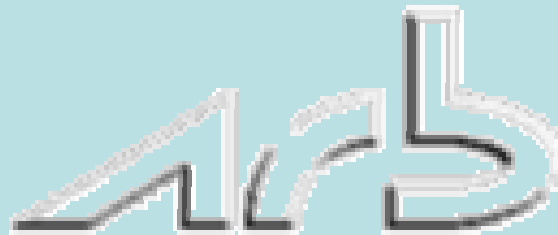


# The ARB Project

---

**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

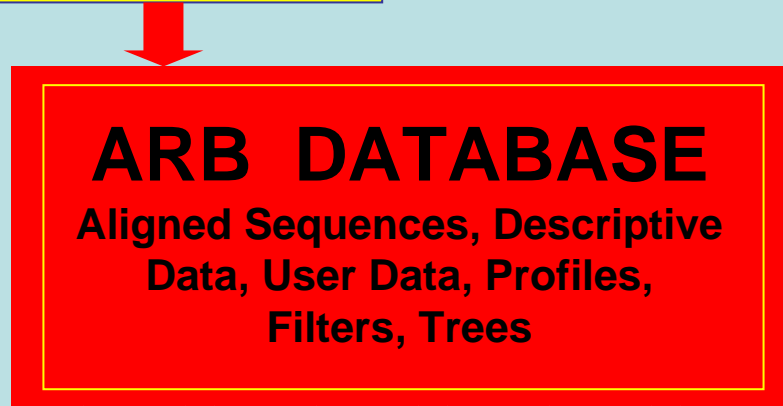
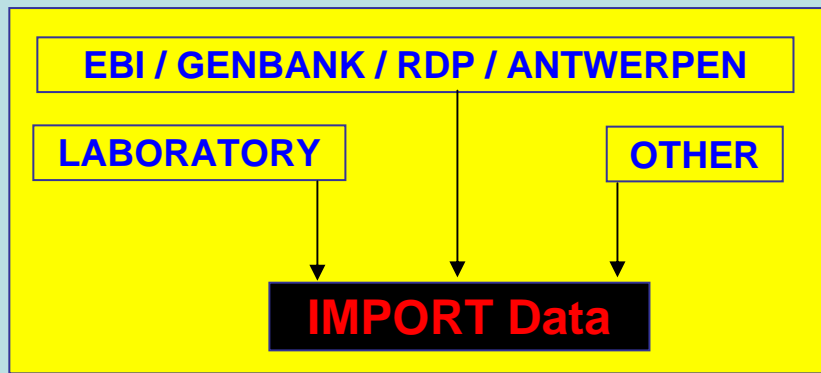
- ◆ **Central Database to maintain a structured integrative secondary data in combination with processed primary structures (aligned sequences) and any additional data assigned to the individual sequences.**
- ◆ **Comprehensive selection of software tools directly interacting with one another and as 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**





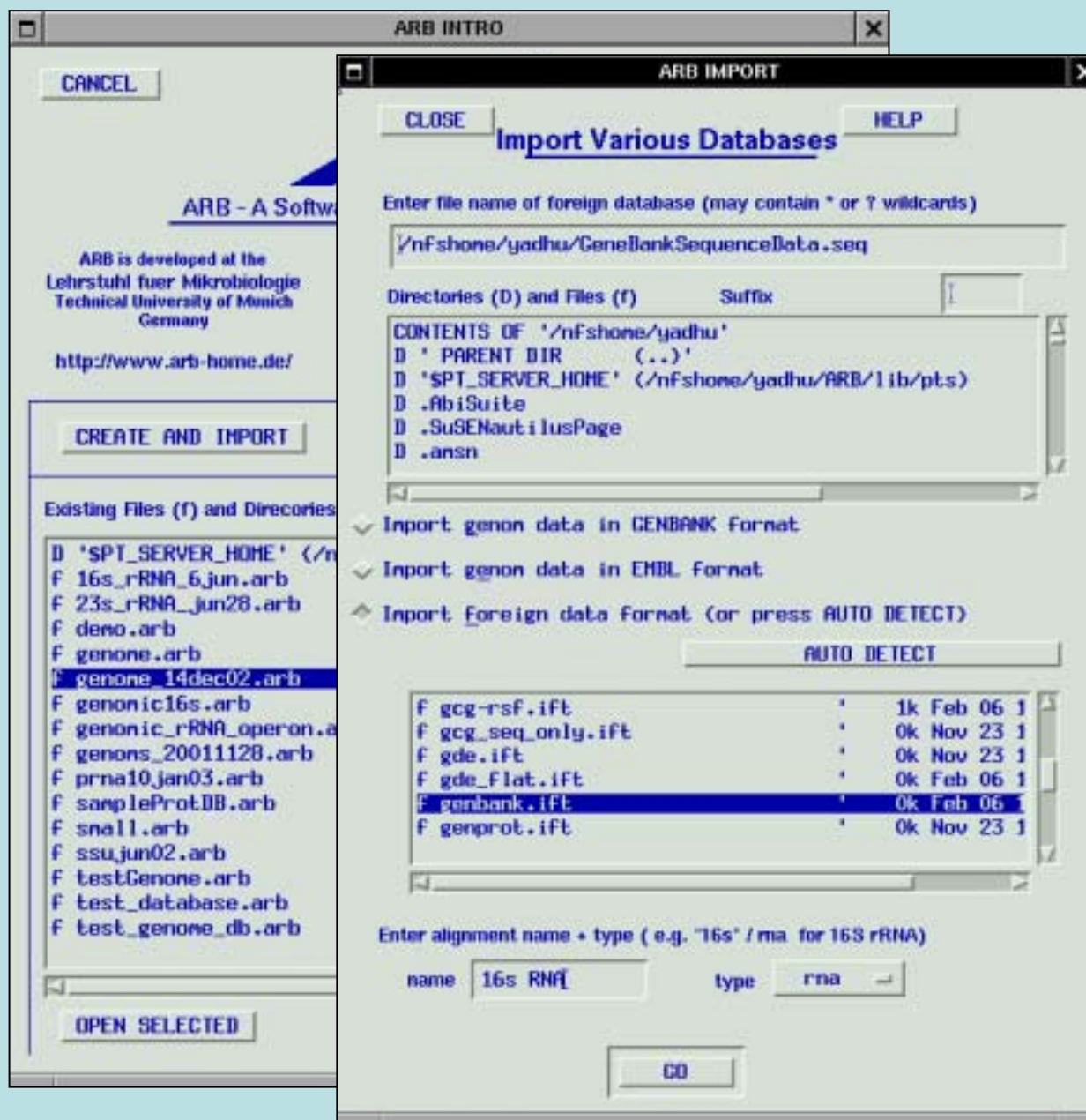


# ARB Main Window & Import Window

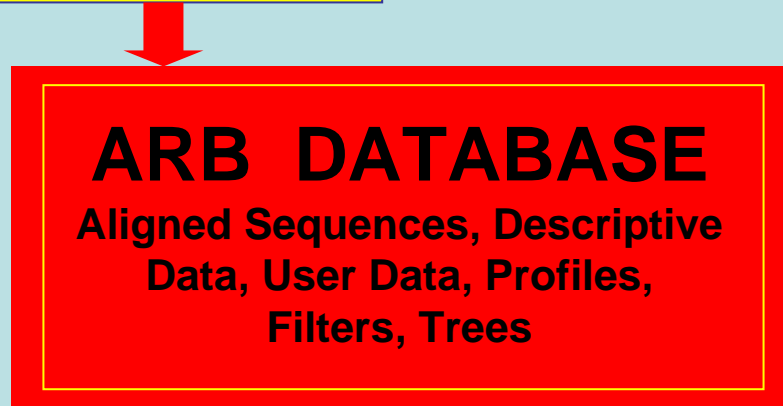
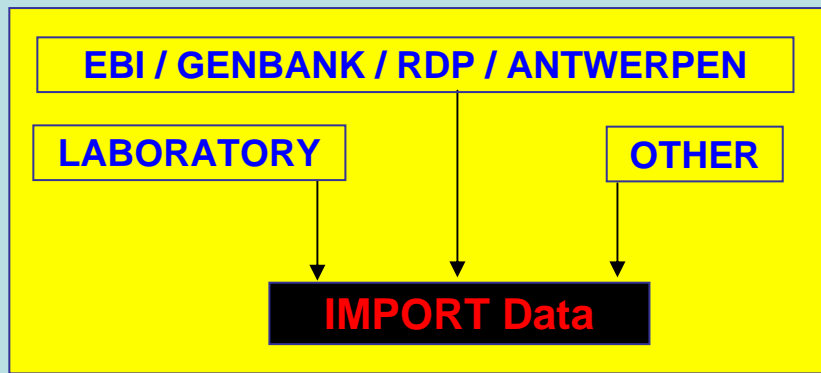




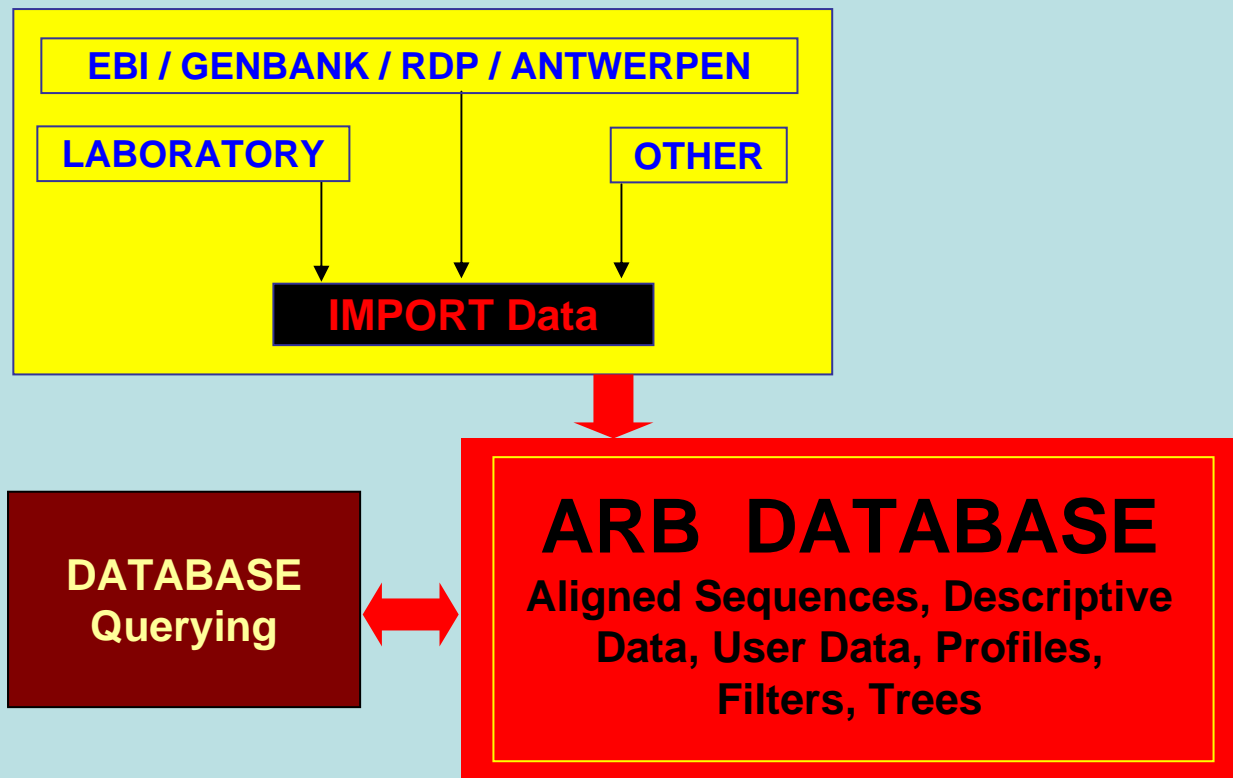
# ARB Main Window & Import Window













# Search, Query & Modify ARB Database

SEARCH and QUERY

More functions More search [HELP](#)

CLOSE

DATABASE SEARCH

HELP

◆ Search species

◆ Add species

◆ Keep species

◆ that match the query

◆ that dont match the q.

◆ that are marked

QUERY

Search fields

Search strings

and

full\_name

=

\*salmonella\*

and

nuc\_term

=

>1500


and

ebi\_check

=

\*

SEARCH



HITLIST Hits: 18

\* SalTyp44 :Salmonella typhi

\* SalPara2 :Salmonella paratyphi A

\* SalEnt22 :Salmonella paratyphi B

\* SalEnt23 :Salmonella paratyphi C

\* SalEnt25 :Salmonella blockley

\* SalEnt30 :Salmonella matopeni

\* SalTyp45 :Salmonella typhimurium

\* SalEnt31 :Salmonella weltevreden

MARK LISTED  
UNMARK REST

UNMARK LISTED  
MARK REST

DELETE LISTED

WRITE TO FIELDS  
OF LISTED

REFRESH



# Search, Query & Modify ARB Database

SEARCH and QUERY

More functions More search [HELP](#)

CLOSE

◆ S

◆ A

◆ K

QUERY

Search f

fu

and

and

SEARCH

HITLIST

Hits: 18

\* SalTyp44

\* SalPara2

\* SalEnt22

\* SalEnt23

\* SalEnt25

\* SalEnt30

\* SalTyp45

\* SalEnt31

REFRESH

SPECIES INFORMATION

CLOSE DETACH SEARCH [HELP](#)

Edit enabled ? ☒ Marked ? ☒

Edit box (select a field and edit it in this box)

[ [DEV] Salmonella typhi U88545 [EBI,RDP] Salmonella typhi

DATABASE FIELDS

full_name	S0: Salmonella typhi
strain	S0: [DEV;EBI] St111 [RDP] St111. Former RDP Sids: U88545 Corresponding GenBank entry: U88545
next_rel	S0: 384 'SalTyp47'
lib_name	S0: [DEV] Salmonella typhi U88545 [EBI,RDP] Salmonella typhi
acc	S0: U88545
db_acc	S0: [DEV;EBI;RDP] U88545
version	S0: [EBI] U88545.1
aligned	S0: 1apr90UL 12feb00UL 10dec00UL 21jul01UL 28dec01UL
author	S0: [DEV] Subramanian G. [EBI] # Subramanian G. [RDP] Subramanian G.
title	S0: [EBI] (1):
journal	S0: [DEV;EBI] Submitted (05-FEB-1997) to the EMBL/GenBank/DBJ databases. Institute for
seqcheck	S0: ARB_09049490
nuc	S0: 1541
nuc_tern	S0: 1541
id	S0: [EBI] STU88545 [RDP] S.typhi3
trp	S0: [RDP] up_inFo
name	S6: SalTyp44
ali_16s/data	S5: .....
tax	S0: [DEV] Bacteria Proteobacteria (gamma subdivision) Enterobacteriaceae Salmonella [EBI]
gene	S0: [DEV] SSU [EBI] .
description	S0: [EBI] Salmonella typhi 16S ribosomal RNA gene; complete sequence. [RDP] typhoid Fever
ebi_comment	S0: <A HREF=http://srs6.ebi.ac.uk/srs6bin/cgi-bin/vgetz?e+[EMBL_features-id:STU88545_1]>
nuc_gene_03sep01	S0: 1541



# Search, Query & Modify ARB Database

The screenshot displays two overlapping windows from the ARB database software. The background window is titled "SEARCH and QUERY" and contains a search interface with a "QUERY" section for building search criteria and a "HITLIST" section showing 18 results. The foreground window is titled "MODIFY DATABASE FIELD of listed species" and provides options to modify field contents, including a list of destination fields and a list of predefined programs.

**SEARCH and QUERY Window:**

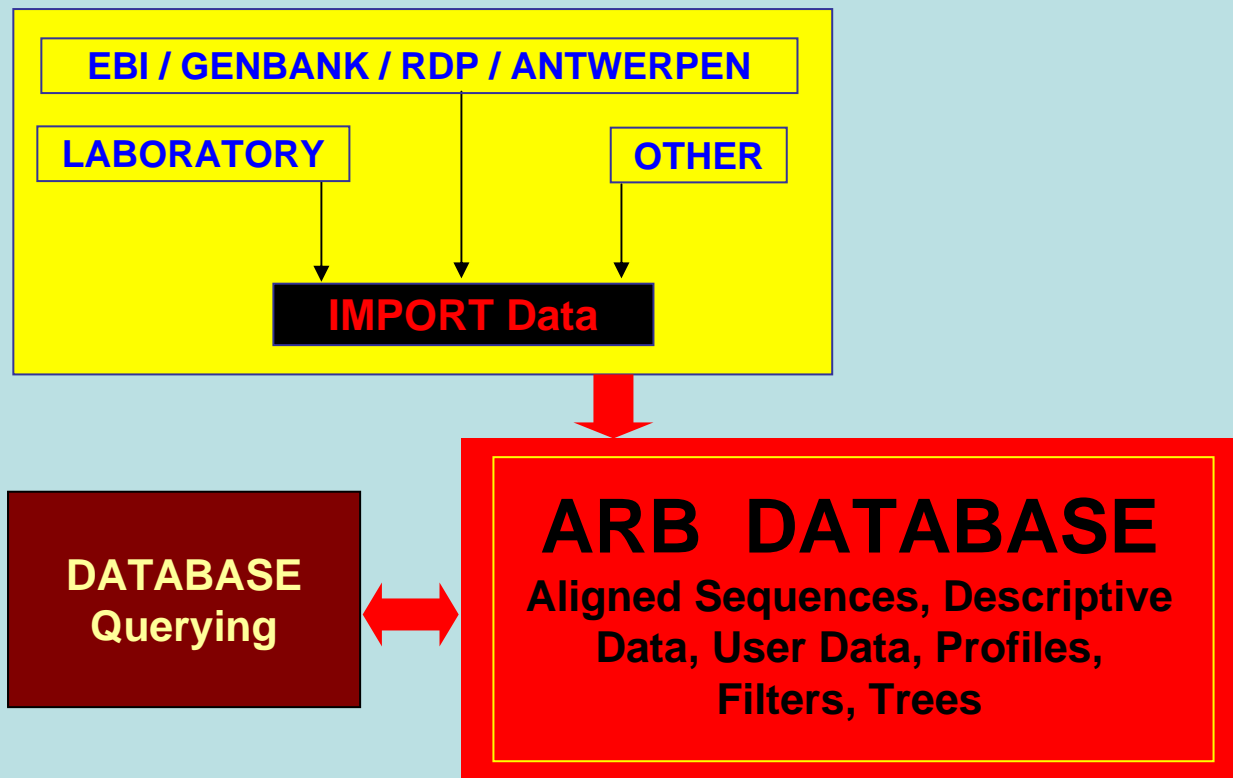
- More functions** | **More search**
- CLOSE** button
- SPECIES FIELDS** section:
  - CLOSE** | **DETACH** | **SEARCH** buttons
  - Checkboxes: **Edit enabled ?** (checked) | **Marked ?** (unchecked)
  - Text input: **Edit box (select a field and edit it)** with value "[DEU] Salmonella typhi U885"
- QUERY** section:
  - Search for** text input
  - and** | **or** | **not** operators
  - SEARCH** button
- HITLIST** section:
  - Hits: 18**
  - List of search results (e.g., \* SalTyp44, \* SalPara2, etc.)
  - REFRESH** button
- DATABASE FIELDS** table:

Field Name	Value
full_name	S0: Salmo
strain	S0: [DEU]
next_rel	S0: 384
lib_name	S0: [DEU]
acc	S0: U885
db_acc	S0: [DEU]
version	S0: [EBI]
aligned	S0: 1apr
author	S0: [DEU]
title	S0: [EBI]
journal	S0: [DEU]
seqcheck	S0: ARB
nuc	S0: 1541
nuc_term	S0: 1541
id	S0: [EBI]
tmp	S0: [RDB]
name	S6: SalTy
ali_16s/data	S5: ....
tax	S0: [DEU]
gene	S0: [DEU]
description	S0: [EBI]
ebi_comment	S0: <A H
nuc_gene_03sep01	S0: 1541

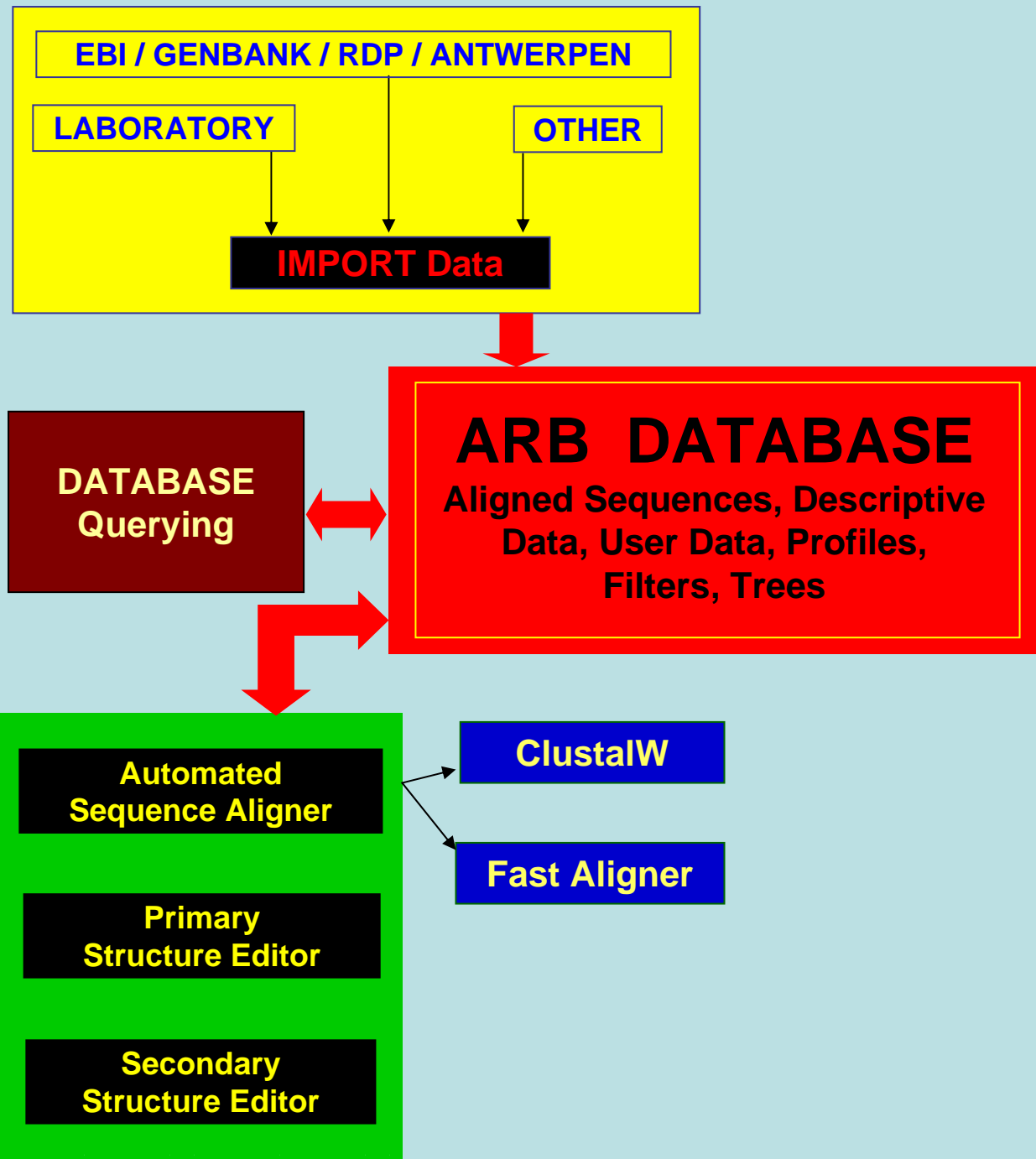
**MODIFY DATABASE FIELD of listed species Window:**

- CLOSE** | **HELP** buttons
- Text: "This module modifies the contents of fields. You can:"
- List of actions:
  - \* substitute substrings
  - \* copy one field to another
  - \* extract and calculate sequence information
- Destination Field** list:
  - journal
  - medline
  - seqcheck
  - nuc** (selected)
  - nuc\_term
  - id
  - tmp
  - name
- Use Tags** ☐ | **Default Tag** [ ] | **Tag to Modify** [ ] | **HELP TAGS**
- Double Pars** ☐ (Experts only, read help text carefully)
- Command ('a=b' or 'ACI')** text input: "|sequence|count("agctuAGCTU")"
- Or select a predefined program:** list:
  - remove all ''
  - copy (full\_name) to this field
  - append (full\_name) to this field
  - calculate sequence checksum
  - count ambiguities (nry. ...) (filter: ECOLI)
  - truncate seq. ranges outside ecol range
  - increase G+C content in helical regions by 5%
  - SIMPLPRO Simplified AminoAcid
  - Count nucleotides** (selected)
  - Count nucleotides (incl. IUPAC)
- GO** button











# Primary & Secondary Structure Editors

The screenshot displays the ARB Edit4 software interface, titled "ARB\_EDIT4 \*1\*". The menu bar includes File, Create, Edit, View, Block, Properties, and HELP. The toolbar contains icons for Quit, Undo, Jump, Help, Redo, Get, Position, E.coli, Base, IUPAC, Helix #, Protect, and a search icon. The status bar at the top right indicates "This is ARB Edit4 (Build 03\_01\_21)".

The main window shows a sequence alignment of 16S rRNA genes. The alignment is displayed in a table format with columns for sequence identifiers and sequence data. The sequences are grouped into three main sections:

- SAI:tail (usually)**: Contains sequences SAI:HELIX and SAI:HELIX\_NB.
- SAI:Maingroup (3)**: Contains sequences SAI:SAI's (3) and SAI:POS\_VAR\_BY\_PARSIMONY.
- New Sequences (16)**: A list of bacterial species names, each followed by a sequence identifier (e.g., STREPTOMYCES GRISEUS, STREPTOMYCES RIMOSUS, CORYNEBACTERIUM AQUATICUM, etc.).

The sequence data is displayed in a grid format, with each row representing a sequence and each column representing a position in the sequence. The sequences are color-coded to show matches and mismatches. The alignment is displayed in a table format with columns for sequence identifiers and sequence data.



# Primary & Secondary Structure Editors

The image displays two windows from the ARB (Arbitrary Requester) software, used for editing RNA sequences and their secondary structures.

**Left Window: ARB\_EDIT4 "1"**  
This window shows the primary structure editor. It includes a menu bar (File, Create, Edit, View, Block, Properties) and a toolbar with icons for various editing functions. Below the toolbar is a list of sequences and their properties:

Sequence Name	Properties
SAT:cali (default)	0da
SAT:HELIX	0da
SAT:HELIX_NB	0da
SAT:Malegroup (3)	-CO TTTTTTTTT
SAT:SAI'S (3)	-CO TTTTTTTTT
SAT:POS_VAR_BY_PARSIMONY	0da
New Sequences (0)	-CO TTTTTTTTT
STREPTOMYCES GRISEUS	0da
STREPTOMYCES RIMOSUS	0da
CORYNEBACTERIUM AQUATICUM	0da
AUREOBACTERIUM TESTACEUM	0da
CURTIBACTERIUM CITREUM	0da
SAPROSPIRA GRANDIS	0da
BREVIBACTERIUM LINENS	0da
MICROCOCCUS LUTEUS	0da
CELLULOMONAS BIAZOTEA	0da
RHODOCOCCLUS ERYTHROPOLIS	0da
ARTHROBACTER POLYCHROMOGENES	0da
ARTHROBACTER GLOBIFORMIS	0da
ARTHROBACTER GLOBIFORMIS	0da
ARTHROBACTER OXYDAN	0da
BREVIBACTERIUM HELVOLIUM	0da
MICROCOCCUS LYSODEIKTICUS	0da
RHODOCOCCLUS FASCIANS	0da
MYCOBACTERIUM SMEGMATIS	0da
MYCOBACTERIUM PHLEI	0da
MYCOBACTERIUM LEPRAE	0da
MYCOBACTERIUM BOVIS	0da
MYCOBACTERIUM FLAVESCENS	0da
ARTHROBACTER LUTEUS	0da
PINELABACTER SIMPLEX	0da

**Right Window: ARB: SECONDARY STRUCTURE EDITOR WINDOW**  
This window shows the secondary structure editor. It includes a menu bar (File, Properties) and a toolbar with icons for various editing functions. The main display area shows a complex RNA secondary structure diagram with various loops, stems, and bulges. The diagram is color-coded (blue, red, green) and includes numerical labels (e.g., 2700, 2580, 2500, 2450, 2400, 2100, 2050, 2250, 2150, 2800, 2750, 2600, 2550, 2400, 2350, 2300, 2250, 2200, 2150, 2100, 2050, 2000, 1950, 1900, 1850, 1800, 1750, 1700, 1650, 1600, 1550, 1500, 1450, 1400, 1350, 1300, 1250, 1200, 1150, 1100, 1050, 1000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, 100, 50, 0) indicating sequence positions. The diagram is set against a black background with white and colored lines representing the RNA backbone and base pairs.



# Primary & Secondary Structure Editors

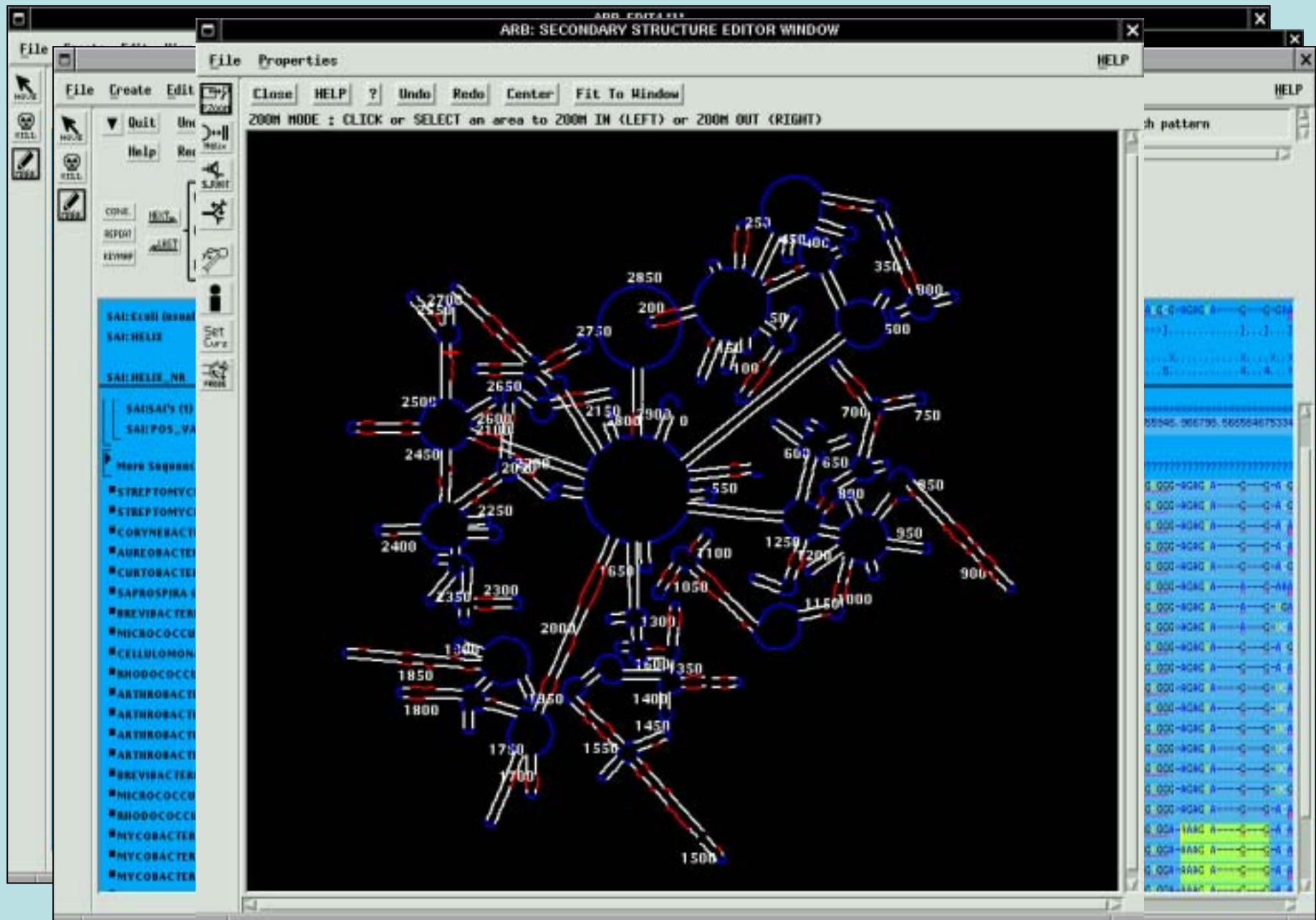
The screenshot displays the ARB Secondary Structure Editor window, titled "ARB\_EDIT4 '1'". The interface includes a menu bar (File, Create, Edit, View, Block, Properties), a toolbar with icons for navigation and editing, and a status bar at the bottom.

The main editing area is divided into several sections:

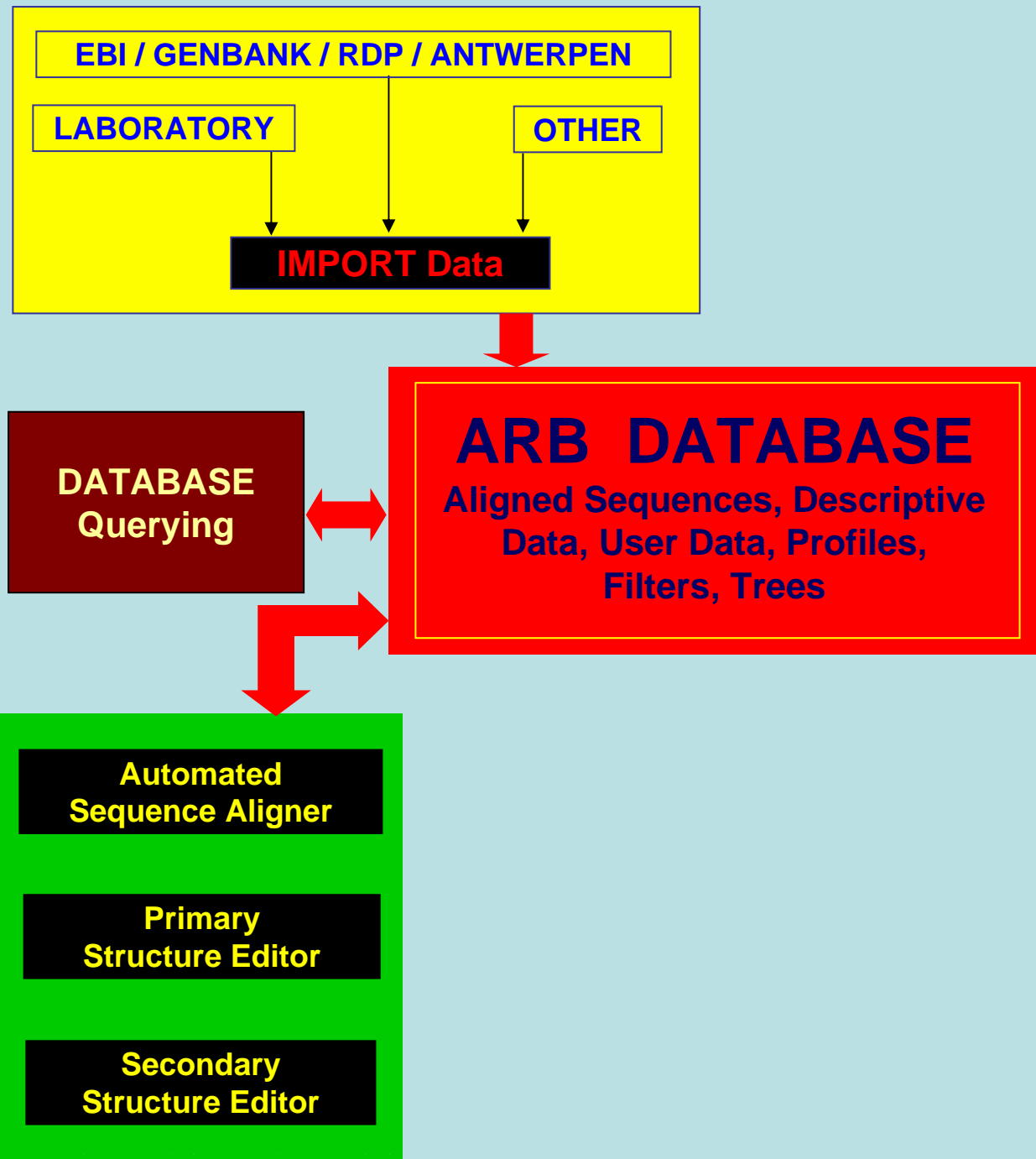
- Sequence Information:** Displays the current sequence name "ARB\_EDIT4 '1'", position "132", and other parameters like "99", "97", "B", "D", "0", "Align", "Insert", "Q", and a search pattern "1: This is the only occurrence of the search pattern".
- Primer and Signature Settings:** A table with columns for "Primer (local)", "Primer (region)", "Primer (global)", "Signature (local)", "Signature (region)", and "Signature (global)". Each column has checkboxes for "LAST", "NEXT", and "ALL" settings.
- Sequence Alignment:** A table showing sequence alignment for "User1", "User2", and "Probe". The sequences are: User1: CCGGUNCUUCGGA, User2: GAUGGUNCUGACCCG, and Probe: CU-A-GCC-UUG-CC.
- Sequence List:** A list of sequences under the heading "More Sequences ID". The list includes: SAISAI's (1), SAIP05\_VAR\_BY\_PARSIMONY, STREPTOMYCES GRISEUS, STREPTOMYCES RIMOSUS, CORYNEBACTERIUM AQUATICUM, AUROBACTERIUM TESTACEUM, CURTOBACTERIUM CITREUM, SAPROSPIRA GRANDIS, BREVIBACTERIUM LINENS, MICROCOCCUS LUTEUS, CELLULOMONAS BIAZOTEA, RHODOCOCCLUS ERYTHROPOLIS, ARTHROBACTER POLYCHROMOGENES, ARTHROBACTER GLOBIFORMIS, ARTHROBACTER GLOBIFORMIS, ARTHROBACTER OXYDAN, BREVIBACTERIUM HELVOLIUM, MICROCOCCUS LYSODEIKTICUS, RHODOCOCCLUS FASCIANUS, MYCOBACTERIUM SMEGMATIS, MYCOBACTERIUM PHLEI, and MYCOBACTERIUM LEPRAE.



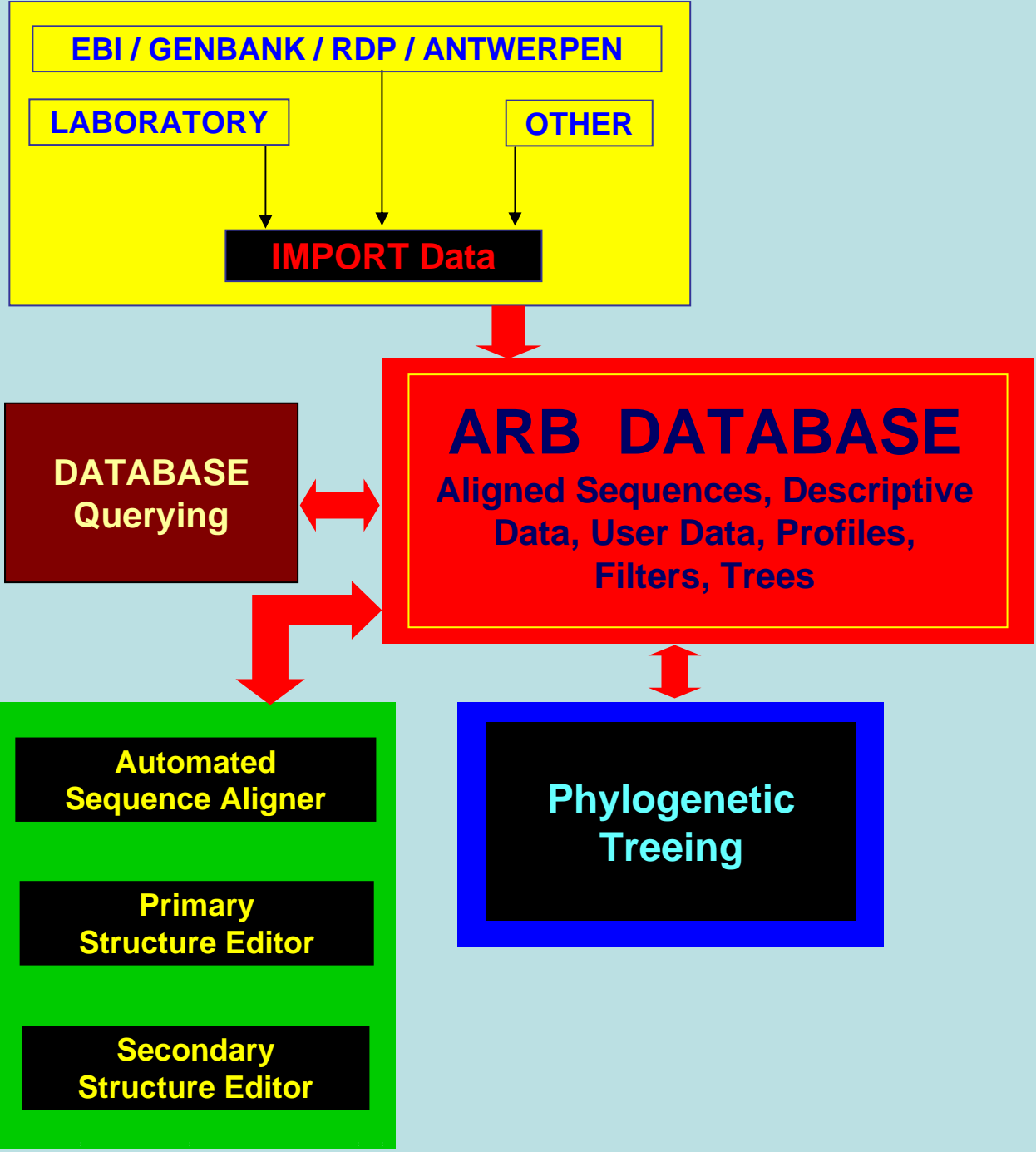
## Primary & Secondary Structure Editors



















## 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

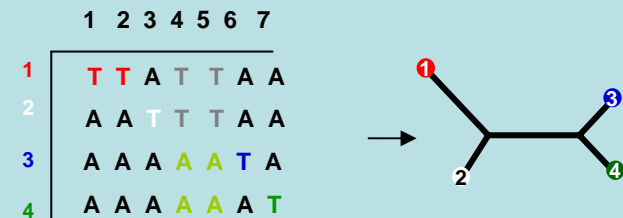
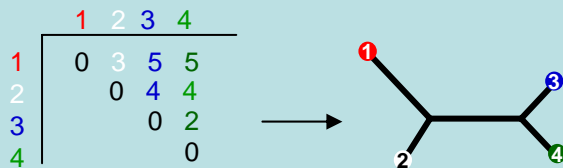
FastDNAmI Method

Distance Based

Character Based

Phylip Distance Matrix Method

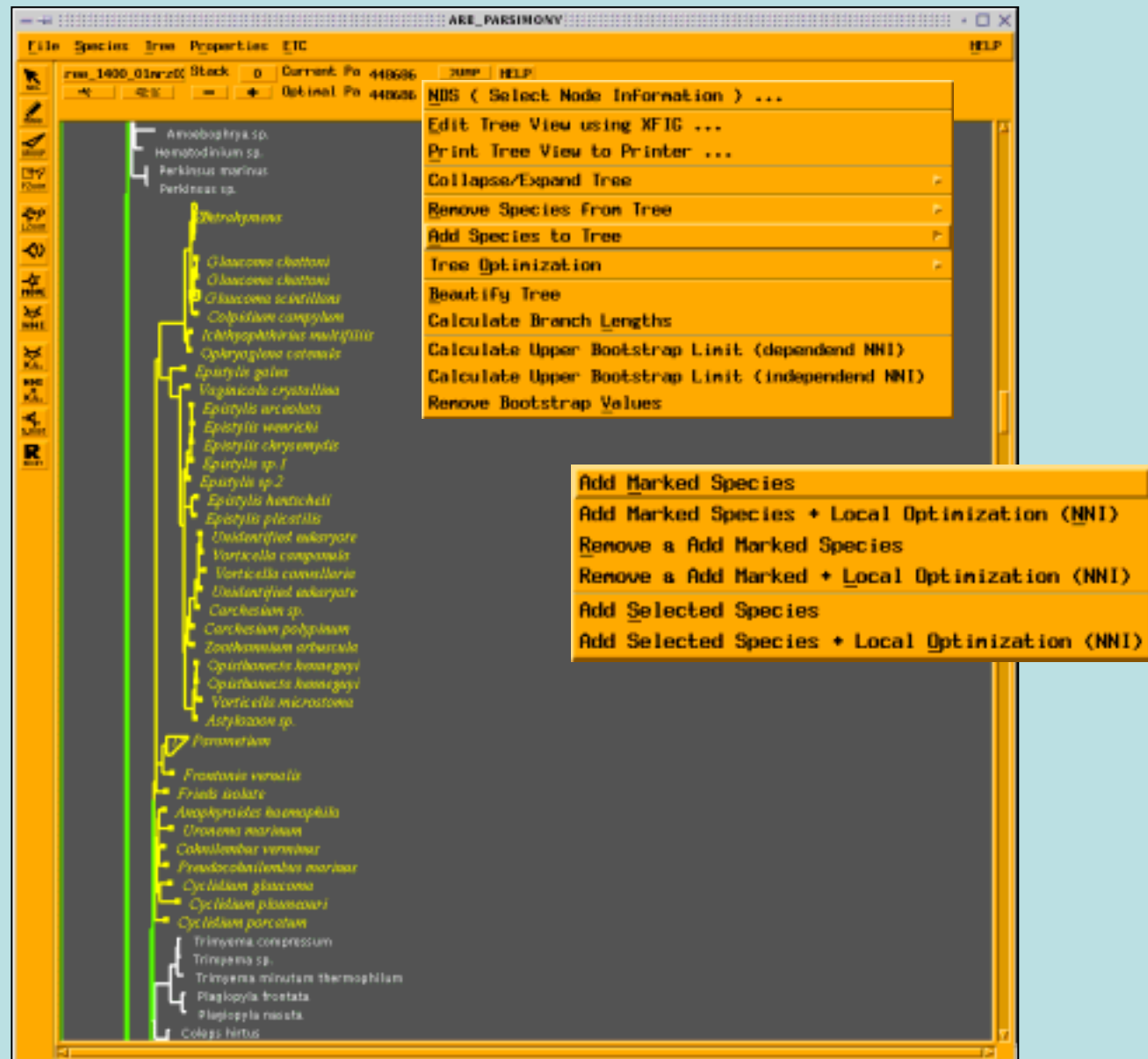
Parsimony Method





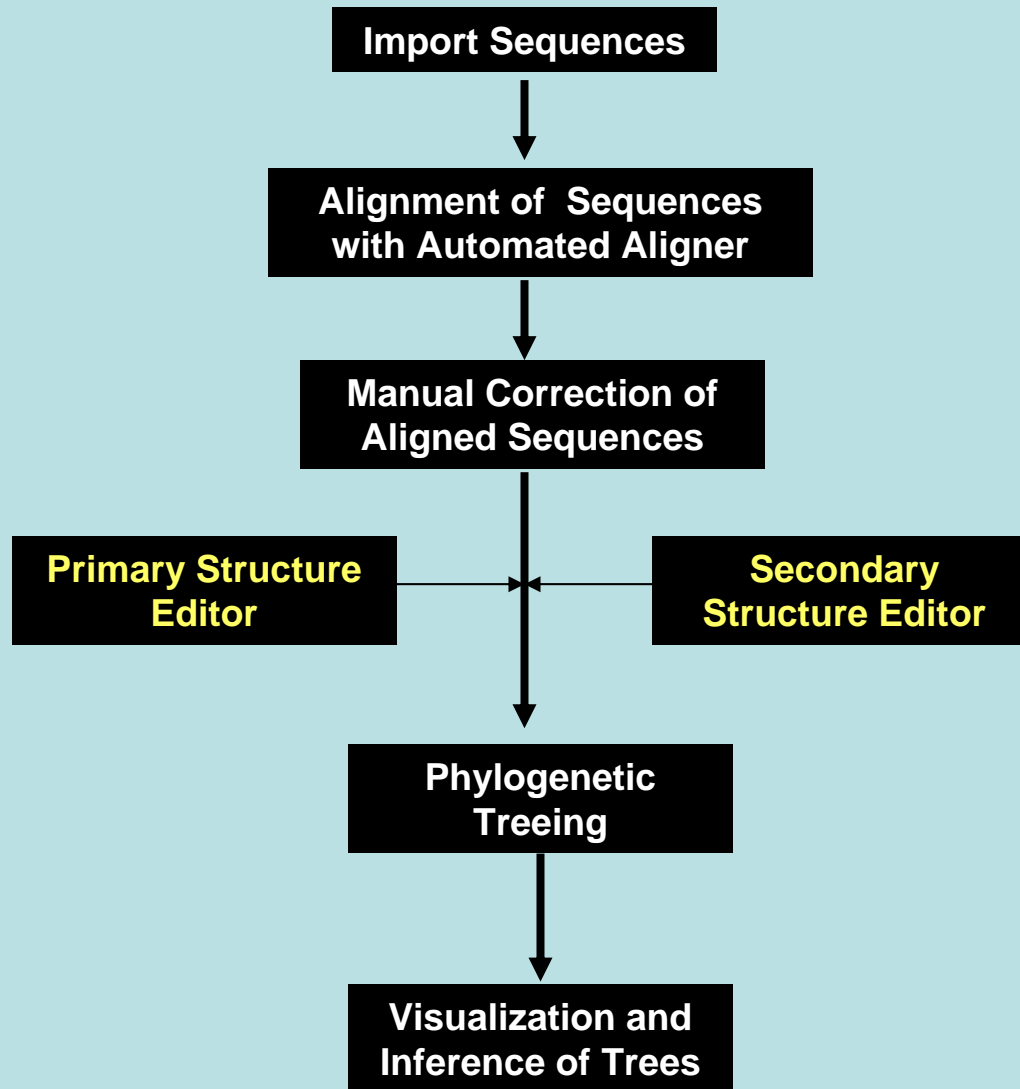
# The ARB Parsimony Tool

- ❑ 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.

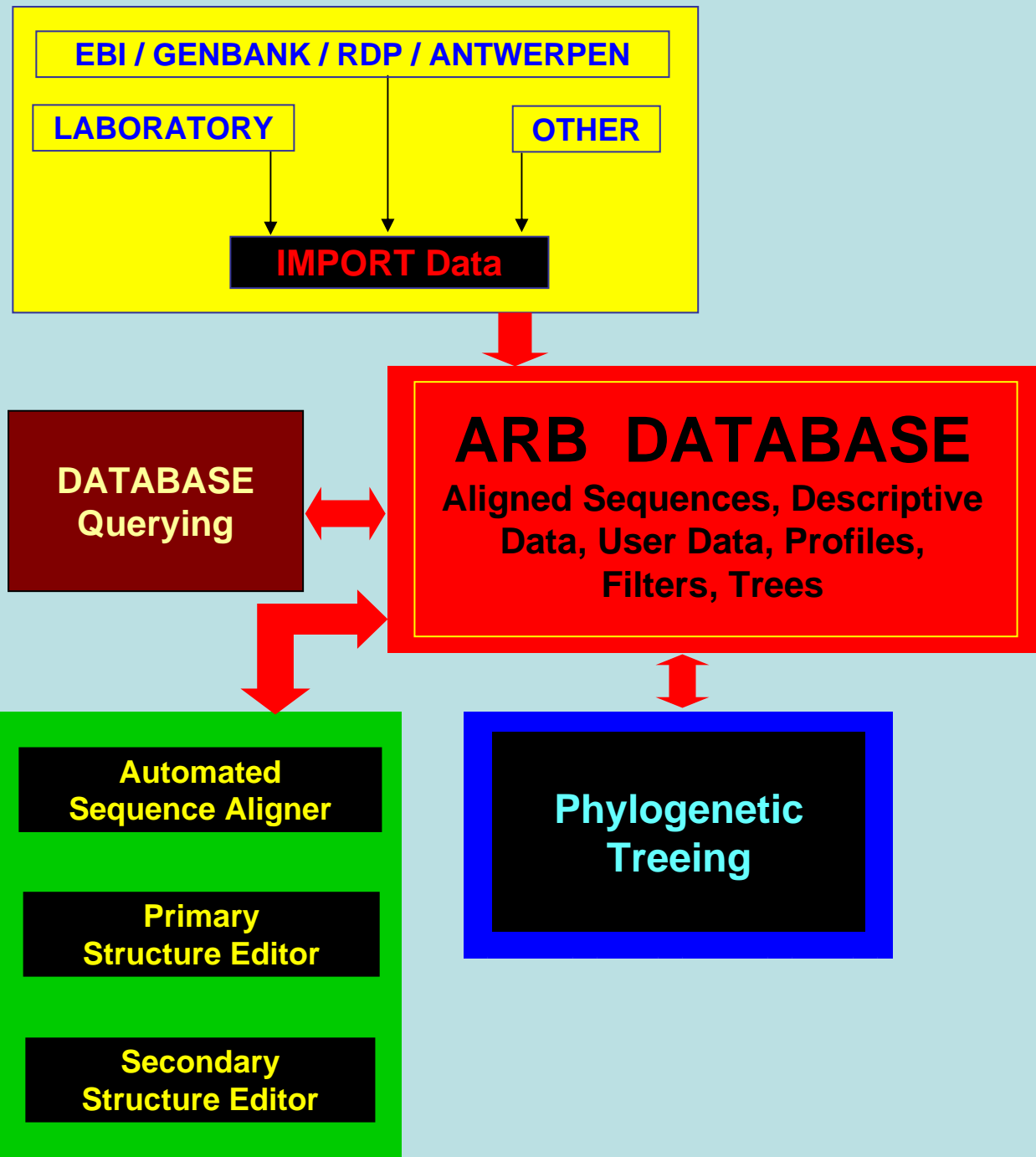




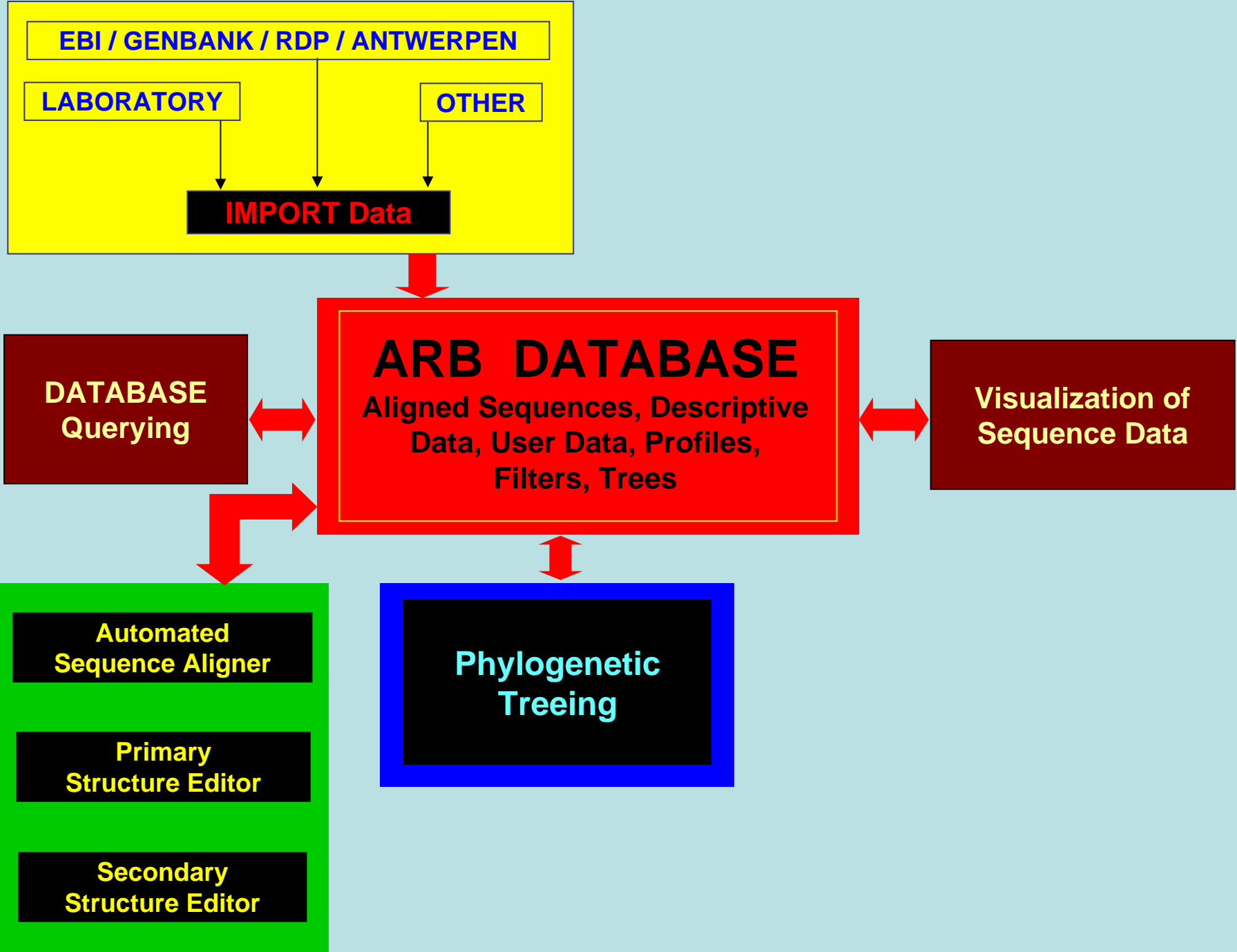
# Phylogenetic Treeing using ARB Software





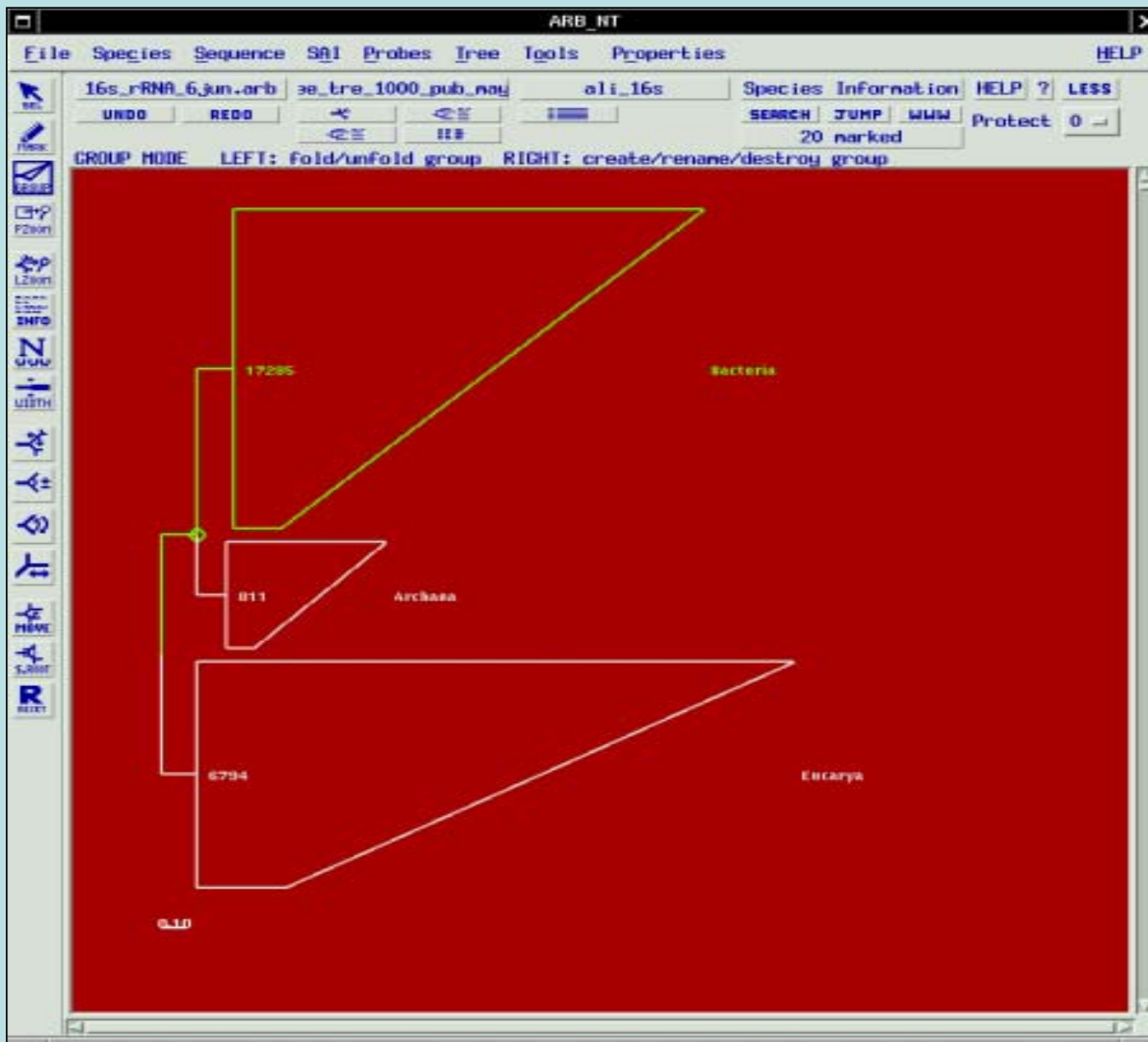






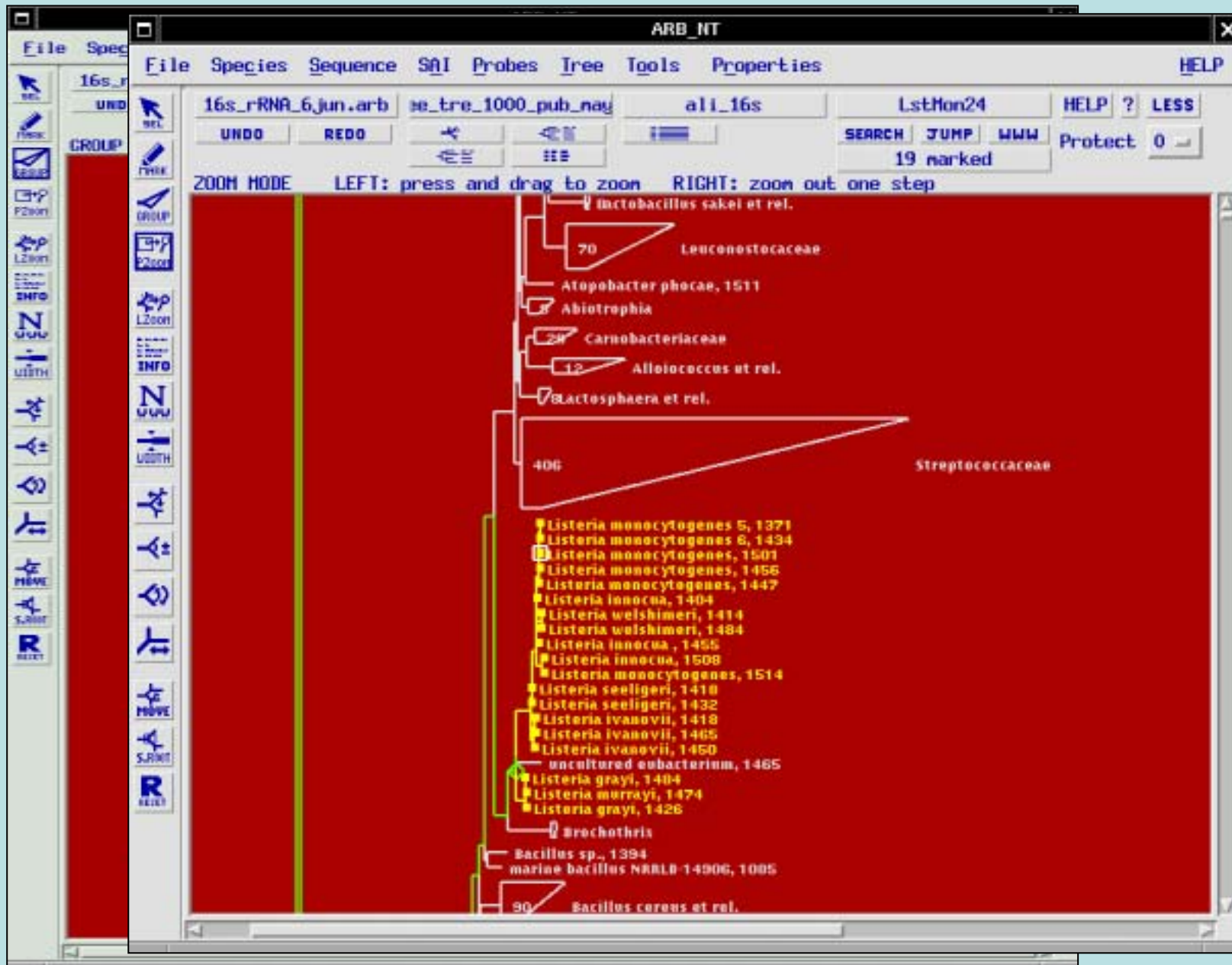


# Visualization of Sequence Data



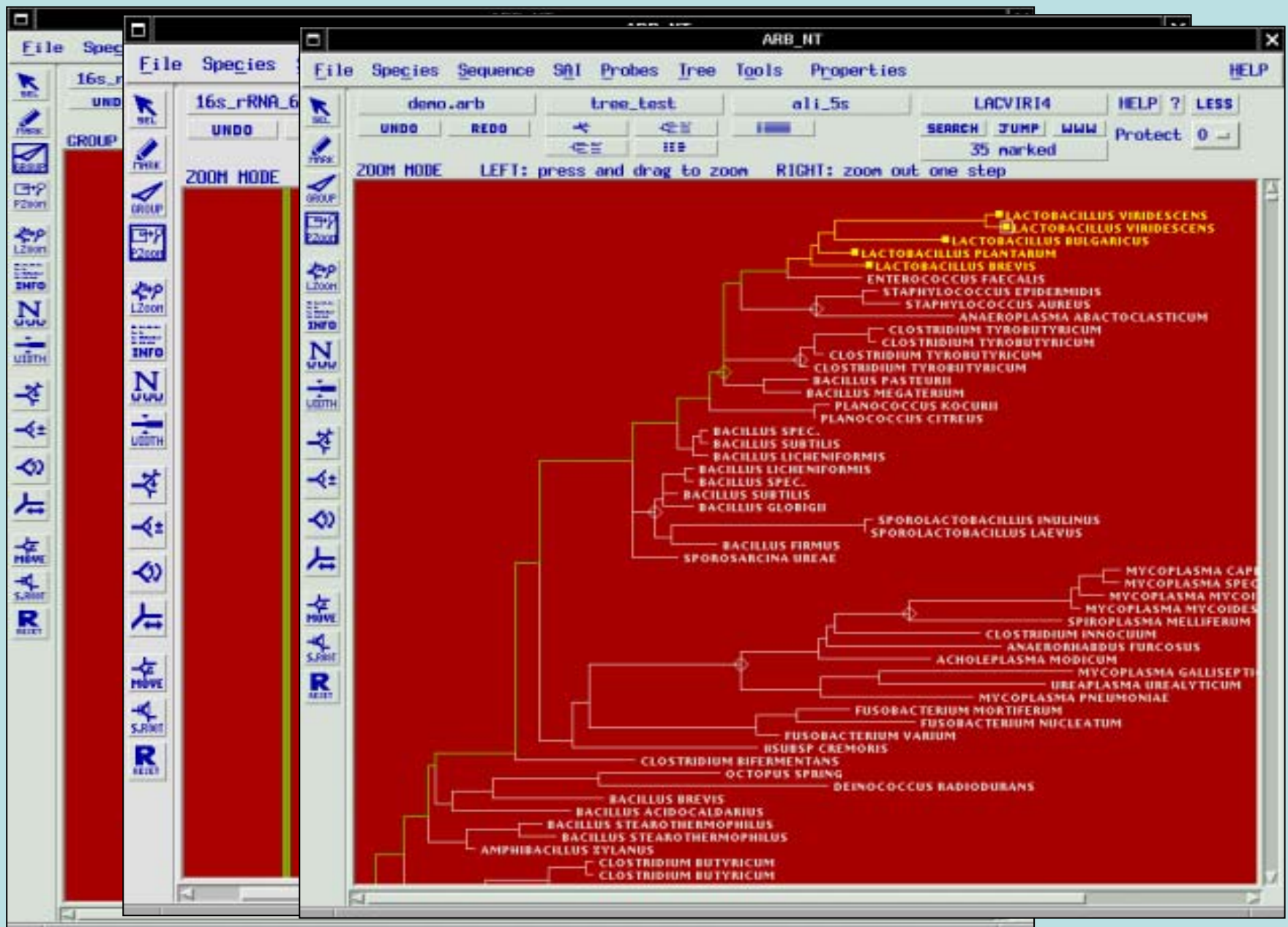


# Visualization of Sequence Data





# Visualization of Sequence Data



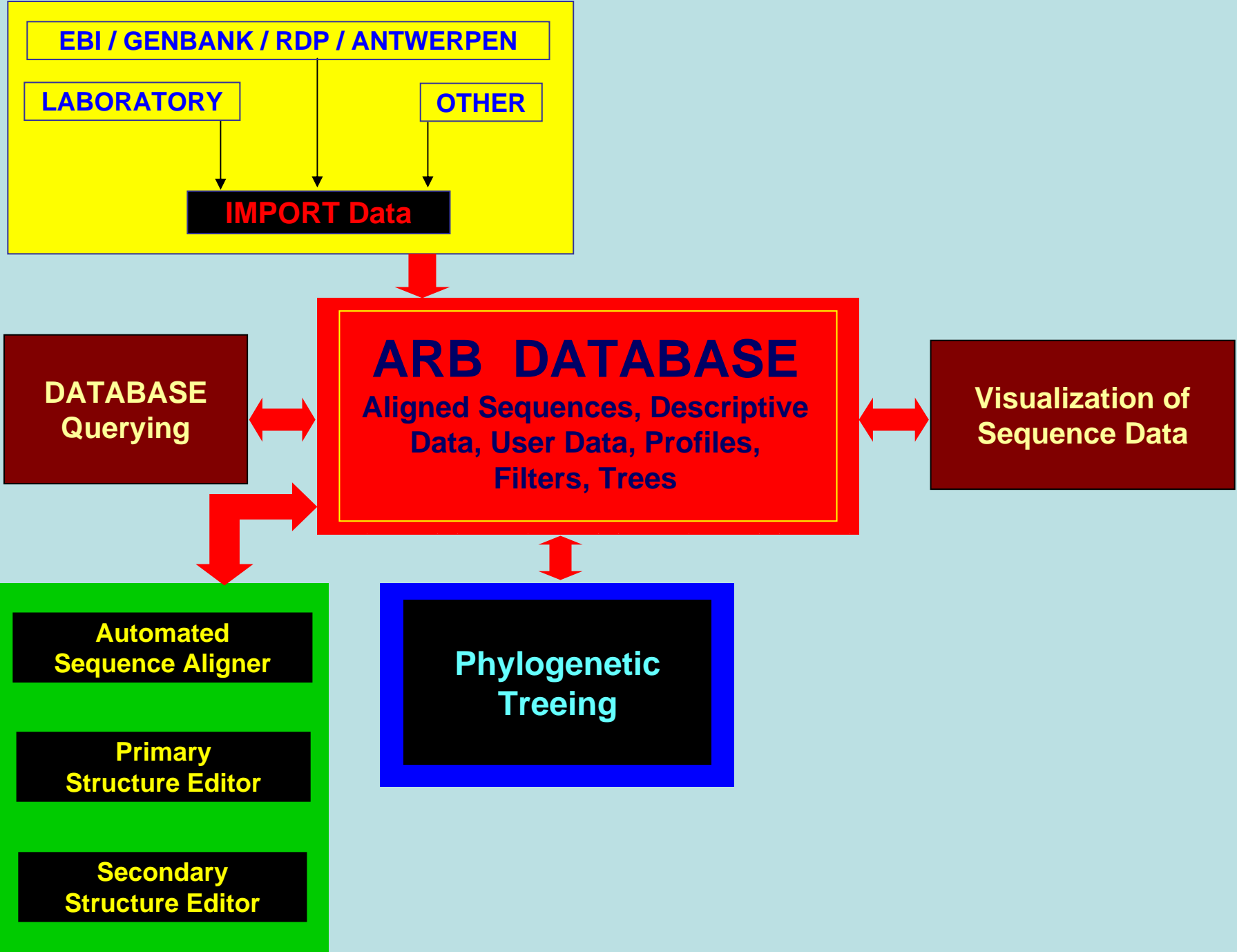


# Visualization of Sequence Data

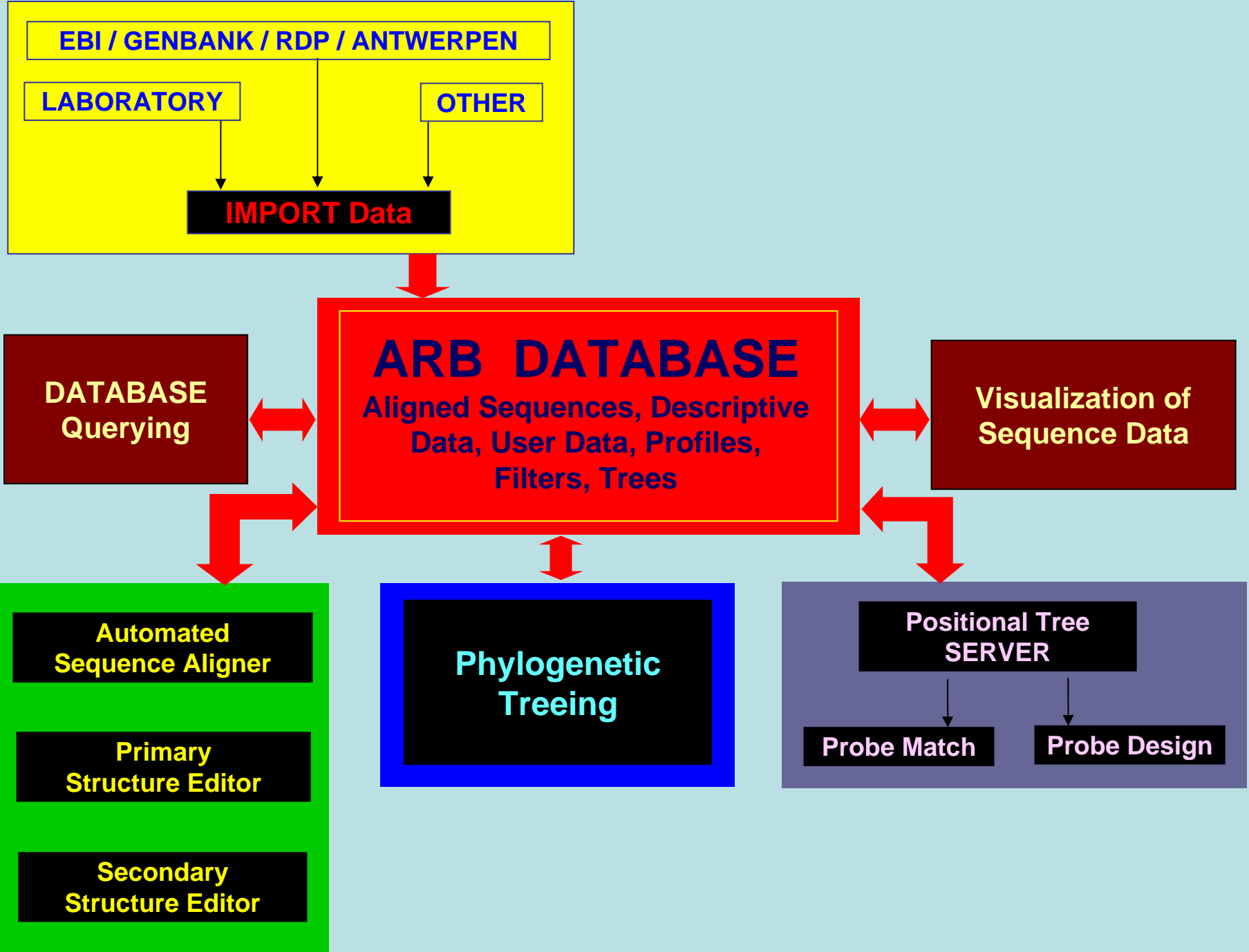
The screenshot displays a bioinformatics software interface with multiple windows. On the left, a vertical toolbar contains icons for various functions: selection (arrow), zoom (magnifying glass), pan (hand), and other navigation tools. The main window is titled 'ARB\_NT' and contains a 'SPECIES INFORMATION' panel. This panel has a 'SPECIES FIELDS' section with buttons for 'CLOSE', 'DETACH', 'SEARCH', and 'HELP'. Below these buttons, there are checkboxes for 'Edit enabled?' and 'Marked?'. The 'Edit box' contains the text: '[DEV] Salmonella typhi U88545 [EBI,RDP] Salmonella typhi'. The 'DATABASE FIELDS' section lists various fields and their corresponding values for the selected sequence.

Field	Value
full_name	S0: Salmonella typhi
strain	S0: [DEV;EBI] St111 [RDP] St111. Former RDP Sids: U88545 Corresponding GenBank entry: U88545
next_rel	S0: 384 'Saltyp47'
lib_name	S0: [DEV] Salmonella typhi U88545 [EBI,RDP] Salmonella typhi
acc	S0: U88545
db_acc	S0: [DEV;EBI;RDP] U88545
version	S0: [EBI] U88545.1
aligned	S0: 1apr90UL 12feb00UL 10dec00UL 21jul01UL 28dec01UL
author	S0: [DEV] Subramanian G. [EBI] # Subramanian G. [RDP] Subramanian G.
title	S0: [EBI] (1):
journal	S0: [DEV;EBI] Submitted (05-FEB-1997) to the EMBL/GenBank/DBJ databases. Institute for
seqcheck	S0: ARB_09049490
nuc	S0: 1541
nuc_tern	S0: 1541
id	S0: [EBI] STU88545 [RDP] S.typhi3
trp	S0: [RDP] up_inFo
name	S6: Saltyp44
ali_16s/data	S5: .....
tax	S0: [DEV] Bacteria Proteobacteria (gamma subdivision) Enterobacteriaceae Salmonella [EBI]
gene	S0: [DEV] SSU [EBI] .
description	S0: [EBI] Salmonella typhi 16S ribosomal RNA gene; complete sequence. [RDP] typhoid Fever
ebi_comment	S0: <A HREF=http://srs6.ebi.ac.uk/srs6bin/cgi-bin/vgetz?e+[EMBL_features-id:STU88545_1]>
nuc_gene_03sep01	S0: 1541



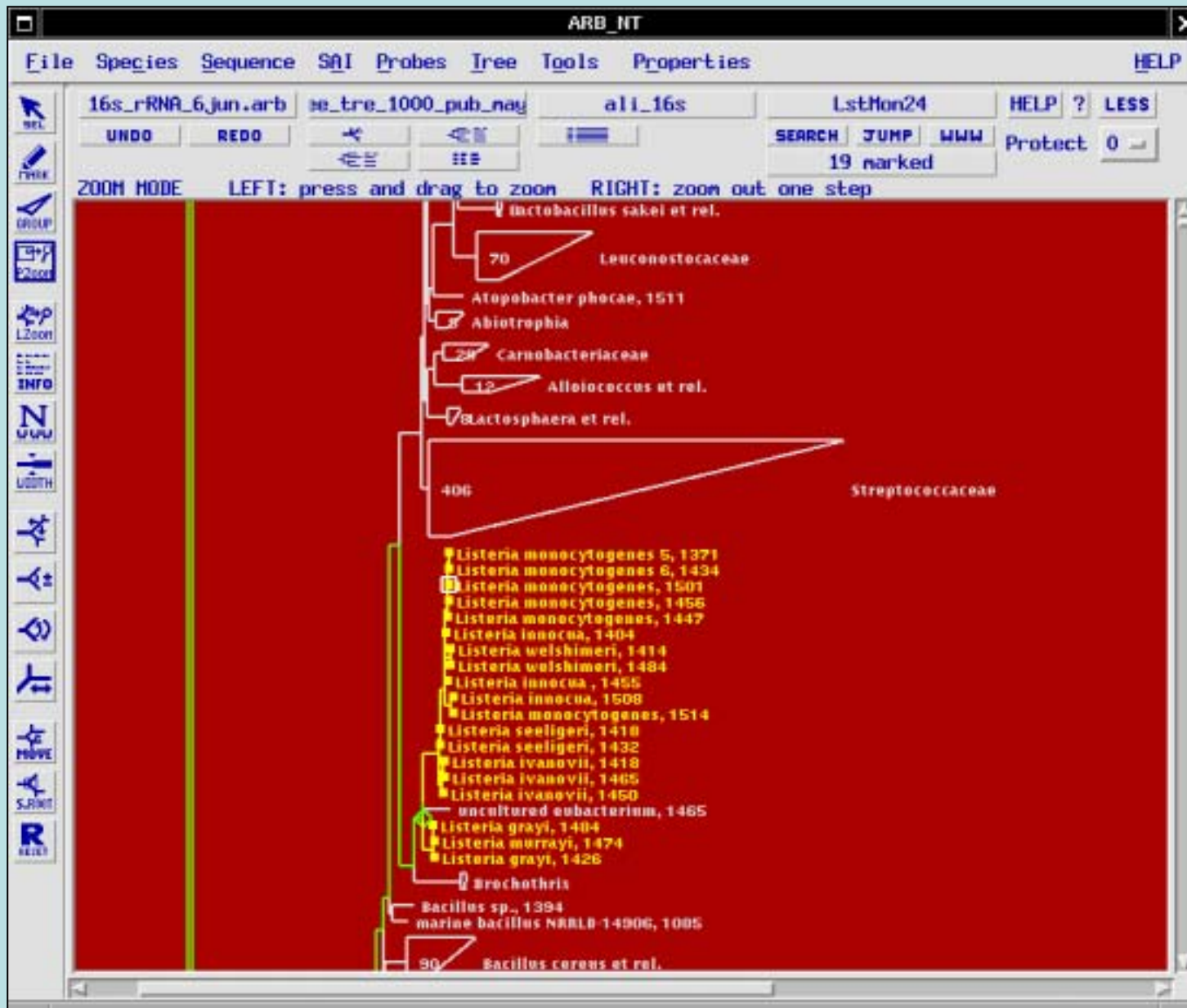








# Probe Design and Probe Match





# Probe Design and Probe Match

The screenshot displays the ARB\_1T software interface. The main window has a menu bar (File, Species, Sequence, SAI, Probes, Tree, Tools, Properties) and a toolbar with various icons. The 'Probes' menu is active, and the 'PROBE DESIGN' sub-window is open. This sub-window contains a 'CLOSE' button, a 'HELP' button, and instructional text: 'This module searches for specific oligonucleotids in the database. The PT\_SERVER's (not the current) database is used searching probe targets'. Below this, the 'PT\_SERVER' is set to 'localhost: LSU\_rRNA.arb'. A section titled 'Enter some parameters (press help to get more information)' contains two columns of input fields with 'MIN' and 'MAX' values. The parameters are: Length of output (50), Max. non group hits (0), Max. hairpin bonds (4), Min group hits (%) (50), Length of probe (18), Temperature (50 to 100), G+C-content (50 to 100), and ECOLI-position (0 to 10000). At the bottom of the sub-window are 'GO', 'RESULT', and 'EXPERT' buttons. The background shows a phylogenetic tree with labels like 'Bacillus sp., 1394' and 'Bacillus cereus et rel.'.

ARB\_1T

File Species Sequence SAI Probes Tree Tools Properties HELP

16s\_rRNA\_6jun.arb se\_tre\_1000\_pub\_nay all\_16s LstMon24 HELP ? LESS

UNDO REDO SEARCH JUMP WWW Protect 0

19 marked

ZOOM MODE LEFT: press and drag to zoom RIGHT: zoom out one step

PROBE DESIGN

CLOSE HELP

This module searches for specific oligonucleotids in the database.  
The PT\_SERVER's (not the current) database is used searching probe targets

PT\_SERVER localhost: LSU\_rRNA.arb

Enter some parameters (press help to get more information)

	MIN	MAX
Length of output	50	
Length of probe	18	
Max. non group hits	0	
Temperature	50	100
Max. hairpin bonds	4	
G+C-content	50	100
Min group hits (%)	50	
ECOLI-position	0	10000

GO RESULT EXPERT

90

Bacillus sp., 1394  
marine bacillus NABLD-14906, 1005  
Bacillus cereus et rel.



# Probe Design and Probe Match

ARB\_IT

PD RESULT

CLOSE SAVE PRINT HELP

Probe design Parameters:

Length of probe 18  
 Temperature [50.0 -100.0]  
 GC-Content [50.0 -100.0]  
 E.Coli Position [ 0 -100000]  
 Max Non Group Hits 0  
 Min Group Hits 50%

Target	le apos	ecol grps	G+C	4GC+2AT	Decrease T by n*.3C	probe matches	n non group species	Probe sequence	
GAACGACGAGGAGAGAC	18 A=10376	63	15	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 3; 5; 7; 9;	1	GUCUCCUCCGUUCGL	
AACGAGGAGAGAGCUUC	18 A+	8	68	15	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 3; 5; 6; 8; 8; 9;	1	GCAAGCUCUCCUCCGL
GAACGAGGAGAGAGCUUC	18 A+	6	67	15	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 4; 5; 5; 5;	1	CAGCUCUCCUCCUCCGL
UCCAGCAGCAGGAGGAG	18 A-	3	61	14	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 2; 2; 2; 2; 2; 4; 5;	1	CUCCUCCGUCCGUUCGL
ACGAGGAGAGAGCUUCGU	18 B=10394	68	15	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 2; 2; 4; 4; 6; 8; 10;	1	AGCAAGCUCUCCUCCGL	
CGGAGGAGAGAGCUUCGUC	18 B+	2	69	15	61.1	58.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 2; 2; 4; 7; 15; 16; 17; 47;	1	GAGCAAGCUCUCCUCCGL
AGUCCAGCAGCAGGAGGA	18 A-	6	59	14	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 2; 4; 29; 30; 36; 82;	1	UCCUCCGUCCGUCCGAC
GUCGACGAGCAGGAGGAA	18 A-	4	60	14	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 2; 4; 5; 8; 9;	1	UCCUCCGUCCGUCCGAC
AACGAGCAGGAGAGAGC	18 A+	1	64	15	55.6	56.0	1 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 2; 2; 2; 2; 4; 4; 5; 10;	1	GCUCUCCUCCGUUCGL

Min group hits (%) 50 ECOLI-position 0 10000

GO RESULT EXPERT

Phylogenetic tree showing relationships between *Bacillus* species and *Archothrix*.



# Probe Design and Probe Match

ARB\_IT

PD RESULT

CLOSE

SAVE

PRINT

HELP

Probe design Parameters:  
Length of probe 18

PROBE MATCH

Target String

GAACGAACCGAGGAAGAG

Use weighted mismatches

☐

PT\_SERVER

localhost: LSU\_rRNA.arb

Check complement too

☒

Search depth

SEARCH UP TO 1 MISMATCHES

Mark in database

☒

Write Result to field 'tmp'

☐

PRINT

EXPERT

MATCH

Number of Hits: 15

Searched for

GAACGAACCGAGGAAGAG

name	fullname	mis	N_mis	wnis	pos	ecoli	rev	'GAACGAACCGAGGAAGAG'
* LstMon24	Listeria monocytogenes	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstMon19	Listeria monocytogenes	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstInn15	Listeria innocua	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstMon23	Listeria monocytogenes 6	0	0	0.0	10376	63	0	AUNCAAGUC-----CUUGCUCUU
* LstSeel3	Listeria seeligeri	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstWels2	Listeria welshimeri	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstIvan2	Listeria ivanovii	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstInn16	Listeria innocua	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstMon21	Listeria monocytogenes	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstInn17	Listeria innocua	0	0	0.0	10376	63	0	.....-CUUGCUCUU
* LstIvano	Listeria ivanovii	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU
* LstIvan3	Listeria ivanovii	0	0	0.0	10376	63	0	AUGCAAGUC-----CUUGCUCUU



# Probe Design and Probe Match

Probe design Parameters:  
Length of probe 18

Target String

PT\_SERVER

Search depth

**MATCH** Number of Hits: 15

Searched for

	name	fullname
*	LstMon24	Listeria monocytogenes
*	LstMon19	Listeria monocytogenes
*	LstInn15	Listeria innocua
*	LstMon23	Listeria monocytogenes
*	LstSee13	Listeria seeligeri
*	LstWels2	Listeria welshimeri
*	LstIvan2	Listeria ivanovii
*	LstInn16	Listeria innocua
*	LstMon21	Listeria monocytogenes
*	LstInn17	Listeria innocua
*	LstIvano	Listeria ivanovii
*	LstIvan3	Listeria ivanovii

ARB: SECONDARY STRUCTURE EDITOR WINDOW

File Properties

Close HELP ? Undo Redo Center Fit To Window

ROTATE MODE LEFT: Rotate SUBhelix RIGHT: rotate strand

2700

2550

2500

2650

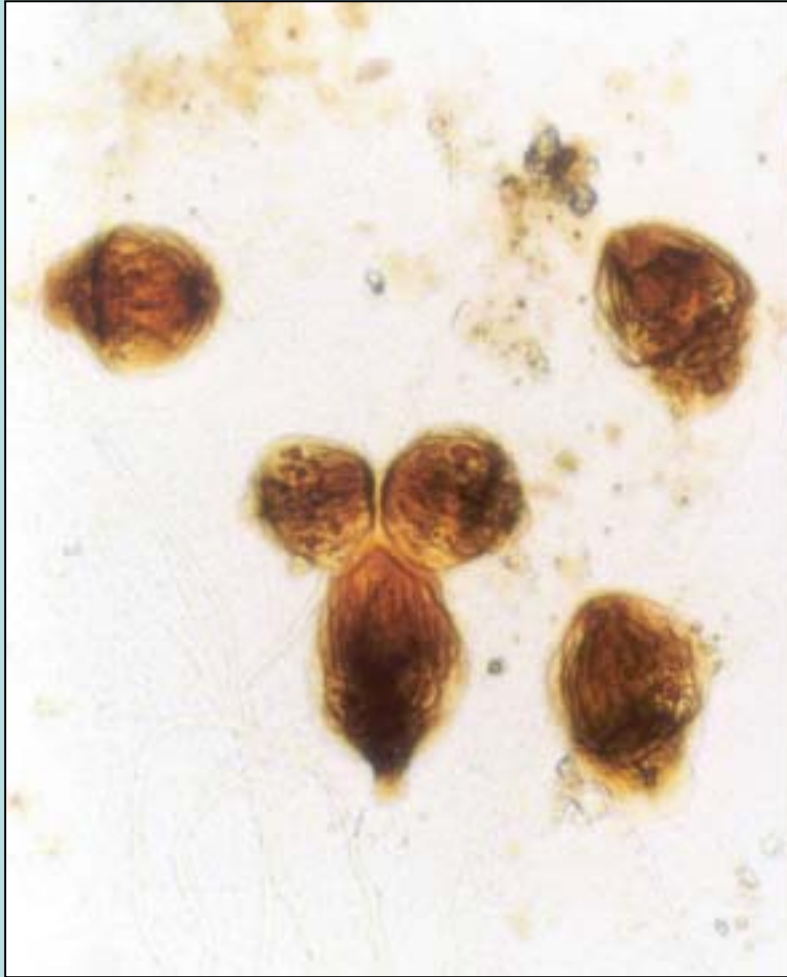
2600

2100

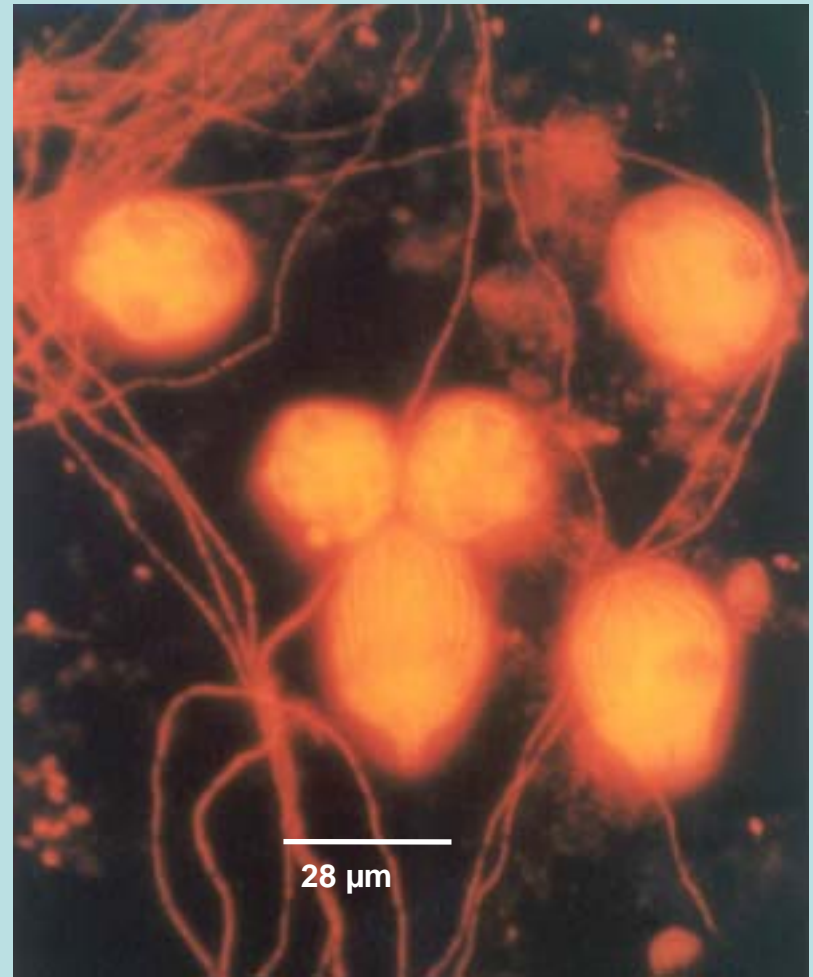
2450



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



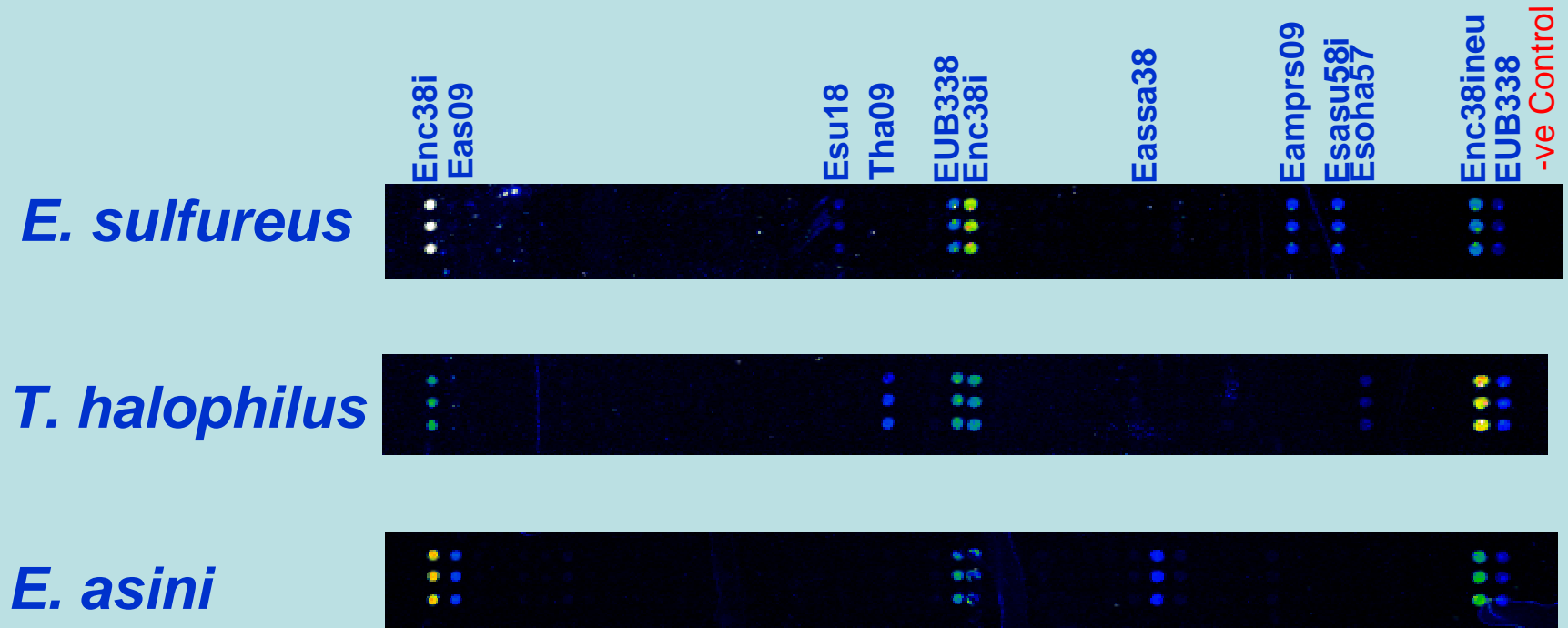
Under Light Microscope



Under Fluorescent Microscope

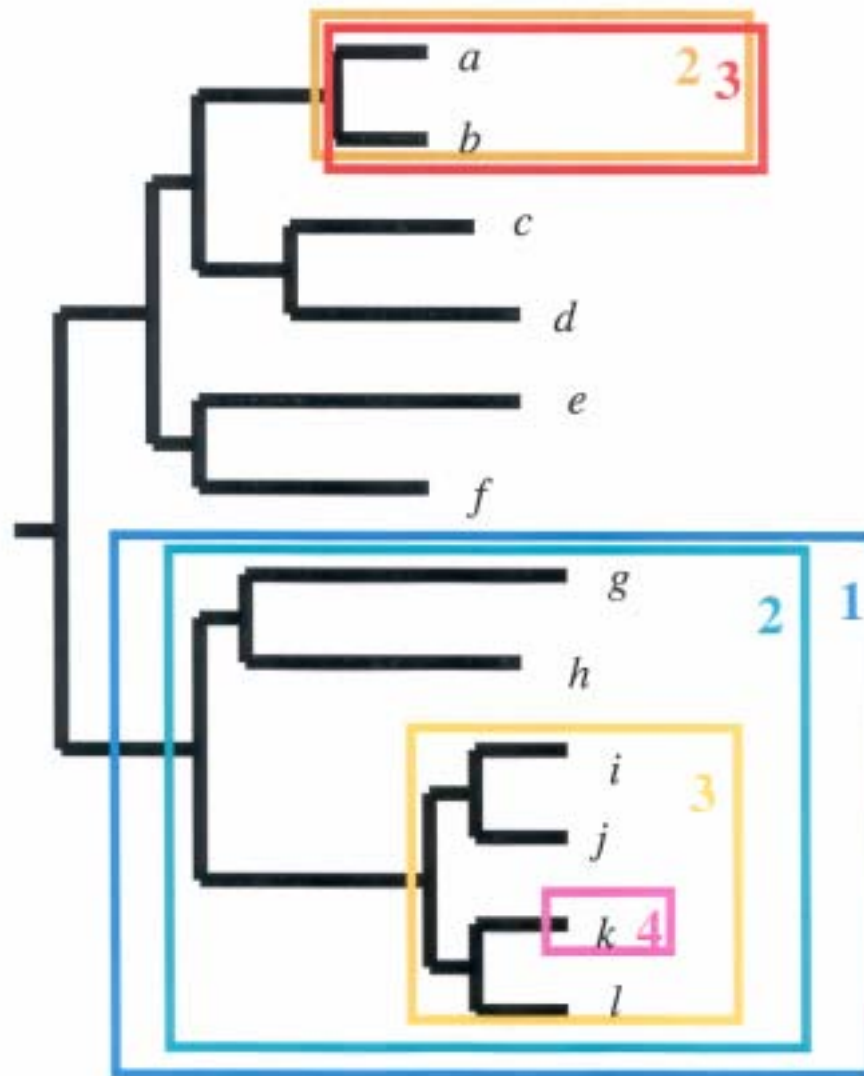


# Hybridised DNA Chips



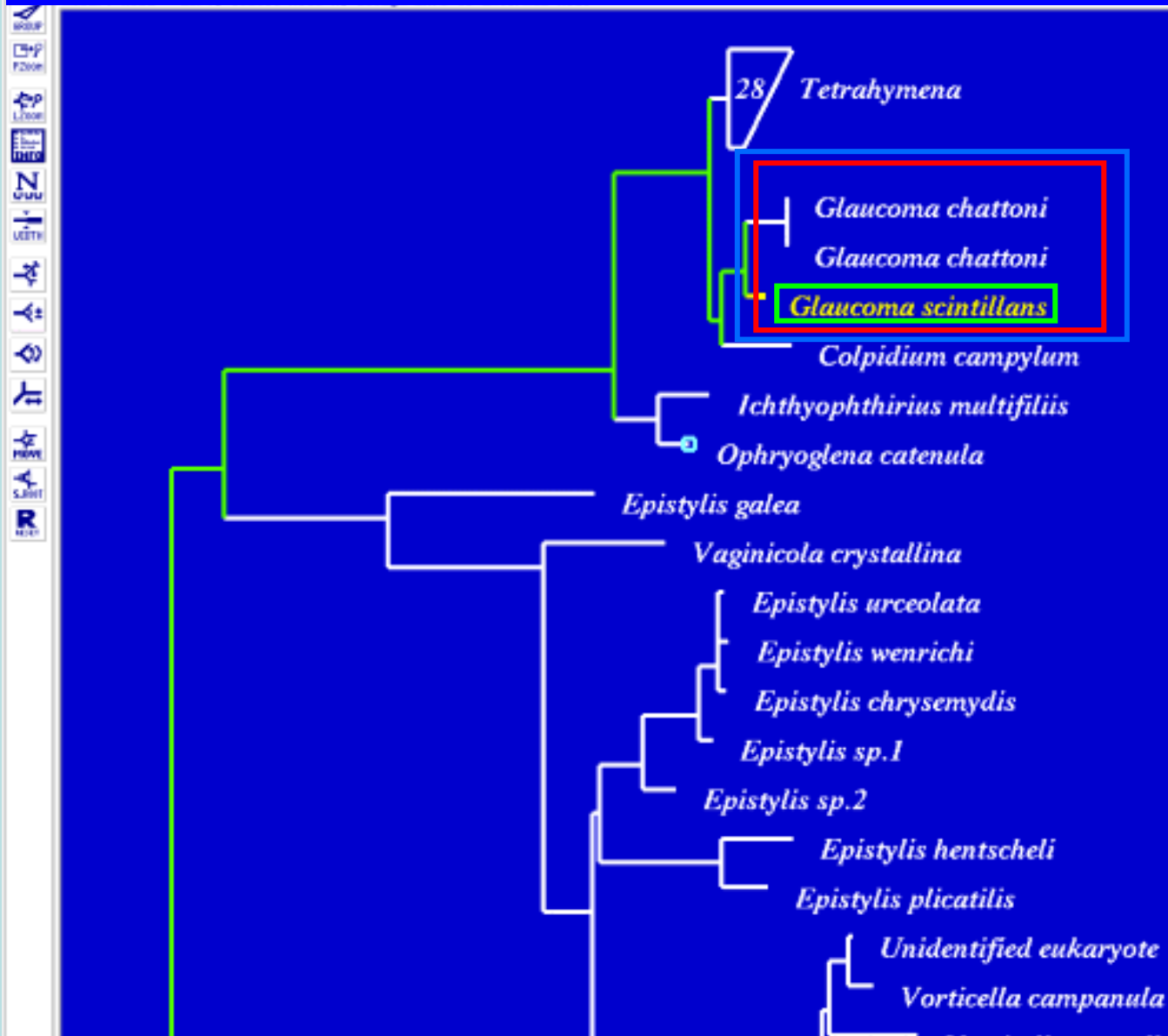


# The Multiple Probe Concept



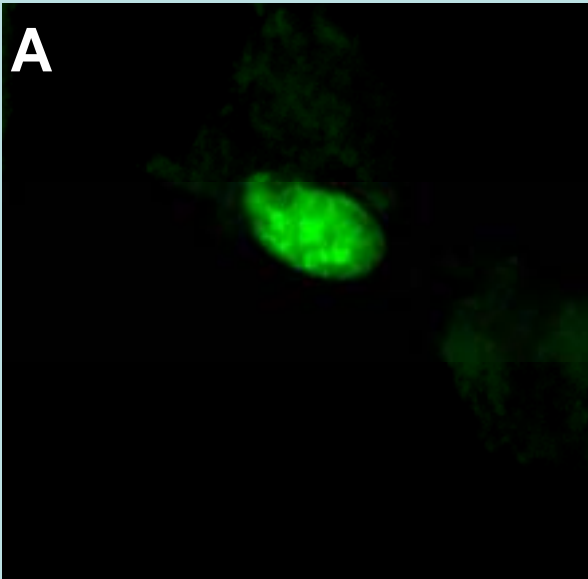


# Design of multiple probes for *Glaucoma scintillans*

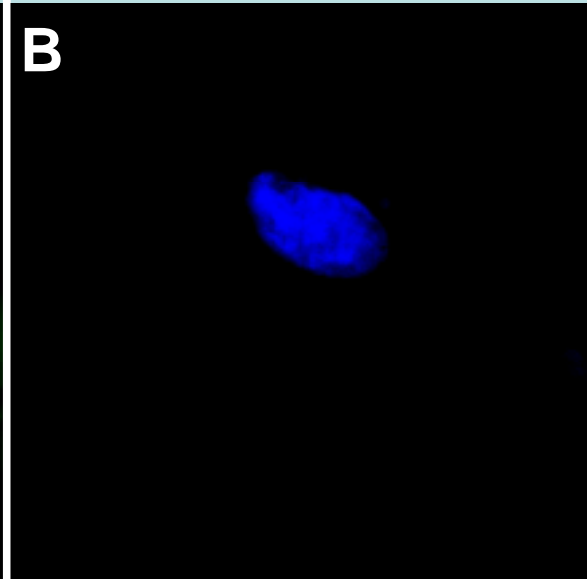




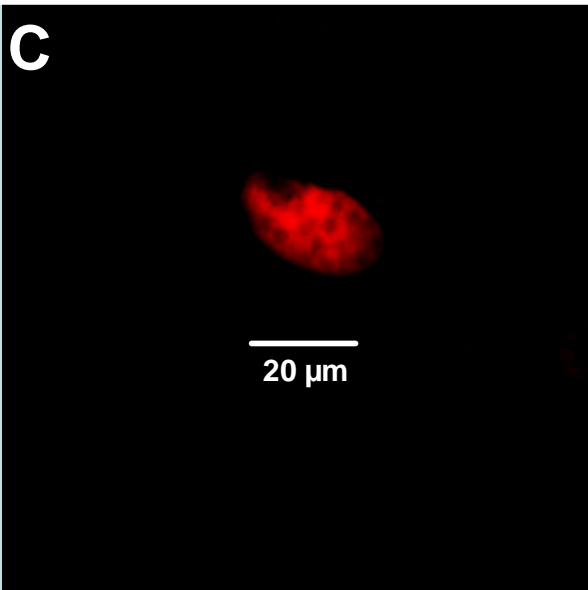
# In situ hybridization of *Glaucoma scintillans* with multiple probes



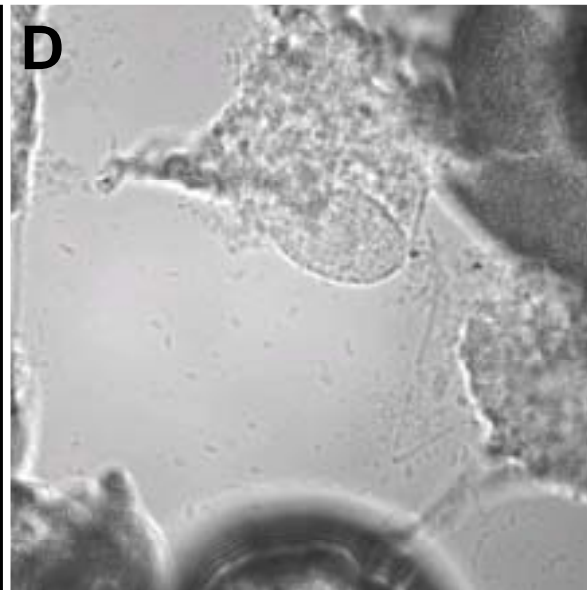
**A : Species  
Specific Probe**



**B : Genus Specific  
Probe 1**

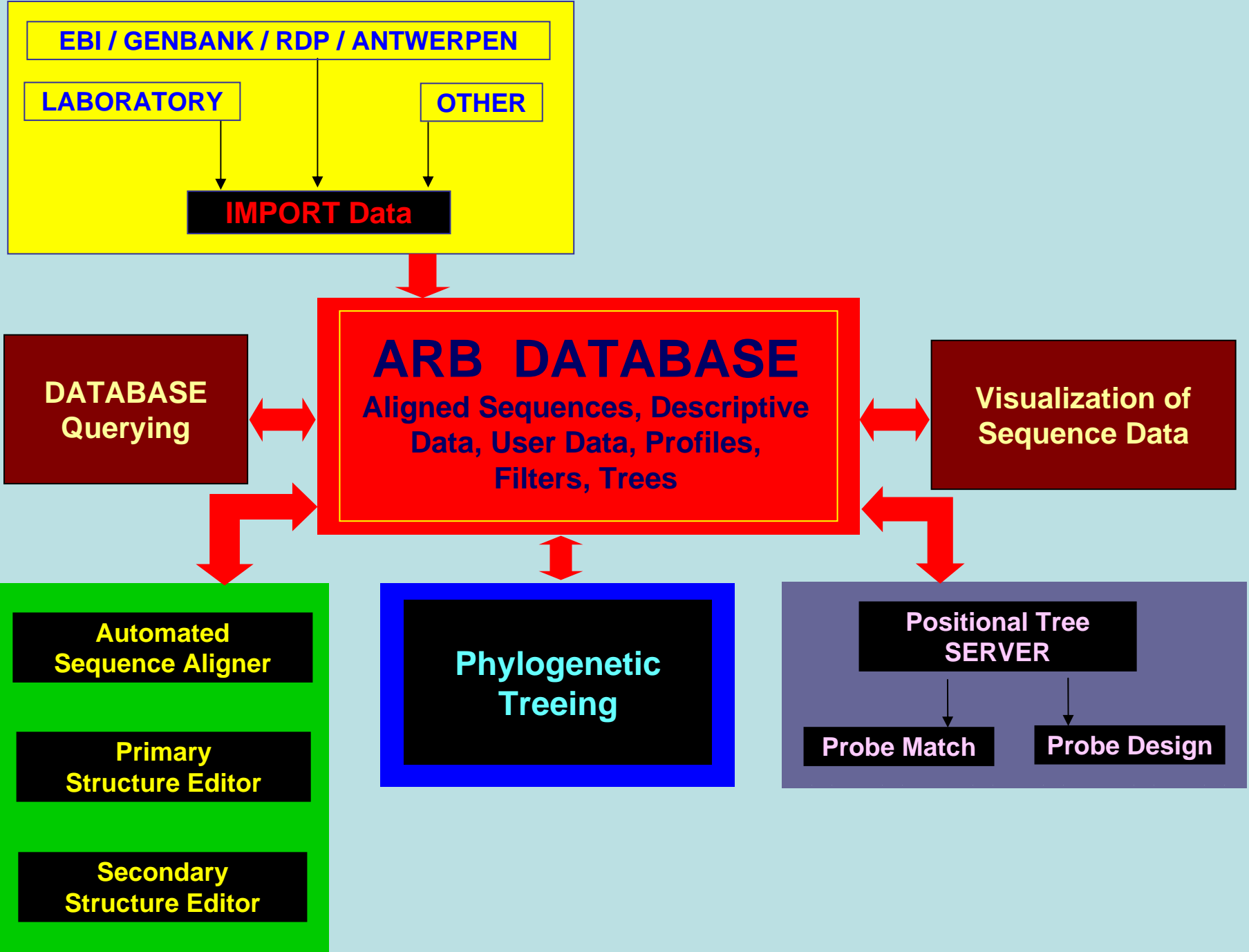


**C : Genus Specific  
Probe 2**

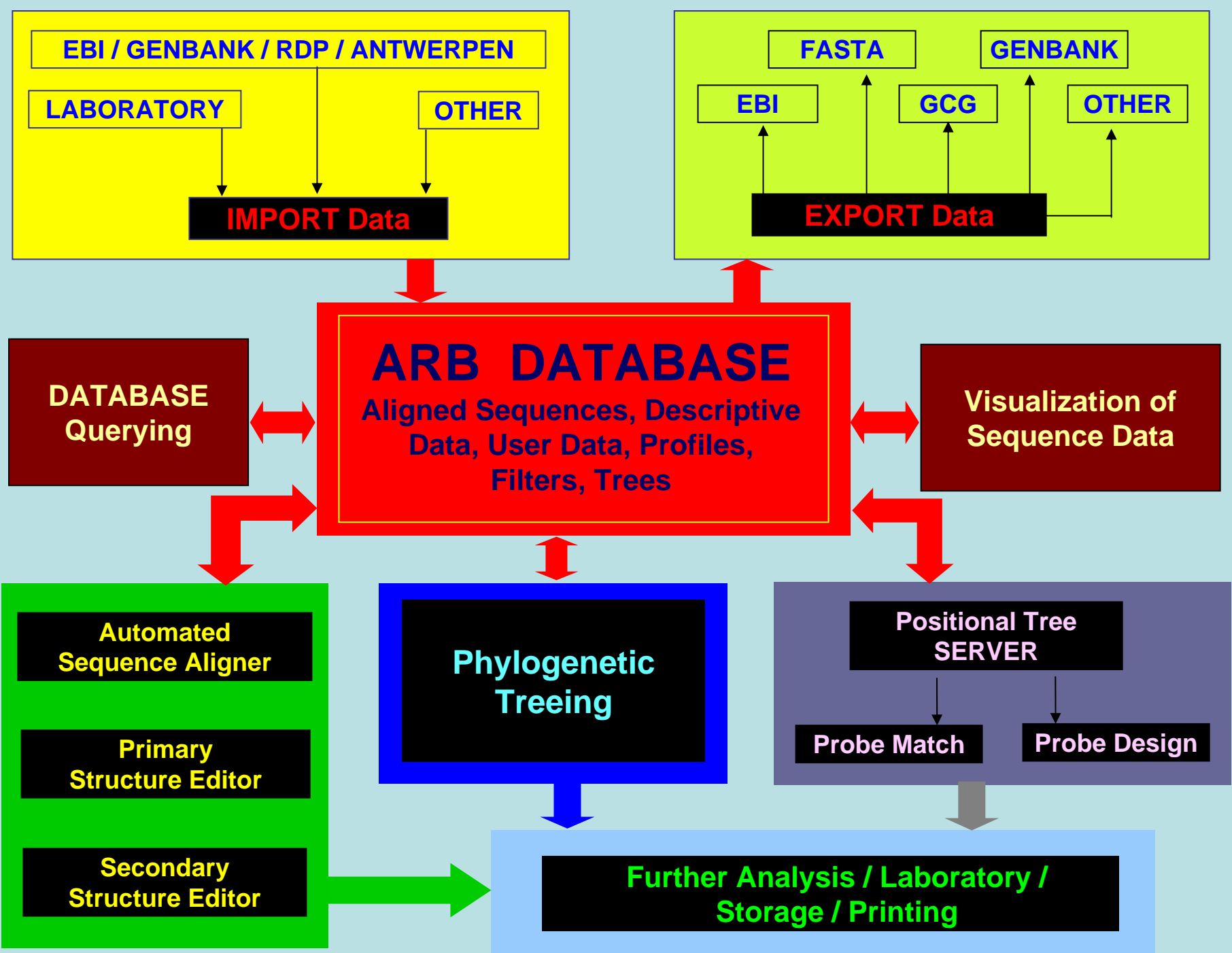


**D : *G. scintillans*  
under light  
microscope**











# The ARB Genome Project

---

**EXPERIMENT**

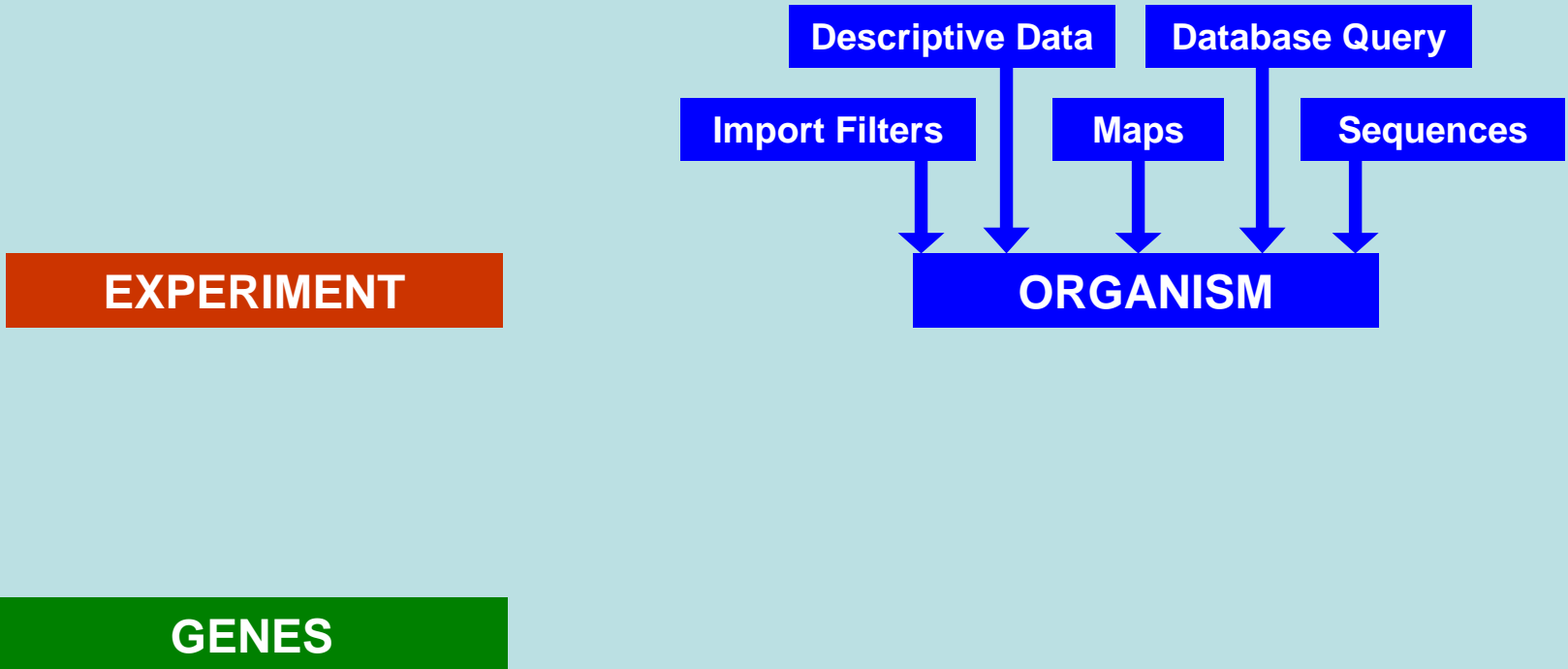
**ORGANISM**

**GENES**



# The ARB Genome Project

---





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

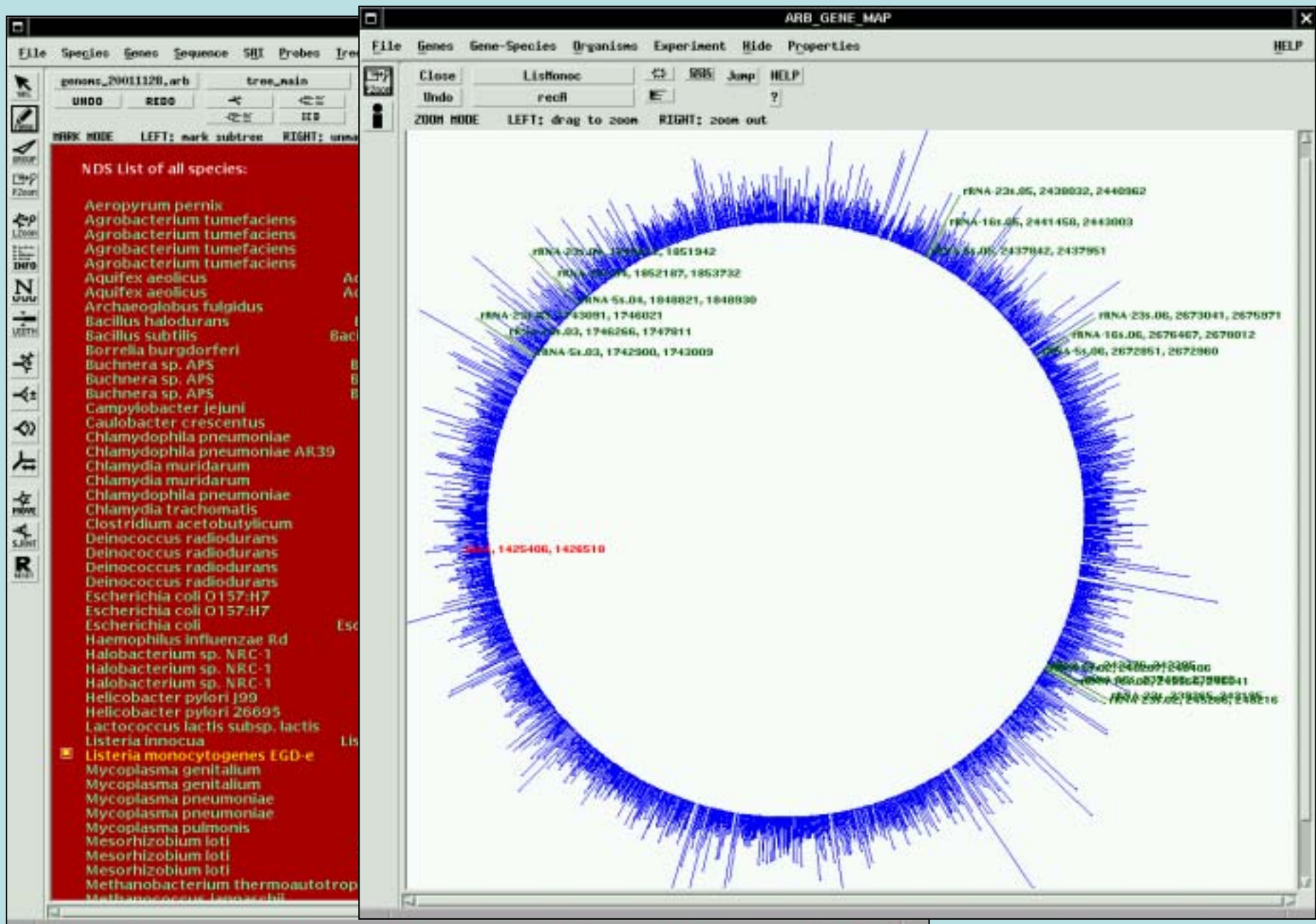
The screenshot displays the ARB NT software window, which is used for managing and viewing genomic data. The interface includes a menu bar (File, Species, Genes, Sequence, SHI, Probes, Tree, Tools, Properties, HELP), a toolbar with various icons for file operations and tree manipulation, and a main display area. The main display area is divided into two panes: a left pane showing a list of species and a right pane showing detailed genomic information for the selected species.

**NDS List of all species:**

Species	Genomic Information
Aeropyrum pernix	Aeropyrum pernix K1 complete genome.
Agrobacterium tumefaciens	Agrobacterium tumefaciens strain C58 linear chromosome, complete genome.
Agrobacterium tumefaciens	Agrobacterium tumefaciens strain C58 plasmid Ti, complete genome.
Agrobacterium tumefaciens	Agrobacterium tumefaciens strain C58 plasmid AT, complete genome.
Agrobacterium tumefaciens	Agrobacterium tumefaciens strain C58 circular chromosome, complete genome.
Aquifex aeolicus	Aquifex aeolicus plasmid ece1, complete plasmid sequence.
Aquifex aeolicus	Aquifex aeolicus complete genome.
Archaeoglobus fulgidus	Archaeoglobus fulgidus complete genome.
Bacillus halodurans	Bacillus halodurans C-125, complete genome.
Bacillus subtilis	Bacillus subtilis complete genome.
Borrelia burgdorferi	Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
Buchnera sp. APS	Buchnera sp. APS plasmid pLeu DNA, complete sequence.
Buchnera sp. APS	Buchnera sp. APS complete genome.
Buchnera sp. APS	Buchnera sp. APS plasmid pTrp DNA, complete sequence.
Campylobacter jejuni	Campylobacter jejuni complete genome.
Caulobacter crescentus	Caulobacter crescentus complete genome.
Chlamydomophila pneumoniae	Chlamydomophila pneumoniae J138 complete genome.
Chlamydomophila pneumoniae AR39	Chlamydomophila pneumoniae AR39, complete genome.
Chlamydia muridarum	Chlamydia muridarum plasmid pMoPn, complete sequence.
Chlamydia muridarum	Chlamydia muridarum, complete genome.
Chlamydomophila pneumoniae	Chlamydia pneumoniae complete genome.
Chlamydia trachomatis	Chlamydia trachomatis complete genome.
Clostridium acetobutylicum	Clostridium acetobutylicum ATCC824 complete genome.
Deinococcus radiodurans	Deinococcus radiodurans R1 complete chromosome 2.
Deinococcus radiodurans	Deinococcus radiodurans R1 megaplasmid MP1, complete plasmid sequence.
Deinococcus radiodurans	Deinococcus radiodurans R1 plasmid CP1, complete plasmid sequence.
Deinococcus radiodurans	Deinococcus radiodurans R1 complete chromosome 1.
Escherichia coli O157:H7	Escherichia coli O157:H7, complete genome.
Escherichia coli O157:H7	Escherichia coli O157:H7, complete genome.
Escherichia coli	Escherichia coli K-12 MG1655 complete genome.
Escherichia coli	Escherichia coli K-12 MG1655 complete genome.
Haemophilus influenzae Rd	Haemophilus influenzae Rd complete genome.
Halobacterium sp. NRC-1	Halobacterium sp. NRC-1 plasmid pNRC200 complete genome.
Halobacterium sp. NRC-1	Halobacterium sp. NRC-1 plasmid pNRC100, complete plasmid sequence.
Halobacterium sp. NRC-1	Halobacterium sp. NRC-1 complete genome.
Helicobacter pylori J99	Helicobacter pylori strain J99 complete genome.
Helicobacter pylori 26695	Helicobacter pylori 26695 complete genome.
Lactococcus lactis subsp. lactis	Lactococcus lactis subsp. lactis IL1403 complete genome.
Listeria innocua	Listeria innocua Clip11262, complete genome.
Listeria monocytogenes EGD-e	Listeria monocytogenes strain EGD, complete genome.
Mycoplasma genitalium	Mycoplasma genitalium, complete genome.
Mycoplasma genitalium	Mycoplasma genitalium G37 complete genome.
Mycoplasma pneumoniae	Mycoplasma pneumoniae M129 complete genome.
Mycoplasma pneumoniae	Mycoplasma pneumoniae, complete genome.
Mycoplasma pulmonis	Mycoplasma pulmonis (strain UAB CTIP) complete genome.
Mesorhizobium loti	Mesorhizobium loti plasmid pMLb DNA, complete genome, complete sequence.
Mesorhizobium loti	Mesorhizobium loti complete genome, complete sequence.
Mesorhizobium loti	Mesorhizobium loti plasmid pMLa DNA, complete genome, complete sequence.
Methanobacterium thermoautotrophicum	Methanobacterium thermoautotrophicum delta H complete genome.
Methanococcus jannaschii	Methanococcus jannaschii JF-1 complete genome.

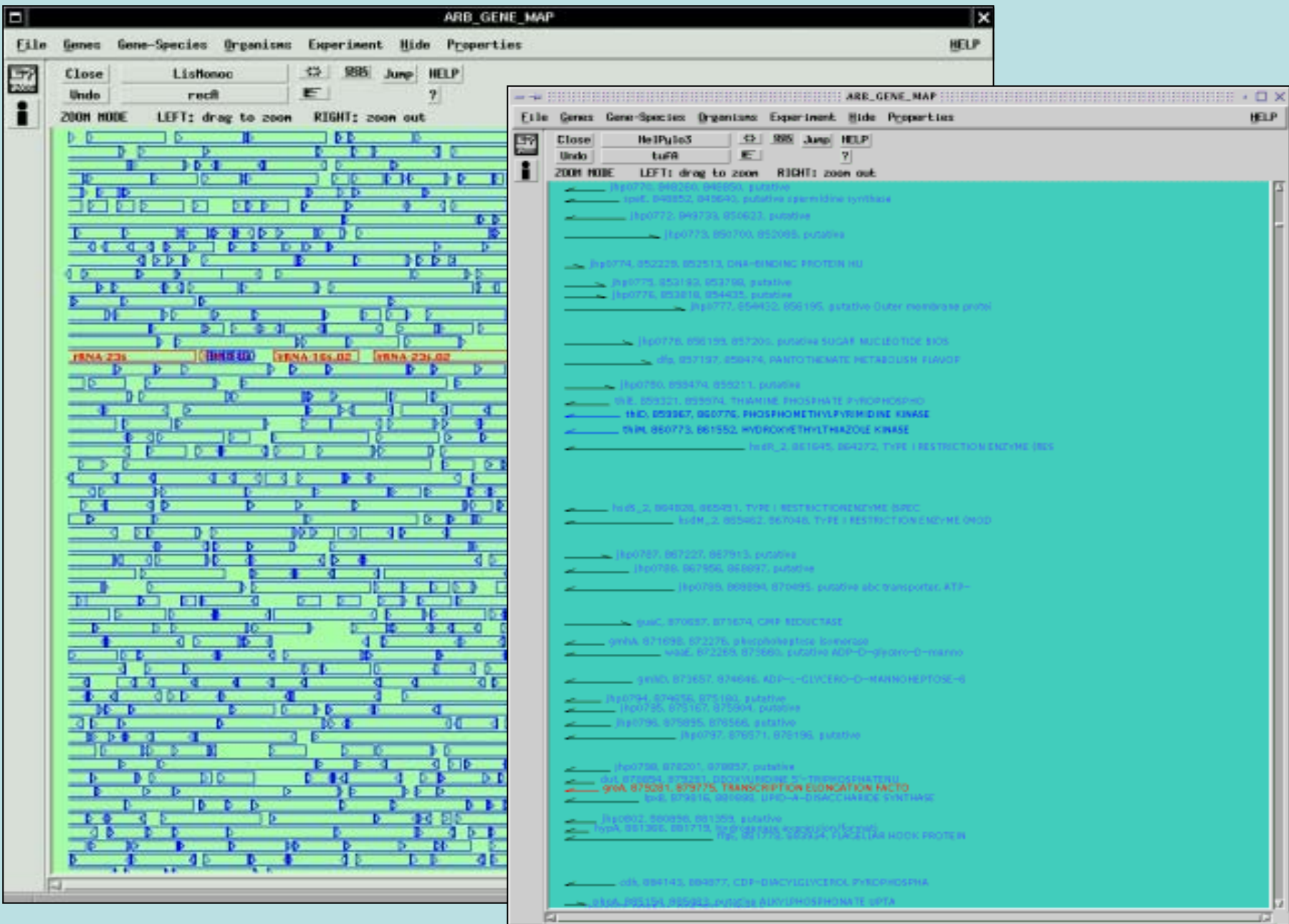


# Genome map of *Listeria monocytogenes* displaying rRNA Operons





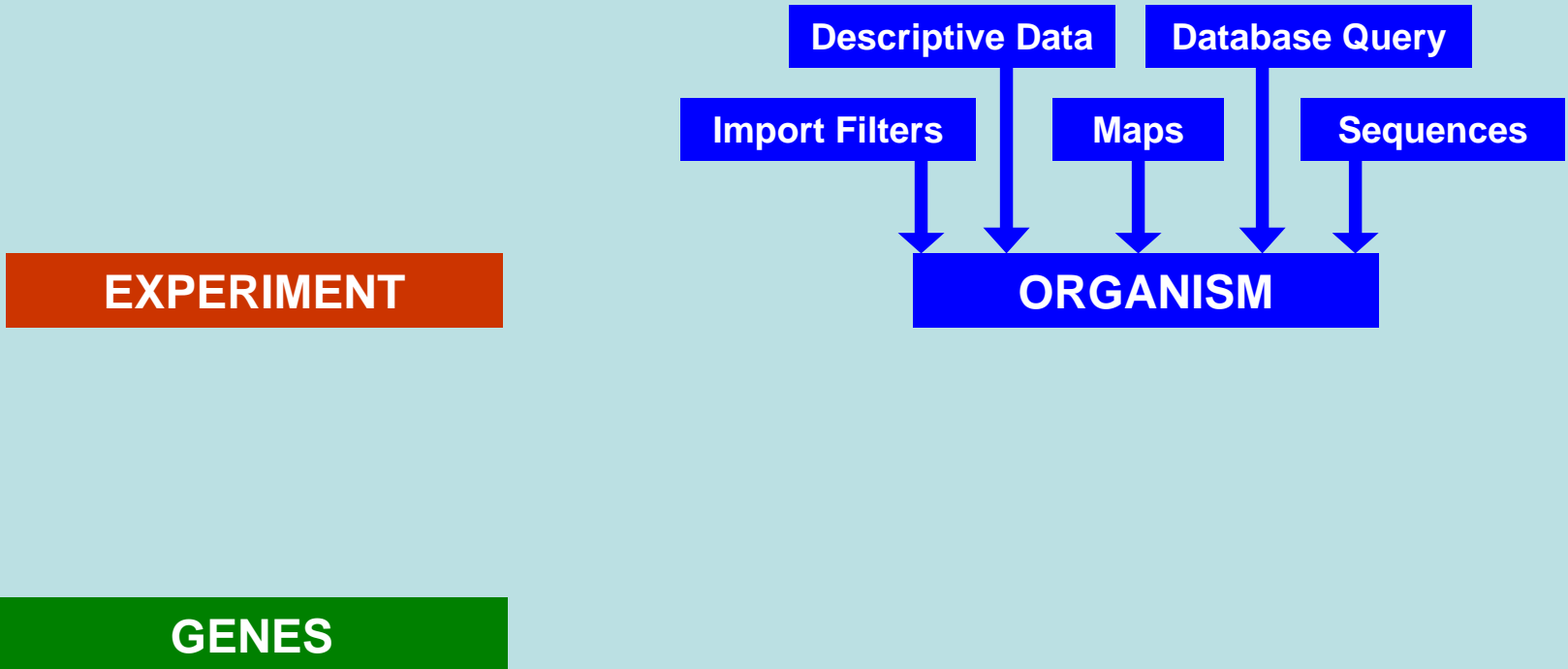
## Genome map in block view and Individual Gene Information display





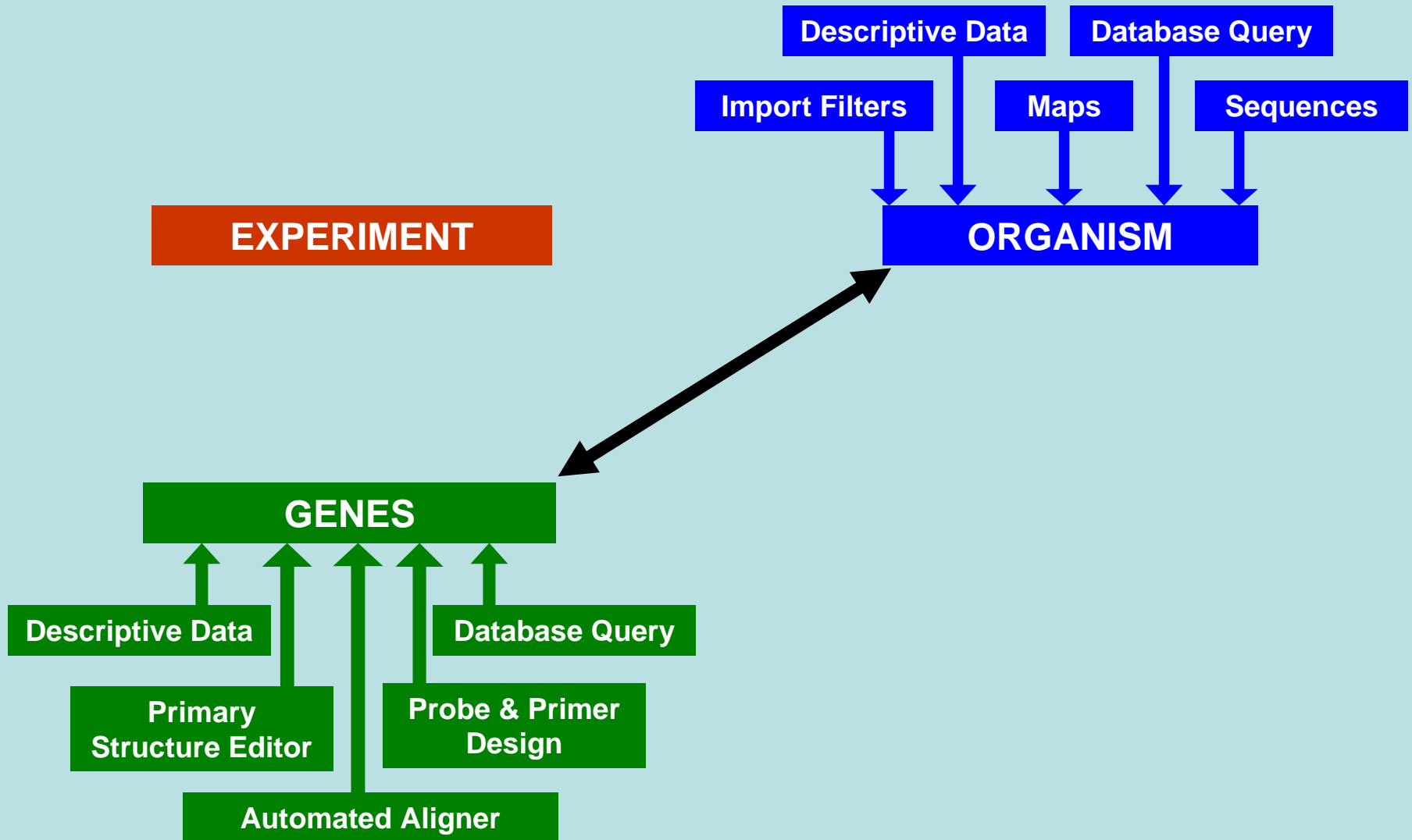
# The ARB Genome Project

---





# The ARB Genome Project





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

Gene SEARCH and QUERY

More functions More search

CLOSE

DATABASE SEARCH

Search genes

of current organism

Add genes

of marked organisms

Keep genes

of all organisms

QUERY

Search fields

Search strings

name

=

\*rRNA\*

ign

CDS/codon\_start

=

\*

ign

product

=

\*Ribosomal\*

SEARCH

HITLIST

Hits: 18

LisMonoc/rRNA-23s.03	:rRNA-23s.03
LisMonoc/rRNA-16s.03	:rRNA-16s.03
LisMonoc/rRNA-5s.04	:rRNA-5s.04
LisMonoc/rRNA-23s.04	:rRNA-23s.04
LisMonoc/rRNA-16s.04	:rRNA-16s.04
LisMonoc/rRNA-5s.05	:rRNA-5s.05
LisMonoc/rRNA-23s.05	:rRNA-23s.05

REFRESH

GENE INFORMATION

GENE FIELDS

CLOSE

DETACH

SEARCH

HELP

Edit enabled ? ☒ Marked ? ☒

Edit box (select a field and edit it in this box)

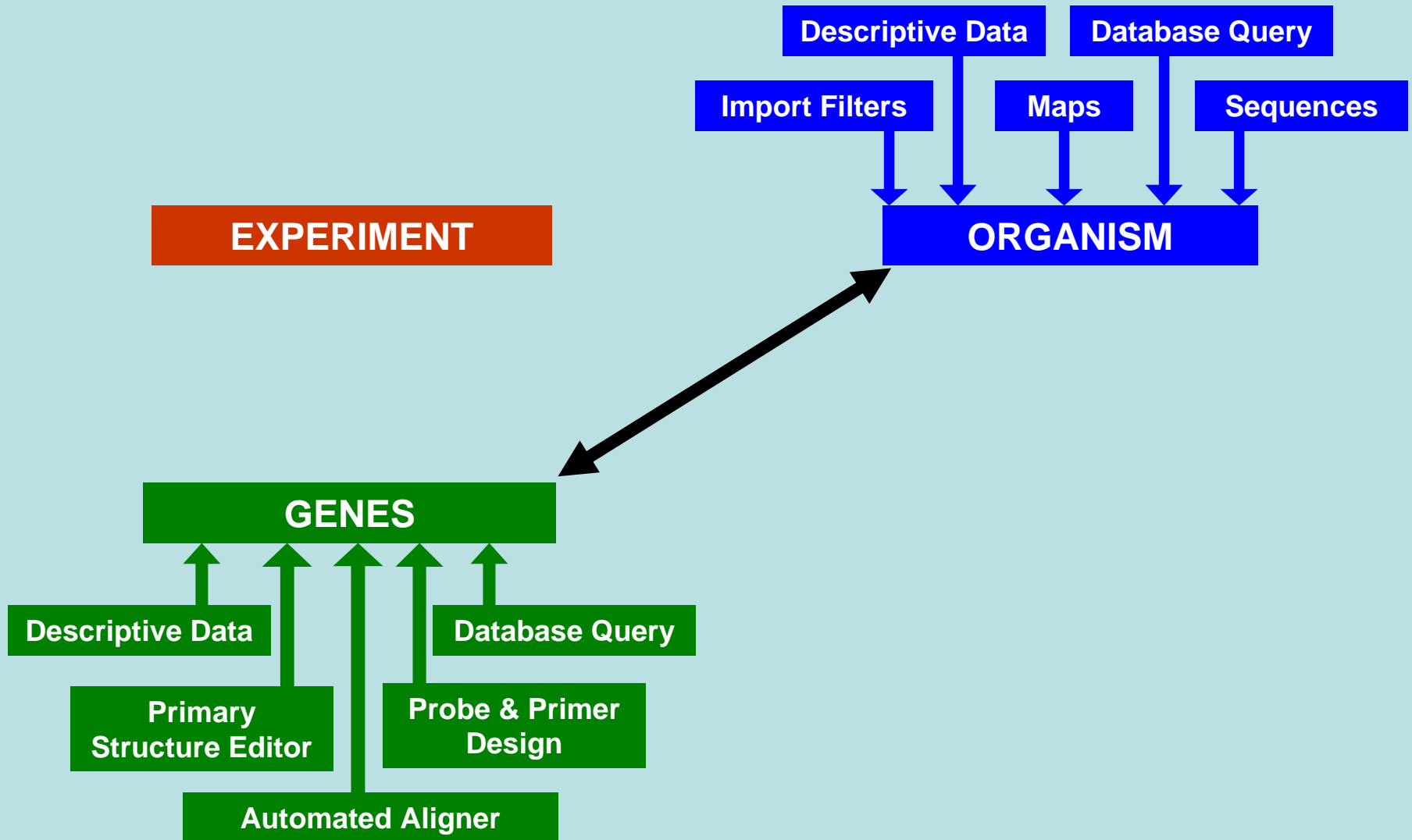
Recombination protein recA

DATABASE FIELDS

name	S0: recA
pos_begin	i0: 1425406
pos_end	i0: 1426518
complement	c0: 0
CDS/codon_start	S0: 1
CDS/complement	c0: 0
CDS/db_xref	S0: GI:16410827
CDS/pos_begin	i0: 1425419
CDS/pos_end	i0: 1426465
CDS/product	S0: Recombination protein recA
CDS/protein_id	S0: CAC99476.1

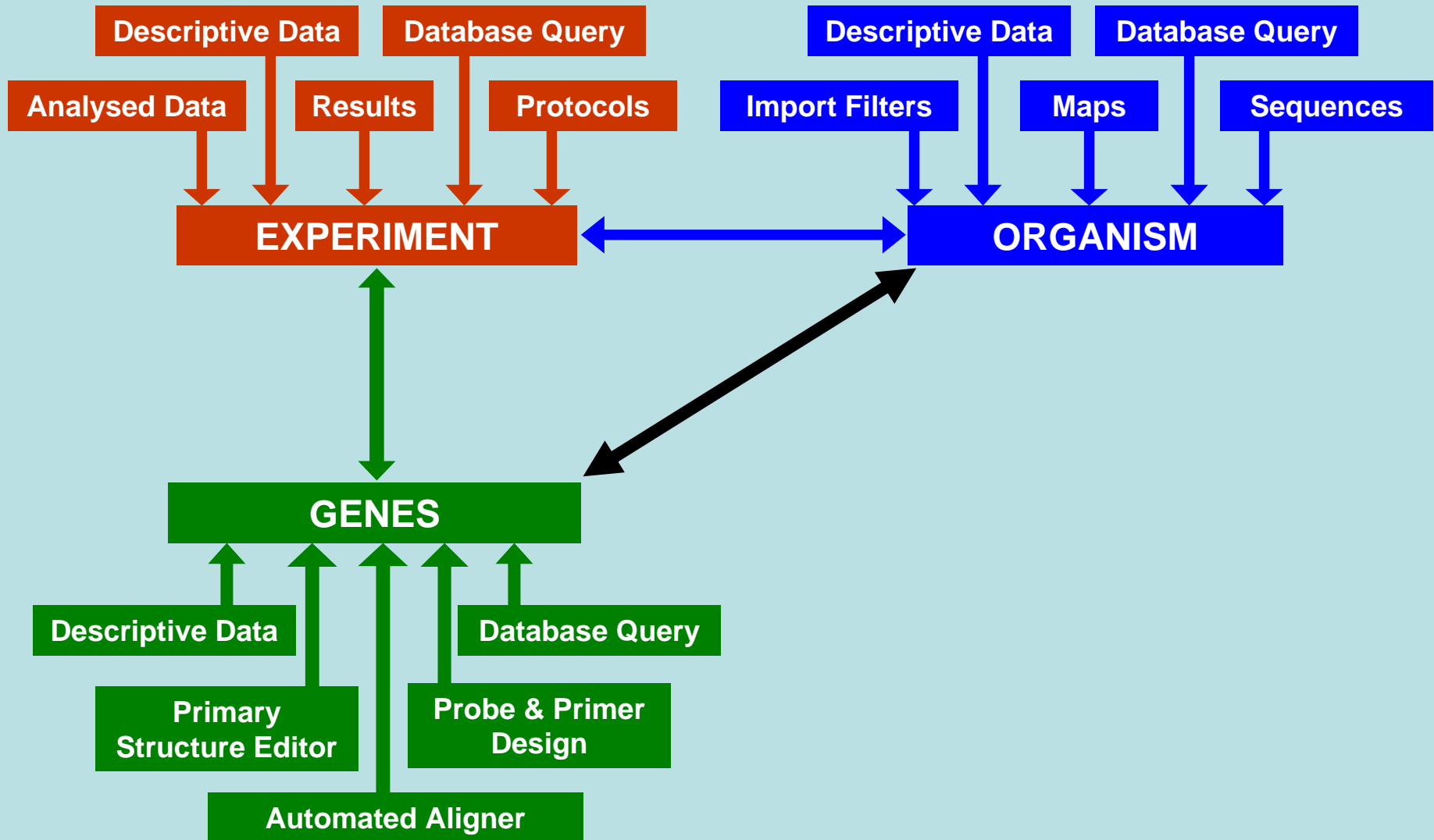


# The ARB Genome Project





# The ARB Genome Project





# Experiment Data Entry Form

Experiment: Proteomics I mask

CLOSE

HELP

EDIT

RELOAD

Enable edit?

☒

Marked

☒

Switch to..

Physiology I

Basic fields

You are editing ' StrPyoge/ '

Sample preparation:

Protocol

Incubation:

Protocol

Stainig:

Protocol

Buffer

Buffer I

Stain

Lysis

%

times

Time

min

Incubation

h

rpn

Centr.

rpn

h

°C

Buffer II

Washing

h

rpn

Protein content

ng/ml

Time

min

Protocol

Storage

☒ No
 ☐ Yes

-5

°C

1th Din.

Protocol

Rehydration buffer

2th Din.

Protocol

Scanning

Supplier

Supplier

Scanner

Protein content

ug

SDS PAGE

%

Res.

pixel

IPG-stripes

length

cm

Gel size

cm

Laser

nm

Isoelctr. Foc.

h

V

Parallels

Standards

Filter

Rehydration

h

V

Conditions

h

M

°C

Temperature

°C

Fixation

Protocol

Buffer

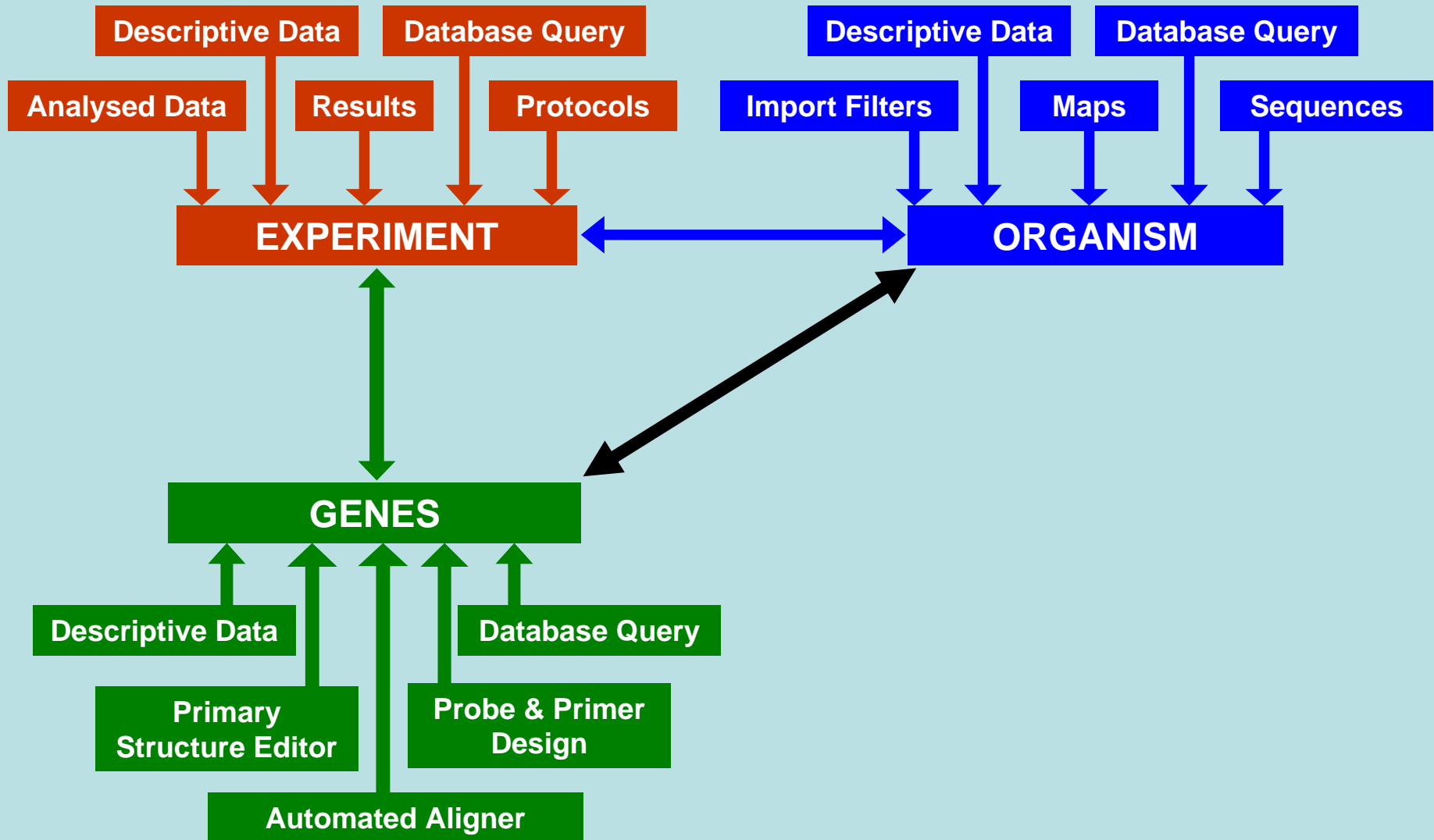
Incubation

h

rpn

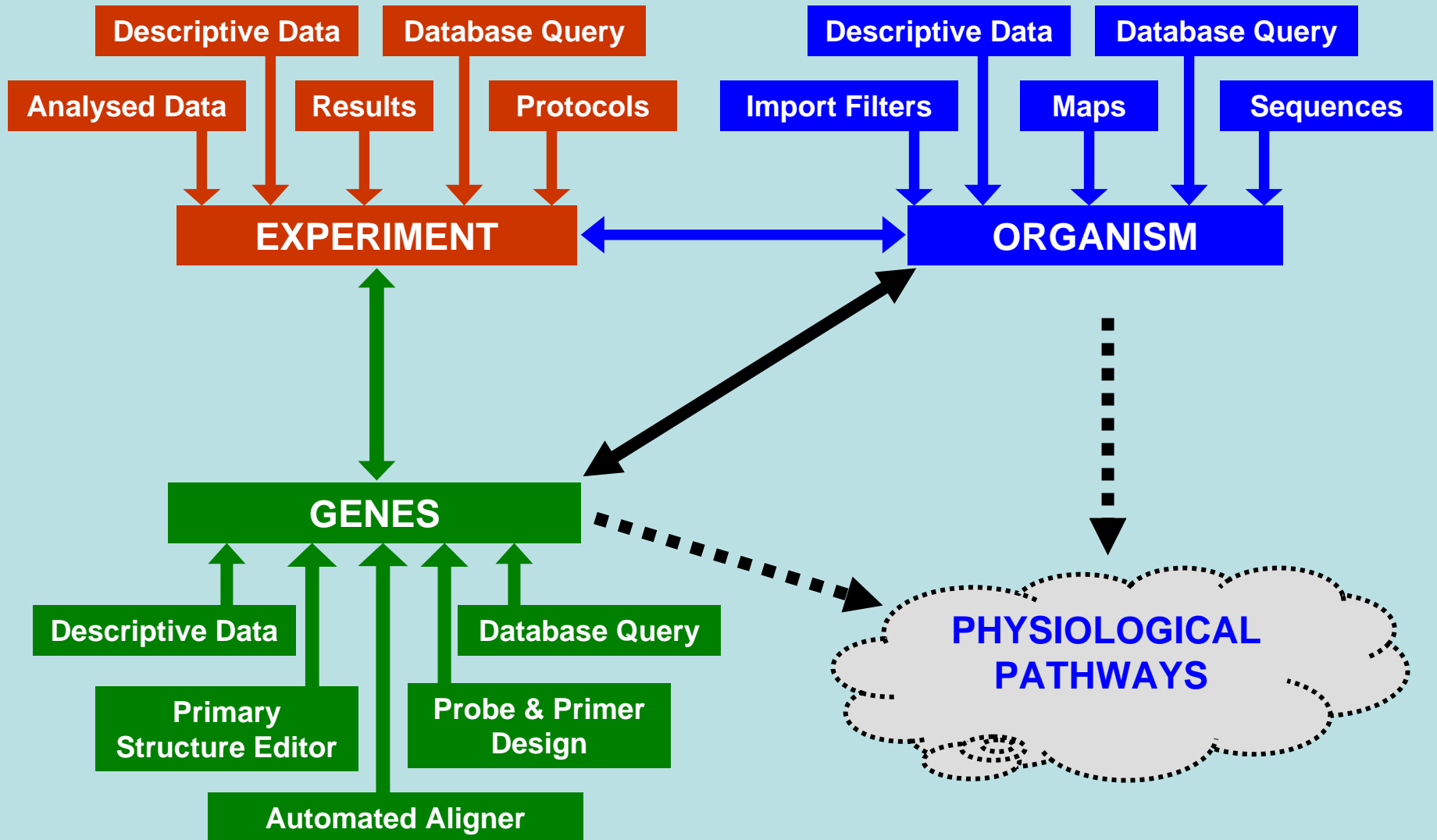


# The ARB Genome Project





# 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



## EMBL European Bioinformatics Institute

Search: Nucleotide sequences for [ ] Go

Site search: [ ] Go

EBI Home

About EBI

Databases

Tools

Submission

Research

Downloads

The [EMBL](#) Outstation - European Bioinformatics Institute (EBI) is a centre for research and services in bioinformatics. The Institute manages databases of biological data including nucleic acid, protein sequences and macromolecular structures.

**EMBL**  
New 23rd 2001  
A new public nucleic acid and protein sequence database



## NCBI PubMed

Search: PubMed for [ ] Go Clear

About Entrez

Entrez PubMed

Help | FAQ

Advanced

Web Information

PubMed Services

Annual Browser

MeSH Browser

Single Citation Matcher

Book Crossing Matcher

Clinical Queries

LinkOut

Collate

Related Resources

Online Resources

NCBI Gateway

Genome Health

Clinical Alerts

Clinical Trials

PubMed Central

Privacy Policy

- Enter one or more search terms, or click [PubMed links](#) for advanced search.
- Enter [author names](#) or [search by title](#) is optional.
- Enter [journal titles](#) in full or as MEDLINE abbreviations. Use the [Journal Browser](#) to find journal titles.

PubMed, a service of the National Library of Medicine, provides access to over 1 million MEDLINE citations back to the mid-1960's and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.

### Books are here!

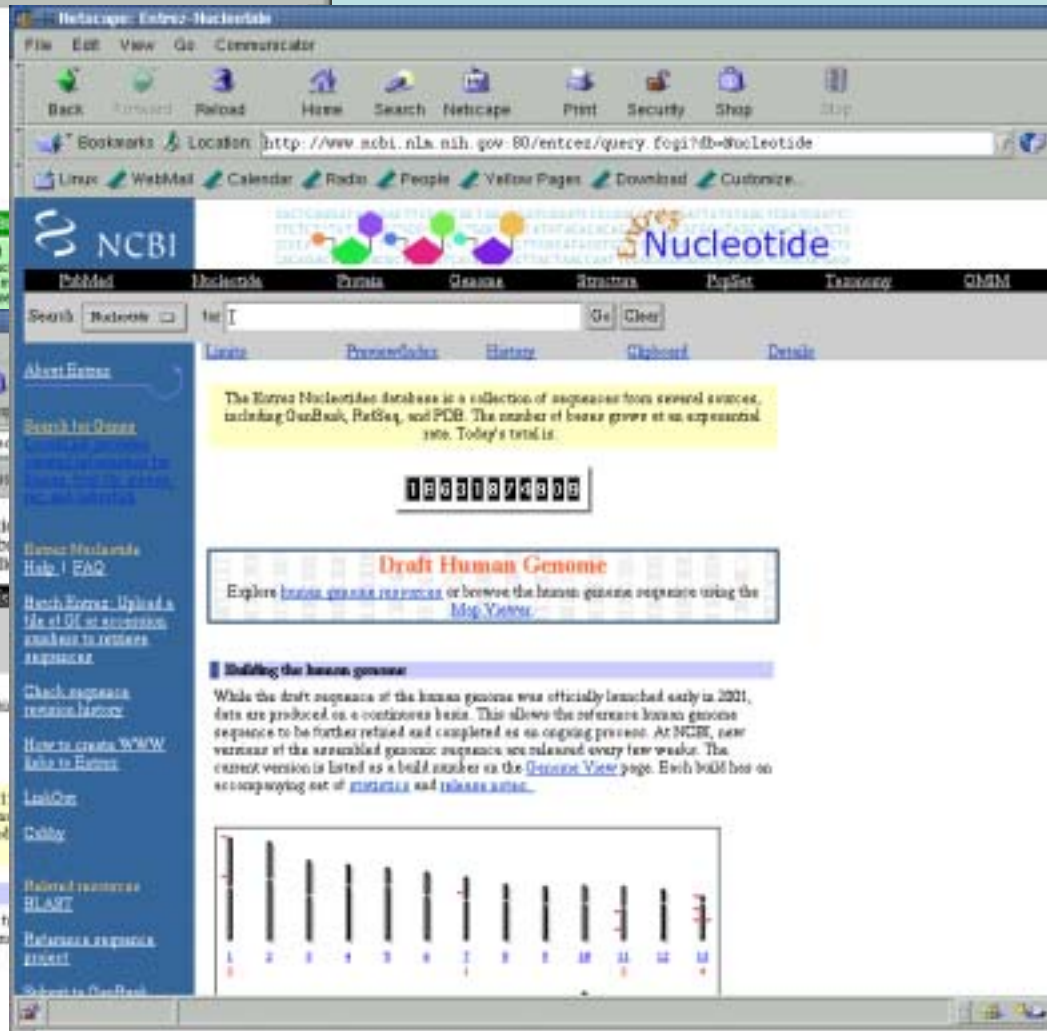


Look for background information or research new topics with freely accessible, online biomedical textbooks.

The growing NCBI Bookshelf can be searched directly or accessed via [phrases](#) in PubMed abstracts by clicking the "Books" link. For a quick start, use the [Bookshelf homepage](#).

### Free Full-text Articles

Try the new [PubMed Central](#), NLM's free and unrestricted online archive of journal articles. Links to full-text in PubMed Central are available from PubMed citations.





## Operating Systems



**LINUX / Unix**  
Operating System



**Mac OS**

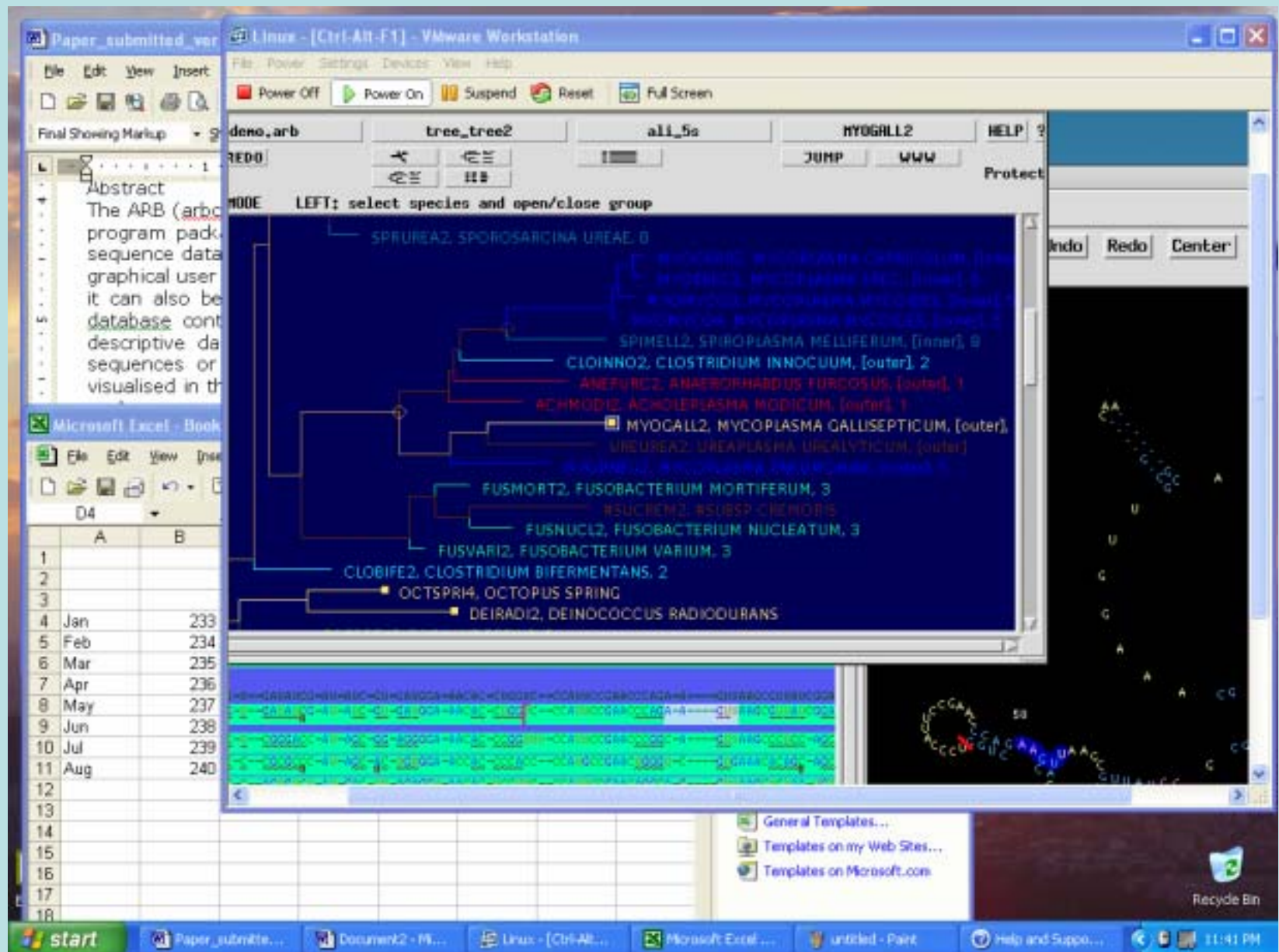


## Programming Languages

- ◆ 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](http://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](mailto:ludwig@mikro.biologie.tu-muenchen.de)

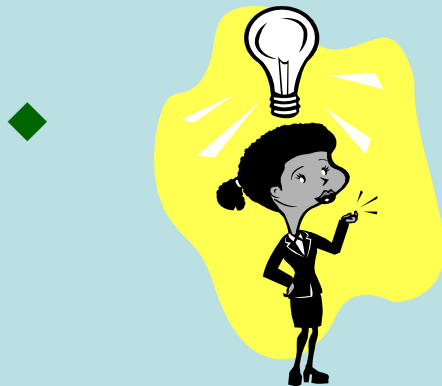
### Programmers and Curators

**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





# Thank You



rRNA-23s.05, 2438032, 2440962

rRNA-16s.05, 2441458, 2443003

rRNA-5s.05, 2437842, 2437951

rRNA-23s.04, 1849012, 1851942

rRNA-16s.04, 1852187, 1853732

rRNA-5s.04, 1848821, 1848930

rRNA-23s.03, 1743091, 1746021

rRNA-16s.03, 1746461, 1747811

rRNA-5s.03, 1742001, 1743003

rRNA-23s.06, 2673041, 2675971

rRNA-16s.06, 2676467, 2678012

rRNA-5s.06, 2672851, 2672960

rRNA-16s.02, 1425406, 1426518

rRNA-5s.02, 2437842, 2437951

rRNA-16s.02, 2441458, 2443003

rRNA-23s.02, 2438032, 2440962