

BioNumerics consists of:

- **6 application modules:** *Fingerprint types, Character types, Sequence types, 2D gel types, Matrix types, and Trend Data types.*
- **4 analysis modules:** *Cluster analysis, Identification & Libraries, Dimensioning techniques, 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 call for details and prices.

## ▪ **General functionality**

---

**Database.** Fully relational multi-user database. Display of up to 150 information fields per database entry. Easy drag & drop linkage of multiple experiments to database entries. Powerful search engine for combined database searches on information fields and experiment presence and/or contents (character values, sequences, ranges, bands, etc.). Storage and management of database queries and external attachments. Multi-database system: each database can contain any combination of different experiment types. Creation of levels and relations preserving your real-life data structure.

## ▪ **Fingerprint types**

---

**Image processing and normalization.** Input of any bitmap images, densitograms, and chromatograms of unlimited file size. Image pre-editing and cleaning. 3D representation of bitmaps. Automatic lane finding for all types of gels. Gelstrip borders and tracking splines adjustable for individual lanes. Automated and manual alignment by pattern recognition using external reference patterns and/or internal reference bands. On-screen normalization of bitmap images with indication of reliability and possible misalignments. Direct processing of sequencer chromatogram files and fragment analysis files with inline reference tracks. Adjustable background subtraction and curve smoothing. Spotremoval. Display of any combination of normalized 2D-bitmap strips, densitograms or reconstructed patterns. Direct comparison of patterns normalized with different reference systems.

**Quantification.** Band-search algorithms with adjustable sensitivity for shoulder and double-band finding. 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. Comparative quantification of bands between groups of patterns. Generation of tables and reports for unlimited numbers of patterns, indicating molecular weight, fragment length, absence/presence or absolute amounts of protein or DNA per band. Search for discriminative bands between selected groups of patterns; search for unique and common bands. Binary and quantitative band matching tables of multiple combined fingerprints. Possibility to define named band classes based upon size and position (e.g. for DGGE/TGGE analysis).

## ▪ **Character types**

---

Flexible import tools and programmable routines for the import of character data from text files and ODBC-compatible sources (MS Excel®, MS Access®,...). Character types may include any existing data type, numerical, binary or continuous, within any range, with fixed or variable number of characters. Character names may be given by the user or automatically imported. Unlimited length of character arrays. Possibility to map character values to categorical names according to predefined criteria. Direct digitization and processing of micro-arrays, test panels, microtiter plates, dot blots, etc. from TIFF files. Character profiles can be displayed in a panel with user-defined representations and color scales or in a list with values. Display of truthful image of any test panel and easy on-screen data input.

## ▪ **Sequence types**

---

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. Project-based contig assembly and consensus editing from sequencer chromatogram files (ABI, Beckman, MegaBace). Full IUPAC code support for consensus naming. Contig projects can be opened from entry editor, comparison and multiple alignments. SNP detection and analysis (beta version).

## ▪ **Matrix types**

---

Import of similarity or distance matrices. Partial matrices accepted (e.g. DNA homology matrices). Unlimited matrix size.

## ▪ **Trend Data**

---

Analyzes series of readings in function of a changing factor (time, temperature, etc.), which define a trend. Examples are bacterial growth curves, kinetics of metabolic and enzymatic activity measurements, real-time PCR, or time-course experiments using microarrays. Mathematical fitting using any of twelve different models, including Logistic growth, Gompertz, Gaussian, Hyperbolic, Power, Exponential, etc. with automatic parameter calculation, useful 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 plotting tools with color and group indications.

## ▪ **2-D gel types**

---

Full proteomics analysis. See BioNumerics 2D for further details.

## ▪ **Comparison and Cluster Analysis**

---

**Methods.** Creation of dendrograms up to 10,000 database entries, using Pearson product-moment correlation, cosine correlation, Dice or Nei and Li, Jaccard, Jeffrey's X, Ochiai. Fuzzy logic and area sensitivity for banding patterns. Gower, Canberra metric, Simple Matching, etc. for character data. Categorical coefficient for multi-state data (VNTR, MLST, AB resistance patterns, etc.). 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, including faithful tree representation ('rendered trees').

**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. Impressive edit and publishing functions. Enhanced 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 can be stored to disk.

**Dendrogram degeneracy calculation.** Tracing of cluster degeneracy. Display of consensus trees. Includes comparison of trees obtained by different algorithms or from different techniques.

**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 3-D scatterplots from database fields and characters. Professional presentation, printing and exporting tools.

## ▪ **Identification**

---

**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 for each experiment type and used for quick and accurate identification of complex groupings.

**Decision Networks.** Allows you to build automated workflows that make decisions, predict features, perform queries, fill in fields, create graphs and plots, etc. Can be used for resistance prediction, in breeding research, for complex reporting or for automated analysis of multilevel and polyphasic data analysis and data sorting.

## ▪ **Dimensioning techniques**

---

**Principal Component Analysis.** Non-hierarchical grouping by PCA. Spatial representation of clouds of entries in userdefinable X-Y-Z coordinate systems. 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 and methodological validations.

**Multi-Dimensioning Scaling.** Non-hierarchical grouping by MDS. Iterative optimization of distances according to similarity matrix. Same presentation features as for PCA.

**Self-Organizing Maps.** Non-hierarchical grouping by the technique of Self-Organizing Maps (Kohonen maps), a sort of neural network approach, extremely useful for large and complex data sets.

**MANOVA.** Advanced statistical analysis of discriminative features between selected groups with indication of confidence and 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, Spearman rank-order test. Automatic display available tests for each input data type. Kolgomorov-Smirnov test for normality. Clear significance reporting.

## ▪ **Database Sharing Tools**

---

**XML export/import.** Creation of XML files from any selection of entries and techniques in the database. 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.

**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.

**Geographical coding.** Perform geocoding and plot database entries on an interactive geographical map based on locations present in the database.

## ▪ **BioNumerics Network Software**

---

**Powerful network solution.** Compatible with Windows NT, Windows 2000, Windows XP, and Windows Vista.

**License limits.** Network versions are available for any number of users. Contact us for details and pricing.