# The ARB Project

**An integrated, non-commercial software solution for Phylogenetic Treeing, Sequence Data Analysis and Molecular Probe Design**

Presentation by **Yadhu Kumar**, ARB Group

**Lehrstuhl für Mikrobiologie**
**Lehrstuhl für Rechnertechnik & Rechnerorganisation**
**Technische Universität München, München, Germany**

# Ideas

- **Central Database to maintain a structured integrative secondary data in combination with processed primary structures (aligned sequences) and any additional data assigned to the individual sequences.**

# Ideas

- **Central Database to maintain a structured integrative secondary data in combination with processed primary structures (aligned sequences) and any additional data assigned to the individual sequences.**

- **Comprehensive selection of software tools directly interacting with one another and as well as with the central database facilitating in depth analysis of molecular data.**

# Ideas

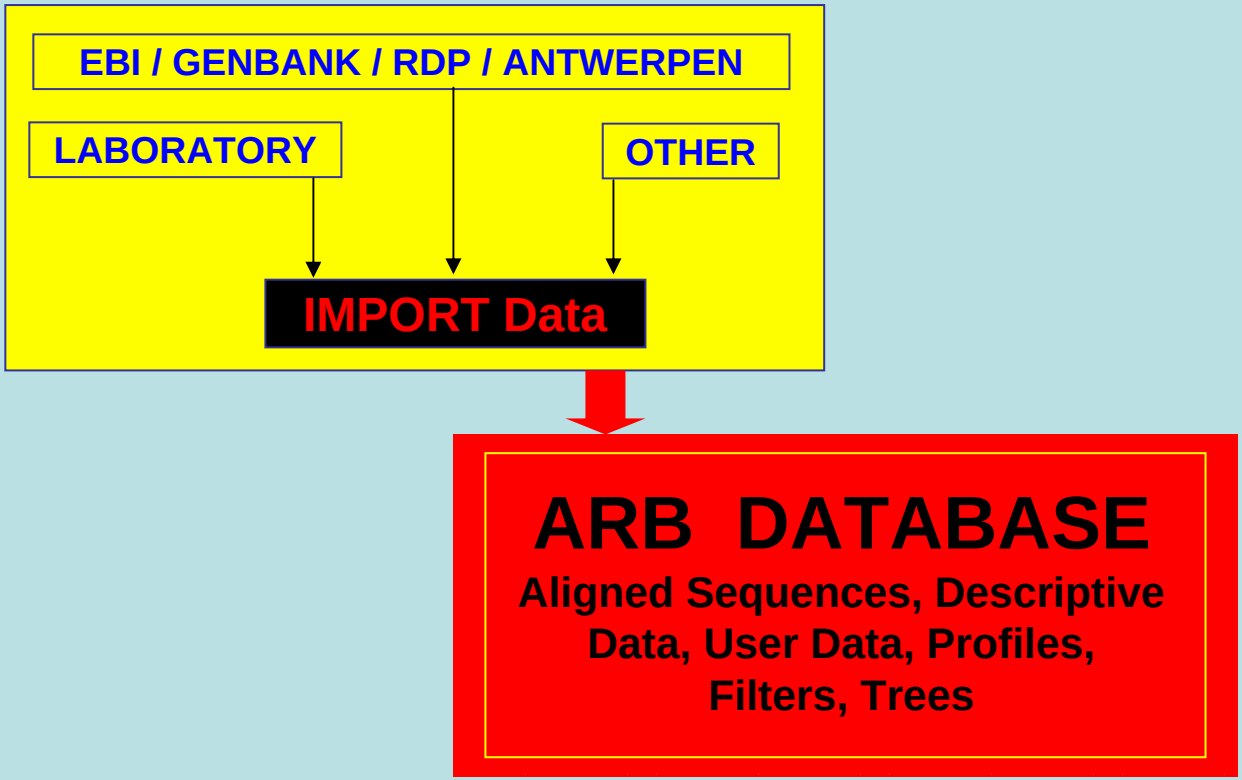
- Central Database to maintain a structured integrative secondary data in combination with processed primary structures (aligned sequences) and any additional data assigned to the individual sequences.

- Comprehensive selection of software tools directly interacting with one another and as well as with the central database facilitating in depth analysis of molecular data.

- Common Graphical User Interface

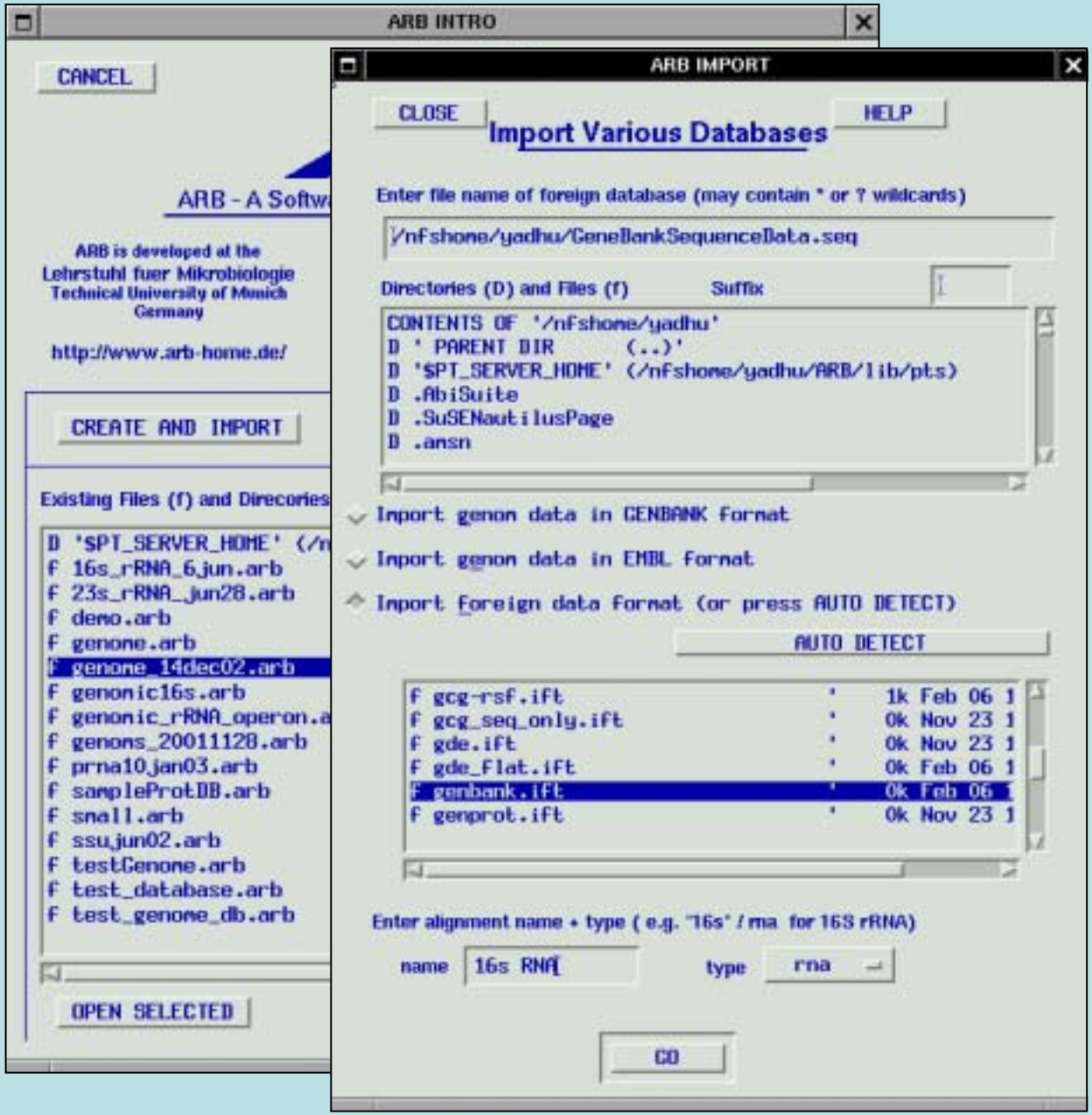
# ARB DATABASE

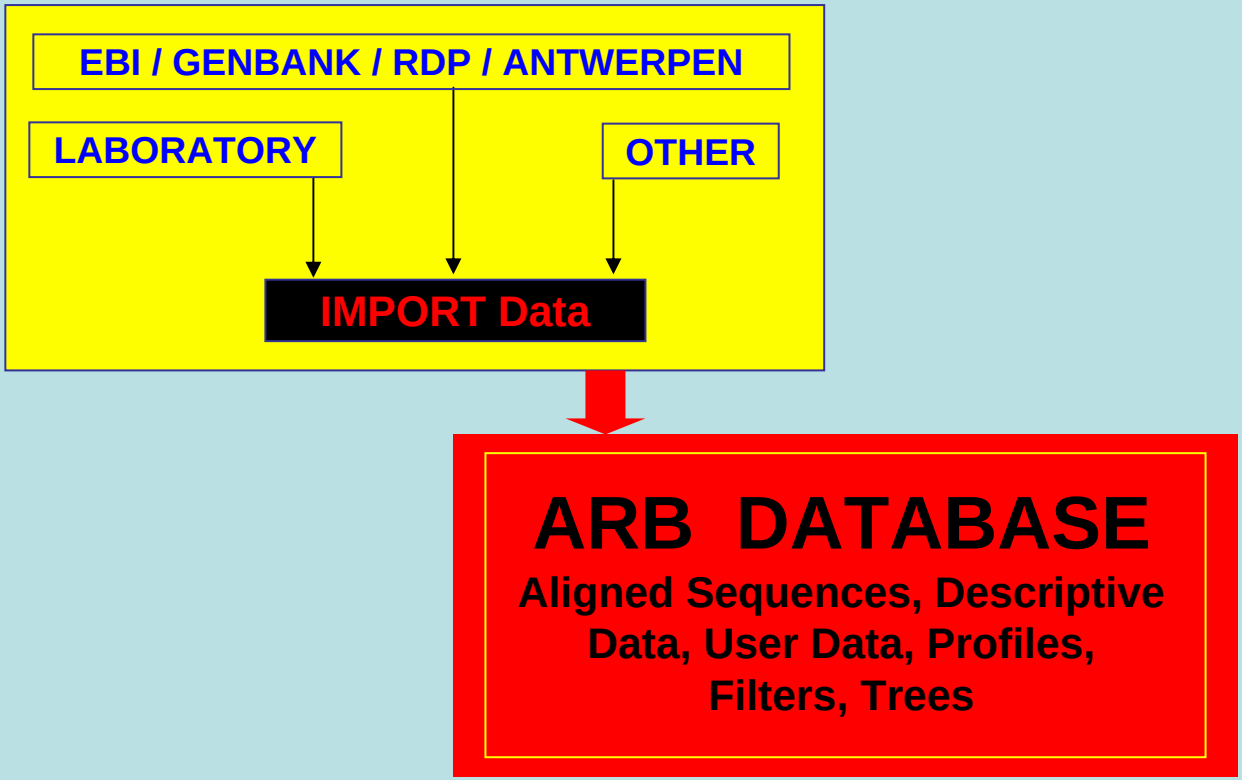
**Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees**

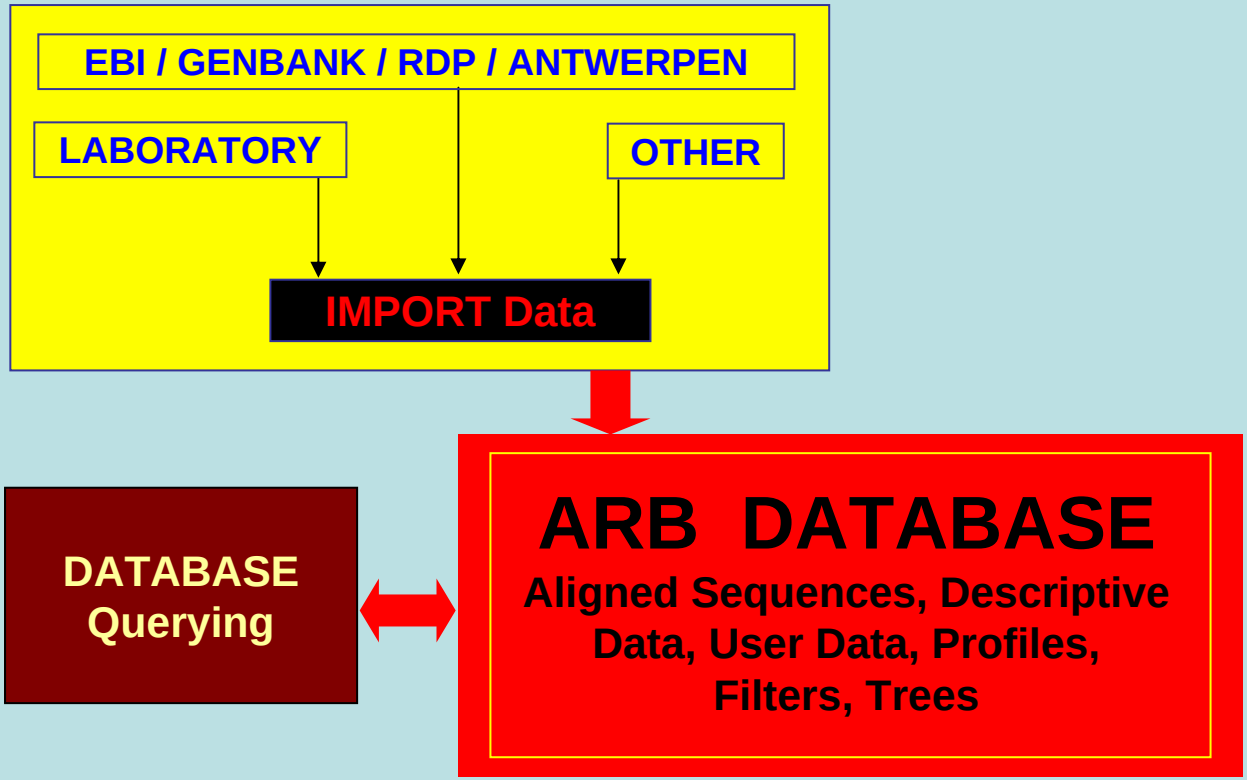

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees

# ARB Main Window & Import Window

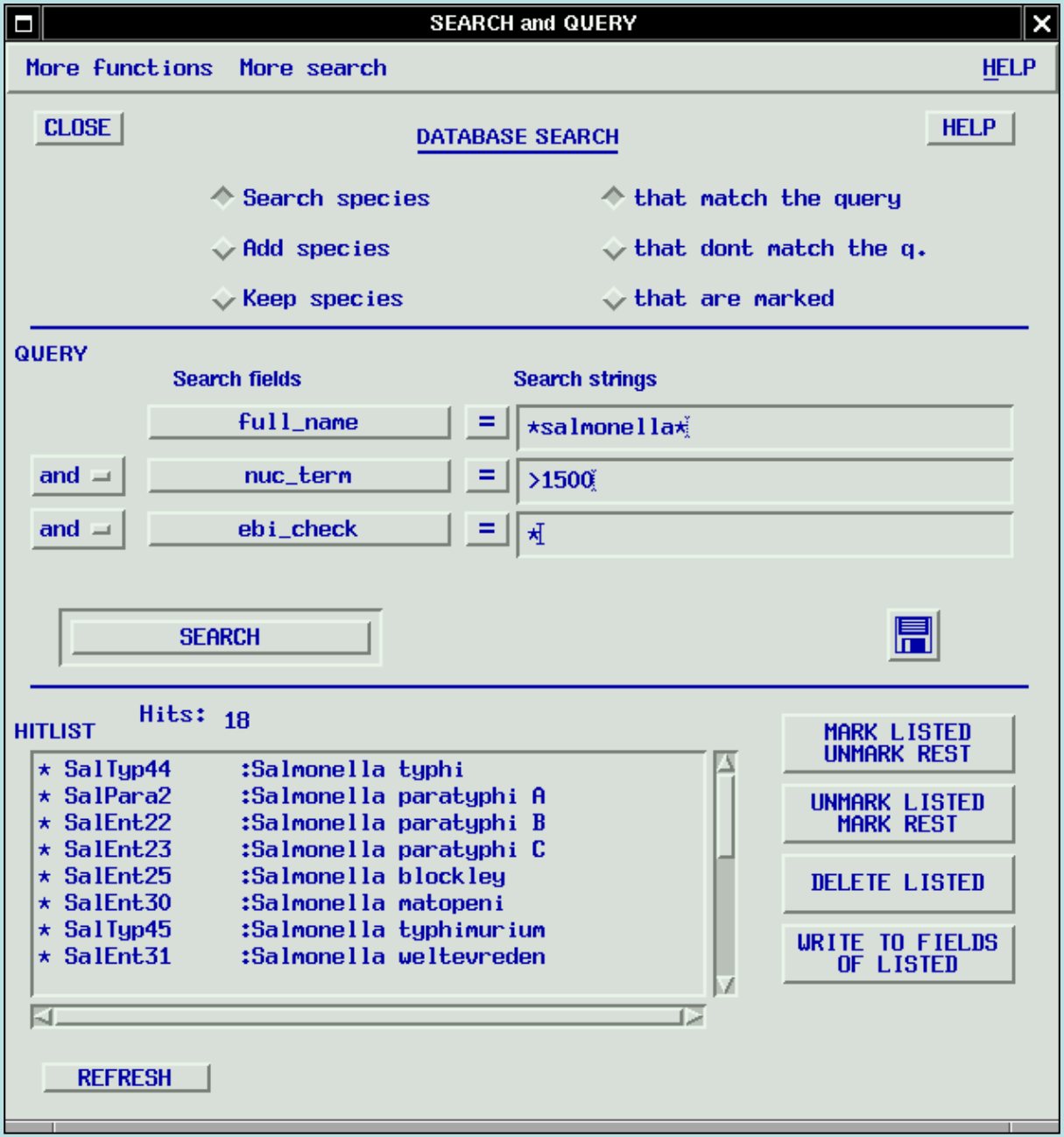
# ARB Main Window & Import Window



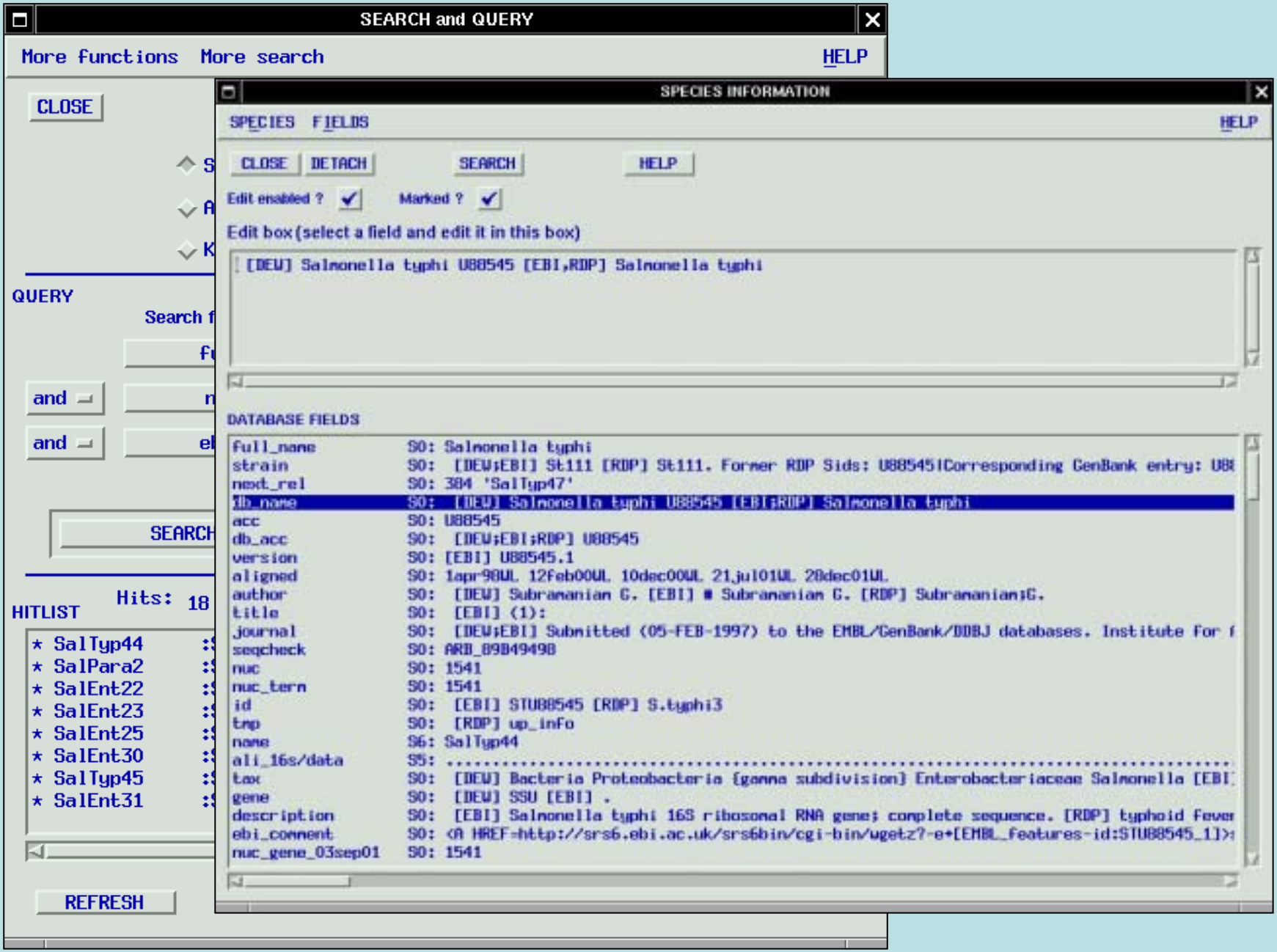

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying

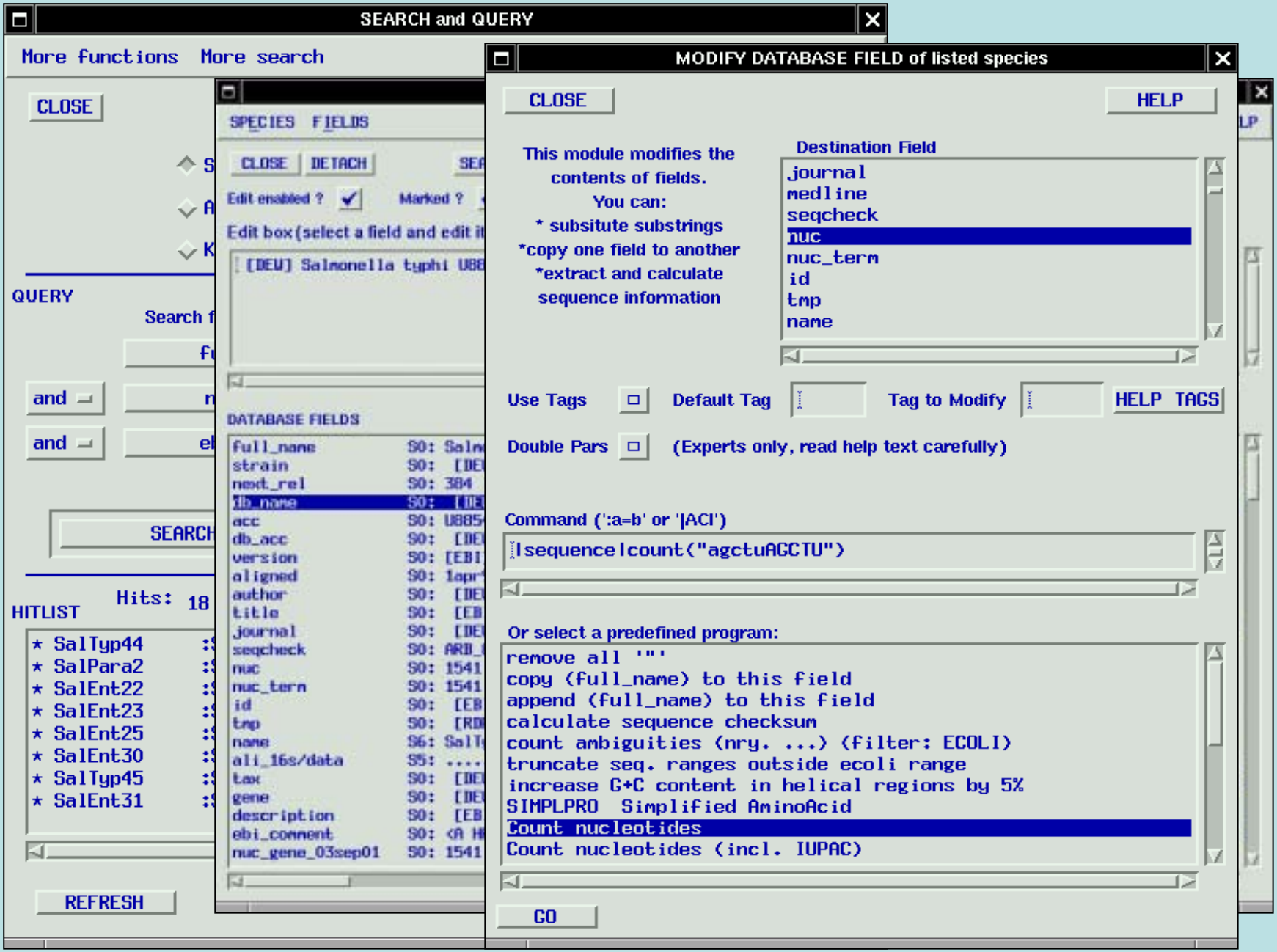
# Search, Query & Modify ARB Database

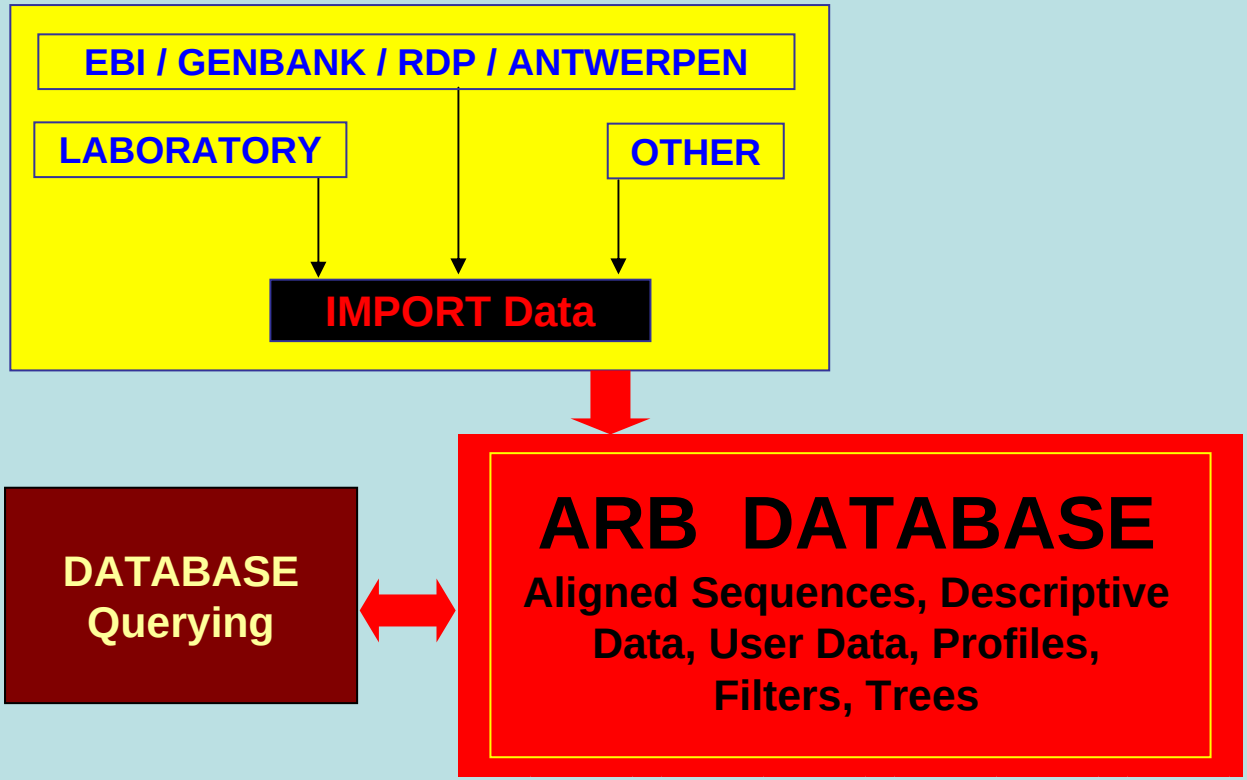


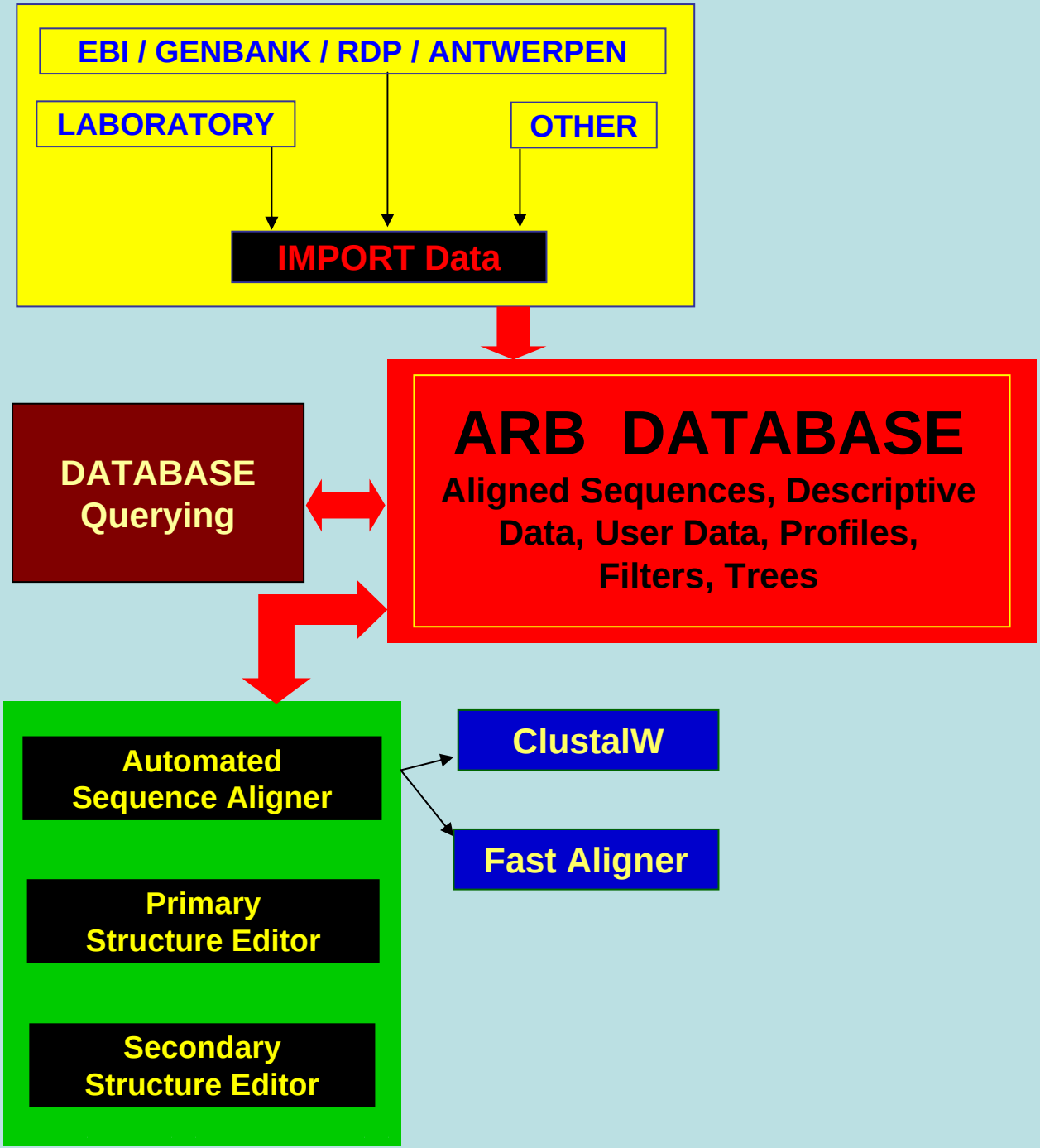
# Search, Query & Modify ARB Database



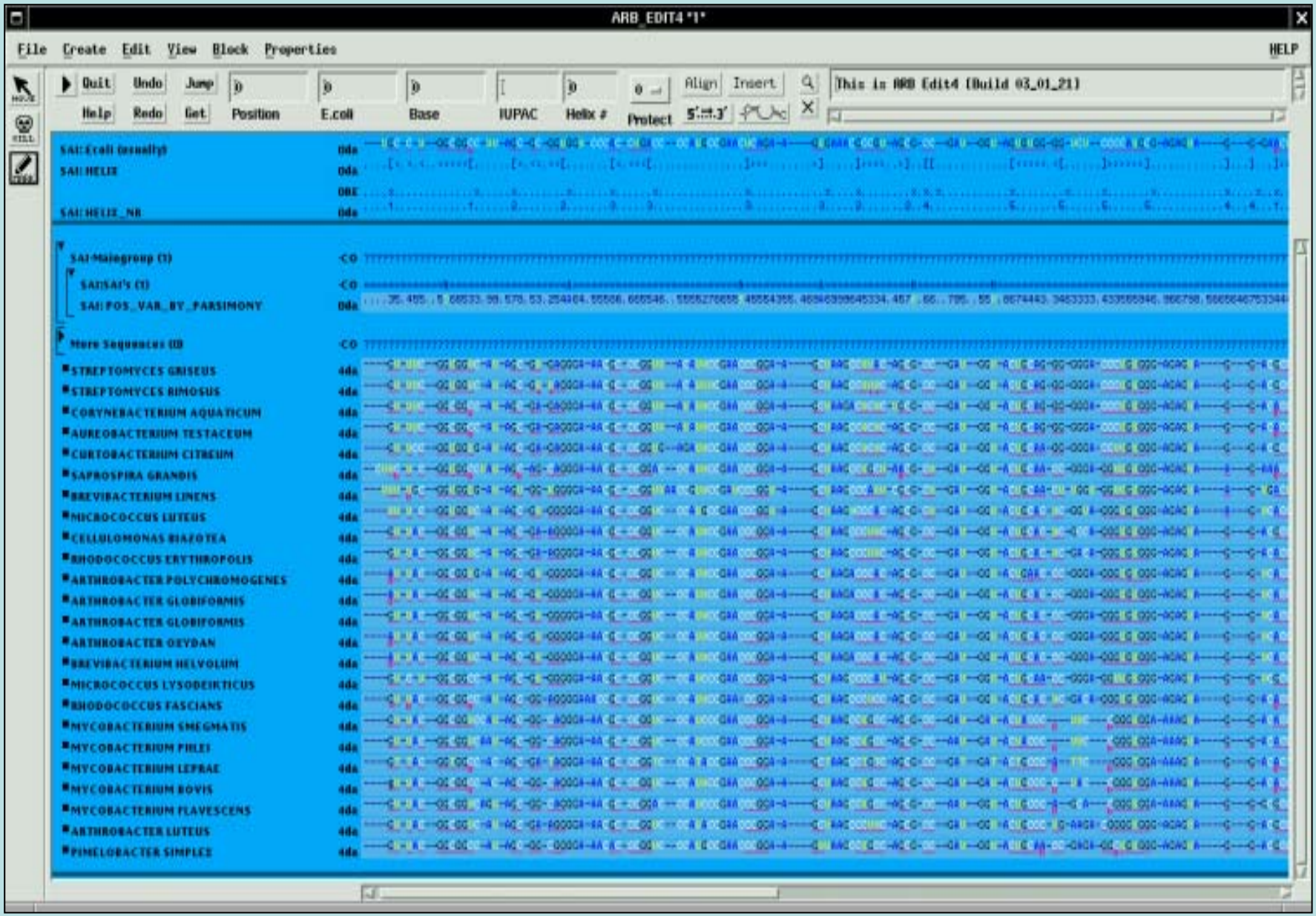
# Search, Query & Modify ARB Database



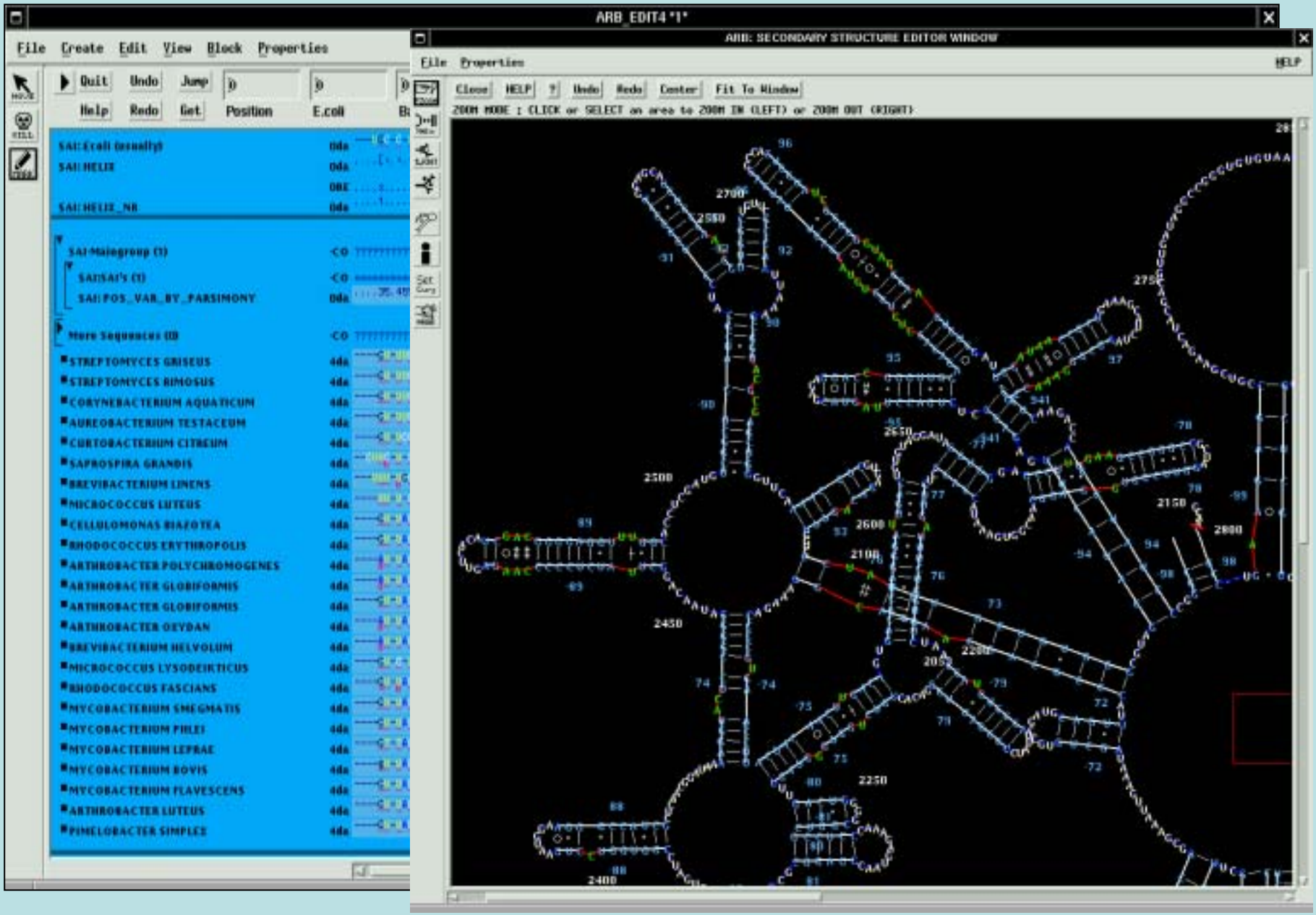

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Automated Sequence Aligner
ClustalW
Fast Aligner
Primary Structure Editor
Secondary Structure Editor

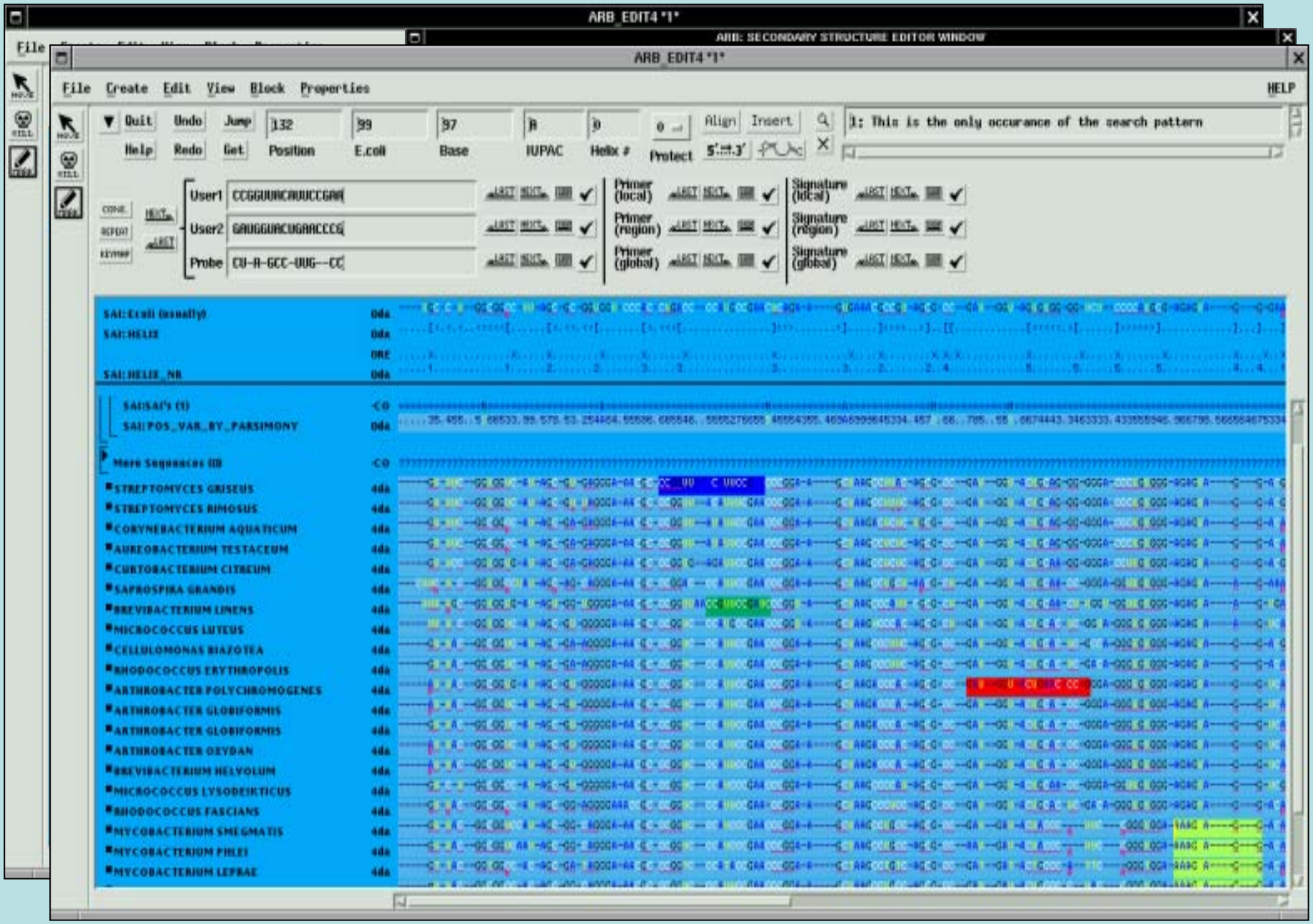
# Primary & Secondary Structure Editors



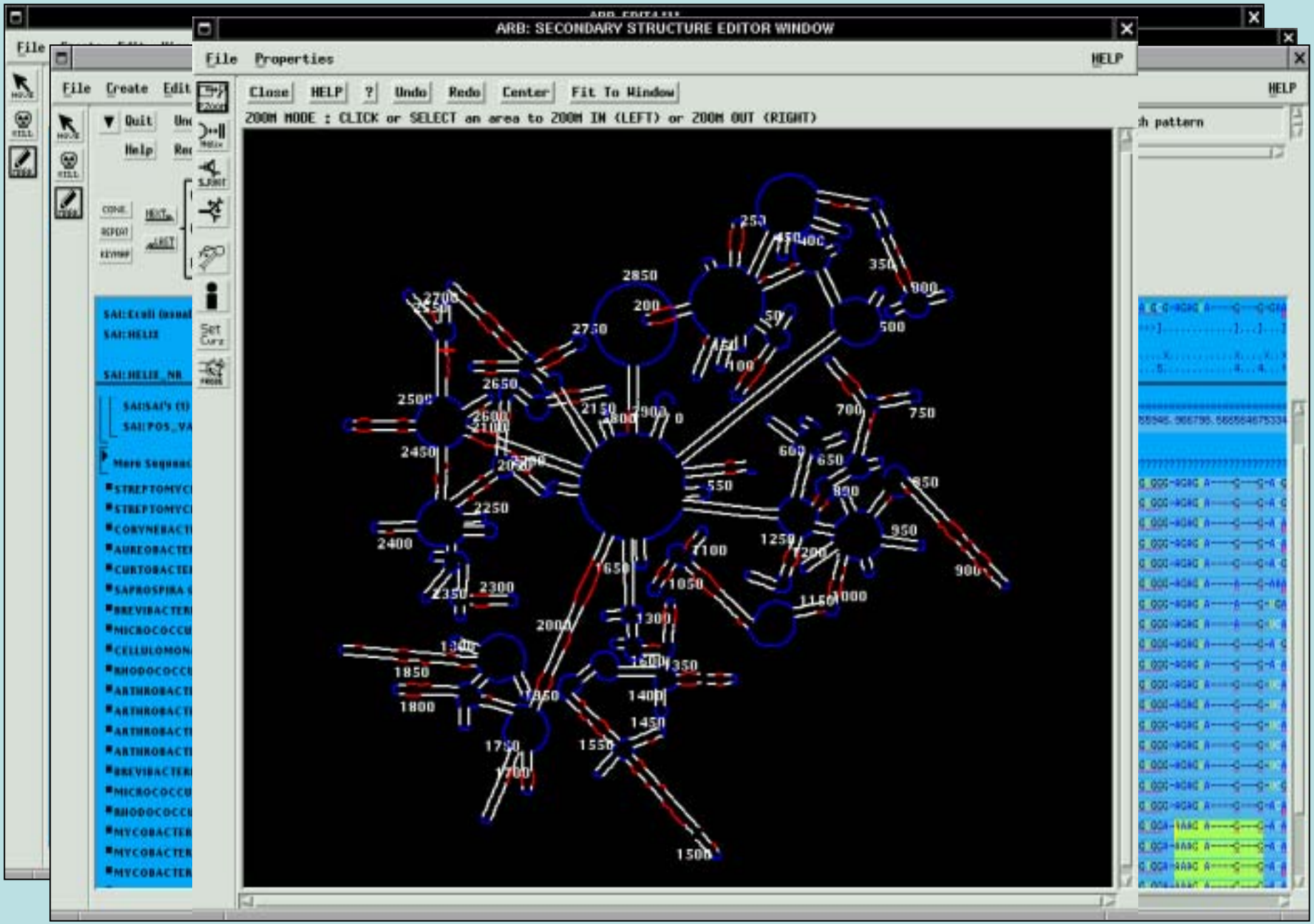
# Primary & Secondary Structure Editors

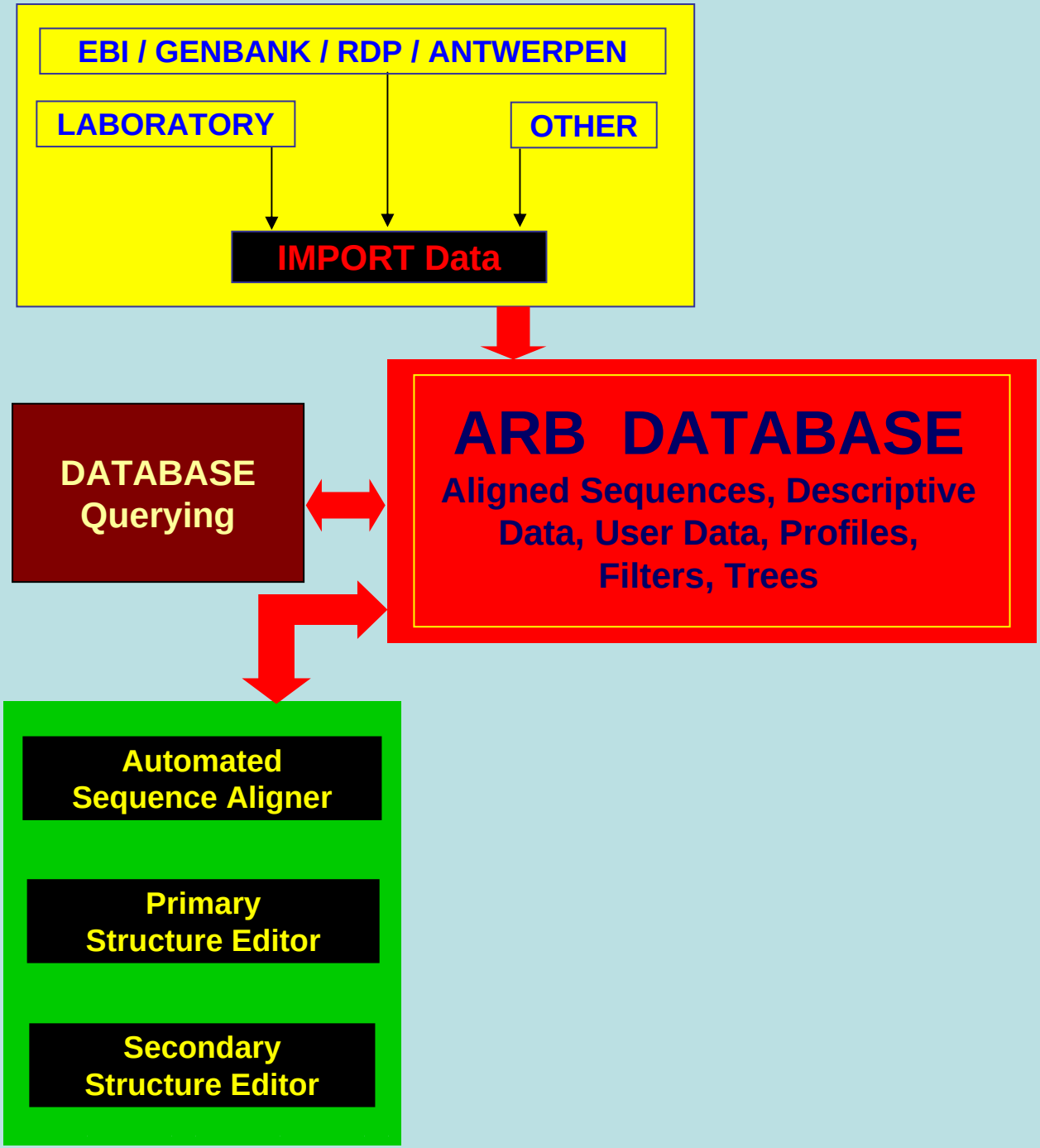


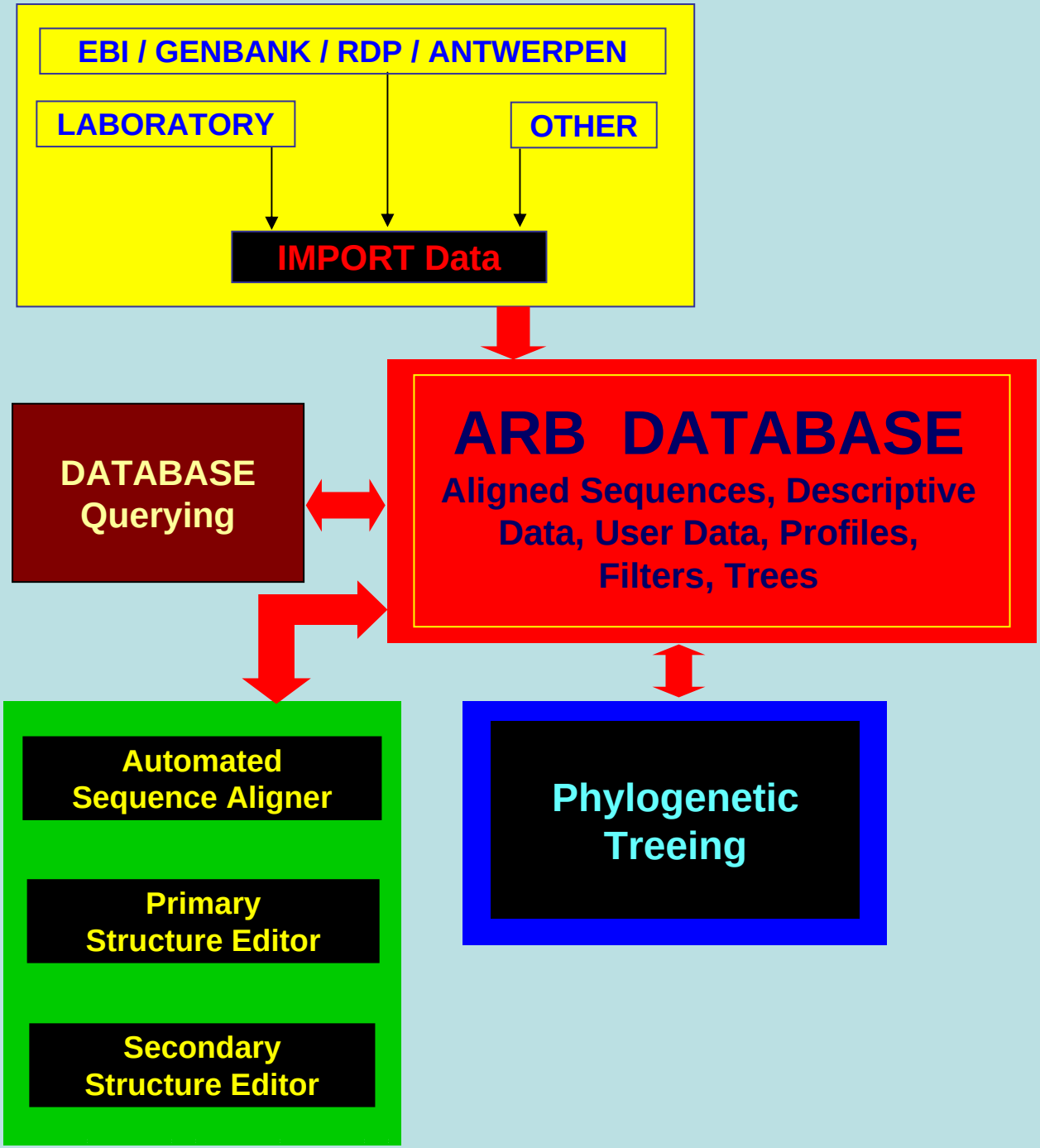
# Primary & Secondary Structure Editors

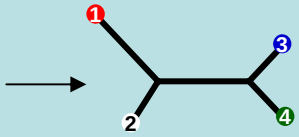


# Primary & Secondary Structure Editors

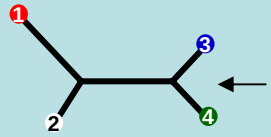


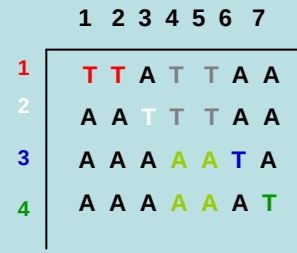

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing




# Phylogenetic Tree Building Methods



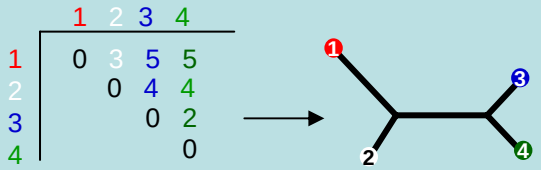

## Phenetic Methods

Distance based

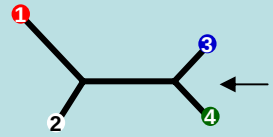
## Cladistic Methods

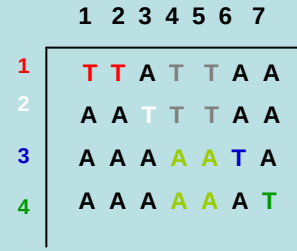
Character Based

- Neighbor Joining Method
- Minimum Evolution Method
- Maximum Likelihood Method
- Maximum Parsimony Method




# Phylogenetic Tree Building Methods




## Phenetic Methods

Distance based

### Neighbor Joining Method

**Clustering Algorithm**

Builds a small tree and keeps adding the sequences to arrive at a full desired tree

### Minimum Evolution Method

**Optimality Criterion**

Selects the tree whose sum of branch lengths is the minimum

## Cladistic Methods

Character Based

### Maximum Likelihood Method

**Optimality Criterion**

Selects the tree that is most likely to have produced the observed data

### Maximum Parsimony Method

**Optimality Criterion**

Selects the tree that require fewer evolutionary changes

# Treeing Methods in ARB

**Neighbor Joining Method**

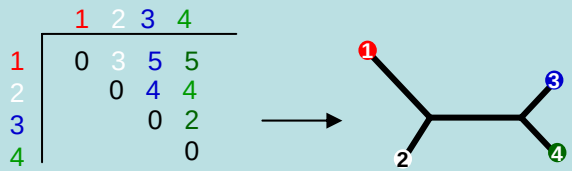
**FastDNAml Method**

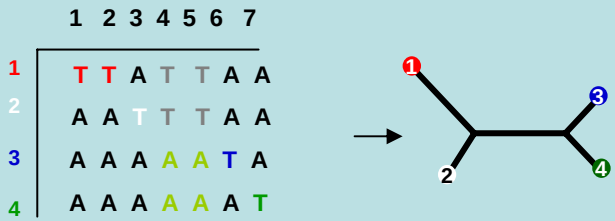
Distance Based

Character Based

**Phylip Distance Matrix Method**

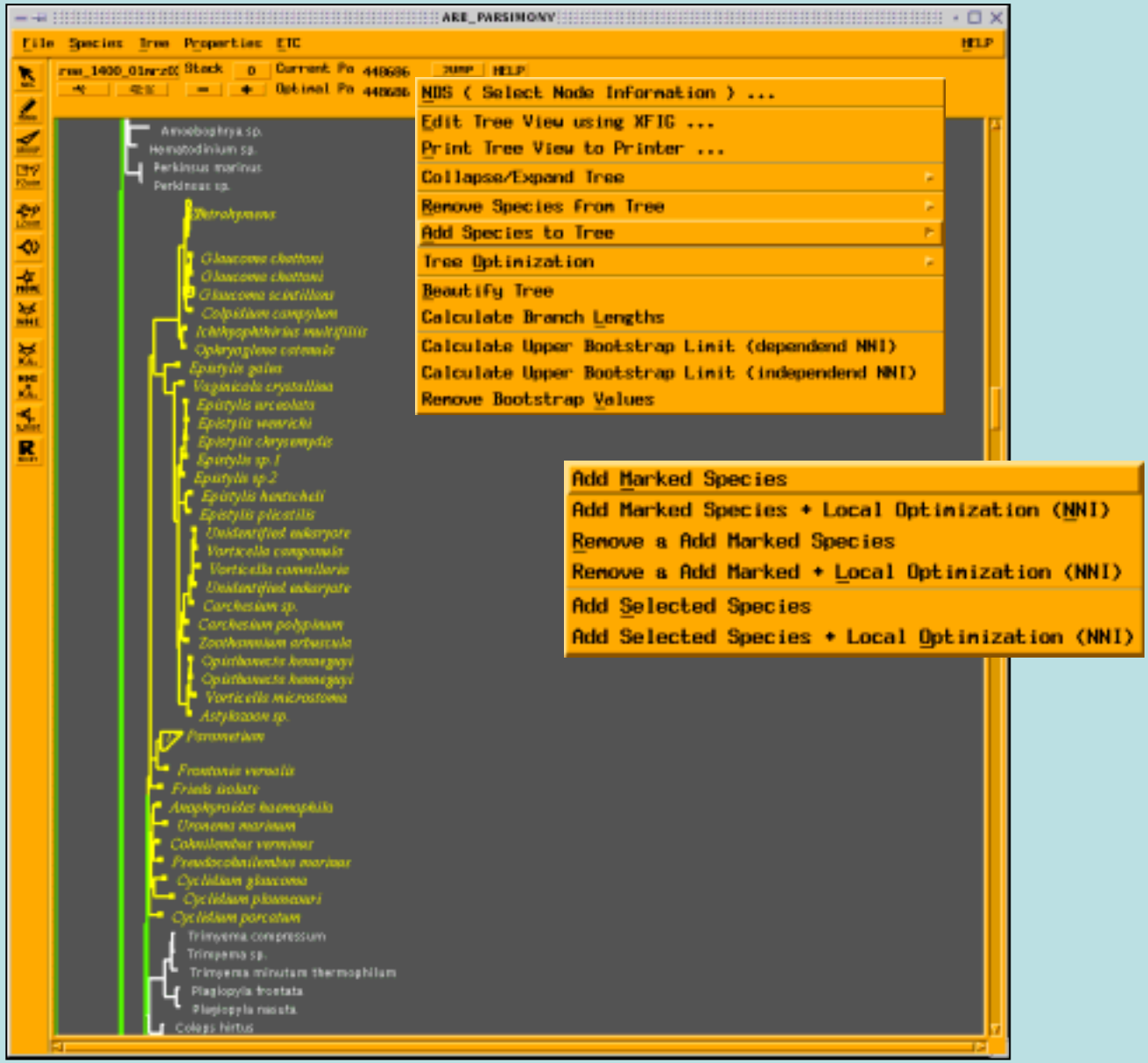
**Parsimony Method**



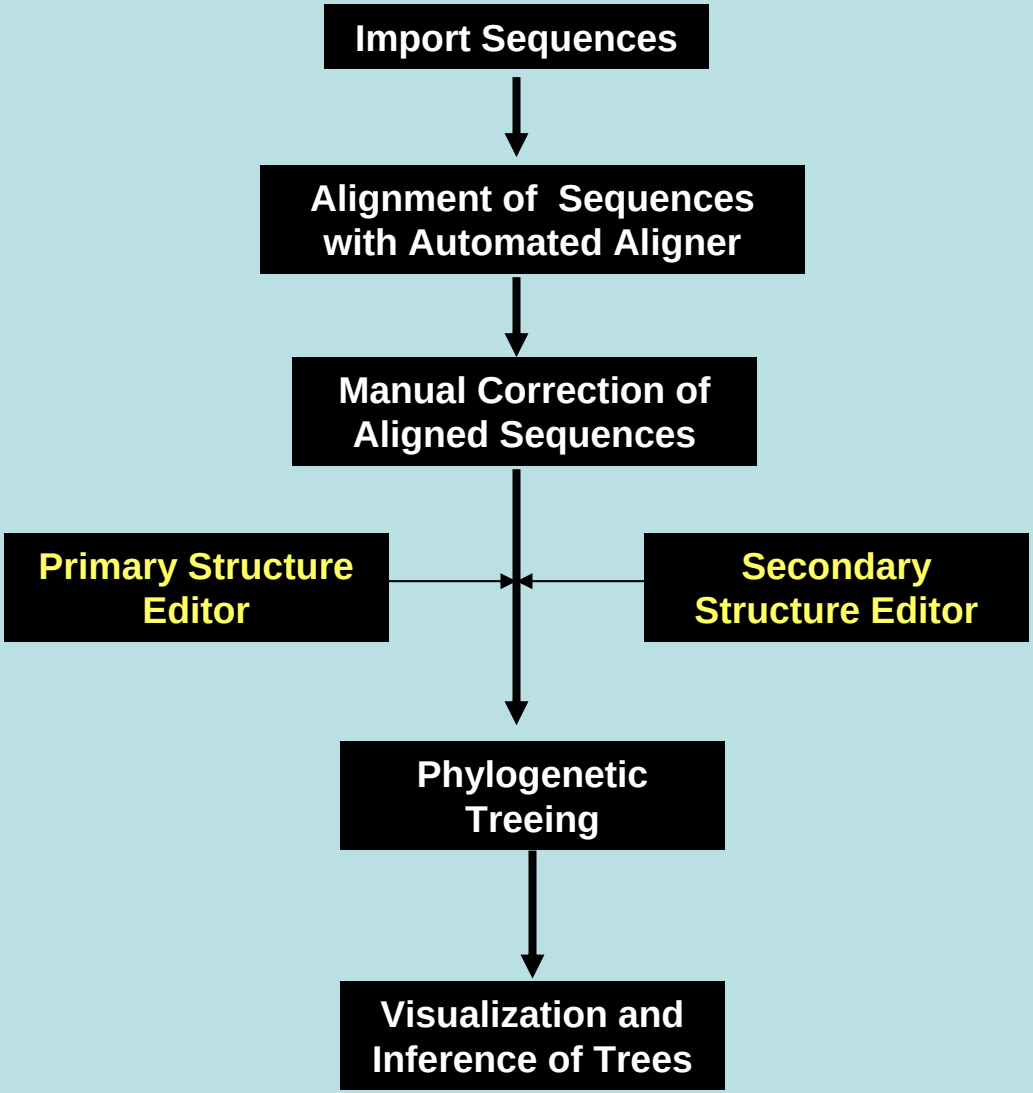



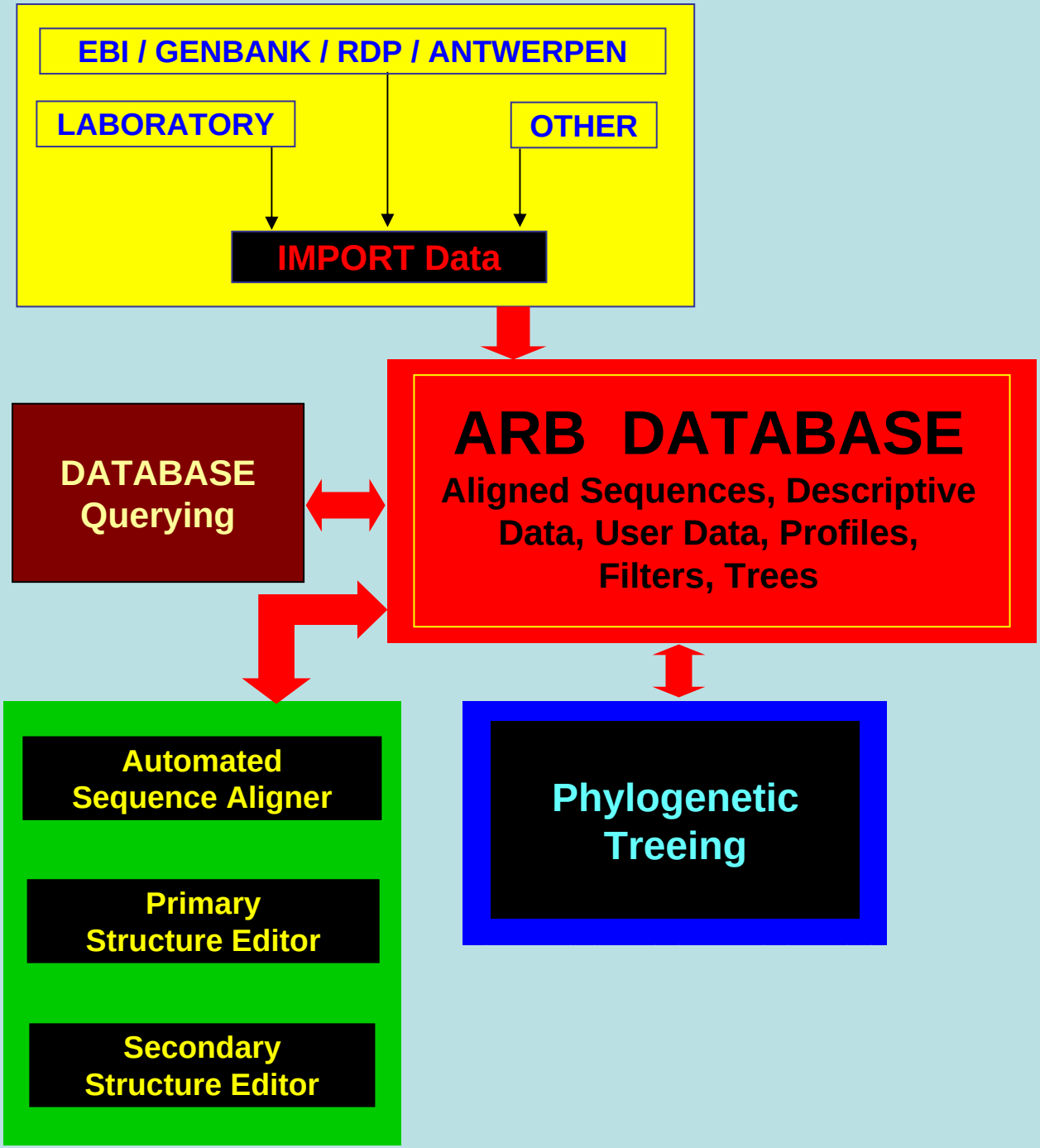
# The ARB Parsimony Tool

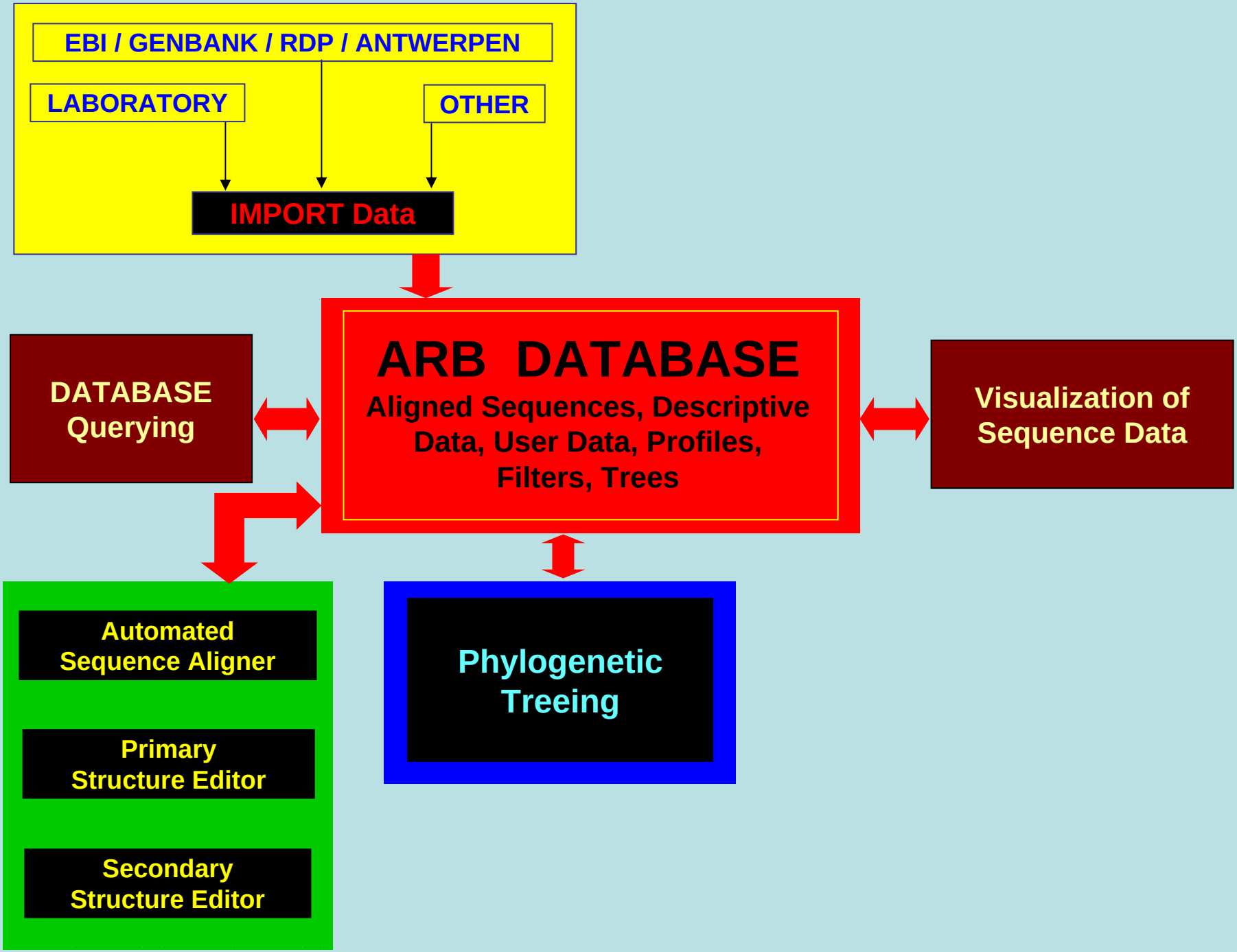
- Able to handle big trees (e.g. >30.000 16S/18S rRNA sequences)
- Allows optimization of trees and sub-trees with different parameters.
- Adding sequences is possible without changing initial topology.



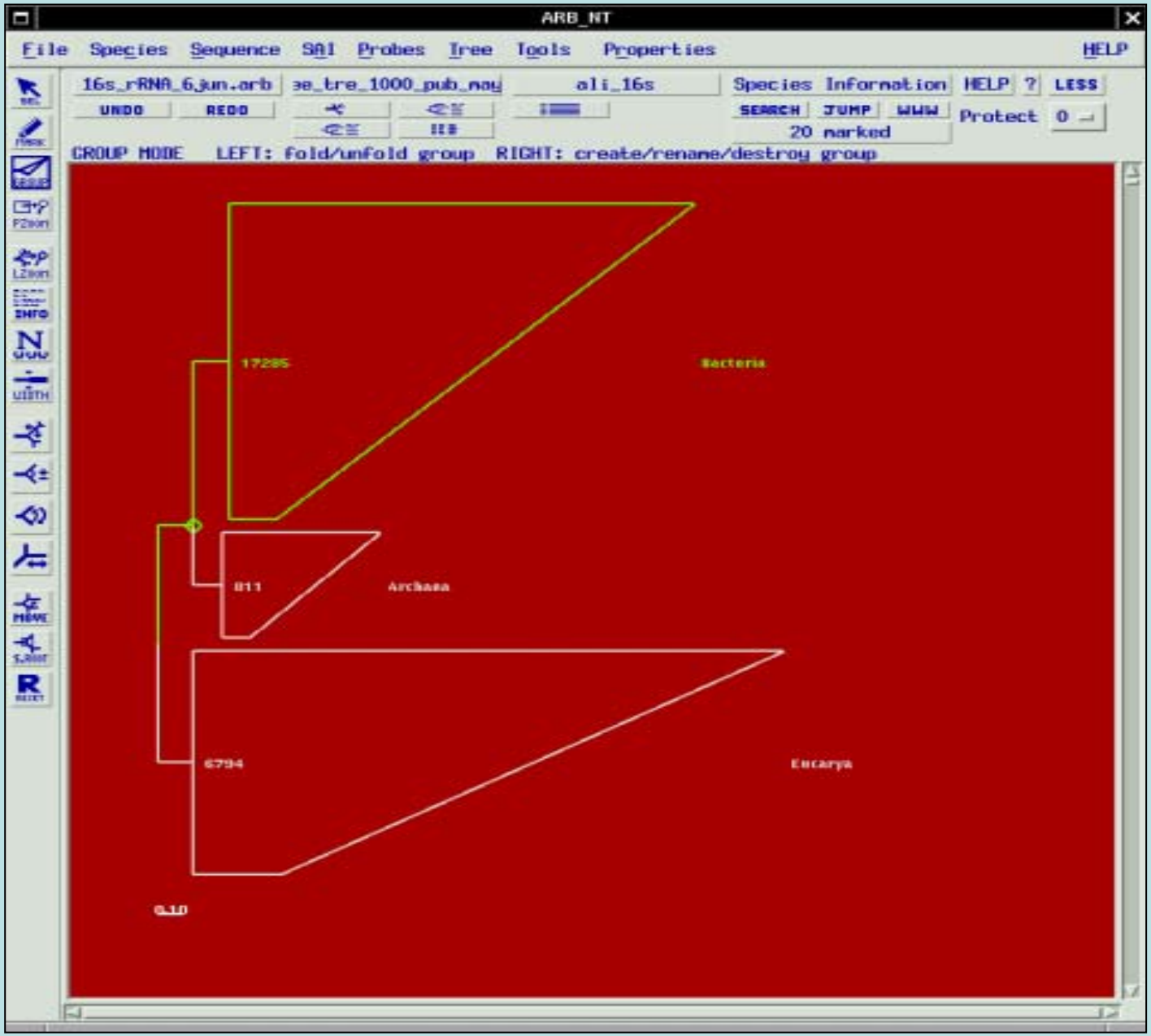
# Phylogenetic Treeing using ARB Software



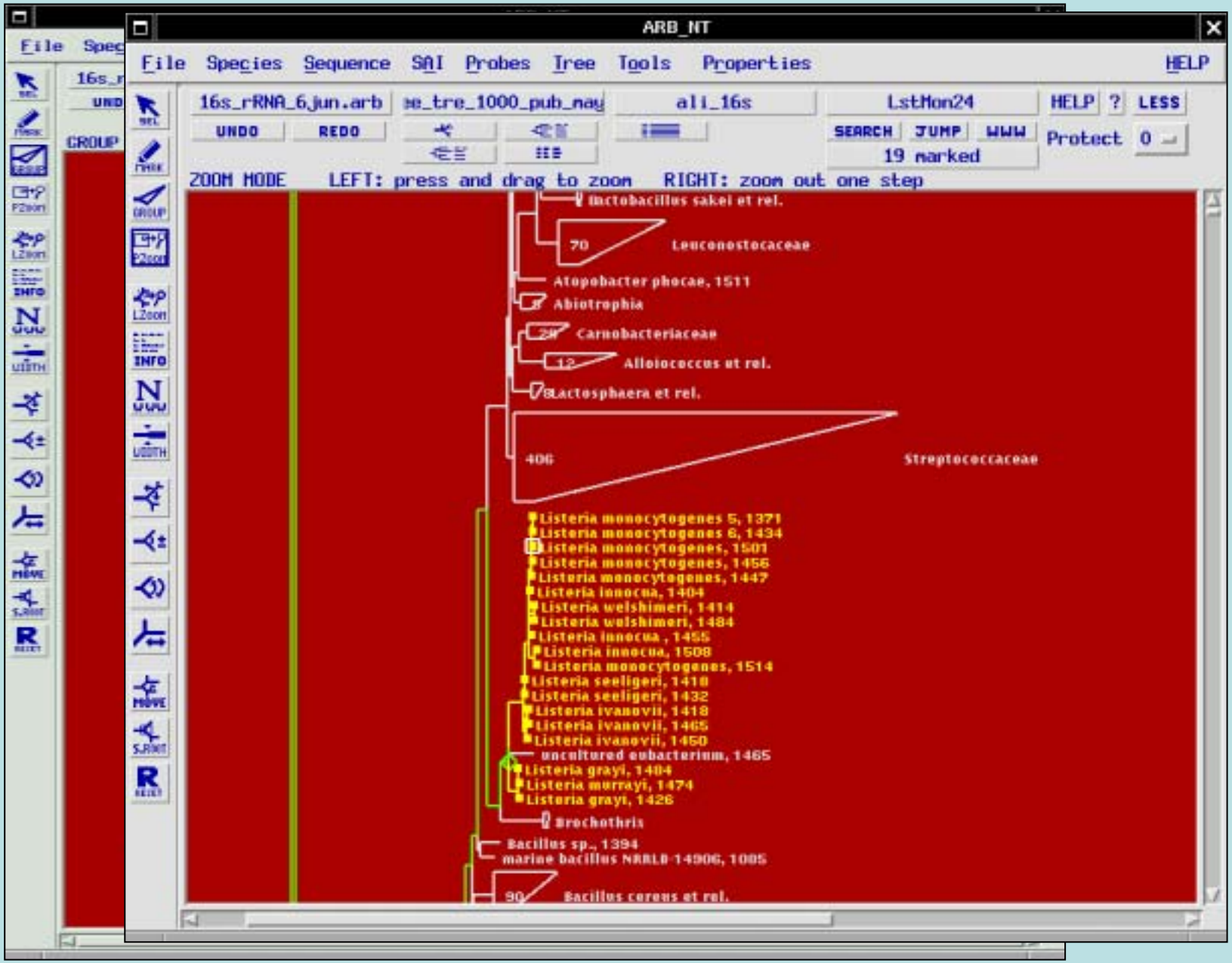

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Visualization of Sequence Data
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing

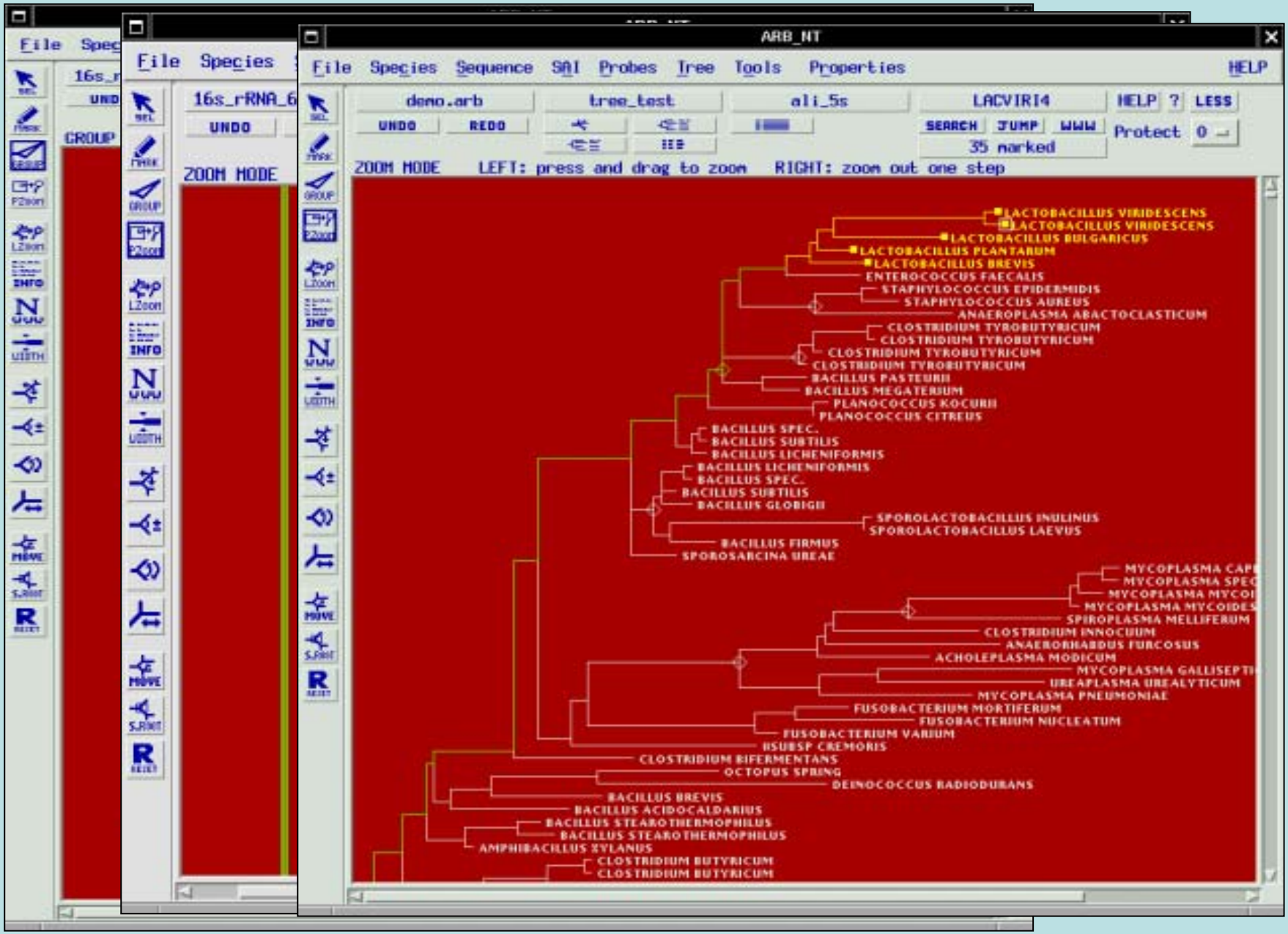
# Visualization of Sequence Data



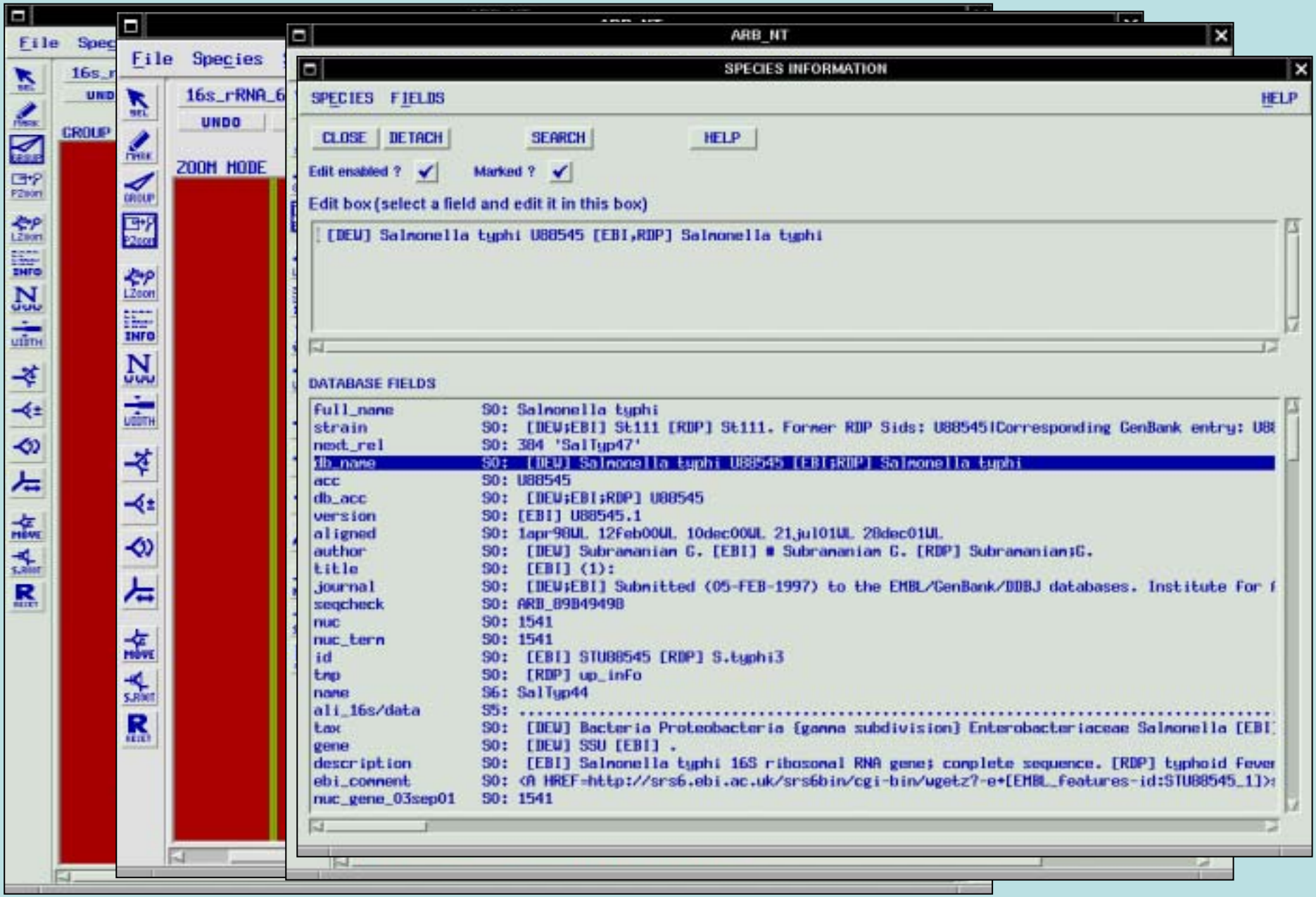
# Visualization of Sequence Data

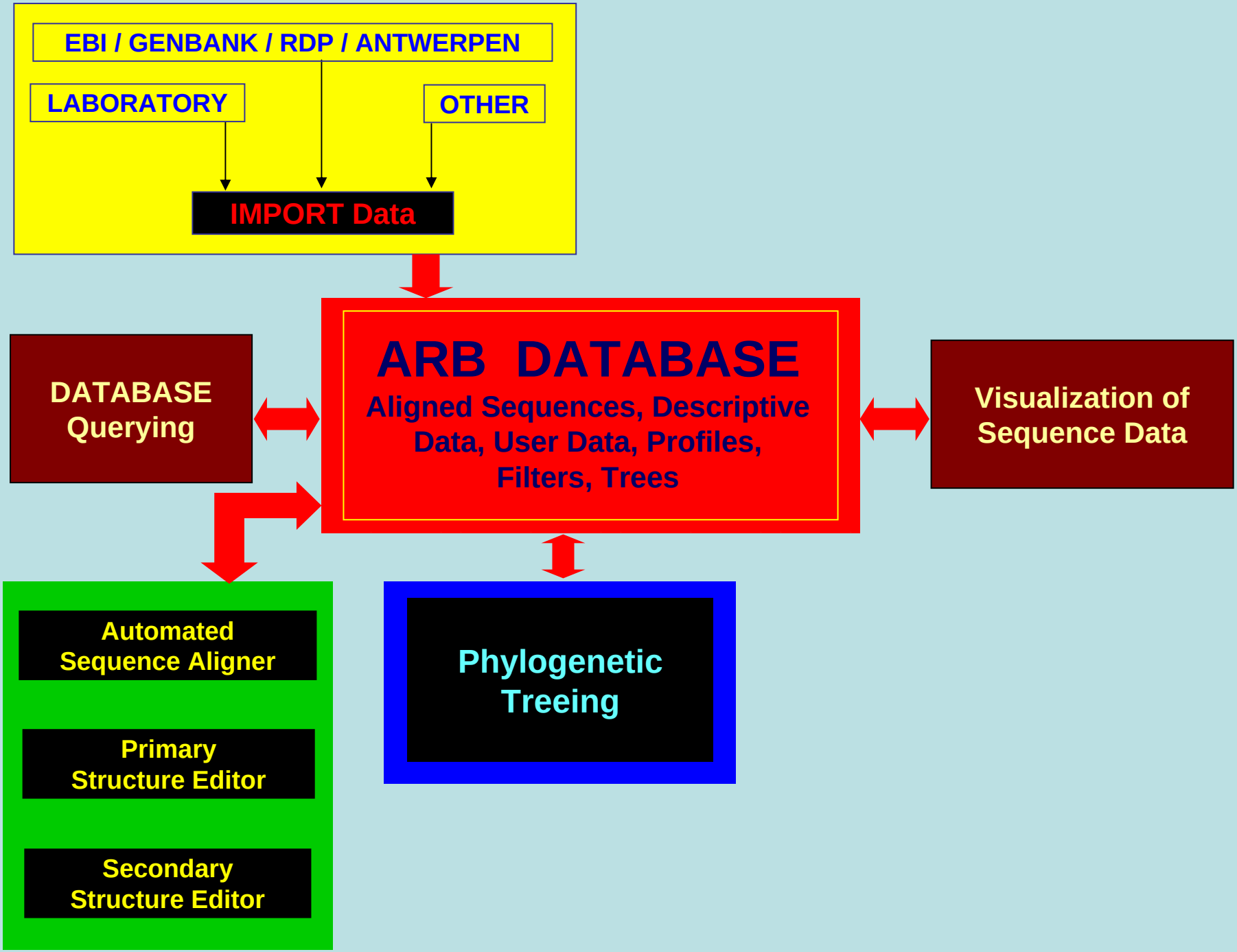


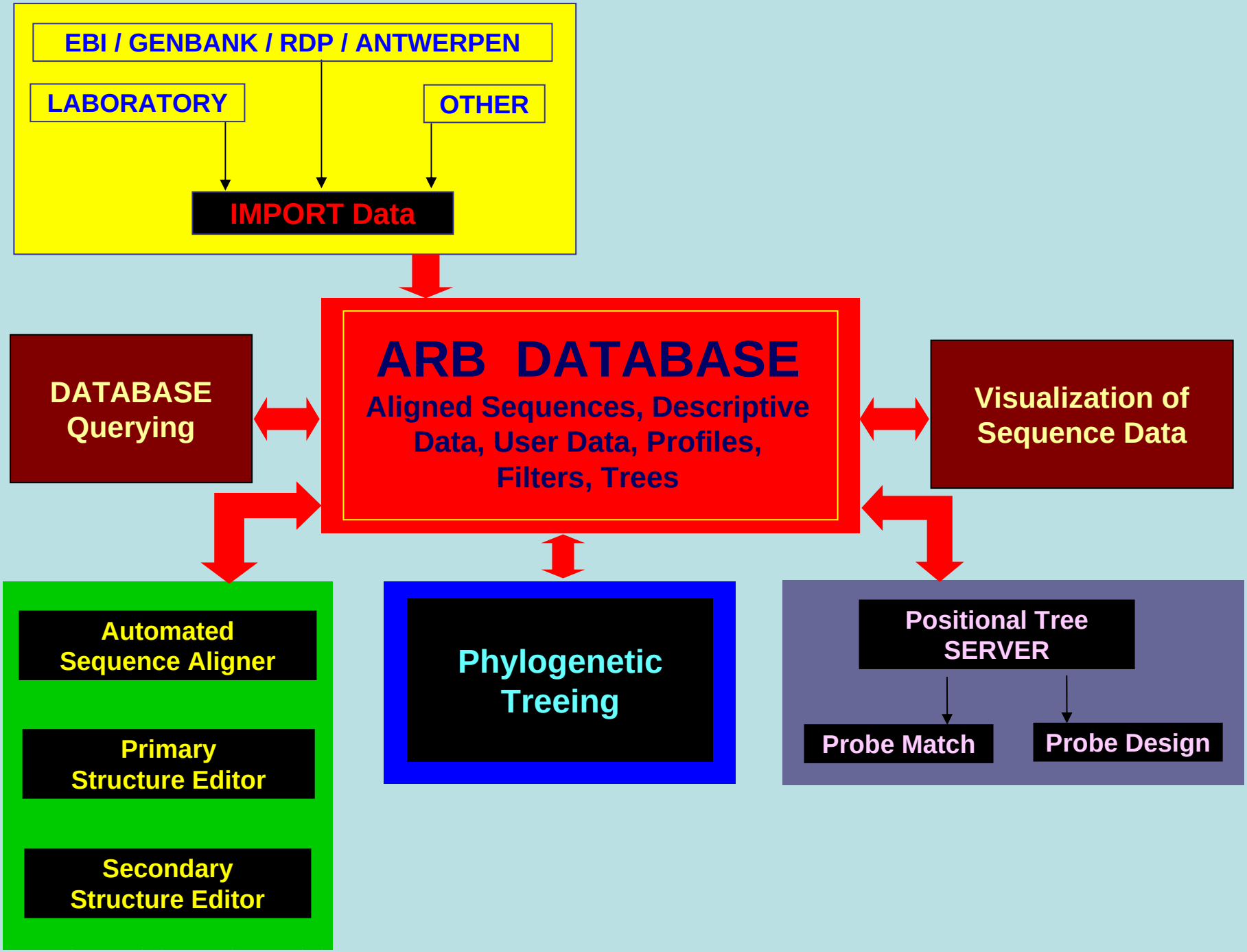
# Visualization of Sequence Data



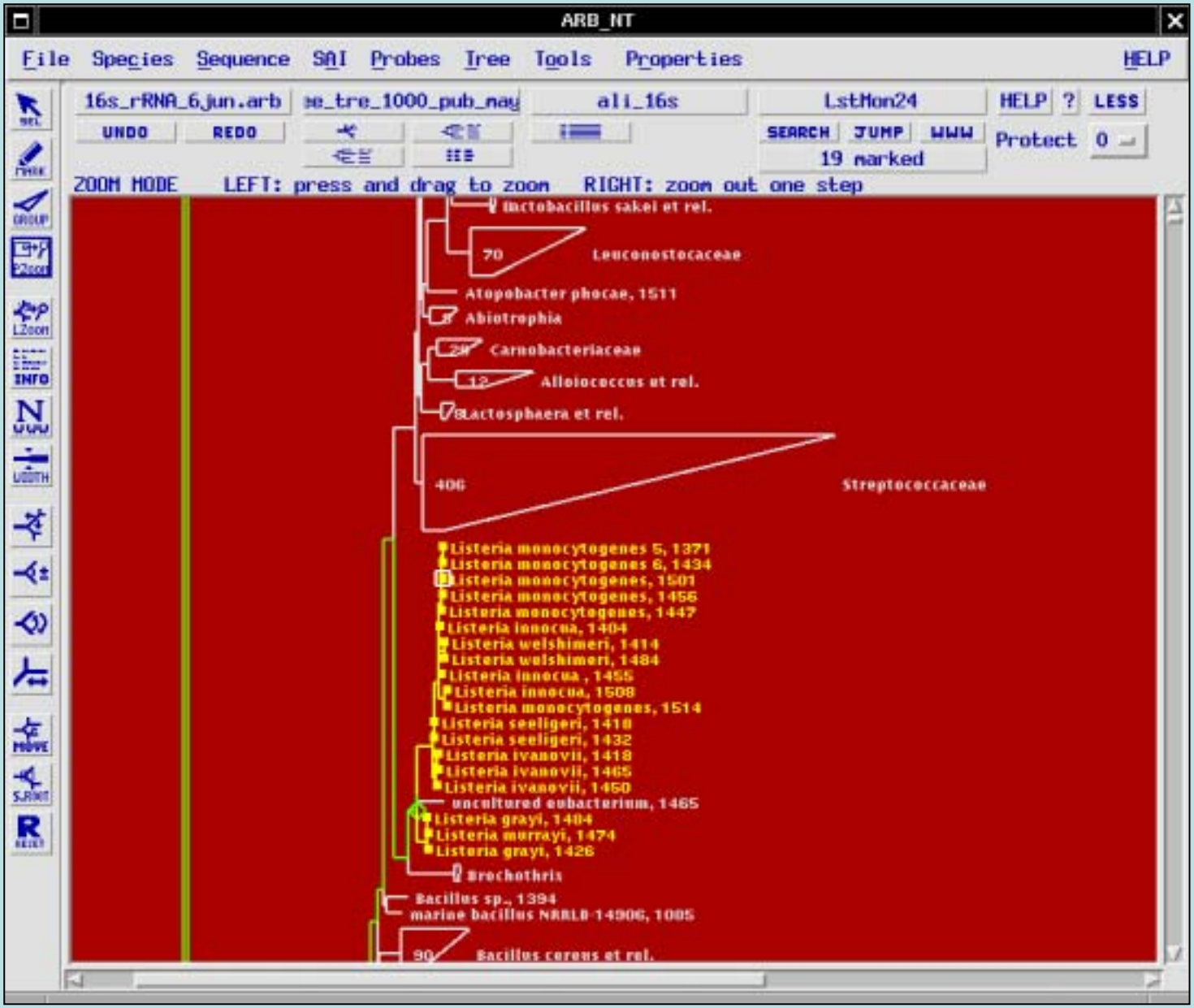
# Visualization of Sequence Data



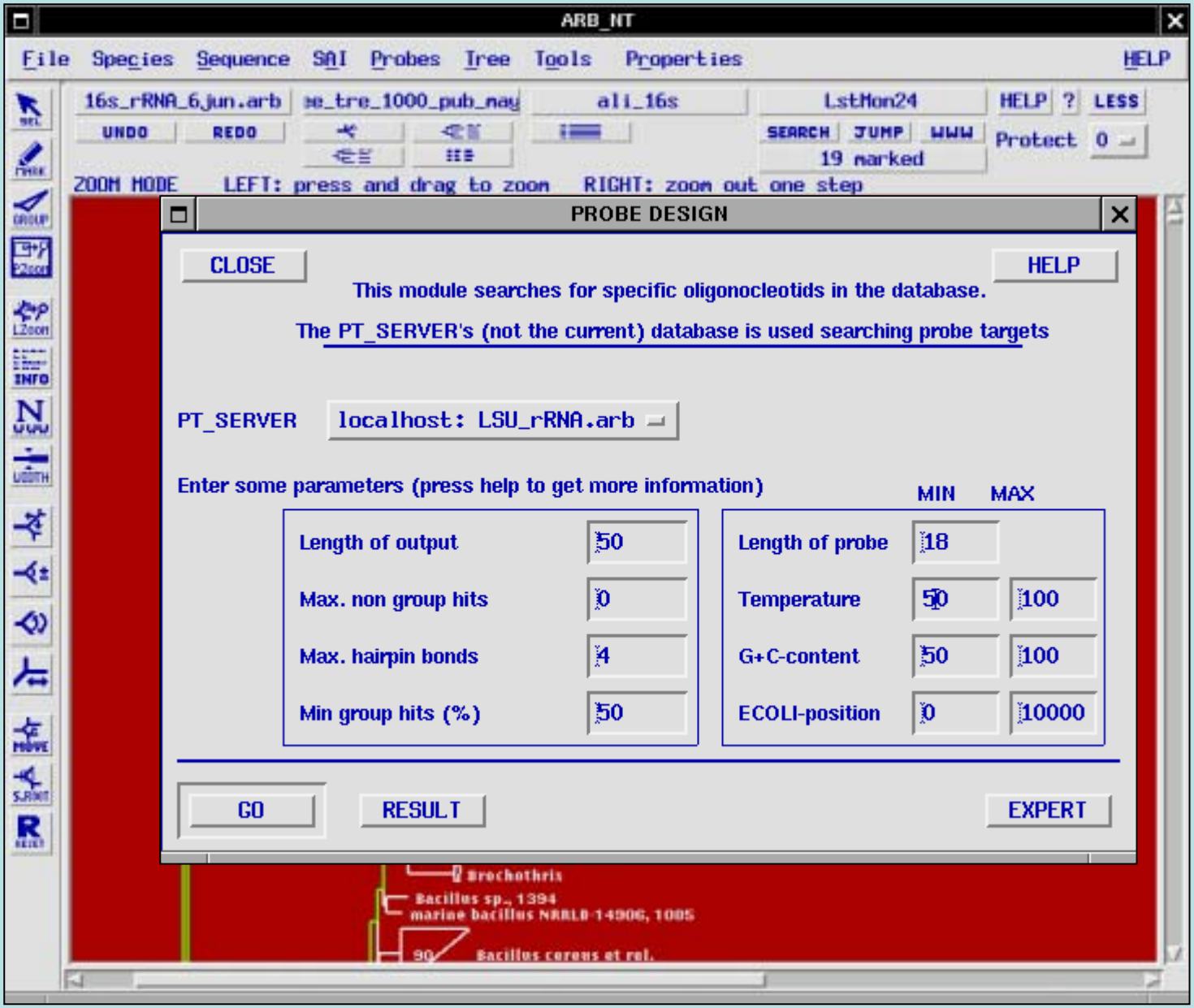

EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Visualization of Sequence Data
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Visualization of Sequence Data
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing
Positional Tree SERVER
Probe Match
Probe Design

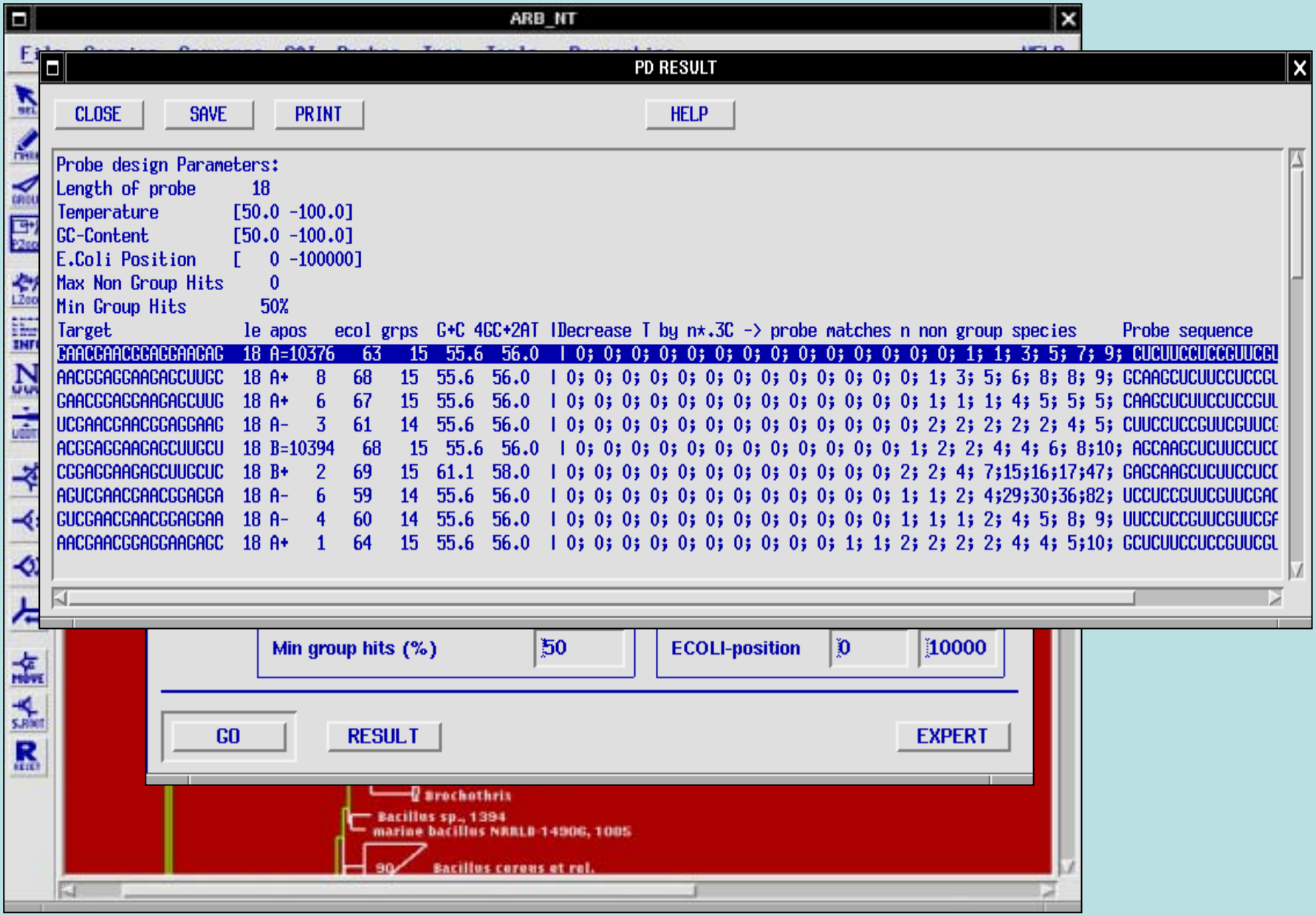
# Probe Design and Probe Match



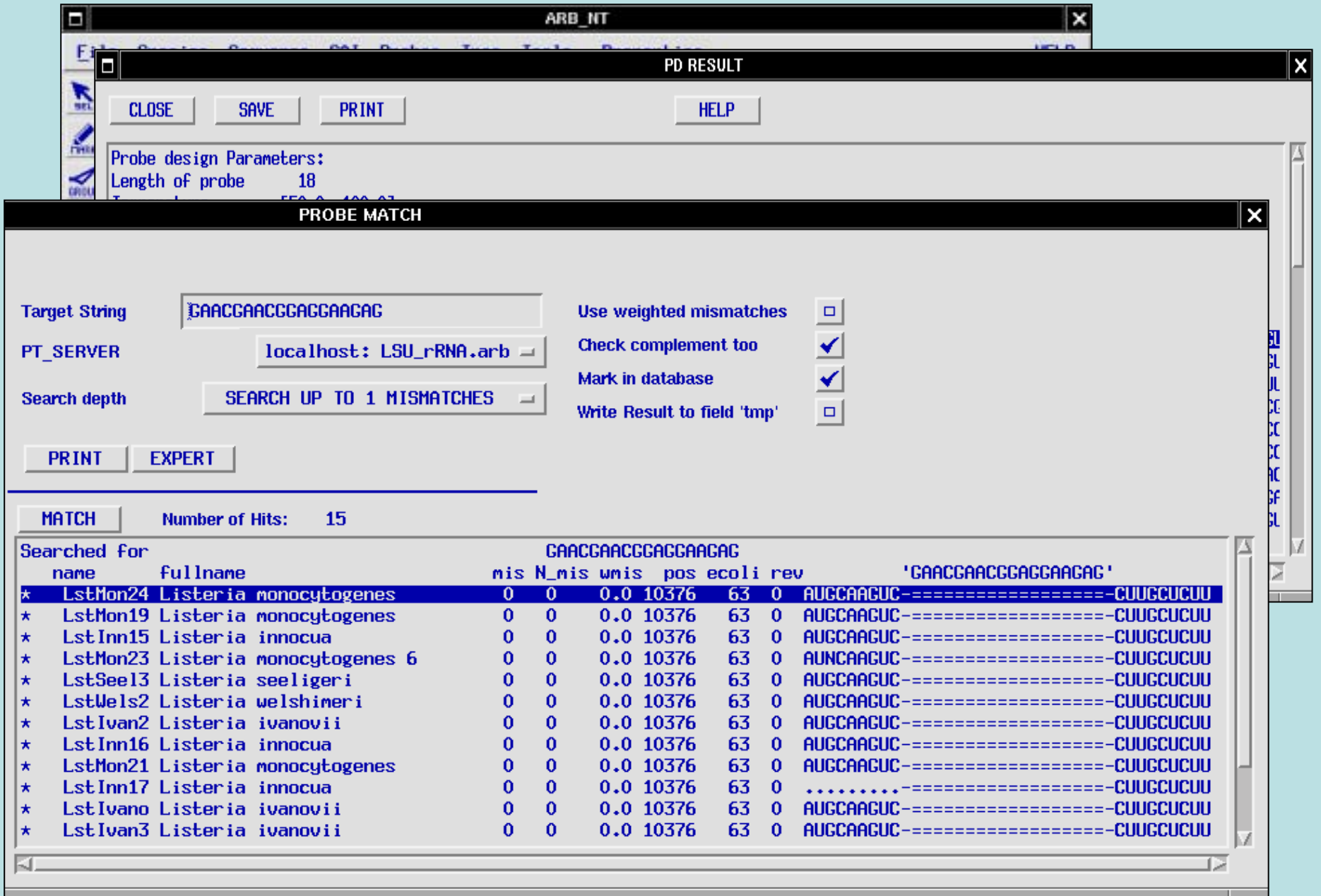
# Probe Design and Probe Match



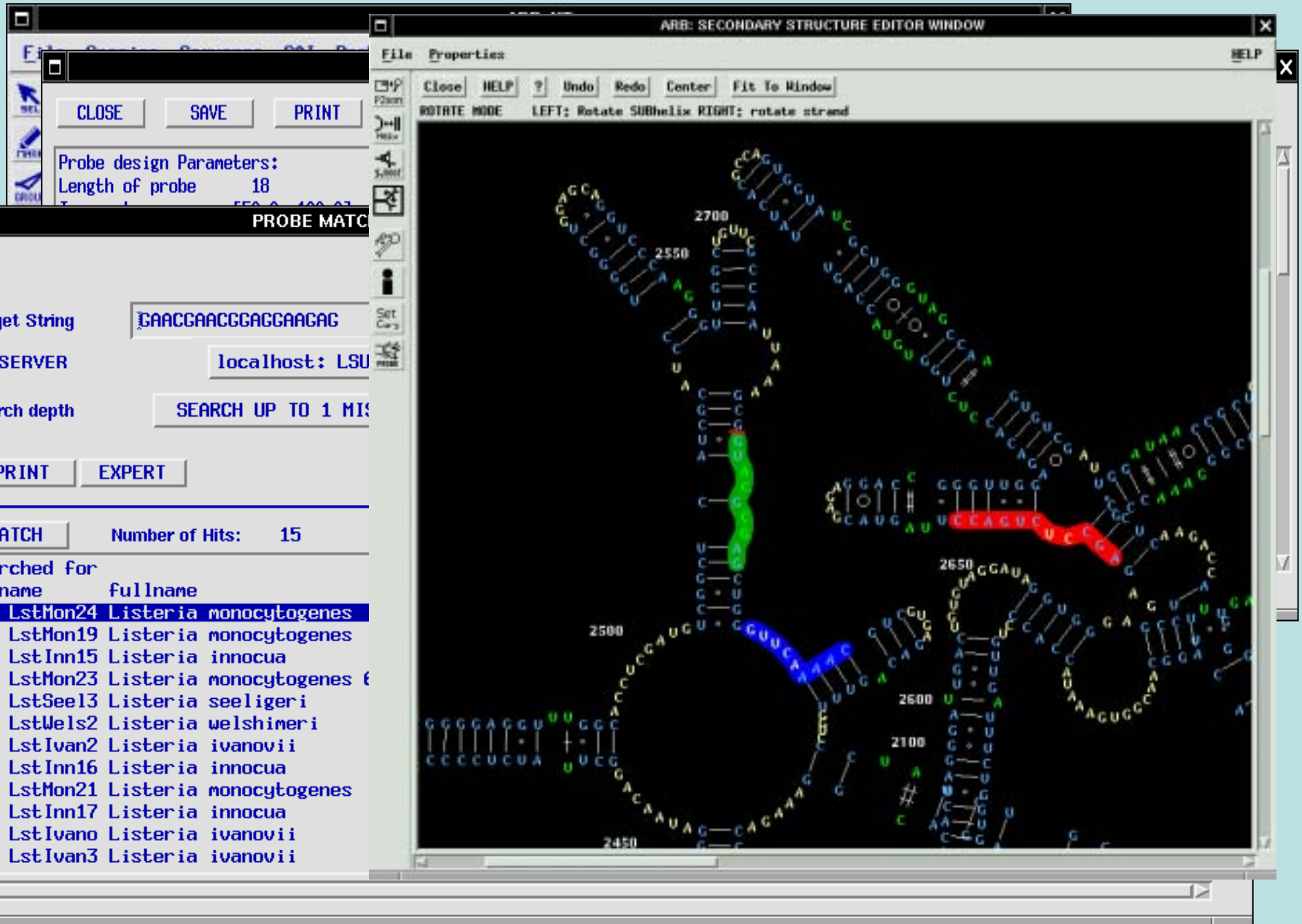
# Probe Design and Probe Match



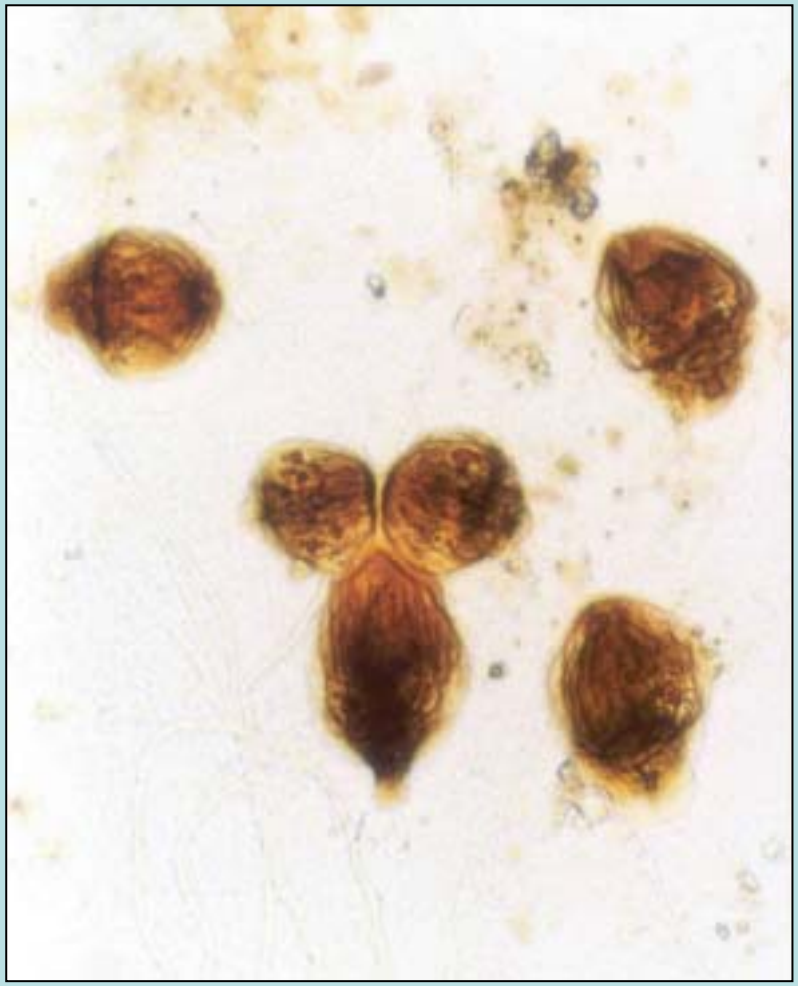
# Probe Design and Probe Match



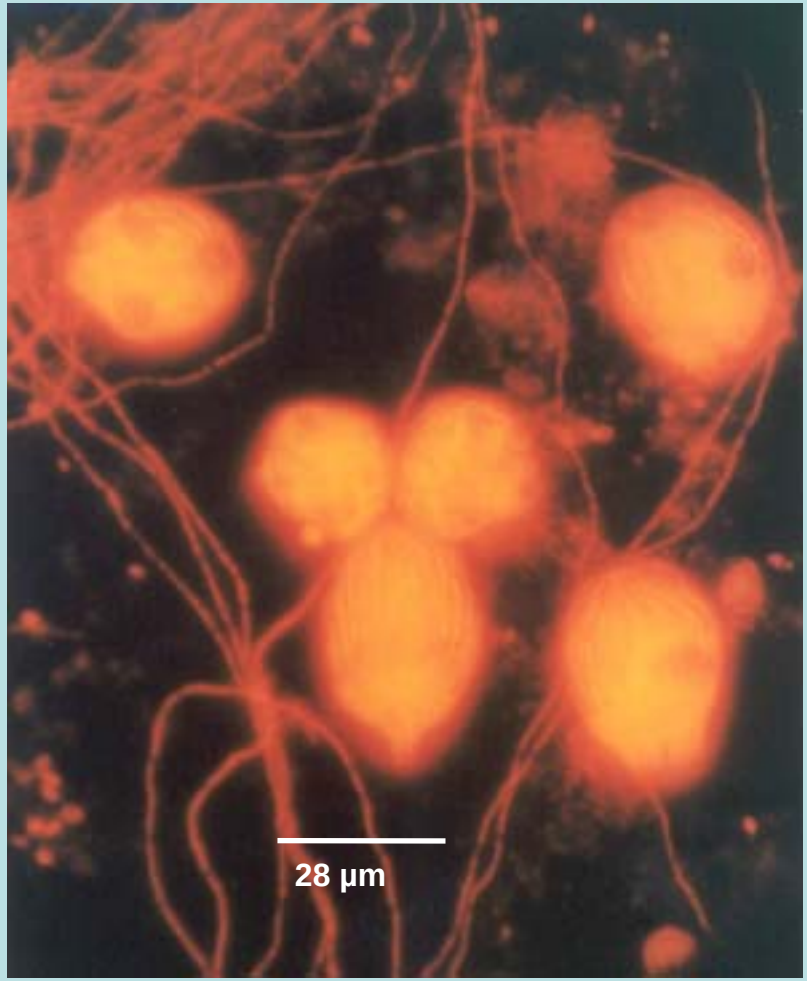
# Probe Design and Probe Match



# Combination of the Protargol Method according to Foissner with FISH on *Epistylis* sp



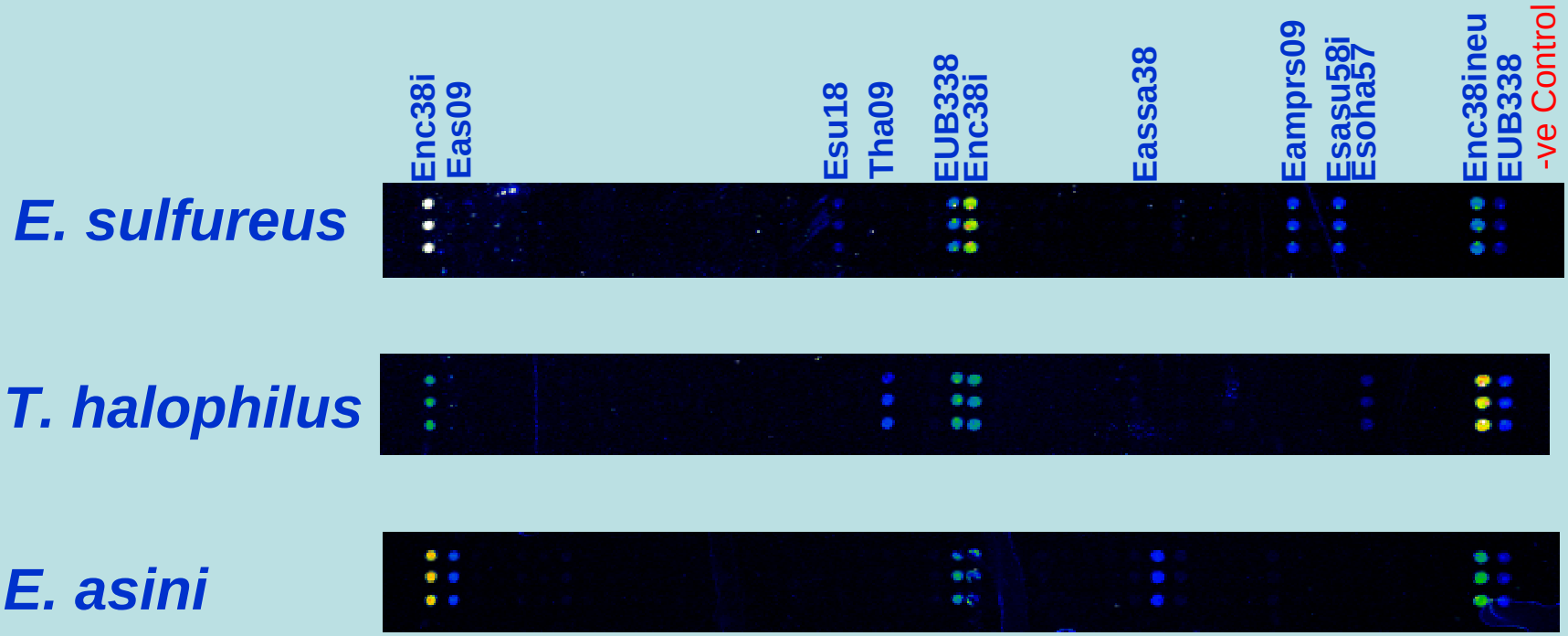
Under Light Microscope




Under Fluorescent Microscope

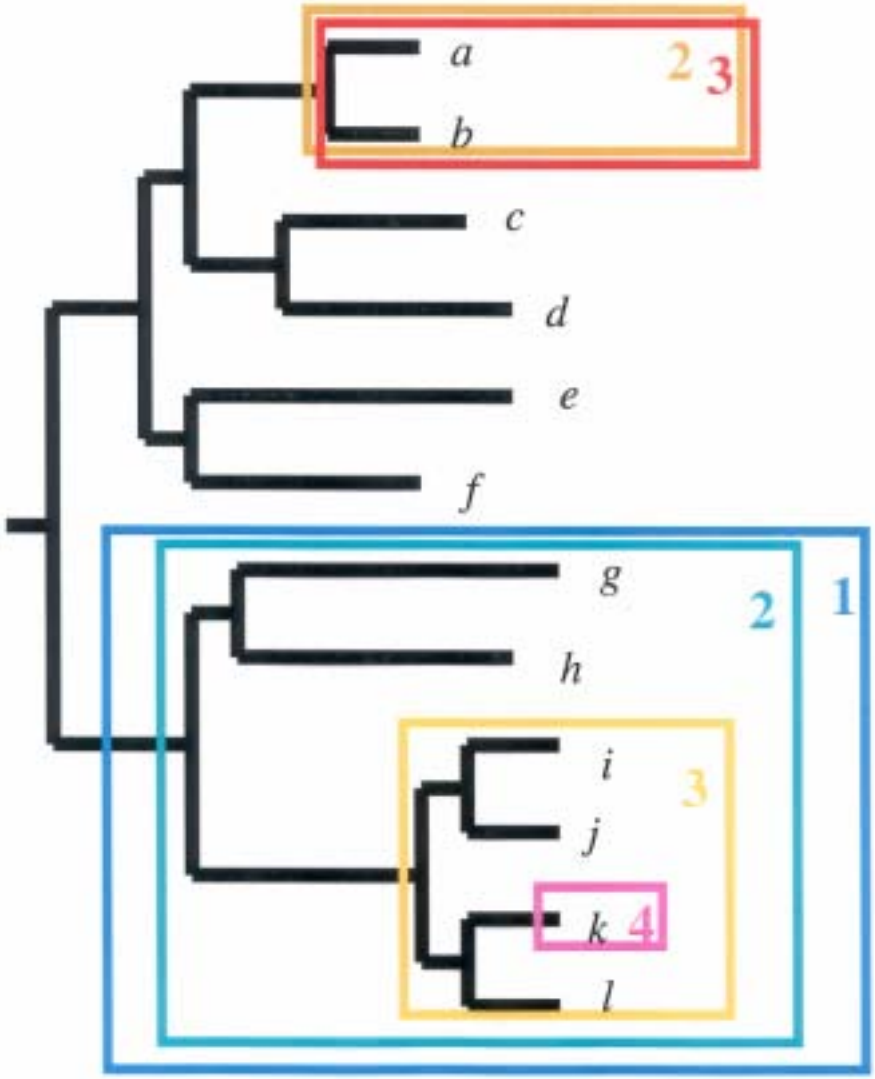
Pictures from Dr. Johannes Fried

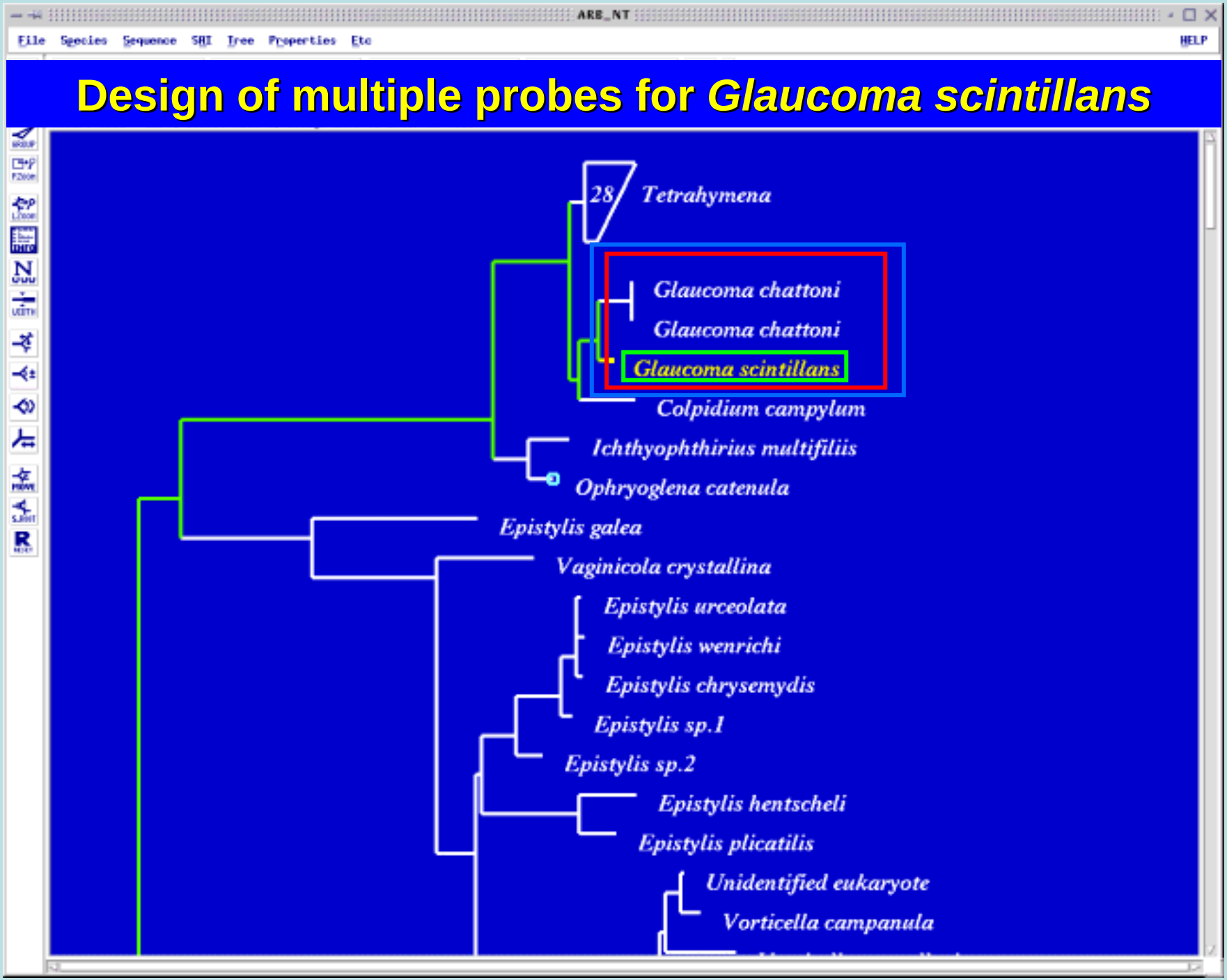
# Hybridised DNA Chips



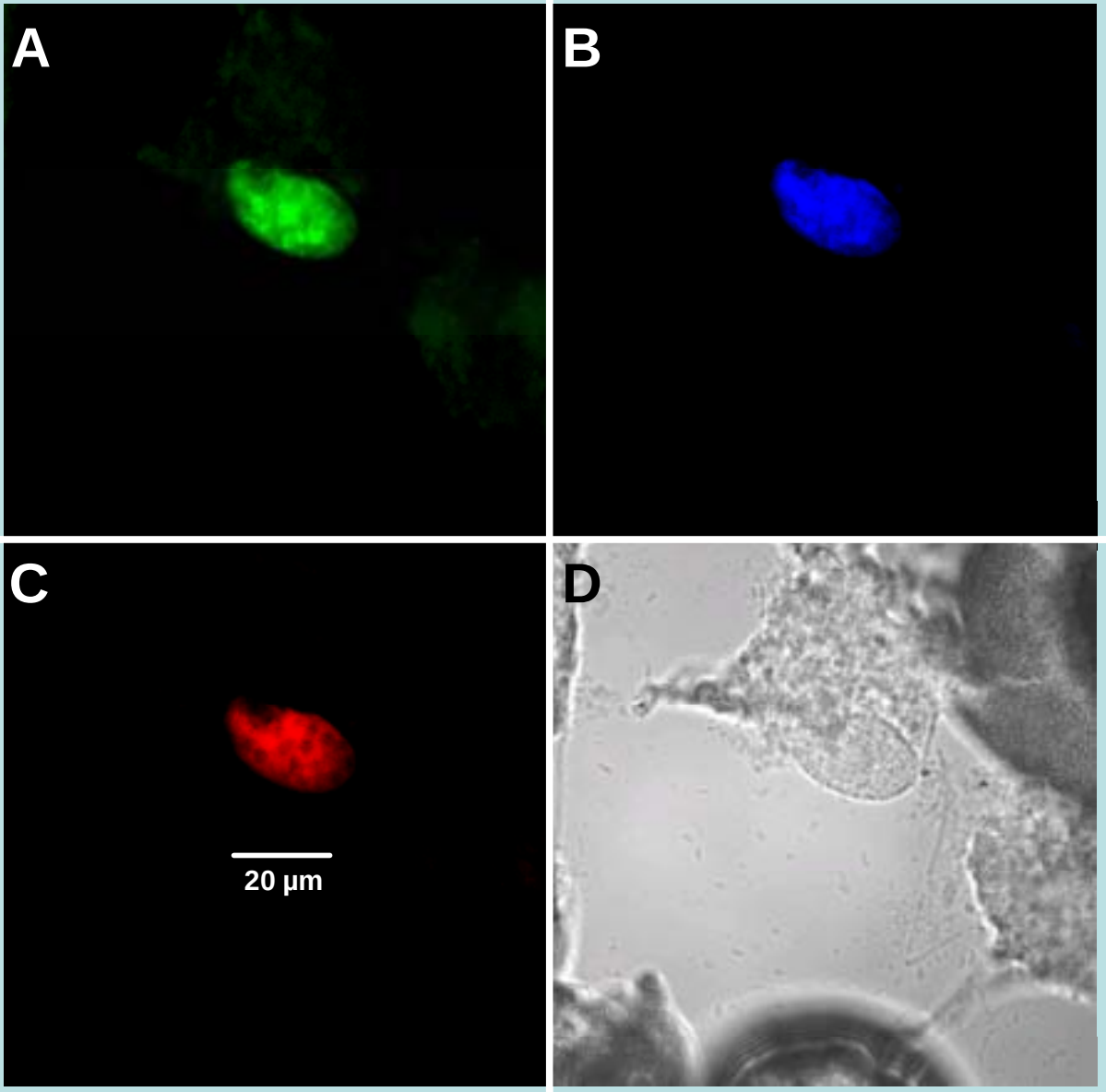

DNA chip images from Dr. Thomas Behr

# The Multiple Probe Concept




ARB_NT
File Species Sequence SAI Tree Properties Etc
HELP
Design of multiple probes for Glaucoma scintillans
28
Tetrahymena
Glaucoma chattoni
Glaucoma chattoni
Glaucoma scintillans
Colpidium campylum
Ichthyophthirius multifiliis
Ophryoglena catenula
Epistylis galea
Vaginicola crystallina
Epistylis urceolata
Epistylis wenrichi
Epistylis chrysemydis
Epistylis sp.1
Epistylis sp.2
Epistylis hentscheli
Epistylis plicatilis
Unidentified eukaryote
Vorticella campanula

# In situ hybridization of *Glaucoma scintillans* with multiple probes



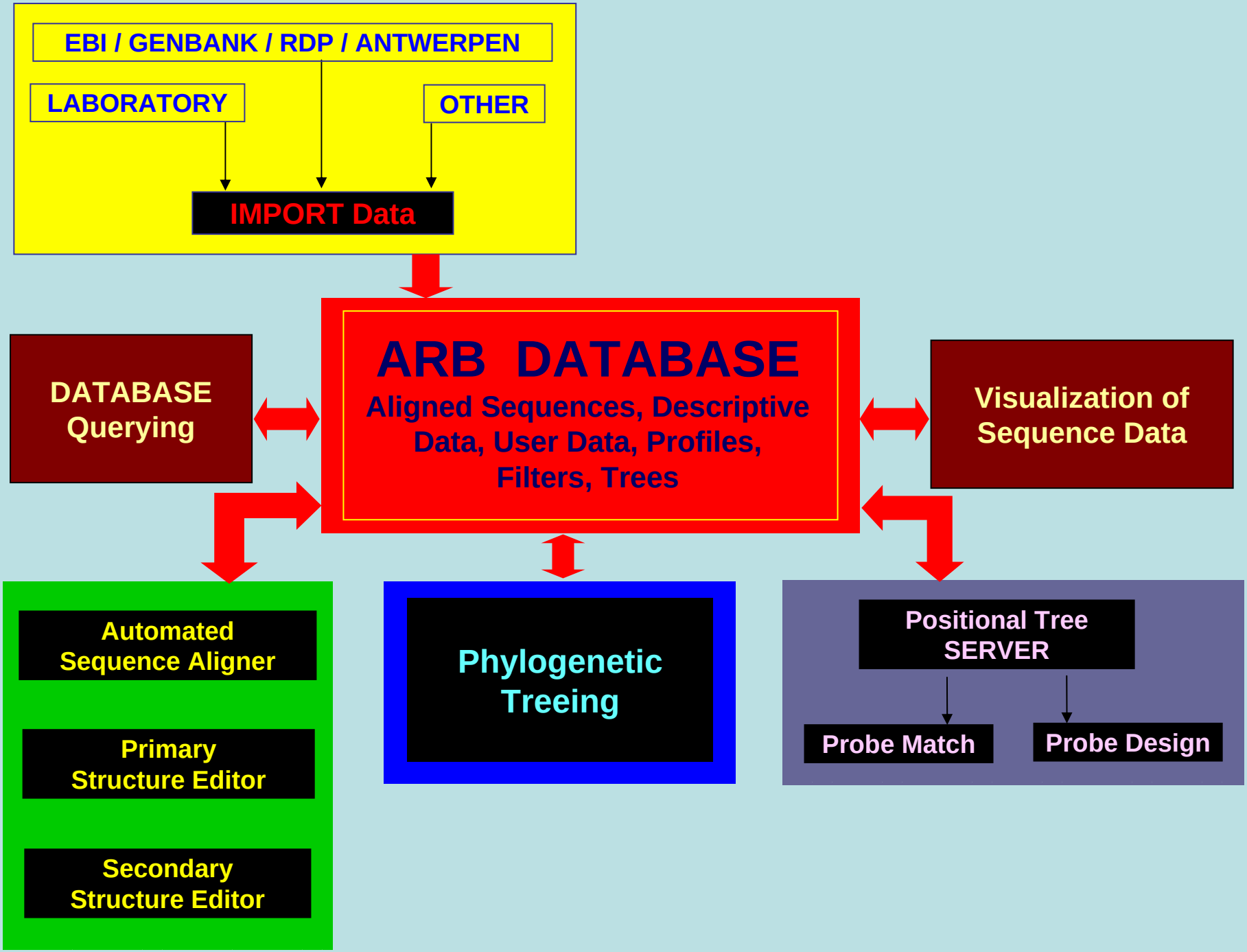

A : Species Specific Probe

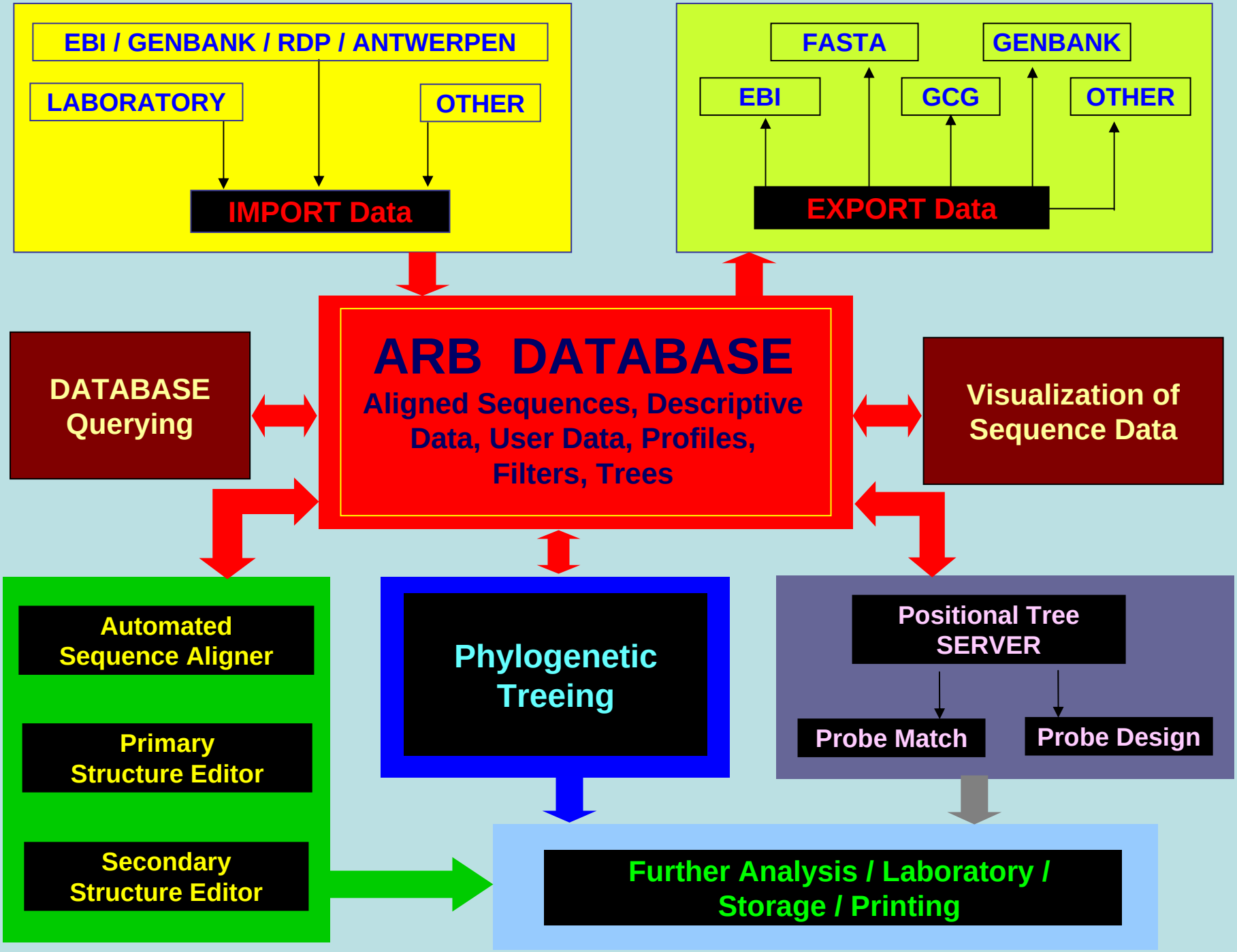
B : Genus Specific Probe 1

C : Genus Specific Probe 2

D : *G. scintillans* under light microscope

Pictures from Dr. Johannes Fried


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
DATABASE Querying
Visualization of Sequence Data
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing
Positional Tree SERVER
Probe Match
Probe Design


EBI / GENBANK / RDP / ANTWERPEN
LABORATORY
OTHER
IMPORT Data
FASTA
GENBANK
EBI
GCG
OTHER
EXPORT Data
DATABASE Querying
ARB DATABASE
Aligned Sequences, Descriptive Data, User Data, Profiles, Filters, Trees
Visualization of Sequence Data
Automated Sequence Aligner
Primary Structure Editor
Secondary Structure Editor
Phylogenetic Treeing
Positional Tree SERVER
Probe Match
Probe Design
Further Analysis / Laboratory / Storage / Printing

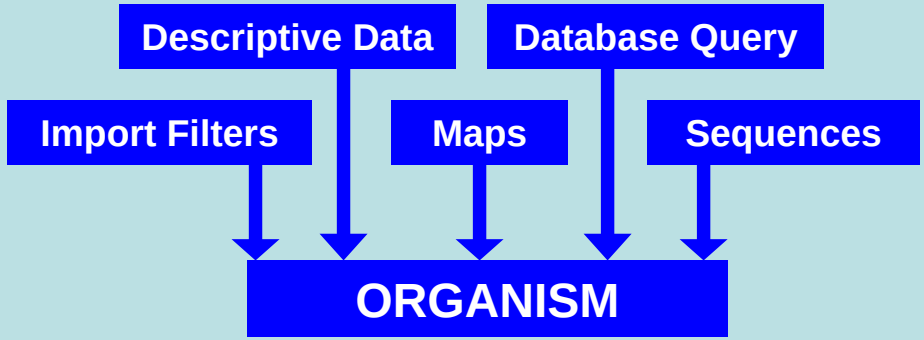
# The ARB Genome Project

EXPERIMENT

ORGANISM

GENES

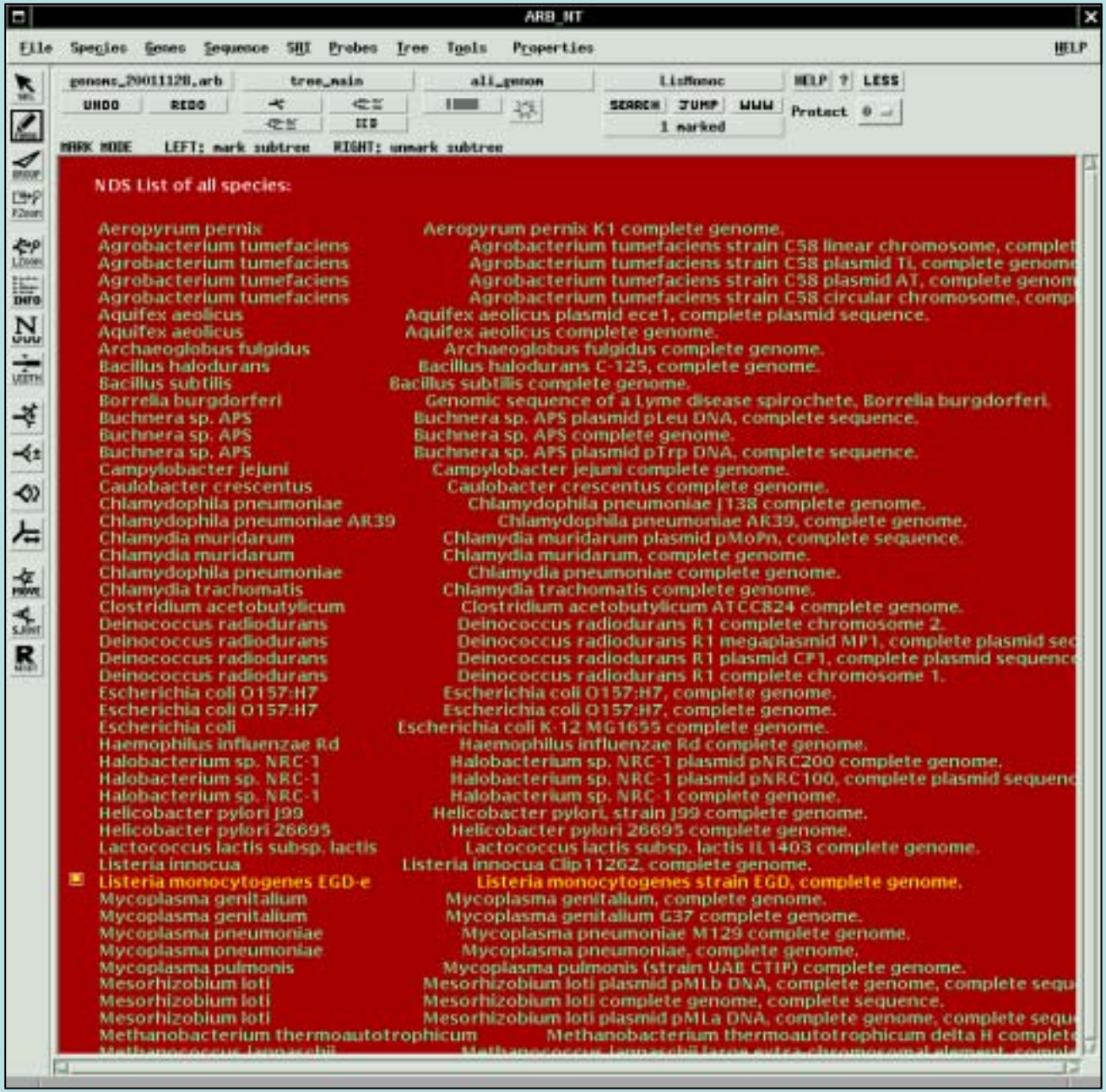
# The ARB Genome Project



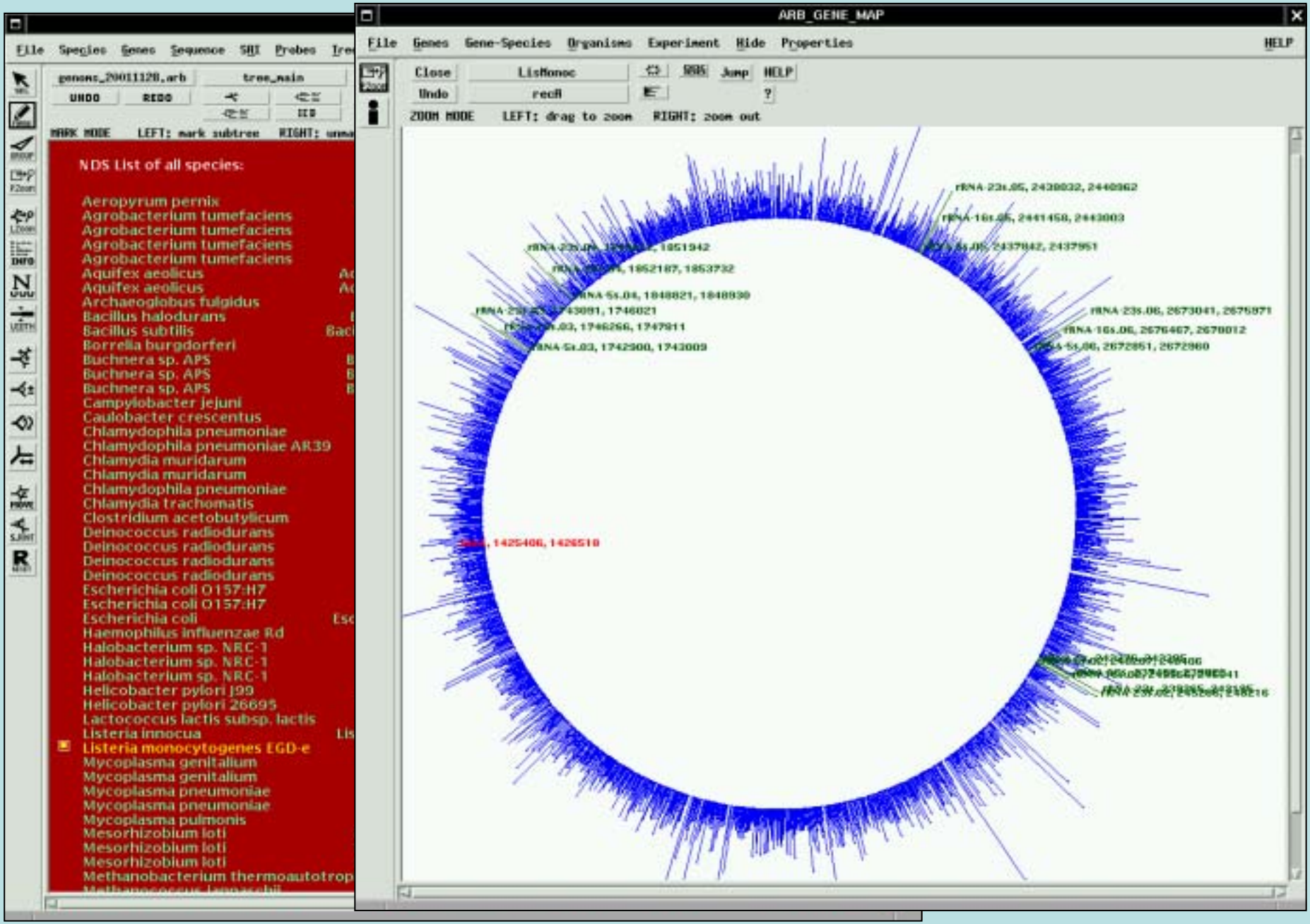

EXPERIMENT

GENES

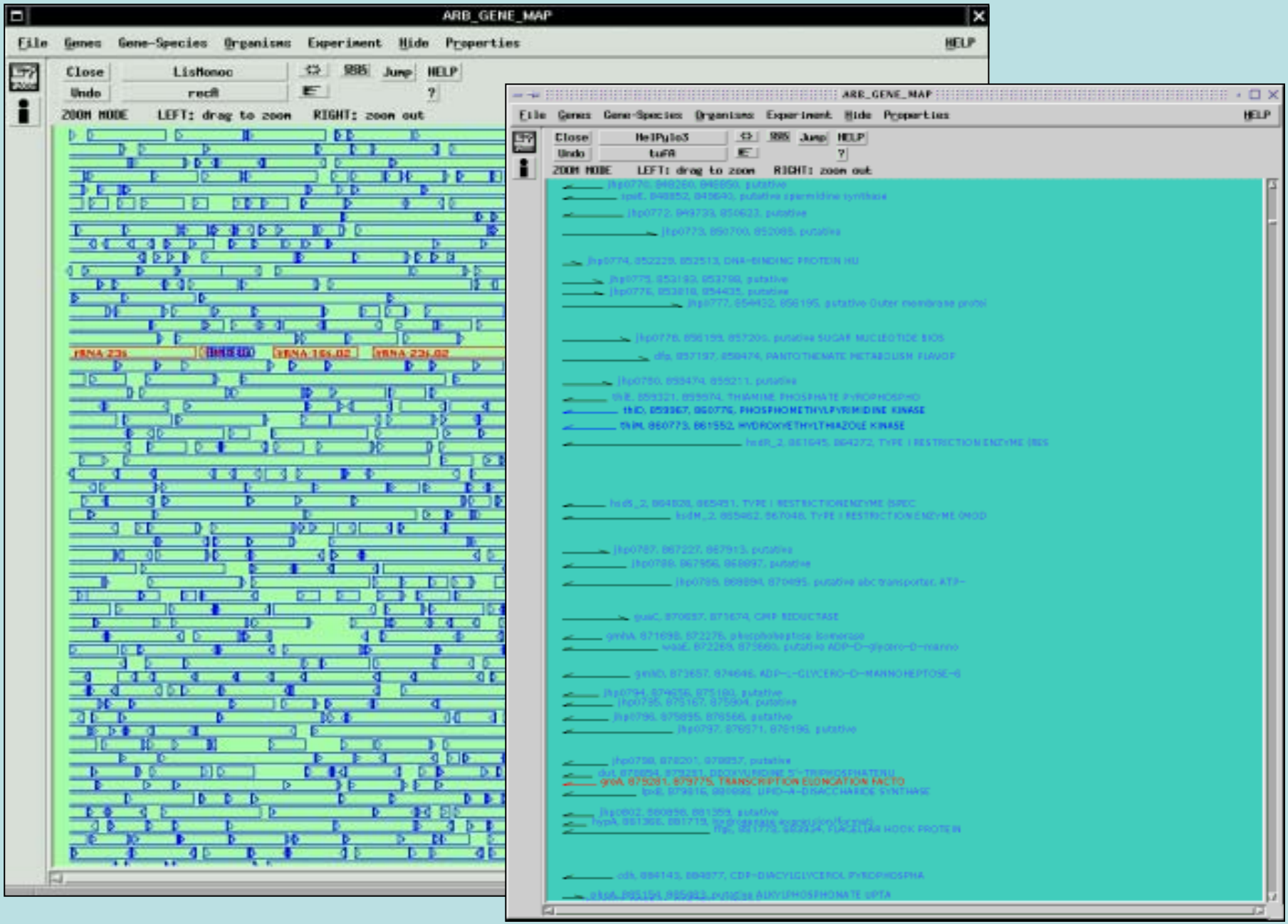
# ARB Genome Window : Displaying List of Organisms and associated information



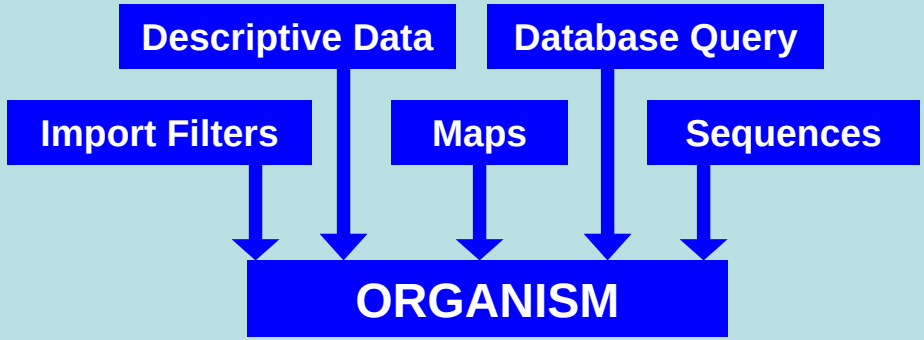
# Genome map of *Listeria monocytogenes* displaying rRNA Operons



# Genome map in block view and Individual Gene Information display



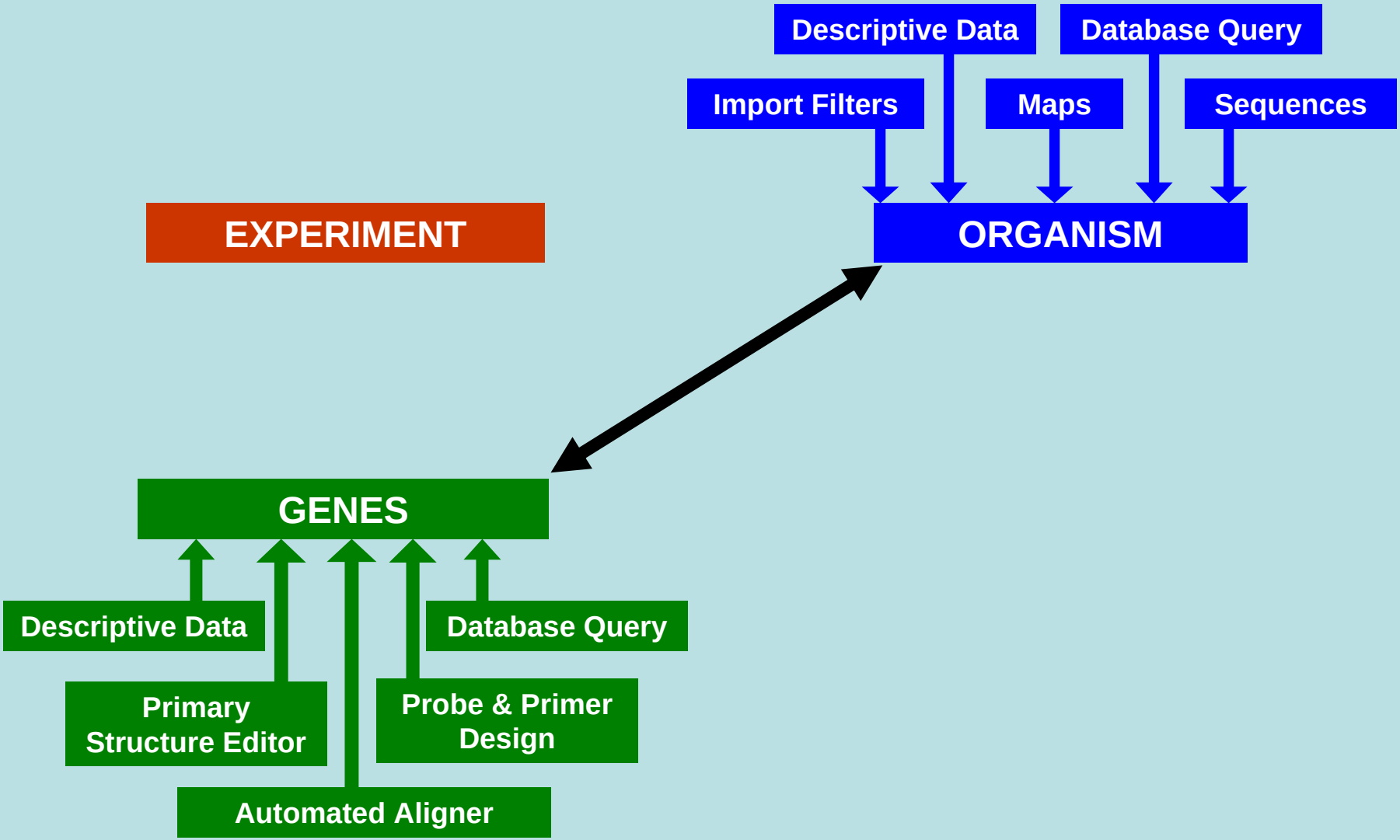
# The ARB Genome Project



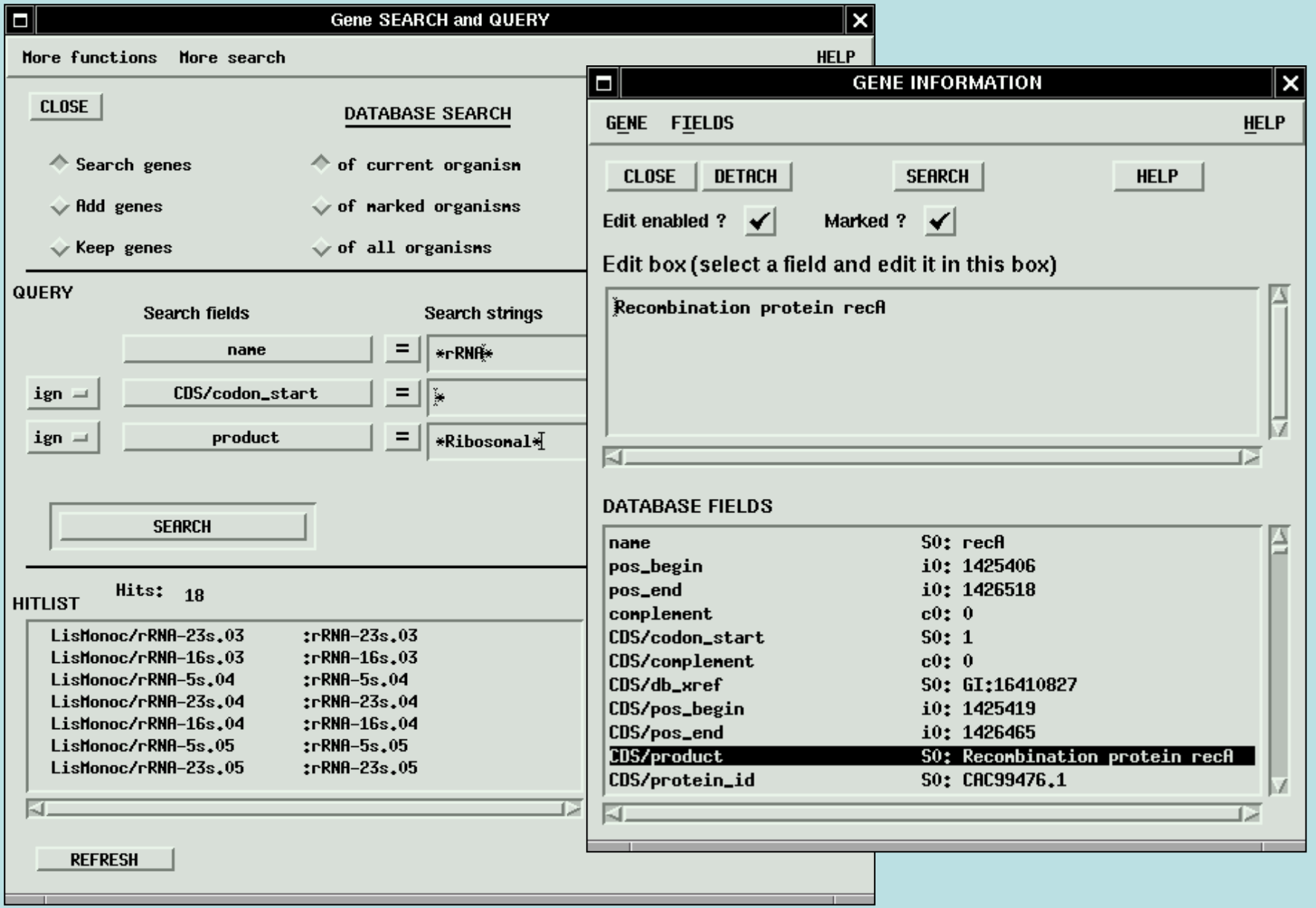

EXPERIMENT

GENES

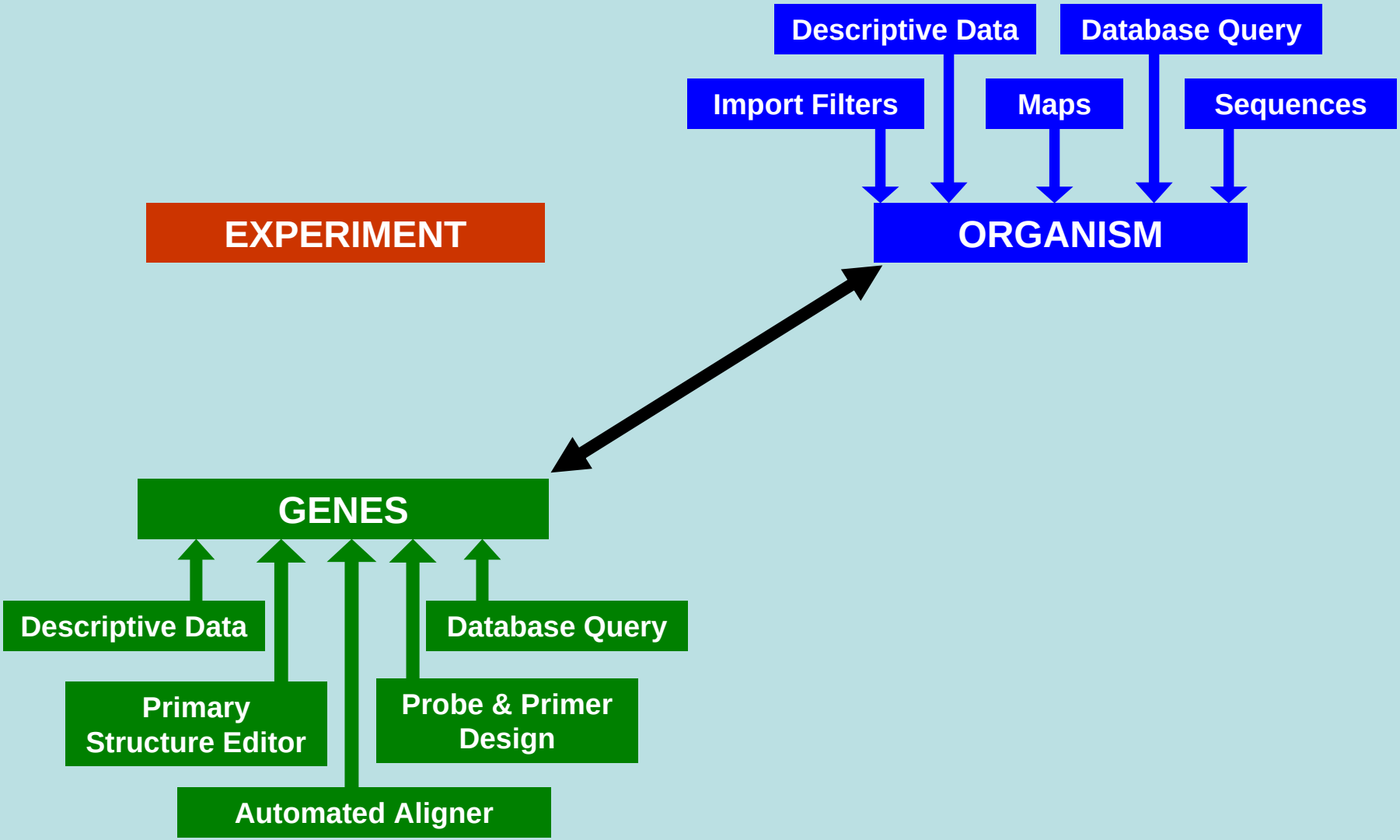
# The ARB Genome Project



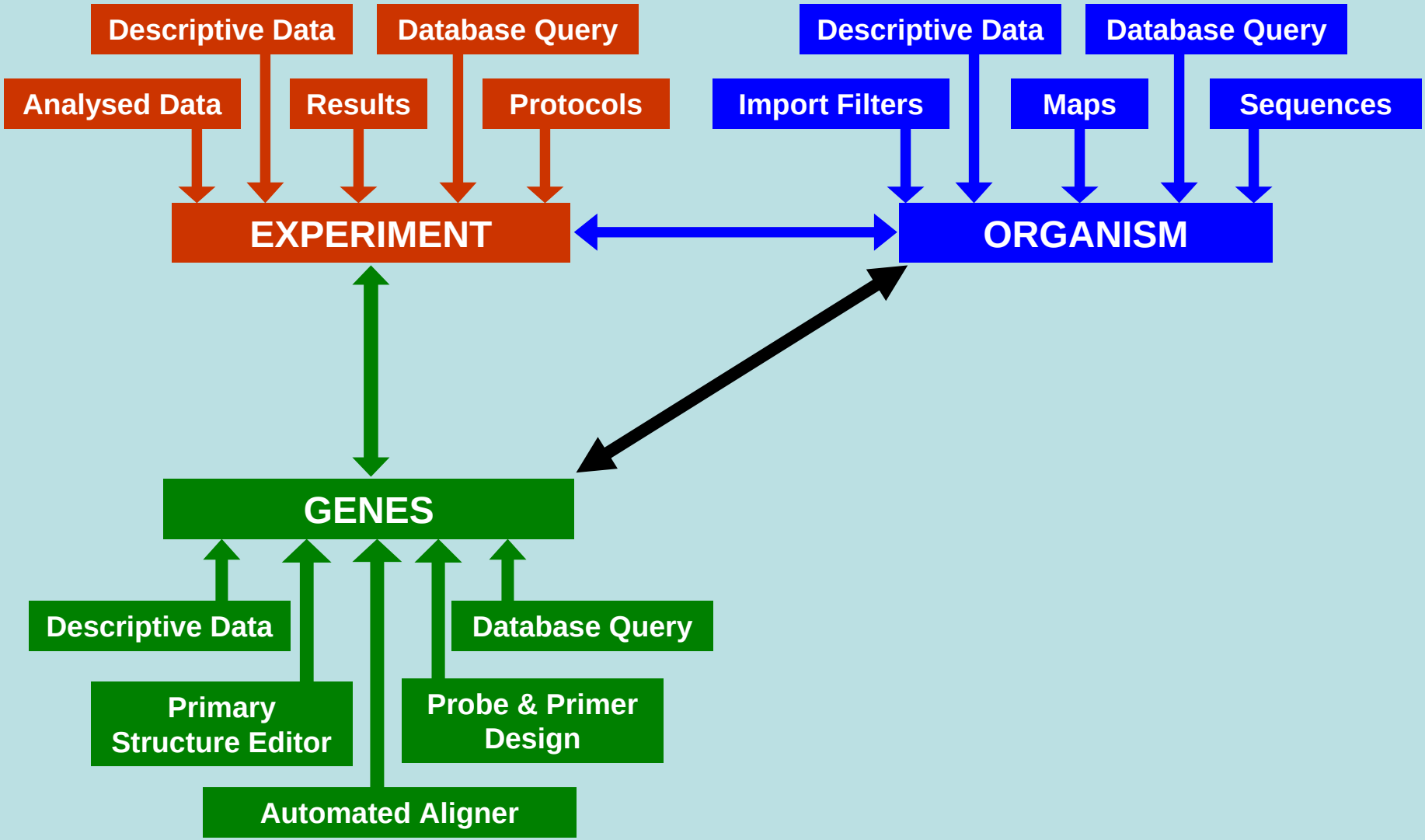
# Searching for Genes in the Genome of Organisms and displaying respective Gene Information



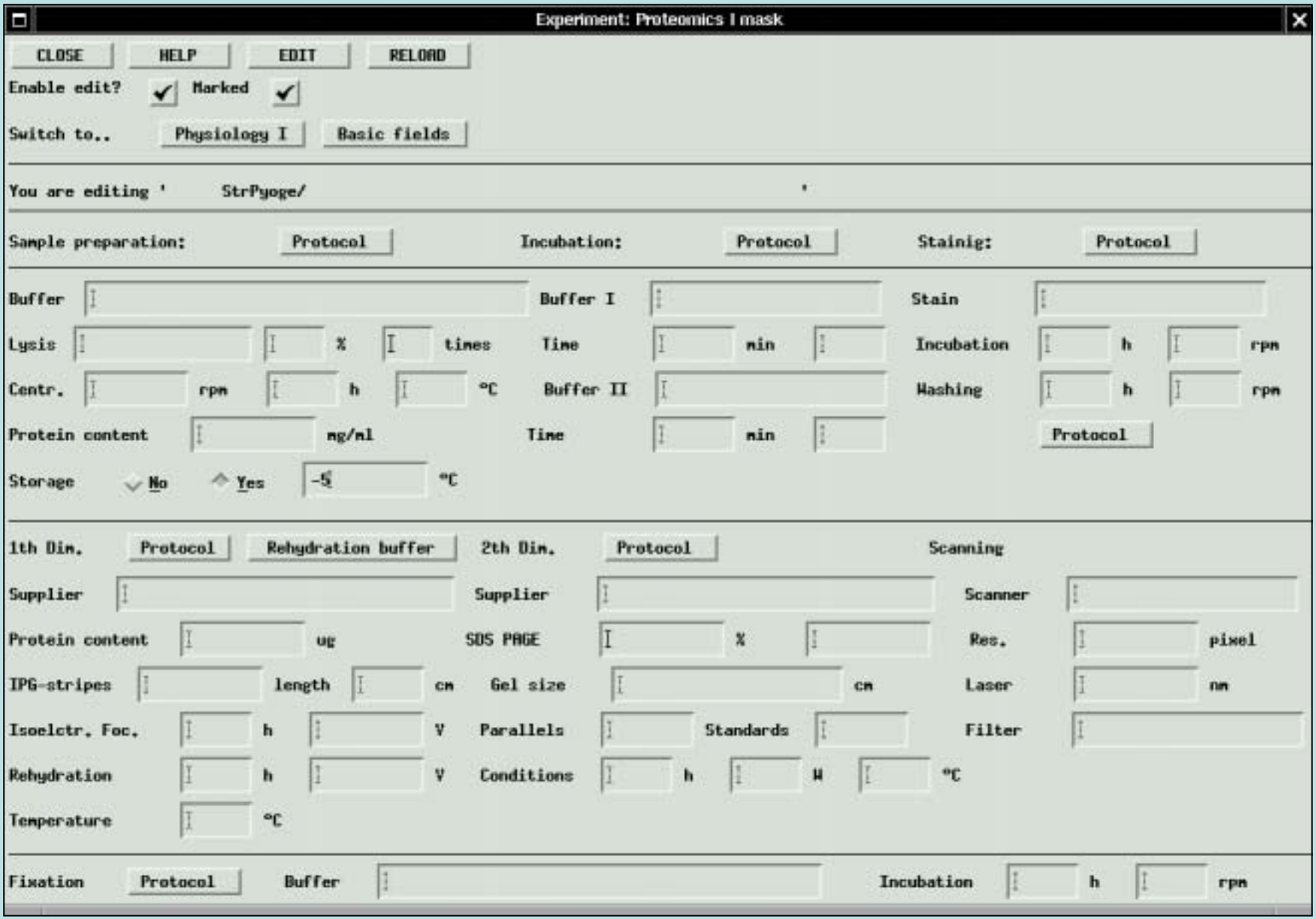
# The ARB Genome Project



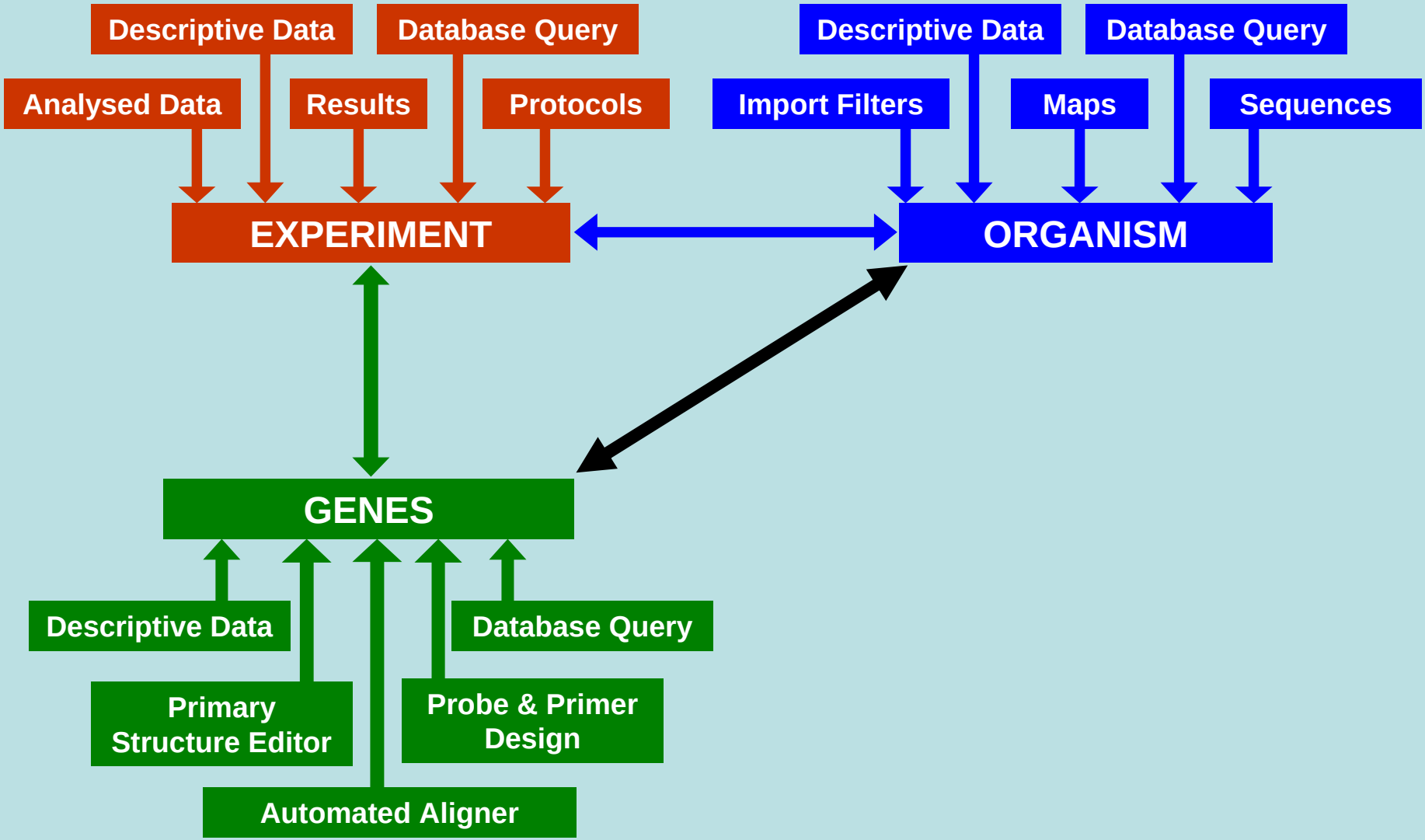
# The ARB Genome Project



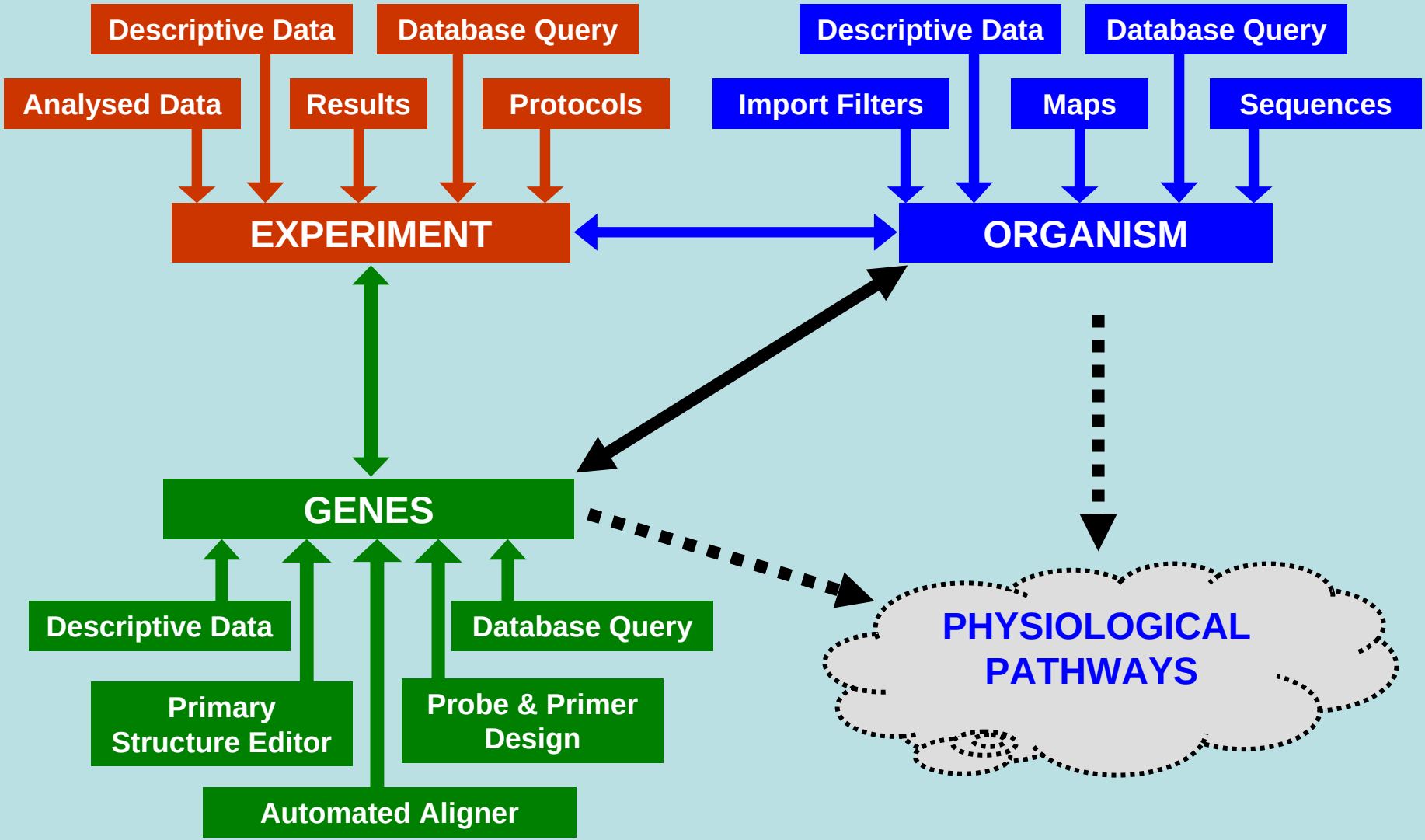
# Experiment Data Entry Form



# The ARB Genome Project



# The ARB Genome Project

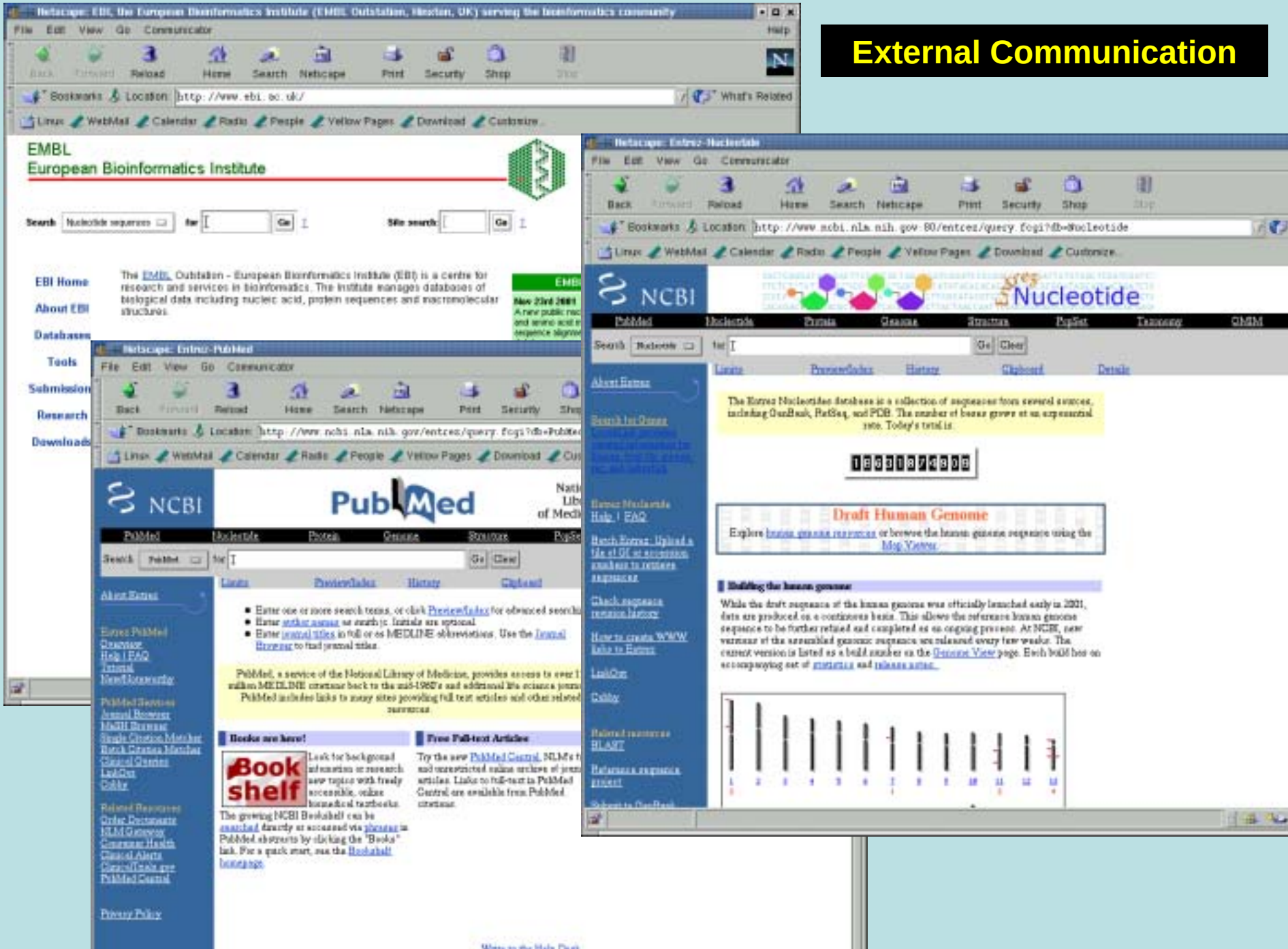


# Currently Maintained ARB Databases

# (Eucarya, Archaea, Bacteria)

- Small subunit rRNA – 16S,18S rRNA (41,737)
- Large subunit rRNA – 23S, 28S rRNA (7,312)
- Elongation – initiation factors
- Proton translocation ATPase subunits
- Heat shock proteins
- recA
- RNA polymerases
- DNA gyrase
- Cytochromeoxidase

# External Communication



## Operating Systems

**LINUX / Unix**
**Operating System**

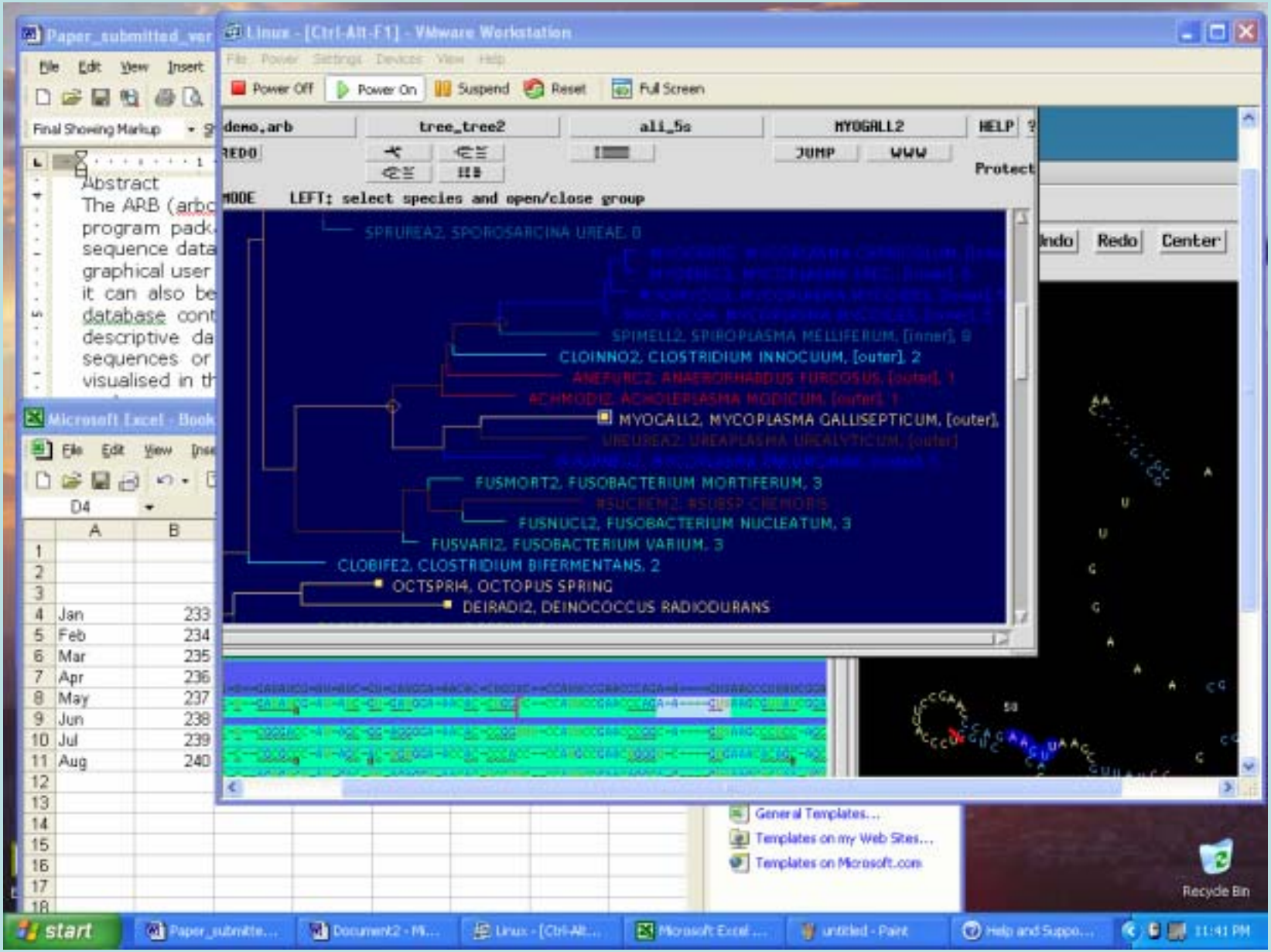
**Mac OS**

**?**

## Programming Languages

- **C, C++, Perl and other scripting languages**
- **GUI is based on X Windows & Open Motif Library**

# ARB running on Vmware, a Linux emulation software under Windows



## Operating Systems

**LINUX / Unix**
**Operating System**

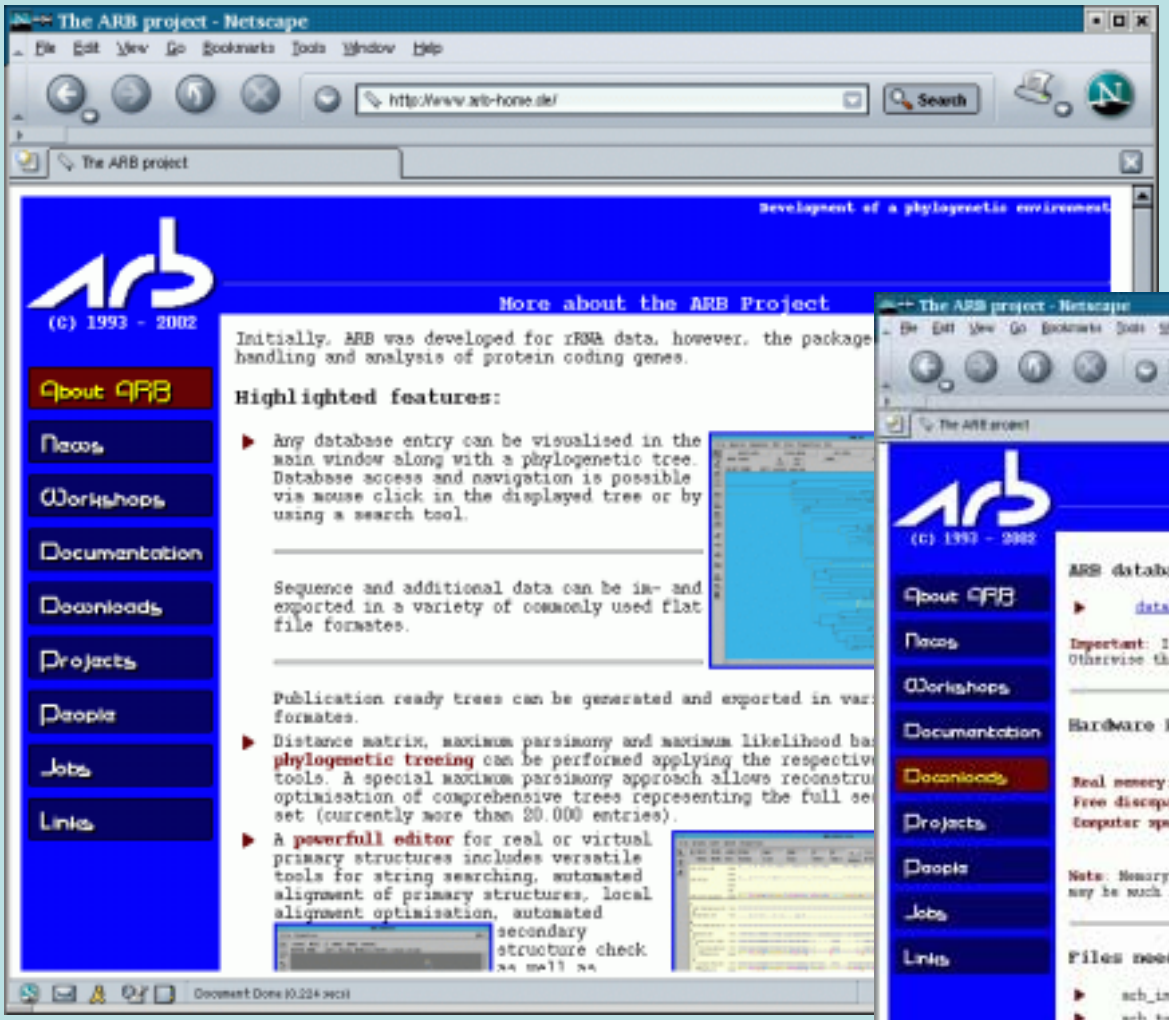
**Mac OS**

**?**

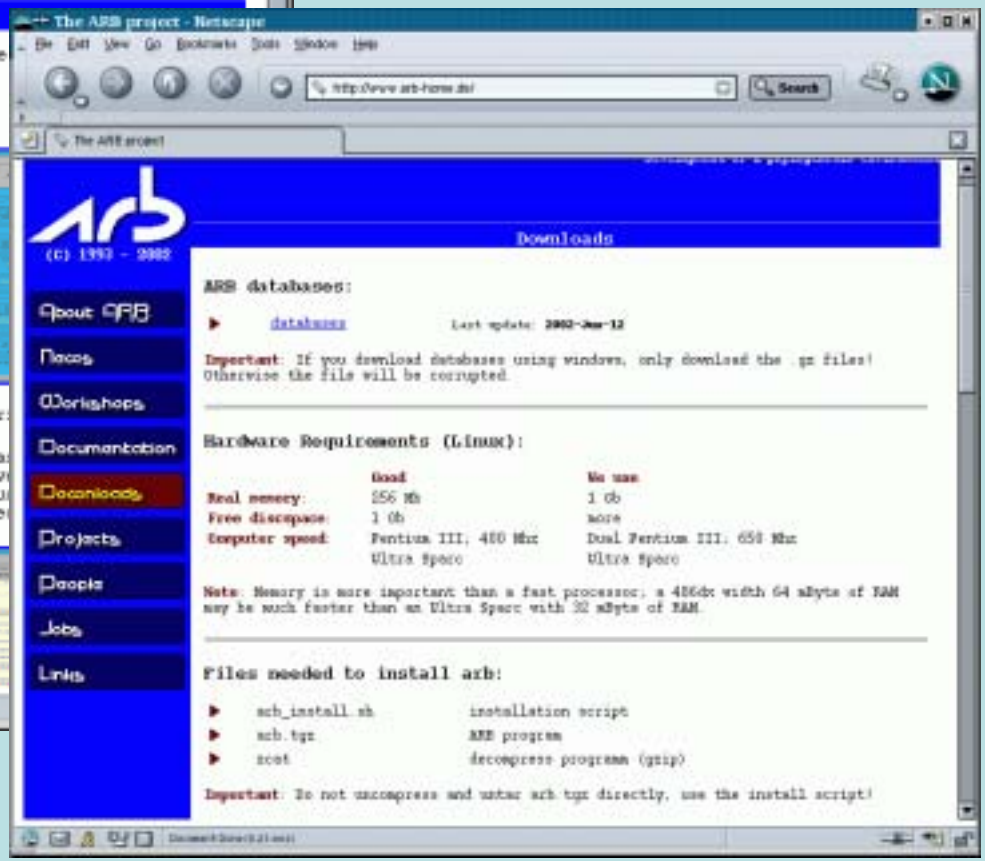
## Programming Languages

- **C, C++, Perl and other scripting languages**
- **GUI is based on X Windows & Open Motif Library**

# Availability & Documentation




> **"As we enjoy great Advantages from the Inventions of others, we should be glad of an Opportunity to serve others by any Invention of ours; and this we should do freely and generously."**
>
> **- Benjamin Franklin**




**www.arb-home.de**

# People Behind The ARB Project

## Group Leader

**Dr. Wolfgang Ludwig**

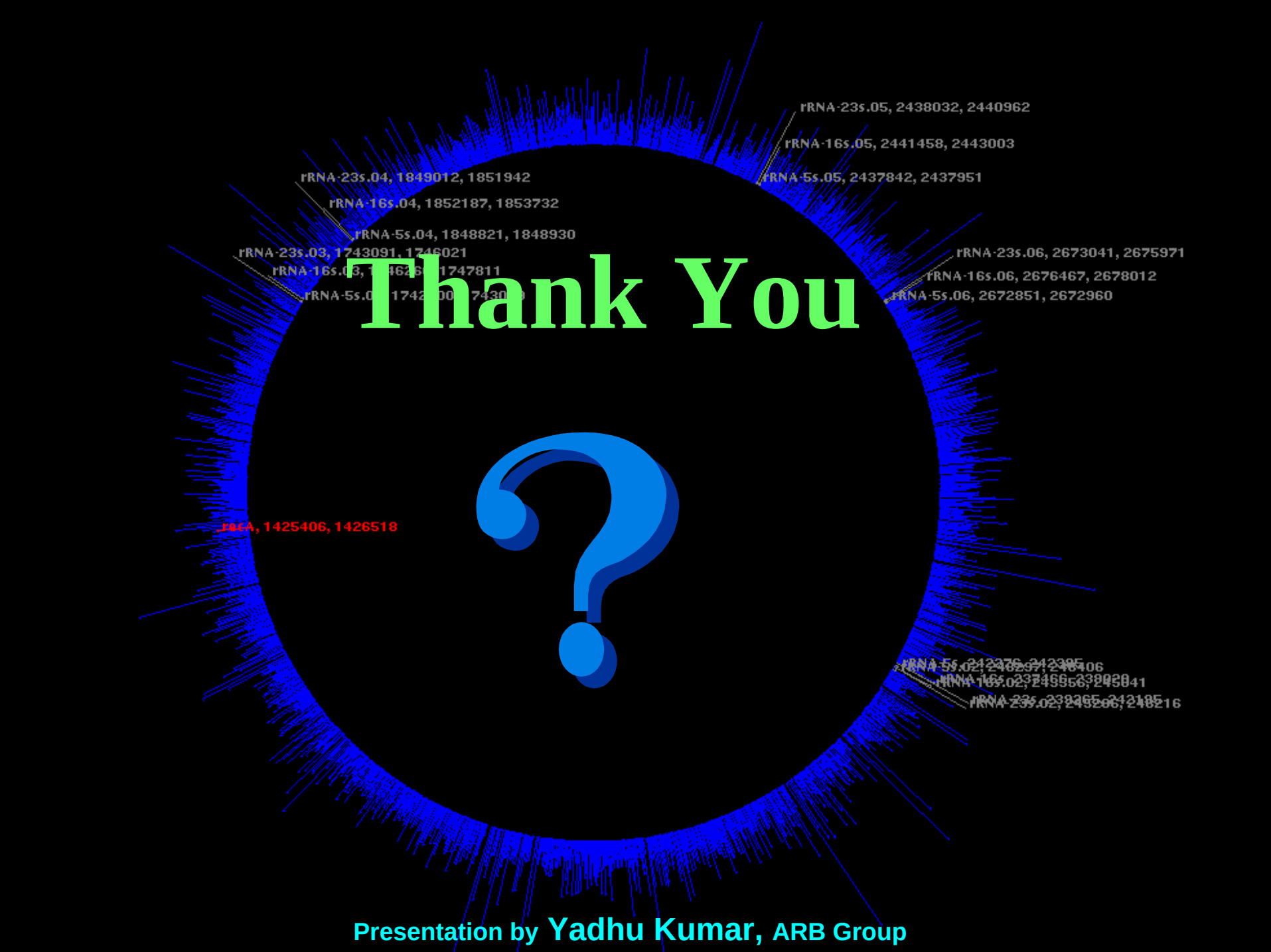
**Lehrstuhl fuer Mikrobiologie**
**Technische Universitaet Muenchen**
**ludwig@mikro.biologie.tu-muenchen.de**

## Programmers and Curators

**O.Strunk, R.Westram, L. Richter, H. Meier, Yadhukumar, A.Buchner, T.Lai, G.Jobb, S.Steppi, W. Förster, H. May, S. Hermann, N. Stuckmann, O. Gross, B. Nonhoff, R. Jost, B. Reichel,T. Ginhart, A. Vilbig, T. Liss, M. Lenke,**

# Future Goals

- Online Probe Design using ARB  Positional Tree server
- Multiple probe sets for selected phylogenetic groups (chip design)
- Chip data analysis and evaluation tool
- Further Development of ARB Genome Analysis Software
-




Presentation by **Yadhu Kumar,** ARB Group