioNumerics® Database
ODBC & SQL compatible DBMS, running on SQL Server®, Oracle®, PostgreSQL, Access®

Fingerprint data

Character data

2D gel data

Microarray data

Sequence data

Public databases
protein pathway and ontology information

Internet

Proteomics
Normalization, quantification, databasing, querying and comparison of protein spots

Transcriptomics
Normalization, data mining, clustering supervised learning (classification), plots and statistics, hypothesis testing

Genomics
Chromosome and gene comparison and mapping, functional gene annotation, multiple alignment and phylogeny, BLAST, sequence analysis tools

Clustering

Querying

Identification

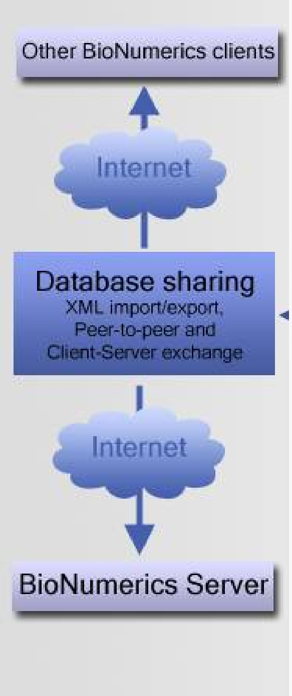
Dimensioning

Statistics

External databases

Printing of reports and analyses, export as metafiles, bitmaps, XML, customized reporting using scripts

What you can do



Other BioNumerics clients

Internet

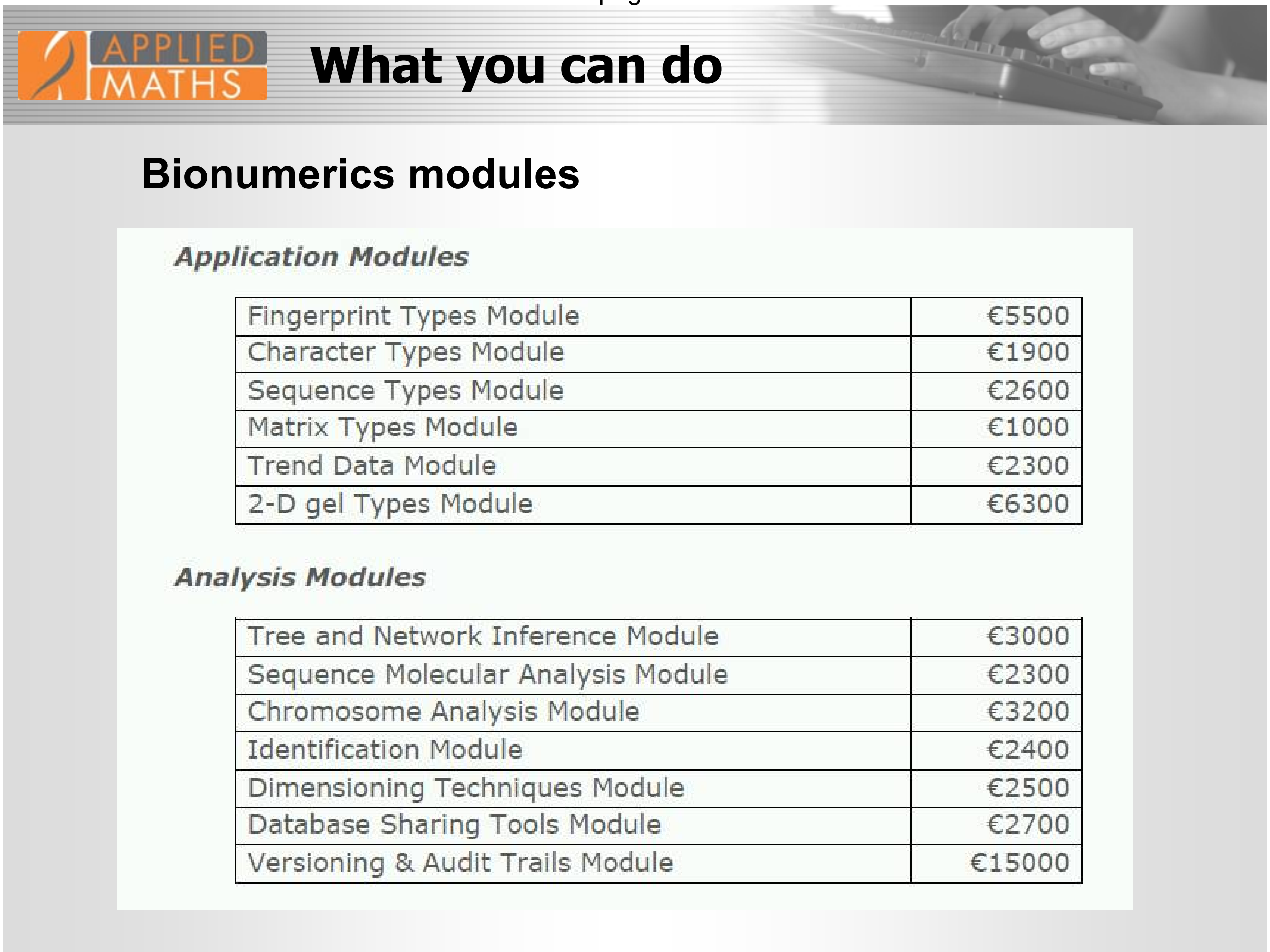
Database sharing
XML import/export,
Peer-to-peer and
Client-Server exchange

Internet

BioNumerics Server

BN works on Windows

Many different users can work on the data at the same time



What you can do

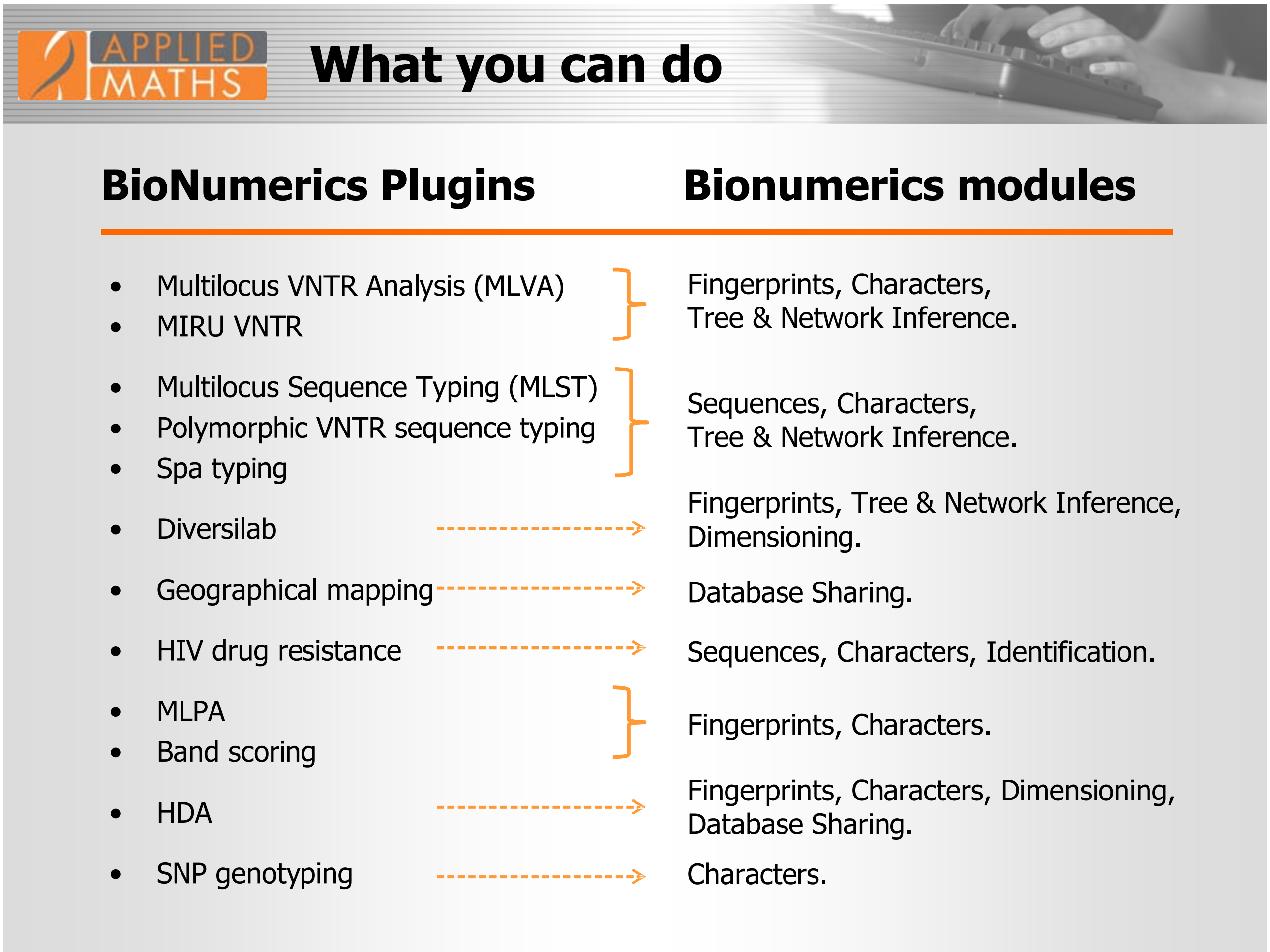
Bionumerics modules

Application Modules

Fingerprint Types Module	€5500
Character Types Module	€1900
Sequence Types Module	€2600
Matrix Types Module	€1000
Trend Data Module	€2300
2-D gel Types Module	€6300

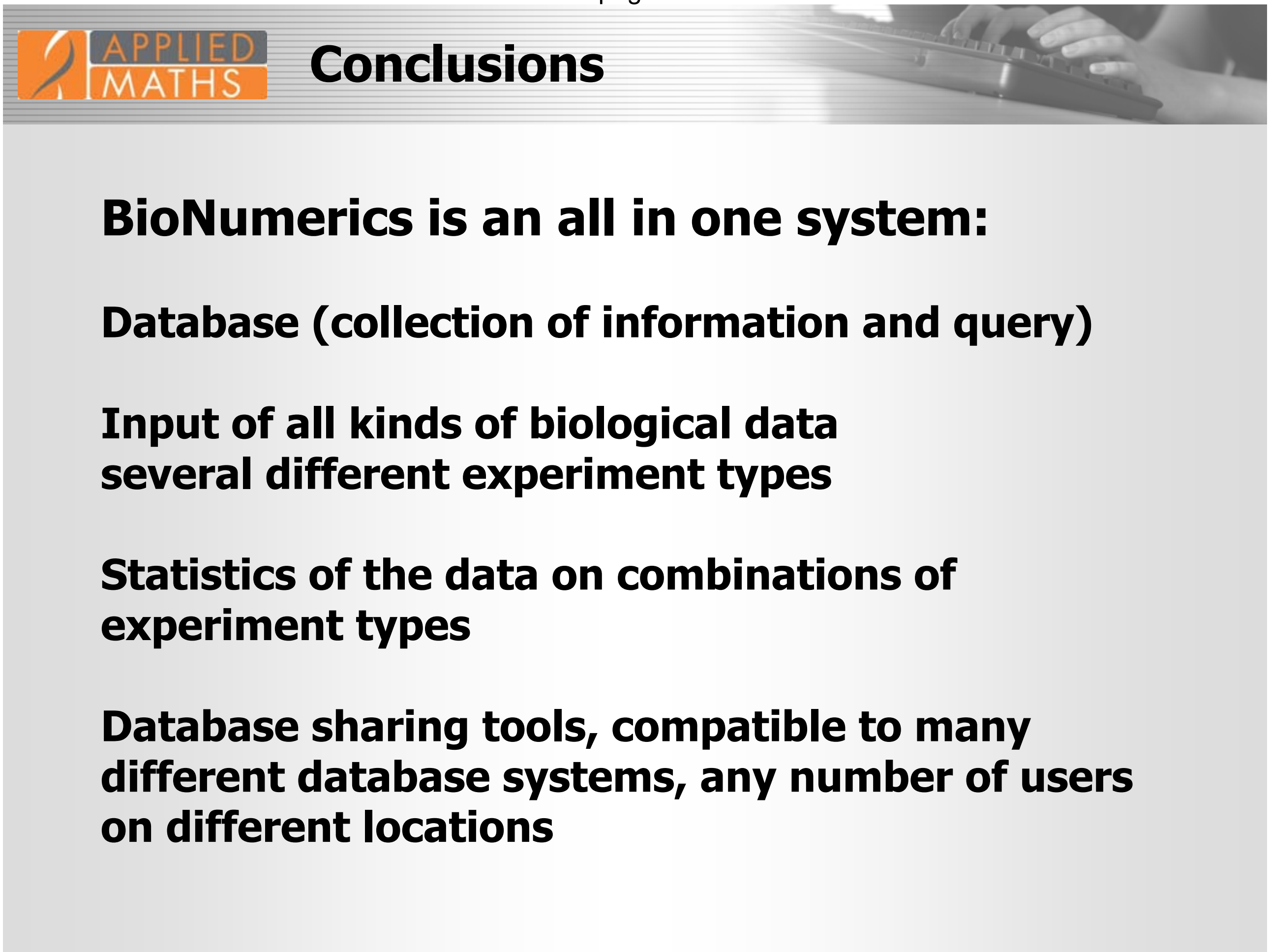
Analysis Modules

Tree and Network Inference Module	€3000
Sequence Molecular Analysis Module	€2300
Chromosome Analysis Module	€3200
Identification Module	€2400
Dimensioning Techniques Module	€2500
Database Sharing Tools Module	€2700
Versioning & Audit Trails Module	€15000



What you can do

BioNumerics Plugins	Bionumerics modules
• Multilocus VNTR Analysis (MLVA) • MIRU VNTR	Fingerprints, Characters, Tree & Network Inference.
• Multilocus Sequence Typing (MLST) • Polymorphic VNTR sequence typing • Spa typing	Sequences, Characters, Tree & Network Inference.
• Diversilab	Fingerprints, Tree & Network Inference, Dimensioning.
• Geographical mapping	Database Sharing.
• HIV drug resistance	Sequences, Characters, Identification.
• MLPA • Band scoring	Fingerprints, Characters.
• HDA	Fingerprints, Characters, Dimensioning, Database Sharing.
• SNP genotyping	Characters.



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]