

Querying Local and National Databases



Erik W. Coleman

May 2009



Overview

- Query Local Databases
 - Query Tools for Local Databases
 - Perform a Simple Query
 - Query by Isolate, Pattern and the Hot List Tool
- Query National Databases
 - Query Tools for National Databases
 - Query by Isolate and the Hot List Tool
 - Narrow Search, Download Selected Entries
- Composite Data Sets
 - What they are and Example of Usage
 - Create a Composite Data Set
 - Using Composite Data Sets in Comparisons

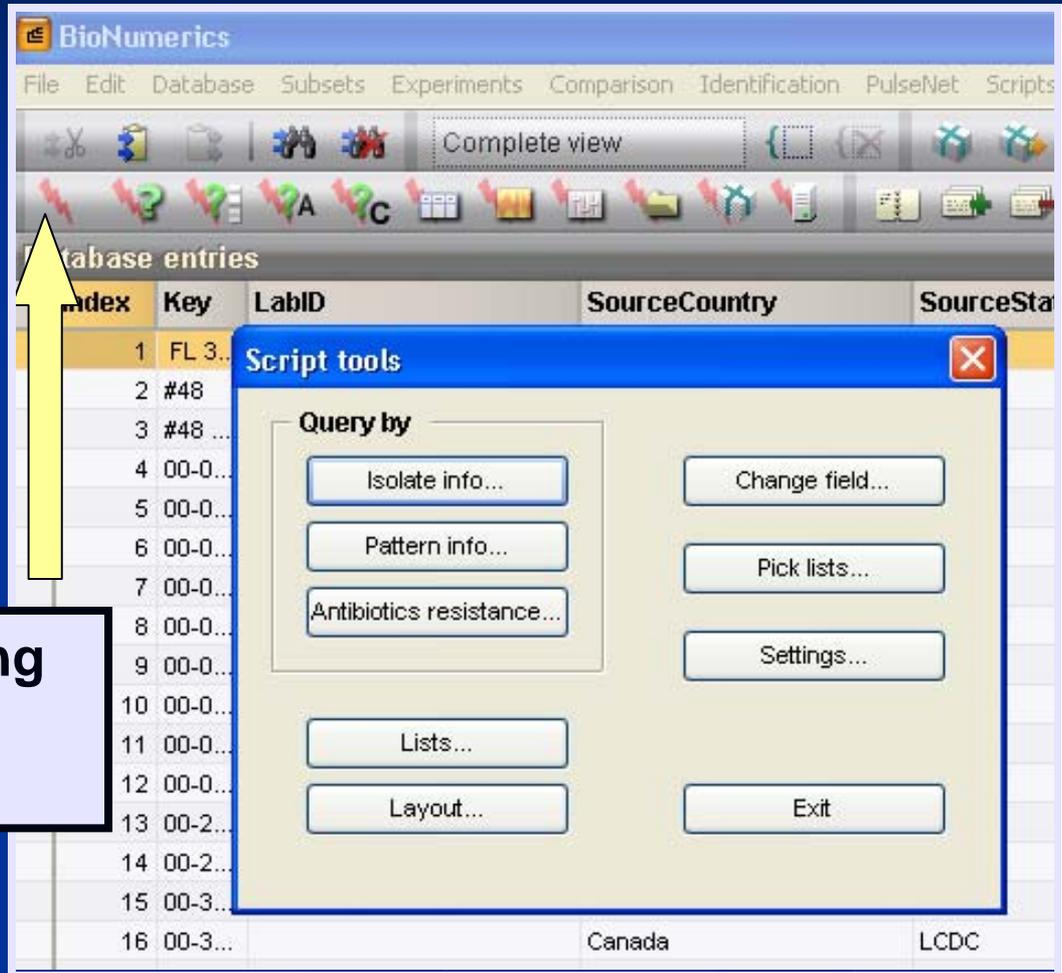
Query Tools for Local Databases

■ Search tools

■ PulseNet tools can be used to:

- ❖ Query by isolate
- ❖ Query by pattern
- ❖ Query by antibiotic resistance

Can click on lightning bolt to show some query options



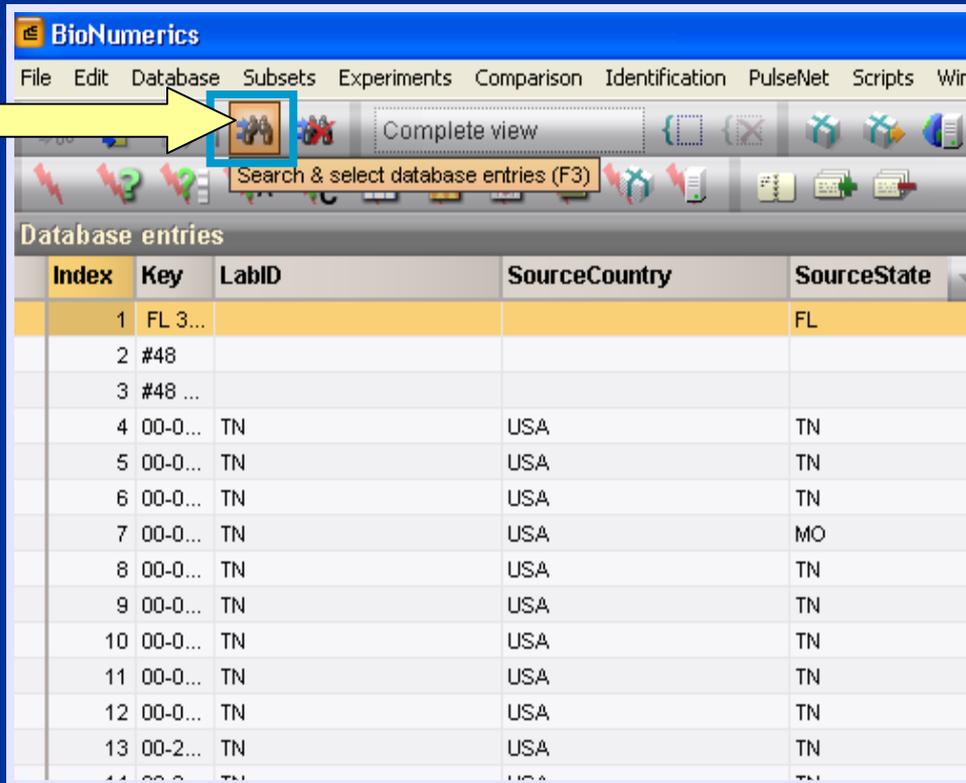
The screenshot shows the BioNumerics software interface. The main window displays a table of database entries with columns: Index, Key, LabID, SourceCountry, and SourceSta. A 'Script tools' dialog box is open, featuring a 'Query by' section with buttons for 'Isolate info...', 'Pattern info...', and 'Antibiotics resistance...'. Other buttons include 'Change field...', 'Pick lists...', 'Settings...', 'Lists...', 'Layout...', and 'Exit'. A yellow arrow points to a lightning bolt icon in the toolbar, which is highlighted in the text box on the left.

Index	Key	LabID	SourceCountry	SourceSta
1	FL 3...			
2	#48			
3	#48 ...			
4	00-0...			
5	00-0...			
6	00-0...			
7	00-0...			
8	00-0...			
9	00-0...			
10	00-0...			
11	00-0...			
12	00-0...			
13	00-2...			
14	00-2...			
15	00-3...			
16	00-3...		Canada	LCDC

Perform a Simple Query

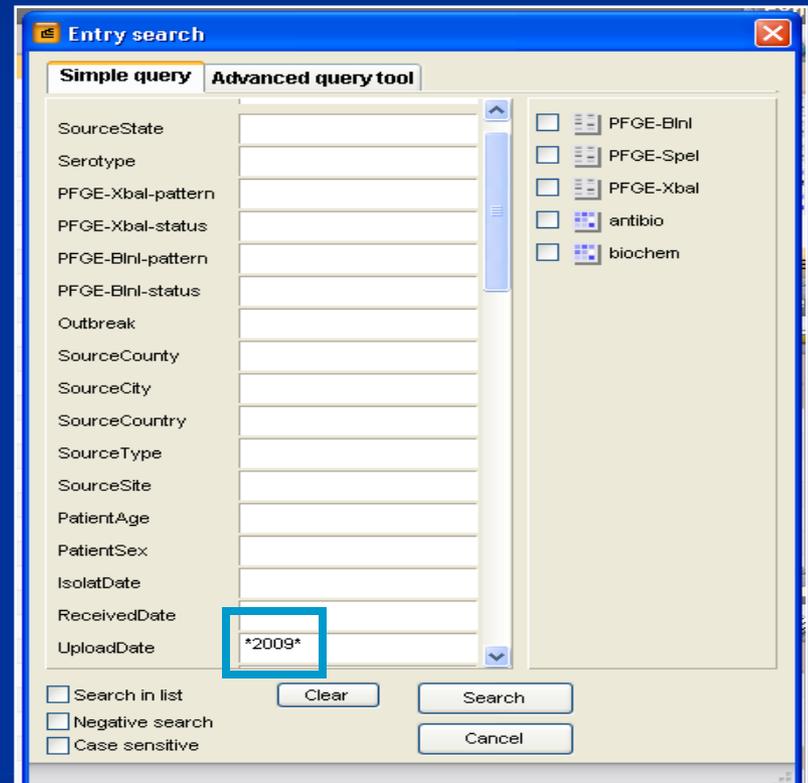
Click on Binocular Icon in Main Window

Use wildcard (*) before and after text



Database entries

Index	Key	LabID	SourceCountry	SourceState
1	FL 3...			FL
2	#48			
3	#48 ...			
4	00-0...	TN	USA	TN
5	00-0...	TN	USA	TN
6	00-0...	TN	USA	TN
7	00-0...	TN	USA	MO
8	00-0...	TN	USA	TN
9	00-0...	TN	USA	TN
10	00-0...	TN	USA	TN
11	00-0...	TN	USA	TN
12	00-0...	TN	USA	TN
13	00-2...	TN	USA	TN
14	00-0...	TN	USA	TN



Simple query Advanced query tool

SourceState
Serotype
PFGE-Xbal-pattern
PFGE-Xbal-status
PFGE-BlnI-pattern
PFGE-BlnI-status
Outbreak
SourceCountry
SourceCity
SourceCountry
SourceType
SourceSite
PatientAge
PatientSex
IsolatDate
ReceivedDate
UploadDate

PFGE-BlnI
 PFGE-Spel
 PFGE-Xbal
 antibio
 biochem

2009

Search in list
 Negative search
 Case sensitive

Clear Search Cancel

Search yields all isolates in your local database that have "2009" as the Upload Date

Query by Isolate

Click on "Query by Isolate" icon

Enter Search Criteria. Ex: Date Range

Click "Replace List"

Isolate query

Source

City: [dropdown]
County: [dropdown]
State: [dropdown]
Country: [dropdown]

Patient

Age betw: [dropdown]
Sex: [dropdown]

Serotype: [dropdown]
Source type: [dropdown]
Type details: [dropdown]
Source site: [dropdown]
Isolate status: [dropdown]

Isolate date

from: Saturday, July 05, 200 [dropdown]
to: Friday, September 05, 200 [dropdown]

Upload date

from: Wednesday, September 24, 200 [dropdown]

Received date

from: Wednesday, September 24, 200 [dropdown]
to: Wednesday, September 24, 200 [dropdown]

Negative

Query by Pattern

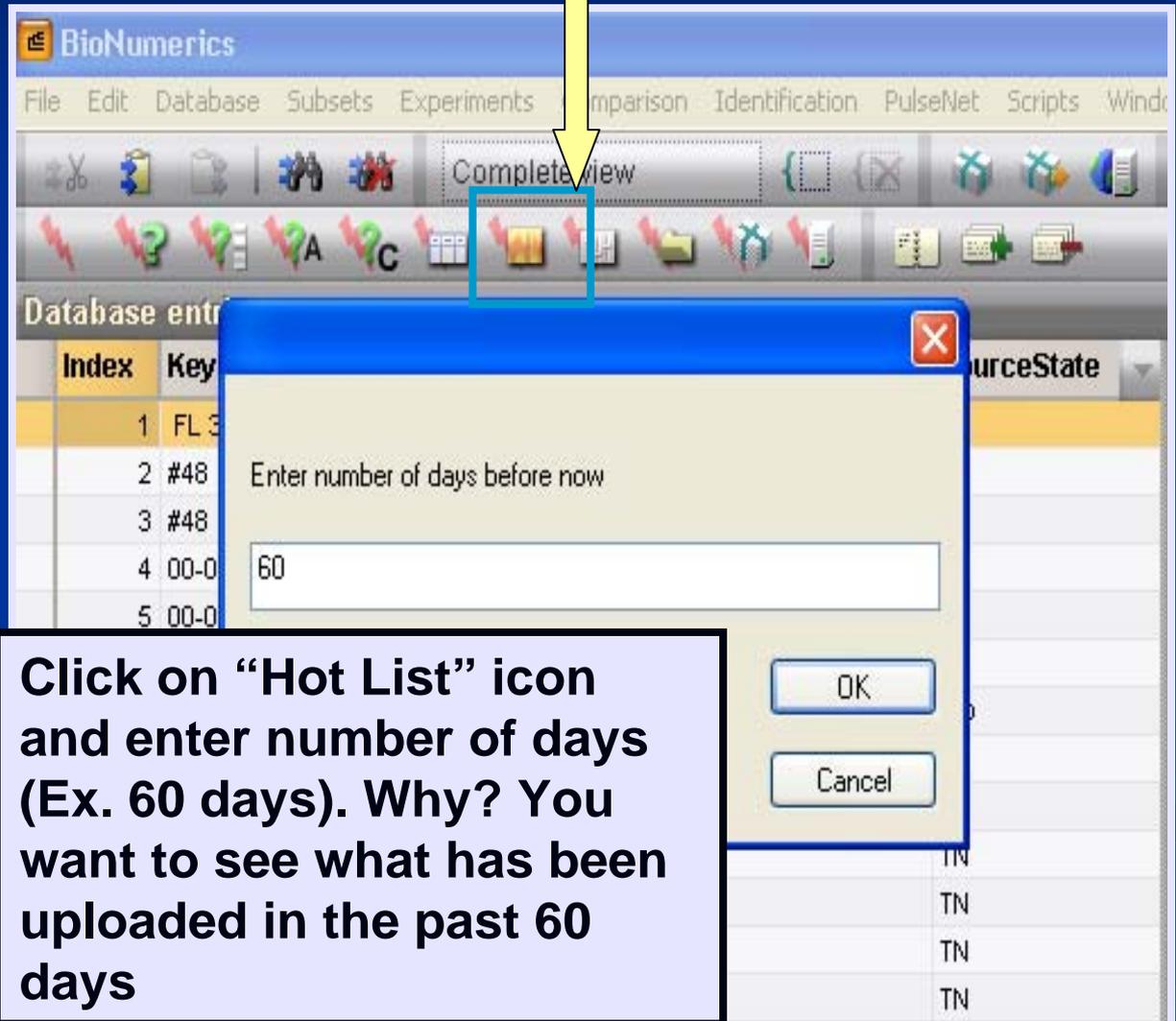
The screenshot shows the BioNumerics software interface. A 'Pattern query' dialog box is open, allowing users to search for specific patterns. The 'Selected enzymes' list includes XbaI, BlnI, and SpeI. The 'Run date' section is set from Saturday, July 05, 2008, to Friday, September 05, 2008. The 'Pattern' field contains the text 'EXHX01.0047'. A yellow arrow points to the 'Pattern' field, and another yellow arrow points to the 'XbaI' enzyme in the list.

Select enzyme, then fill in complete pattern number (i.e. EXHX01.0047). This will search within the field “PFGE-XbaI-Pattern.”
Note: This is especially useful when downloading CDC pattern names

Query Using the Hot List Tool

A Hot List search can be used to show all recently uploaded isolates within your local database.

NOTE: the isolates *must* have an upload date for this query to work



The screenshot shows the BioNumerics software interface. A yellow arrow points to the 'Hot List' icon in the toolbar, which is highlighted with a blue box. A dialog box is open, prompting the user to 'Enter number of days before now' with a text input field containing '60'. The dialog box has 'OK' and 'Cancel' buttons. In the background, a table with columns 'Index' and 'Key' is visible, showing rows with values like '1 FL 3', '2 #48', '3 #48', '4 00-0', and '5 00-0'. The 'Complete view' button in the toolbar is also visible.

Click on “Hot List” icon and enter number of days (Ex. 60 days). Why? You want to see what has been uploaded in the past 60 days

Query Using the Hot List Tool

BioNumerics

File Edit Database Subsets Experiments Comparison Identification PulseNet Scripts Window

Complete view

Database entries

Index	Key	LabID	SourceCountry	SourceSta	1	2	3	4	5	6
1	08EMSO...		USA	MS	•		•			
2	08M-MIS...		USA	ME	•		•			
3	224425		USA	DE	•		•			
4	226868		USA	DE	•		•			
5	231540		USA	DE	•		•			
6	AM0107...						•			
7	AM0114...						•			
8	AM0154...						•			
9	AM0220...						•			
10	AM0237...						•			
11	AM0255...						•			
12	AM0256...						•			
13	AM0264...						•			

Isolates submitted within the past 60 days will be selected.

All levels

Database: Ecoli-client 8114 entries 6 experiments \\cdc\project\CCID_NCZVED_DFBMD_PulseNet\Data\Ecoli-client

Overview

- Query Local Databases
 - Query Tools for Local Databases
 - Perform a Simple Query
 - Query by Isolate, Pattern and the Hot List Tool
- Query National Databases
 - Query Tools for National Databases
 - Query by Isolate and the Hot List Tool
 - Narrow Search, Download Selected Entries
- Composite Data Sets
 - What they are and Example of Usage
 - Create a Composite Data Set
 - Using Composite Data Sets in Comparisons

Query Tools for National Databases

Isolate Query

Hot List

National List

Pattern Query

Delete Selected Entries

Show Recent Matches



Server login

File Database PulseNet Window

Temporarily Download Selected Entries to local database

Status

PulseNet USA
The National Molecular Subtyping Network for Foodborne Disease Surveillance
CDC

Server database

1515 entries

Database fields

- LabID
- SourceCountry
- SourceState
- SourceCounty
- SourceCity
- SourceSite
- SourceType
- PatientAge
- PatientSex
- IsolatDate

Fingerprint types

- PFGE-BlnI
- PFGE-SpeI
- PFGE-XbaI

Server database Identifica

Note: Queries within the national database will only look within the number of days you're connected to

Query by Isolate

Use query tools to search for isolates within national database

Click on the Query by Isolate tool

Enter search criteria

Click on "Replace List"

The screenshot shows the 'Isolate query' application window. The 'Source' section has dropdown menus for 'City', 'County', 'State', and 'Country', and a text input for 'City'. The 'Patient' section has 'Age between' and 'Sex' fields. The 'Isolate date' section has 'from' and 'to' date pickers. The 'Upload date' section has 'from' and 'to' date pickers. The 'Replace list' button is highlighted with a red box. The search criteria 'E. coli O103' is entered in the top right dropdown.

Query by Isolate

The number of entries matching your search criteria is shown

Database entries

Index	Key	LabID
1	08EMSO...	
2	08M-MIS...	
3	224425	
4	226868	
5	231540	
6	AM0107...	
7	AM0114...	
8	AM0154...	
9	AM0220...	
10	AM0237...	
11	AM0255...	
12	AM0256...	
13	AM0264...	
14	AM0266...	
15	AM0272...	
16	AM0279...	
17	AM0285...	
18	AM0286...	
19	AM0294...	
20	AM0295...	
21	AM0317...	
22	AM0318...	

Server login

Status

PulseNet USA
The National Molecular Subtyping Network for Foodborne Disease Surveillance
CDC

Number of entries selected: 10

OK

Server database

1515 entries

Database fields

- LabID
- SourceCountry
- SourceState
- SourceCounty
- SourceCity
- SourceSite
- SourceType
- PatientAge
- PatientSex
- IsolatDate

Fingerprint types

- PFGE-BlnI
- PFGE-Spel
- PFGE-Xbal

Character types

- antibio
- biochem

All levels

Database: Ecoli-client 8114

Server database Identification results

Query Using the Hot List Tool

Click on the Hot List tool to search for all recently uploaded isolates within the national database.

Enter a number of days (Ex. 60 days).

Remember: If you are only connected to a portion of the database (such as 60 days), there's no need to do a 60-day hotlist search – you are already connected to it!

The screenshot shows the 'Server login' application window. The 'Hot List' tool icon is highlighted with a blue box and a yellow arrow. A dialog box titled 'Enter number of days to cover' is open, with the number '60' entered in the text field. The background interface shows the 'Server database' section with '1515 entries' and a list of 'Database fields' including LabID, SourceCountry, SourceState, SourceCity, SourceSite, SourceType, PatientAge, PatientSex, and IsolatDate. There are also checkboxes for PFGE-Spel, PFGE-Xbal, and Character types (antibio, biochem).

Narrow Your Search

The results of your hot-list search will appear.

Important! It may be a LARGE number of entries! You do not need to view or download all of them. Downloading an entire 60-day search wastes time and server space, and may cause your database to time out. Therefore, think about what you REALLY need – and narrow your search!

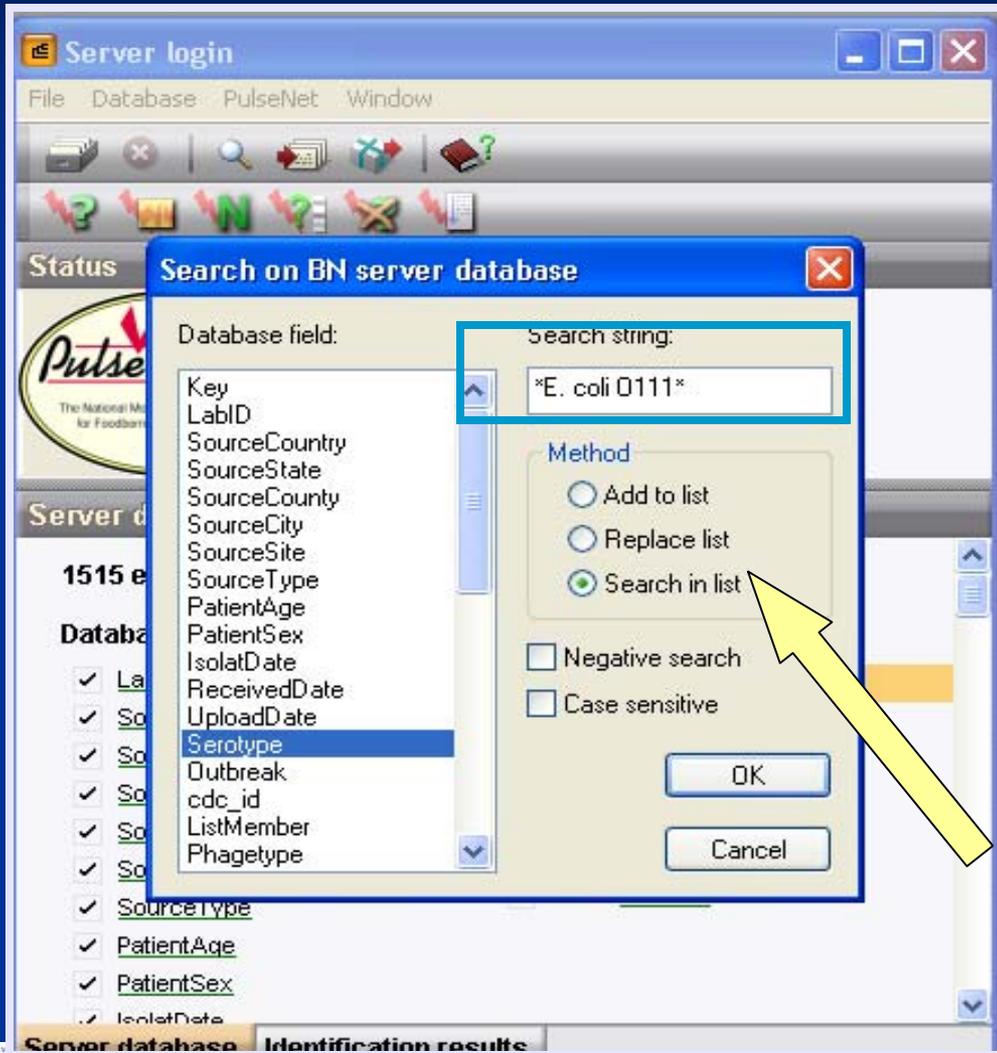
Ex: 60 Day Hot List Search

The screenshot shows the BioNumerics software interface. A table titled 'Database entries' is visible, with columns for Index, Key, and LabID. The table contains 12 rows of data, with the 9th row highlighted. A 'Server login' window is open, showing the 'PulseNet USA' logo and the text 'The National Molecular Subtyping Network for Foodborne Disease Surveillance'. A dialog box in the foreground displays 'Number of entries selected: 1515' and an 'OK' button. A yellow arrow points from a text box on the right to the dialog box.

Index	Key	LabID
1	08EMSO...	
2	08M-MIS...	
3	224425	
4	226868	
5	231540	
6	AM0107...	
7	AM0114...	
8	AM0154...	
9	AM0220...	
10	AM0237...	
11	AM0255...	
12	AM0255...	

Wow! 1515 isolates!
I need to narrow my search before I download anything...

Narrow Your Search



Use query tools to narrow your search. Ex: Narrow your search to include only isolates of the particular serotype that you are looking for.

“Search in List” (your list includes the 1515 that were selected from your original hot-list search)

Narrow Your Search

After narrowing your search, you have a smaller number of isolates that will not take as long to download.

The screenshot displays the PulseNet USA web interface. At the top, there is a 'Status' section with the PulseNet USA logo and text: 'The National Molecular Subtyping Network for Foodborne Disease Surveillance' and 'CDC'. Below this is the 'Server database' section, which shows '1515 entries'. The interface is divided into two columns of search filters. The left column, 'Database fields', includes: LabID, SourceCountry, SourceState, SourceCounty, SourceCity, SourceSite, SourceType, PatientAge, PatientSex, and IsolateDate. The right column, 'Fingerprint types', includes: PFGE-BlnI, PFGE-SpeI, and PFGE-XbaI. Below these is the 'Character types' section, which includes: antibio and biochem. A 'Query' dialog box is overlaid on the interface, showing '1 selected' and an 'OK' button. A yellow arrow points from the right side of the image to the '1 selected' text in the dialog box. At the bottom of the interface, there are two tabs: 'Server database' and 'Identification results'.

Download Entries

The screenshot shows the PulseNet USA software interface. At the top, there is a 'Server login' window with a menu bar (File, Database, PulseNet, Window) and a toolbar. A red box highlights the 'Download the selected entries from the server' icon in the toolbar. Below this, the 'BioNumerics' window is open, showing a menu bar (File, Edit, Database, Subsets, Experiments, Comparison, Identification, PulseNet, Events, Window) and a toolbar. A red box highlights the 'Location' field in the 'Database entries' table. A yellow arrow points from the 'Location' field to a callout box. Another yellow arrow points from the 'Download the selected entries from the server' icon to a callout box. A third yellow arrow points from the 'Database entries' table to a callout box. The 'Database entries' table has columns: Index, Key, Location, LabID, SourceCountry, and SourceState. The first row is highlighted in yellow and has a red arrow pointing to it.

Click on “Download selected entries from server” icon to download to your local database

“Location” field tells you if an isolate is part of a bundle file

Isolate(s) will be downloaded temporarily in the form of a bundle

Index	Key	Location	LabID	SourceCountry	SourceState
1	WV...	Bundle	WV_	USA	
2	FL 552	Local			
3	#48	Local			
4	#48 p...				
5	00-04...				
6	00-06...				
7	00-06...				
8	00-06...				
9	00-07...				
10	00-07...				

Overview

- Query Local Databases
 - Query Tools for Local Databases
 - Perform a Simple Query
 - Query by Isolate, Pattern and the Hot List Tool
- Query National Databases
 - Query Tools for National Databases
 - Query by Isolate and the Hot List Tool
 - Narrow Search, Download Selected Entries
- Composite Data Sets
 - What they are and Example of Usage
 - Create a Composite Data Set
 - Using Composite Data Sets in Comparisons

Composite Data Sets

- BioNumerics allows the user to combine different experiment types (fingerprint, character) in a cluster analysis
 - Takes an average of the similarity matrices
- That is, you could create a “mega” analysis of more than one experiment type
 - Examples:
 - *Xba*I & *Bln*I PFGE data
 - PFGE data and VNTR data
 - PFGE and antibiotic susceptibility data

An interesting example....

- An outbreak of *Salmonella* Heidelberg has the following PFGE results:
- Of 54 PFGE patterns (cases) there are 25 indistinguishable *Xba*I patterns
- *Bln*I fingerprinting was done on the isolates and 14 indistinguishable patterns were observed
- How do the PFGE fingerprint patterns cluster when analyzing **both** *Xba*I and *Bln*I data??
 - Could there be a unique fingerprint with *Xba*I and *Bln*I that is the source of the outbreak?

Create a Composite Data Set

Choose “Create new composite data set...” from the “Experiments” menu

The screenshot shows the BioNumerics software interface. The 'Experiments' menu is open, and the option 'Create new composite data set...' is highlighted with a yellow arrow. The main window displays a table of database entries and a table of experiments.

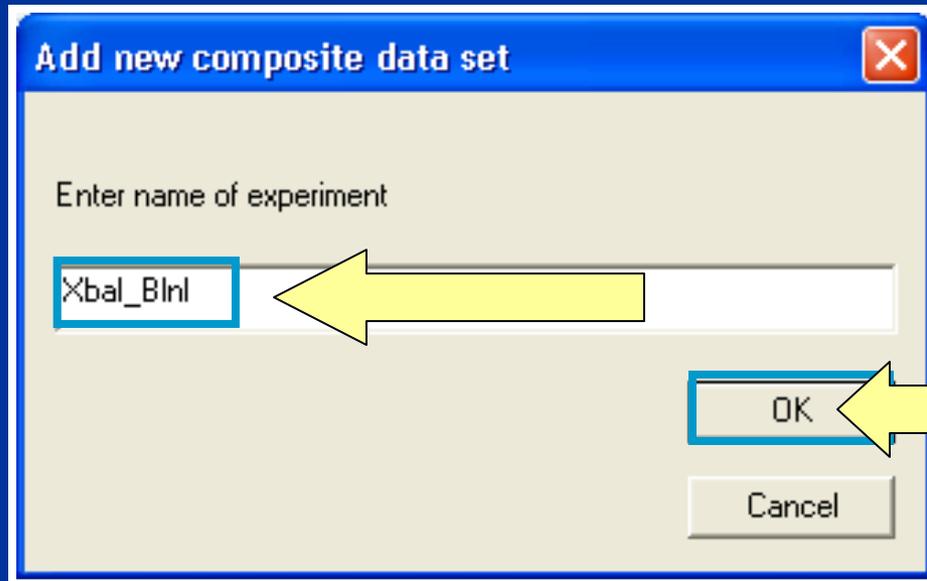
Index	Key	Serotype	PFGE-Xbal-pattern	1	2	3	4	5	6
1	AK_07253	Agona	JABX01.0001	Co					
2	AK_07256	4,[5],12:i-	JPXX01.1192	Co					
3	AK_07256	Montevideo	JIXX01.0825	Co					
4	AK_07261	Newport	JJPX01.0085	Co					
5	AK_07262	Aviana	JGGX01.0966	Co					
6	AK_07269	Myphimurium	JPXX01.0302	Un					
7	AK_07277	Myphimurium	JPXX01.0302	Un					
8	AK_07277	Myphimurium	JPXX01.0302	Un					
9	AK_0728300145	AK	Newport	JJPX01.0085	Co				
10	AK_0728300147	AK	London	TECX01.0004	Co				
11	AK_0728800289	AK	Newport	JJPX01.0418	Co				
12	AK_0729000025	AK	4,[5],12:i-	JPXX01.0206	Un				
13	AK_0729000145	AK	Stanley	JNGX01.0064	Co				
14	AK_0729000147	AK	4,[5],12:i-	JPXX01.0206	Co				
15	AK_0729200310	AK	4,[5],12:i-	JPXX01.0206	Co				

Name	Type
1 PFGE-BlnI	Fingerprint ty
2 PFGE-Spel	Fingerprint ty
3 PFGE-Xbal	Fingerprint ty
4 antibio	Character typ
5 biochem	Character typ

Name	Created
AK__AK05007	2005-04-12 00:00:00

Create a Composite Data Set

Enter the name of your composite data set and click “OK”



Add new composite data set

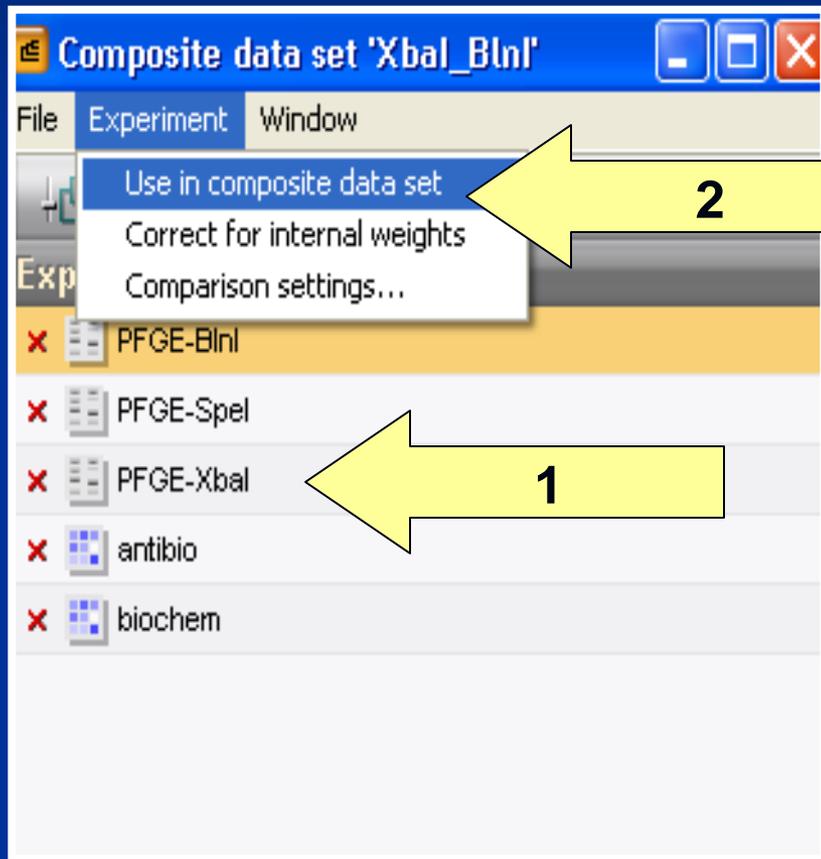
Enter name of experiment

Xbal_BInI

OK

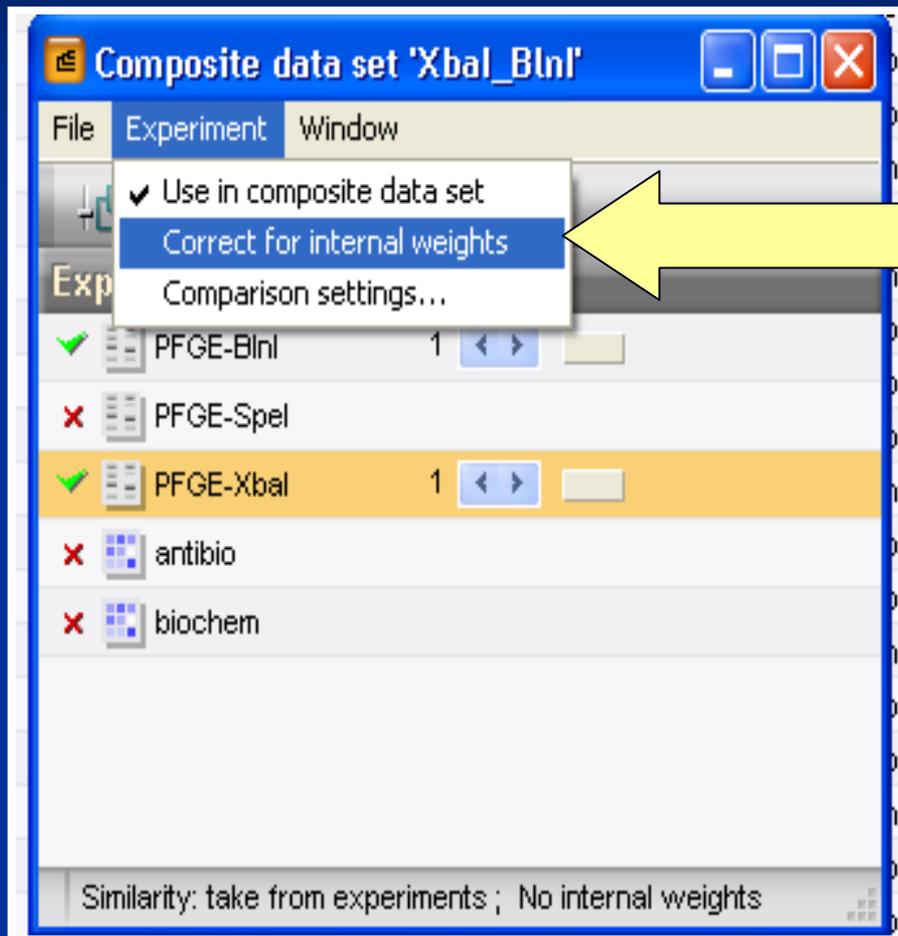
Cancel

Create a Composite Data Set



1. Highlight the experiment/character types you want included
2. Choose “Use in composite data set” from the “Experiment” window

Create a Composite Data Set



Choose "Correct for internal weights" from the "Experiments" window

Create a Composite Data Set

Notice new Composite Data Set in Experiments area

The screenshot displays the BioNumerics software interface. The main window shows a 'Database entries' table with columns for Index, Key, SourceState, Serotype, PFGE-Xbal-pattern, and six comparison columns (1-6). The 'Experiments' panel on the right lists various experiments with columns for Name and Type. A yellow arrow points from the 'Experiments' panel to the 'Database entries' table. A blue circle highlights the 'Xbal_Blnl' entry in the 'Experiments' panel.

Index	Key	SourceState	Serotype	PFGE-Xbal-pattern	1	2	3	4	5	6
1	AK__0725300169	AK	Agona	JABX01.0001	Co					
2	AK__0725600199	AK	I 4,[5],12:i-	JPXX01.1192	Co					
3	AK__0725600200	AK	Montevideo	JIXX01.0825	Co					
4	AK__0726100119	AK	Newport	JJPX01.0085	Co					
5	AK__0726200064	AK	Javiana	JGGX01.0966	Co					
6	AK__0726900180	AK	Typhimurium							
7	AK__0727700112	AK	Typhimurium							
8	AK__0727700113	AK	Typhimurium	JPXX01.0302	Un					
9	AK__0728300145	AK	Newport	JJPX01.0085	Co					
10	AK__0728300147	AK	London	TECX01.0004	Co					

Name	Type	
2	PFGE-Spel	Fingerprint ty
3	PFGE-Xbal	Fingerprint ty
4	antibio	Character typ
5	biochem	Character typ
6	Xbal_Blnl	Composite da

Using Composite Data Sets in Comparisons

The screenshot shows the 'Entry search' window with the 'Advanced query tool' tab selected. The search criteria are as follows:

Field	Value
PFGE-BlnI-status	<input checked="" type="checkbox"/>
Outbreak	<input type="checkbox"/>
SourceCountry	<input type="checkbox"/>
SourceCity	<input type="checkbox"/>
SourceCountry	<input type="checkbox"/>
SourceType	<input type="checkbox"/>
SourceSite	<input type="checkbox"/>
PatientAge	<input type="checkbox"/>
PatientSex	<input type="checkbox"/>
IsolatDate	<input type="checkbox"/>
ReceivedDate	<input type="checkbox"/>
UploadDate	*2009*
OtherStatelolate	<input type="checkbox"/>
Phagetype	<input type="checkbox"/>
Traveled_To	<input type="checkbox"/>
Exposure	<input type="checkbox"/>
Comment	<input type="checkbox"/>

On the right side, the following filters are selected:

- PFGE-BlnI
- PFGE-Spel
- PFGE-XbaI
- antibio
- biochem

At the bottom, the search options are:

- Search in list
- Negative search
- Case sensitive

Buttons: Clear, Search, Cancel

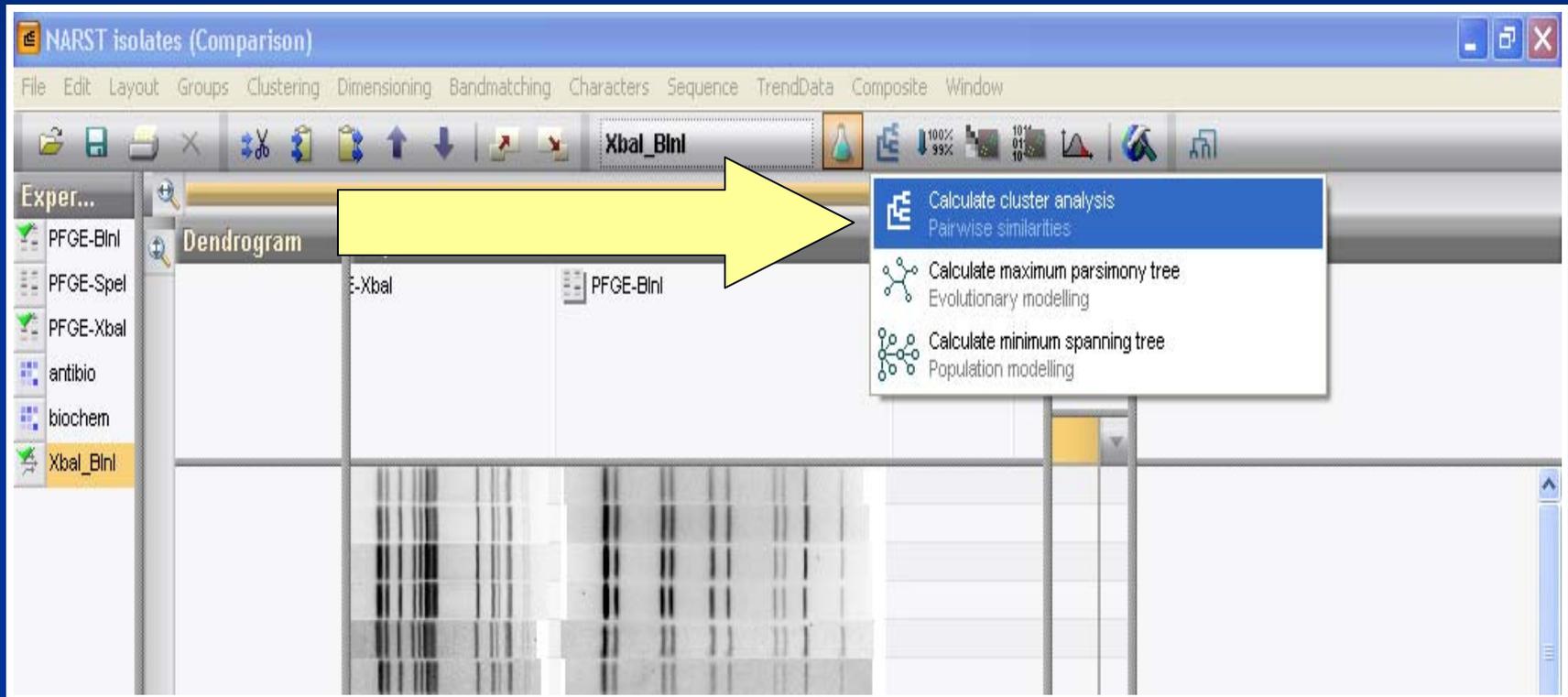
Example: search for all isolates that have both *XbaI* and *BlnI* experiment types uploaded in 2009

Using Composite Data Sets in Comparisons

The screenshot shows the 'NARST isolates (Comparison)' software interface. The window title is 'NARST isolates (Comparison)'. The menu bar includes File, Edit, Layout, Groups, Clustering, Dimensioning, Bandmatching, Characters, Sequence, TrendData, Composite, and Window. The toolbar shows various icons for file operations and analysis. The left sidebar lists experiments: PFGE-Blnl, PFGE-Spel, PFGE-Xbal, antibio, biochem, and Xbal_Blnl. The 'Xbal_Blnl' experiment is highlighted in yellow. The main area is split into 'Dendrogram' and 'Experiment data' panels. The 'Experiment data' panel shows a comparison of 'E-Xbal' and 'PFGE-Blnl' data sets, with a 'Xbal_Blnl' tab selected. The data is visualized as a gel electrophoresis image with multiple lanes and bands.

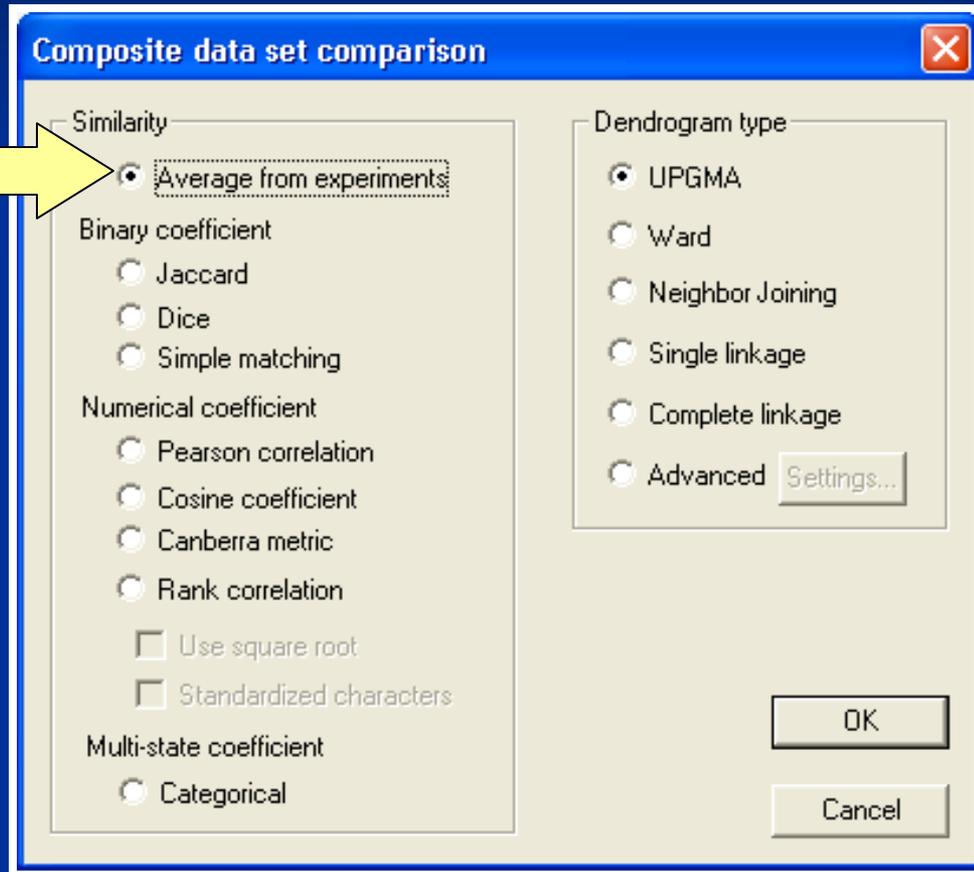
Bring up both sets of images, and then make sure the Composite Data Set is highlighted

Using Composite Data Sets in Comparisons



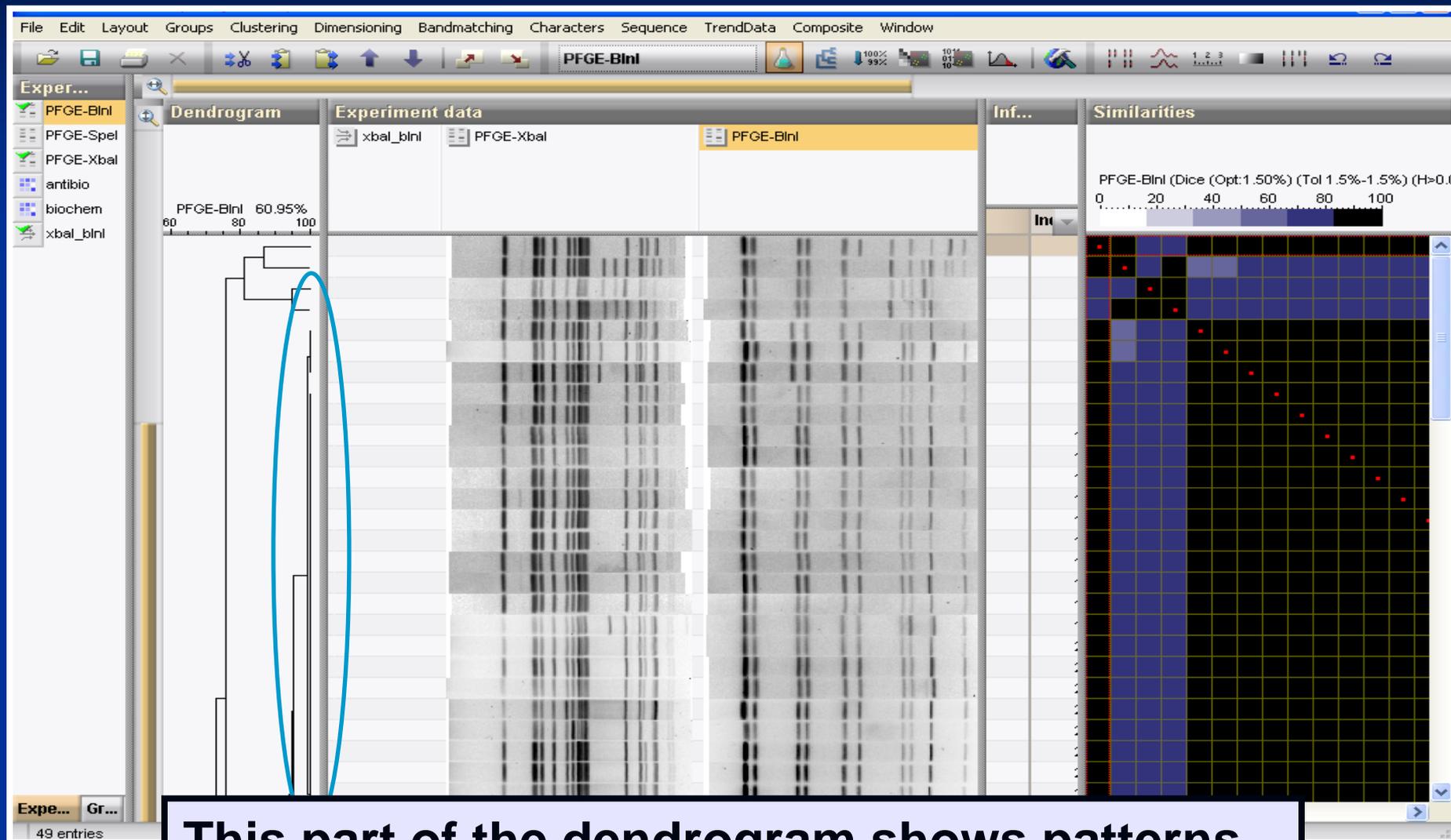
Choose “Calculate cluster analysis” under the dendrogram icon

Using Composite Data Sets in Comparisons



Under "Similarity," choose "Average from experiments"

Resulting Dendrogram



This part of the dendrogram shows patterns of both enzymes that are highly similar

Reminder:

- You should have as much data for both experiment types as possible
- The amount of missing data should be minimal

Summary

- Queried Local Databases
 - Query Tools for Local Databases
 - Performed a Simple Query
 - Queried by Isolate, Pattern and the Hot List Tool
- Queried National Databases
 - Query Tools for National Databases
 - Queried by Isolate and the Hot List Tool
 - Narrowed Search, Downloaded Selected Entries
- Composite Data Sets
 - What they are and Example of Usage
 - Create a Composite Data Set
 - Using Composite Data Sets in Comparisons

Questions?



Thank you for your attention

The findings and conclusions in this presentation are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention