



# BioNumerics



One universal platform for databasing and analysis of 1D patterns, 2D gels, phenotype arrays, and DNA/protein sequences



Powerful databasing, integrated networking and a wide range of data mining, clustering, identification and statistical tools

BioNumerics is the only software platform to offer integrated analysis of all major applications in Bioinformatics: 1D electrophoresis gels, all kinds of chromatographic and spectrometric profiles, 2D protein gels, phenotype characters, microarrays, and sequences. The unique power of BioNumerics lies in its ability to combine information from various genomic and phenotypic sources into one global database and conduct conclusive analyses.

GelCompar is a version of BioNumerics which can only analyse gel data but otherwise has the same functions as BioNumerics. Therefore the descriptions of Sequence, Character & Trend data modules in this brochure are not relevant to GelCompar and only Fingerprint Types and Analysis modules should be considered.

BioNumerics can run on industry leading database engines such as Oracle® and Microsoft® SQL Server™. With its integrated networking and client-server features, the software is the perfect backbone for universal data management and analysis within and between laboratories of any size.

- Unparalleled 1D gel and fingerprint analysis
- Innovative 2D gel analysis with powerful databasing and querying
- Comprehensive sequence and chromosome analysis
- Exploration of high-throughput microarray and genechip expression data
- Combined querying, mining and analysis of all aforementioned data types
- Numerous supervised and unsupervised learning techniques and statistical tests

## Common applications with BioNumerics

BioNumerics is a general platform for databasing and analysis of virtually all types of biological data, featuring three basic concepts:

- **Databasing:** The unique SQL database design offers storage, retrieval and querying tools for 1D profiles, 2D gels, sequences, character data, curves and kinetic readings, similarity matrices, etc. BioNumerics is the perfect backbone for central and lab-wide storage and retrieval of data.
- **Comparison:** The software's main emphasis is on mining large data sets: screening for differences and similarities, revealing and visualizing differential data, grouping and clustering, identification, and statistics. The unique power of BioNumerics lies in its capacity to synthesize the results from different experimental approaches into conclusive answers.

- **Networking:** Thanks to its integrated networking and XML based Internet tools, BioNumerics has been widely implemented in networks for the exchange of biodata in a uniform and standardized way.

Below we list a number of applications for which BioNumerics is commonly used. Thanks to its flexible design and comprehensive scripting possibilities, the software can be used in many more research or diagnostic fields. A number of techniques for which the data is commonly analyzed using BioNumerics are also listed. Please contact us if you would like to learn more about integrating your specific application(s) and/or techniques in BioNumerics.

## Application fields

- Bacterial, fungal and viral epidemiological typing
- Bacterial source tracking
- Microbial community analysis
- Environmental study
- Fermentation monitoring (breweries, dairy products, wine,...)
- Food quality control
- Authenticity testing (food, organic products)
- Variety/cultivar identification
- Managing starter cultures
- Mutation detection and analysis
- Disease and cancer diagnostics
- Plant and animal breeding
- Platform for transcriptomics and proteomics
- Taxonomy and identification
- Phylogenetic inference and evolution
- Population modelling

## Techniques

Data from most techniques can be directly processed and analyzed in BioNumerics. Some techniques require special import, processing and/or automation steps. For these techniques (in bold), a dedicated Plugin is available, developed and supported by Applied Maths. Click on the name for more information.

### **Fingerprint types**

- All 1-D electrophoresis patterns obtained from all kinds of instruments or gels (PFGE, AFLP, RFLP, RAPD, REP-PCR, ARDRA, isoelectric focusing etc.)
- Gradient gel electrophoresis (DGGE, TGGE)
- Microsatellite analysis
- Variable Number Tandem Repeat (VNTR) analysis (see MLVA)
- **Multi-Locus VNTR Analysis (MLVA)**
- Hetero-Duplex Analysis (HDA) (see CSCE analysis)
- **Conformation-Sensitive Capillary Electrophoresis (CSCE) analysis**

- Mass spectrometry (MALDI, SELDI)
- Gas chromatography and HPLC (fatty acid, quinone profiling)
- DHPLC typing
- **Riboprinter® patterns**

#### Character types

- Antibiotic resistance tests
- Fatty Acid Methyl Ester (FAME) analysis
- Biolog® Phenotype arrays
- API® phenotypic test panels (BioMérieux®)
- Microarrays
- Dot blots and probe sets
- Spoligo typing
- Biochemical tests
- Enzymatic and metabolic activity tests

#### Sequence types

- Contig assembly from automated sequencers (Applied Biosystems, Beckman, Amersham)
- Ribosomal RNA gene sequence analysis (identification, phylogeny)
- Housekeeping gene sequence analysis (typing, population study)
- **Multilocus Sequence Typing (MLST)**
- **Spa-typing of multidrug resistant Staphylococcus aureus (MRSA)**

#### Trend data types

- Real-time PCR
- Kinetic readings of metabolic/enzymatic activity (e.g. Biolog®, PhenePlate®)

#### Matrix types

- DNA hybridization
- Complete or partial similarity/distance matrices

#### 2D gel types

- Multigel matching and comparison from all common 2D gel techniques
- **Multiplex 2D gels (DIGE)**

# Modules and Features of BioNumerics

BioNumerics consists of:

- **6 application modules:** *Fingerprint types, Character types, Sequence types, Trend data types, and 2D gel types and Matrix types.*
- **4 analysis modules:** *Cluster analysis, Identification & Libraries, Dimensioning techniques & Statistics, and Database Sharing Tools.*

Each analysis module can be combined with any or all application modules. The descriptions below list most, but not all functions and possibilities of the BioNumerics modules. Please contact us for details.

## ▶ **BioNumerics Fingerprint types**

- **Image processing and normalization.** Input of 8-bit, 12-bit and 16-bit TIFF images, densitograms, and chromatograms of unlimited filesize. Automatic lane finding on images for all types of gels. Gelstrip borders and tracking splines adjustable for each individual lane.

Automated and manual alignment by pattern recognition using external reference patterns and/or internal reference bands. On-screen normalization of TIFF images with indication of reliability and possible misalignments. Direct processing of ABI, Beckman and Amersham chromatogram files with inline reference tracks. Adjustable background subtraction using *rolling disk* method. Adjustable curve smoothing by *Arithmetic, Median* and *Mode* filters and *Least Square* filtering algorithms. Adjustable *Richardson-Lucy* deconvolution.

- **Database.** Entry description by up to 60 information fields, up to 80 characters long each. Easy drag & drop linkage of multiple experiments to database entries. Powerful search engine for combined database searches on different information fields and experiment presence, character values, and ranges. Storage and management of composite database queries. Imaging of any selection of patterns by normalized 2D-bitmap strips, densitograms or reconstructed patterns. Multi-database system, each database can contain many different fingerprint types. Direct comparison of differently normalized patterns through automatic MW remapping functions. Advanced database exchange functions using the Bundles concept and XML exchange packets. Compatibility with external databases via ODBC.

- **Quantification.** Approved band-search algorithms with adjustable sensitivity for shoulder and double-band finding based on curve. Possibility to find and mark *uncertain bands*. Quantification of molecular sizes or any other metric unit using linear, logarithmic, combined logarithmic-third power regression, cubic spline or pole functions. Accurate expression of protein or nucleic acid quantities or concentrations based on cubic spline regression using known calibration bands. Extensive anatomy and comparative quantification of bands between groups of patterns. Generation of characteristics tables and extensive comparative reports between unlimited numbers of patterns, with indication of molecular weight, fragment length, and absence/presence or absolute protein or DNA amount per band. Search for discriminative bands between selected groups of patterns; search for unique and common bands within selections. Binary and quantitative band matching tables of multiple combined fingerprints.

## ▶ **BioNumerics Character types**

Universal import tools and programmable routines (scripts) for all kinds of text-oriented character data. Character types may include any existing test panel, binary, multi-state or continuous within any range, with fixed or variable number of characters. Character names may be entered by the user or automatically imported. Unlimited length of character arrays.

Direct processing of densitometric arrays (spots), test panels, microtiter plates, dot blots, etc. from TIFF files. Character profiles are displayed in a panel with user-defined color scales or in a list with values. Display of truthful image of any test panel and easy data input on-screen.

## ▶ **BioNumerics Sequence types**

Project-based contig assembly and consensus editing from sequencer chromatogram files (ABI, Beckman, MegaBace). Direct import of EMBL, GenBank, Flat A, and FASTA formats. Import of nucleic acid and amino acid sequences. Easy paste from clipboard, and manual editing. Contig projects can be opened directly from entry editor, comparison and multiple alignment

## ▶ **BioNumerics Trend Data types**

Analyzes series of readings in function of a changing factor, which define a trend. Examples are the kinetic analysis of metabolic and enzymatic activity, real-time PCR, or time-course experiments using microarrays. Although multiple readings per experiment are mostly done in function of time, they can also depend on another factor, for example in function of different concentrations.

Twelve different curve fit models, including Logistic growth, Gompertz, Gaussian, Hyperbolic, Power, Exponential, etc. from which specific parameters can be derived and used for analysis and comparison. User can add custom parameters such as statistic parameters, slopes, and values at fixed X.

Comparison and clustering can be done on a selected parameter or a combination of multiple parameters. Comprehensive curve plotting tools.

#### ▶ **BioNumerics 2D Gel types**

Please enquire for full details.

#### ▶ **BioNumerics Matrix types**

Import of similarity or distance matrices. Partial matrices accepted (e.g. DNA homology matrices); special clustering algorithms for incomplete matrices.

#### ▶ **Comparison and Cluster Analysis**

- **Methods.** Creation of dendrograms including up to 10,000 database entries using product-moment Pearson correlation, cosine correlation, Dice or Nei and Li, Jaccard, Jeffrey's X, Ochiai and area sensitive relatives for banding patterns, Gower, Canberra metric, Simple Matching, etc. Categorical coefficient for multi-state character data such as MLST or VNTR. Unweighted pair-grouping (UPGMA), complete linkage (furthest neighbor), single linkage (nearest neighbor), Ward or Neighbor Joining clustering. Adjustable trace-to-trace optimization and tolerance settings for banding patterns. Statistical determination of most justified tolerance settings for banding patterns.

Phylogenetic inference methods: *Generalized Parsimony, Maximum Likelihood.* Population modelling: Analysis of categorical data such as MLST or VNTR (MLVA) using *Minimum Spanning Trees* to reconstruct evolutionary models. Advanced presentation and editing tools.

- **Interpretation.** Combined display of character images, sequences, normalized pattern images, with similarity matrices and sorted according to dendrogram(s). Indication of statistical error at all linkage levels and calculation of co-phenetic correlation. "Seaweed" and pseudo-rooted representation for unrooted trees. Bootstrap analysis for single or composite datasets. Display of sorted similarity matrices, shaded or with numerical similarity values. Comprehensive edit and publishing functions. Professional presentation and printing facilities, in a WYSIWYG environment. Direct interaction between database and dendrogram. Incremental and decremental clustering: new entries can be added to or deleted from existing cluster analyses, without having to recalculate the complete analysis. All features of a comparison are stored to disk.

- **Congruence between techniques.** Calculation of global similarity or congruence between different techniques as matrix or dendrogram. Easy visualization of taxonomic depth or level of each technique by pairwise regression plots of similarities.

- **Composite cluster analysis.** Different data sets of the same type and of different types (fingerprint, character, sequence and matrix) can be combined into one consensus clustering. Calculation of global similarity by merging characters or by averaging experiment-related similarities. Optional weighting based on number of characters or defined by the user.

- **Plots and graphs.** Creation of 2-D and 3-D bar graphs, contingency tables, 2-D and spatial 3-D scatterplots or feature plots from database fields and characters. Professional presentation, printing and exporting tools.

#### ▶ **Identification & Libraries**

- **Database screening.** Fast identification of batches of entries with entire databases or selections from databases, using all available coefficients.

- **Libraries.** Creation of highly characteristic identification libraries using the open unlimited multi-library system. Specific similarity measures and settings can be defined for specific experiment types. Comprehensive identification reports showing results for each available experiment. Many different viewing options and statistical tools to facilitate interpretation.

- **Neural Networks.** Neural Networks can be trained per experiment type and used for quick and accurate identification of complex groupings.

#### ▶ **Dimensioning Techniques & Statistics**

- **Principal Components Analysis.** Non-hierarchic grouping by PCA. Spatial representation of clouds of entries in X-Y-Z coordinate system. Indication of total discrimination of axes. Real-time rotation of coordinate system to enhance perception of 3-D structures. Advanced Open-GL presentation and layout for publication. Delineation of populations using colors and/or codes. Plotting of dendrogram branches on PCA for advanced grouping comparisons.
- **Multi-Dimensional Scaling.** Non-hierarchic grouping by MDS. Same presentation features as for PCA. Iterative optimization of distances according to similarity matrix.
- **Self-Organizing Maps.** Non-hierarchic grouping by the technique of SOM (Kohonen maps), a sort of neural network.
- **MANOVA.** Advanced statistical analysis of discriminative features between selected groups with indication of confidence based on multivariate analysis of variance.
- **Statistics.** A number of parametric and non-parametric statistical tests can be performed in an easy and intuitive environment (Chi-square test, T-test, Wilcoxon signed-ranks test, Kruskal-Wallis test, ANOVA, Pearson correlation test, Spearmann rank-order test. Automatic display available tests for each input data type. Kolgomorov-Smirnov test for normality of character data.

#### ▶ **Database Sharing Tools**

- **Connected Databases.** Possibility to plug BioNumerics onto ODBC compatible database systems such as Oracle, Microsoft Access, SQL Server, MySQL, PostgreSQL.
- **Bundles.** Any selected information and experiment data for selections of entries from the database can be condensed into a *Bundle*. Bundles can be exchanged over internet and opened in a recipient database. Automatic *remapping* makes full comparison between different fingerprint systems possible.
- **Client-Server setup.** Client functions come with the Database Sharing Tools. Functions include querying and downloading entries from the central Server database; Upload of data to the Server for identification; Receipt of detailed identification report from Server. Call for prices of BioNumerics Server package.
- **XML export/import.** Creation of XML files from any selection of entries and techniques from the database. Received XML files can be imported as fully editable database entries. XML exchange is the preferred way of exchanging database entries in a peer-to-peer network.

#### ▶ **BioNumerics Network Licenses**

- **Powerful network license solution!** Compatible with Windows NT, Windows 2000 and Windows XP. See also the [detailed description](#) of the network licensing system.
- **License limits.** Network licenses are available for any number of users. Contact us for details and pricing.

All material © 2008 Applied Maths NV

For further information, a demonstration analysis of your own data and pricing please contact:

## **BioSystematica**

Mountain Hall · Sarnau · Llandysul  
Ceredigion · Wales · SA44 6QA · UK

Tel +44-(0)-845-3283461  
[mail@biosystematica.com](mailto:mail@biosystematica.com)

[www.biosystematica.com](http://www.biosystematica.com)