BioSim2
User’s Manual

Version 2.0.03

EPA/600/R-05/150
November 2005
www.epa.gov

BioSim2
User's Manual

Version 2.0.03

A Program that Applies the Coefficient of
Biotic Similarity, B, to Complex Data Matrices

by

J. Gareth Pearson
U.S. Environmental Protection Agency
Environmental Sciences Division
P.O. Box 93478
Las Vegas, Nevada 89193-3478

Carlos F.A. Pinkham
Visiting Professor, Department of Biology
Norwich University
Northfield, Vermont 05663

Brian P. Reid
DMS Computing
Dartmouth Medical School
1 Rope Ferry Road
Hanover, New Hampshire 03755

Victor T. Chevalier
431 Isom Road, Suite 125,
San Antonio, TX 78216

November 2005

Notice

The information in this document has been funded in part by the United States Environmental Protection Agency in collaboration with Norwich University. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation by the EPA for use.

Contents

List of Figures

Number	Page

List of Tables

Acknowledgement

This project was supported in part by EPSCoR Baccalaureate College Summer Research Program under NSF Grant Number, EPS-0236976, and by the generous donation of Dr. Reid's time for the Java program from the Academic Computing Team, Information Technology Department of Norwich University.

Abstract

The Pinkham-Pearson index of similarity has been evaluated by EPA as one of the more powerful tools for comparing community structure in its rapid bioassessment protocol. However, its use has been limited because the program that ran it, BioSim1, was only available in DOS format. A user-friendly version of BioSim2 is now available in a Java format that can run on Windows, Mac OS, Linux, or any computer operating system that supports Java v1.4 or higher.

Section 1

Introduction

The index of similarity, *B* (Pinkham and Pearson, 1976), was proposed as a means for determining the impact of pollution on communities. Pinkham and Pearson showed that the index overcame many of the shortcomings inherent in other indexes used for the same purpose and was more versatile. Its use was coded in a DOS program (BioSim) (Pinkham *et al.,* 1975).

Since its publication, it has been widely used for diverse investigations. In 1989, Pfalkin *et al*. included it in EPA's rapid bioassessment protocols for use in streams and rivers. In 1990, it was identified by EPA as one of six commonly used community similarity indexes in a manual describing guidelines and standardized procedures for using benthic macroinvertebrates to evaluate the biological integrity of surface waters (Klemm *et al.,* 1990). At least one state, Vermont, has adopted *B* as a legal requirement for assessing surface water quality (Vermont Department of Environmental Conservation, 1990). Barbour *et al.* (1992), in a systematic comparison of the metrics proposed in EPA's rapid bioassessment protocol (Pfalkin *et al.,* 1989), concluded that *B* "may be the most appropriate metric to serve as a measure of community similarity."

In almost all published cases of its use, however, it was not being used to its full potential. Recognizing this, Pearson and Pinkham (1992) published a strategy for using it in an improved DOS version (BioSim1) (Gonzales *et al.,* 1993). However, this strategy also failed to encourage a widespread use of its full capabilities. It soon became apparent that the major reason for this shortcoming was the DOS platform of BioSim1. This manual describes the use of BioSim2, a Java program that finally overcomes that shortcoming.

Section 2

Use of This Manual

It is assumed that users of this manual are familiar with the index of similarity, *B* (Pinkham and Pearson, 1976), and the use of dendrograms in the analysis of complex environmental data sets. If not, users should refer to previous publications on the topic (Bonham-Carter, 1967; Pinkham and Pearson, 1976; Kaesler, 1970; Klemm *et al*., 1990; Pearson and Pinkham, 1992; and Gonzales *et al*., 1993).

Section 3

Installing and Starting BioSim2

Running BioSim2 requires that a Java Runtime Environment (JRE) of version 1.4.1 or higher be installed. Sun Microsystems, Inc. offers the Java Runtime Environment® for Windows®, Linux®, and Solaris®. Apple® offers Java® for the Mac OS. Links to download the appropriate JRE can be found at:

http://www.java.com

BioSim2 will run on most modern computers with Windows 98 SE or higher or Apple OS X or higher.

Once the JRE is installed, biosim2.jar is the only file needed to run BioSim2. The biosim2.jar file is a Java® archive which uses the same format as a compressed Zip file. To install, copy the biosim2.jar file to any convenient location or network directory. The most recent version of the program and its associated User's Manual are available at:

http://www2.norwich.edu/pinkhamc/

Start BioSim2 by double-clicking on the biosim2.jar file. The file menu at the top of the opening screen (see page 8) is used for most commands. The data are imported, or entered manually. The program is run using the "Process Data" command under the "Action" menu. Results are accessed through the various tabs at the bottom of the window.

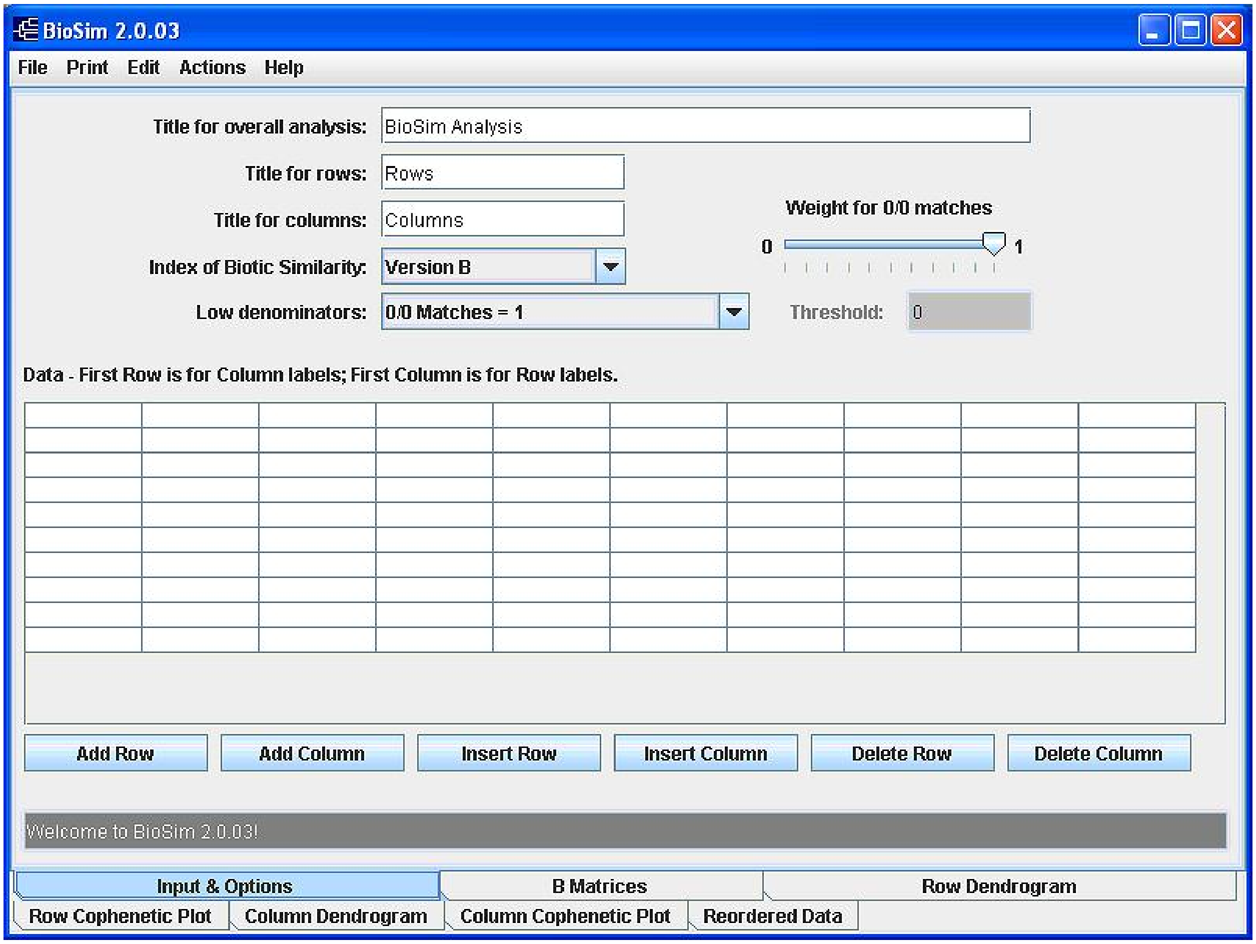
Section 4

Using BioSim2

BioSim2 was written with an entirely new appearance and a simpler approach while retaining most of the features discussed in Pearson and Pinkham (1992). A major change was to drop the agglomerative clustering method for forming dendrograms used in former versions of BioSim (Bonham-Carter, 1967) in favor of the simpler, average link method (Pankhurst, 1991). This method searches through each possible pair of unlinked parameters and an average B-value is calculated for the pair [see Pearson and Pinkham (1992) for a definition of these terms]. This average B-value is determined in three possible situations. 1) Both parameters are not found in any other cluster formed already. This normally happens toward the beginning of the process. In this case, the average is their single B-value. 2) One of these parameters is already part of another cluster. All B-values involving the unlinked parameter and every member of this other cluster are averaged. 3) Both parameters belong to existing clusters. All B-values between the two clusters are averaged. After searching the complete set of unlinked pairs, the pair with the highest average B-value is linked at that average value.

In addition to the above, the authors decided 1) to reduce the presentation of the program to a single screen that would sequentially display the original data and then the results; 2) to make the conditions of the original data matrix flexible enough that most presently used spreadsheet formats would be acceptable; 3) to provide a plot of the actual B-values used to calculate each average B-value found on the dendrogram, to visualize how well each joining branch of the dendrogram represents the actual distribution of B-values that formed the joining branch; 4) to reorient the dendrogram 180° so that the plot resulting from number 3, above, could be easily compared with the dendrogram; 5) to eliminate the matrix of cophenetic correlation coefficients (Kaesler, 1970) in favor of the simpler and more valuable cophenetic correlation coefficient for the entire dendrogram; 6) to enable BioSim2 to accept and analyze data matrices with missing data points (incomplete data matrices); and 7) to eliminate many of the choices that needed to be made in BioSim1 by making most options automatic in BioSim2.

Figure 1 represents the opening blank screen of BioSim2. The screen consists of four basic parts: a standard menu bar along the top; a data input / results display area; a message window near the bottom; and a series of “tabs” at the very bottom that provide access to panels for data input and, after running the program with data, to display results with a different array of “tabs.”


BioSim 2.0.03

Title for overall analysis: BioSim Analysis

Title for rows: Rows

Title for columns: Columns

Index of Biotic Similarity: Version B

Low denominators: 0/0 Matches = 1

Weight for 0/0 matches: 0 – 1

Threshold: 0

Data - First Row is for Column labels; First Column is for Row labels.

Welcome to BioSim 2.0.03!

Figure 1. Opening Screen of BioSim2.

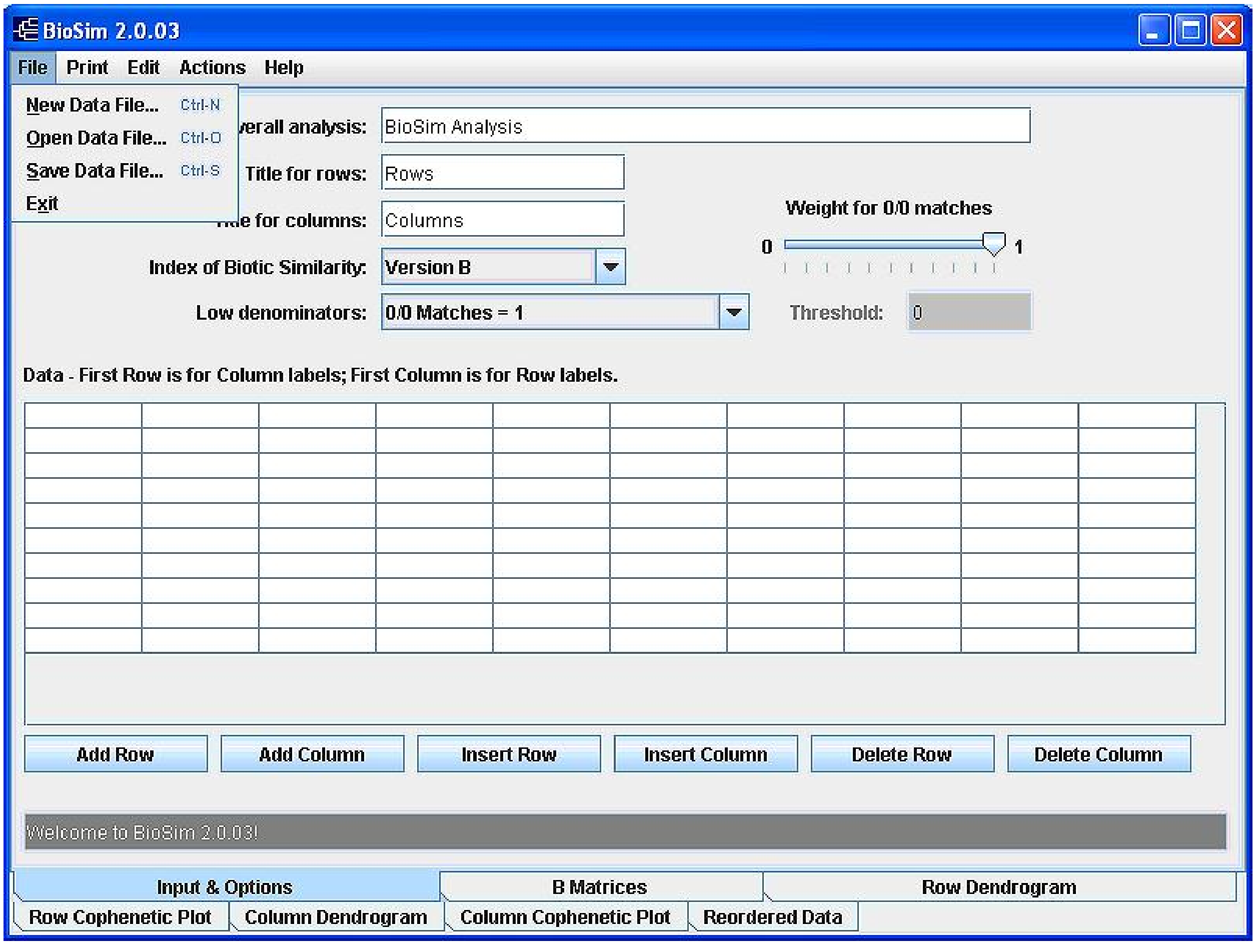
4.1 Standard Menu Functions

4.1.1 "File" Drop-down Menu

Figure 2 represents the opening screen with the "File" drop-down menu opened. This menu contains four options: New Data File, Open Data File, Save Data File, and Exit. The first three options are also available using the keyboard shortcut in the drop-down menu.

New Data File: If an existing file is open, the "New Data File" will clear the data in the spreadsheet portion of the screen. Currently, it does not clear the title data or the options selected. These data require manual entry.

Open Data File: The "Open Data File" option opens existing data files that are in CSV format (comma separated values; also referred to as comma delimited data). Most of the modern spreadsheet programs allow you to save files in this CSV format. Titles and options data may be entered manually using the *Input & Options* tab and its associated drop-down menus or by opening a CSV spreadsheet containing these data. When ***importing all*** of the data (titles, options, and numerical data) from a spreadsheet file, the data must follow the format shown in Table 1 and must be saved as a CSV file.



verall analysis: BioSim Analysis

Title for rows: Rows

e for columns: Columns

Index of Biotic Similarity: Version B

Low denominators: 0/0 Matches = 1

Weight for 0/0 matches: 0 – 1

Threshold: 0

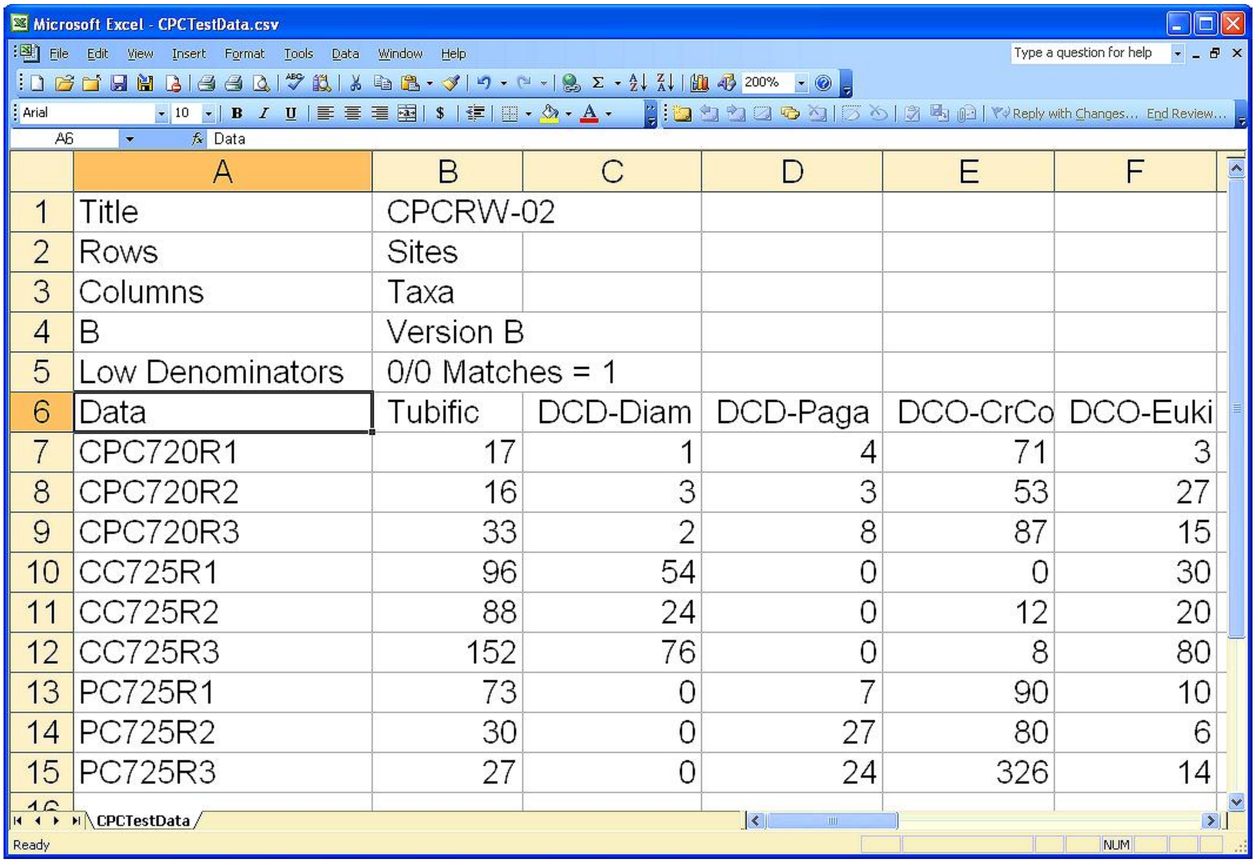
Data - First Row is for Column labels; First Column is for Row labels.

Welcome to BioSim 2.0.03!

Figure 2. Opening Screen with "File" Drop-down Menu Opened.

The first five rows of an external spreadsheet must contain the input titles and program options data selected by the user. The sixth row must contain the titles to be used for the columns and the remaining rows contain the actual data set with the first column being the titles for the rows. The values in column A, rows 1-5 are fixed and must appear as in Table 1. The values in column B, rows 1-3 are user-entered titles and the values in rows 4 and 5 are options for calculating the similarity index (see Pearson and Pinkham (1992) for a complete explanation of these options). B is the standard similarity index and B1 is the same index, but the raw data values are converted to relative abundances. The "Low denominators" field allows the user to define how 0/0 matches and other matches with low denominators will be dealt with during the calculation of the similarity index. The 0/0 matches can be set to equal 1 (highest weight), 0 (lowest weight), any value between 0.0 and 1.0 (using the "Weight for 0/0 matches" slider bar), or ignored altogether in the calculations. Matches with low denominators can be ignored by setting the "threshold" to the value of the denominator at or below which matches should not be used in calculating B. This value is entered in the "Threshold" box on the opening screen.

Table 1. Example of CSV Spreadsheet as an Input File (partial data set *with* titles and options).



	A	B	C	D	E	F
1	Title	CPCRW-02				
2	Rows	Sites				
3	Columns	Taxa				
4	B	Version B				
5	Low Denominators	0/0 Matches = 1				
6	Data	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki
7	CPC720R1	17	1	4	71	3
8	CPC720R2	16	3	3	53	27
9	CPC720R3	33	2	8	87	15
10	CC725R1	96	54	0	0	30
11	CC725R2	88	24	0	12	20
12	CC725R3	152	76	0	8	80
13	PC725R1	73	0	7	90	10
14	PC725R2	30	0	27	80	6
15	PC725R3	27	0	24	326	14

Important: There must be a value in column A, row 6 (the word "Data" is written here by BioSim2 when saving the file).

As indicated above, the titles and options data may be entered directly on the screen accessed with the *Input & Options* tab. The maximum number of characters for these titles is 256, but long titles can cause some problems with the data displays. Data may also be entered directly using the table portion of this screen and the buttons (Add Row, Add Column, Insert Row, Insert Column, Delete Row, and Delete Column) directly below the data input table. You can also just import the data using the Open Data File option. This file must follow the format as shown in Table 2. The first row always contains the titles for the columns and the first column always contains the titles for the rows.

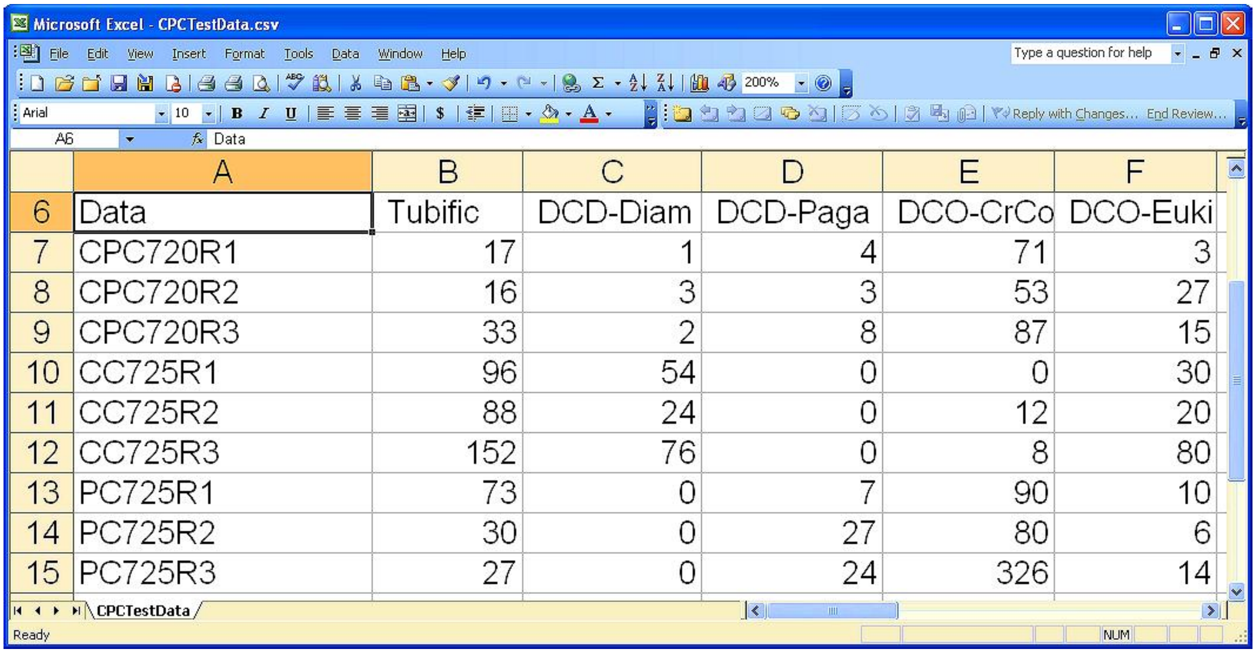
Important: There must be a value in column A, row 1 (the word "Data" is written here by BioSim2 when saving the file).

It is also possible to use the standard copy (Ctrl-C), cut (Ctrl-X), and paste (Ctrl-V) features of Windows to move data from a spreadsheet directly into the table on the *Input & Options* tab. When doing this, make sure that cell A-1 has something in it, otherwise, the column titles in row 1 shift to the left one column.

Save Data File: The "Save Data File" option saves all of the values contained in the spreadsheet under the *Input & Options* tab as a CSV file. The results look like the file in Table 1 when opened in Excel and contain the user input titles, options selected, and the numerical data in the table.

Exit: “Exit” option closes BioSim2. All data are lost if they have not been previously saved.

Table 2. Example of CSV Spreadsheet as an Input File (partial data set *without* titles or user options).

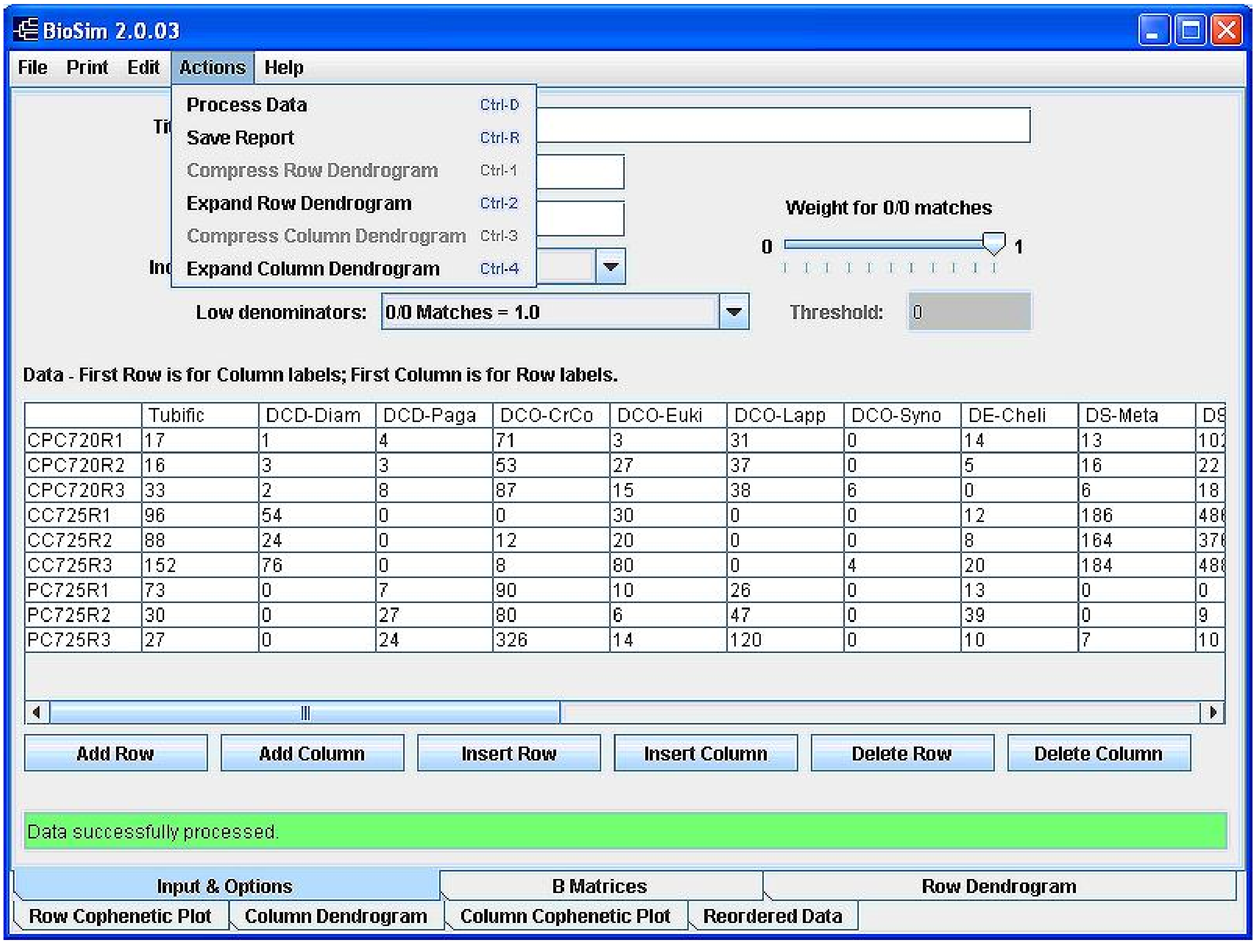


	A	B	C	D	E	F
6	Data	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki
7	CPC720R1	17	1	4	71	3
8	CPC720R2	16	3	3	53	27
9	CPC720R3	33	2	8	87	15
10	CC725R1	96	54	0	0	30
11	CC725R2	88	24	0	12	20
12	CC725R3	152	76	0	8	80
13	PC725R1	73	0	7	90	10
14	PC725R2	30	0	27	80	6
15	PC725R3	27	0	24	326	14

4.1.2 “Actions” Drop-down Menu

Figure 3 represents the opening screen with all the data fields completed and the “Actions” drop-down menu opened.

Process Data: To run the program, select “Process Data” from the Actions menu or via the keyboard shortcut, “Ctrl-D.” If the program executes properly, the following message will appear in green in the message window, “Data successfully processed” (see Figure 3). If the program does not execute properly, the message window will be red and an error message will appear to assist in trouble-shooting problems with the input data. Although the program processes small data sets almost instantaneously, data sets with 50 or more rows or columns take somewhat longer. When you select “Process Data,” an hourglass appears to let you know that the program is running smoothly. You can maximize the performance of BioSim2 by minimizing the number of other applications running on your computer at the time you process the data.



Low denominators: 0/0 Matches = 1.0

Threshold: 0

Data - First Row is for Column labels; First Column is for Row labels.

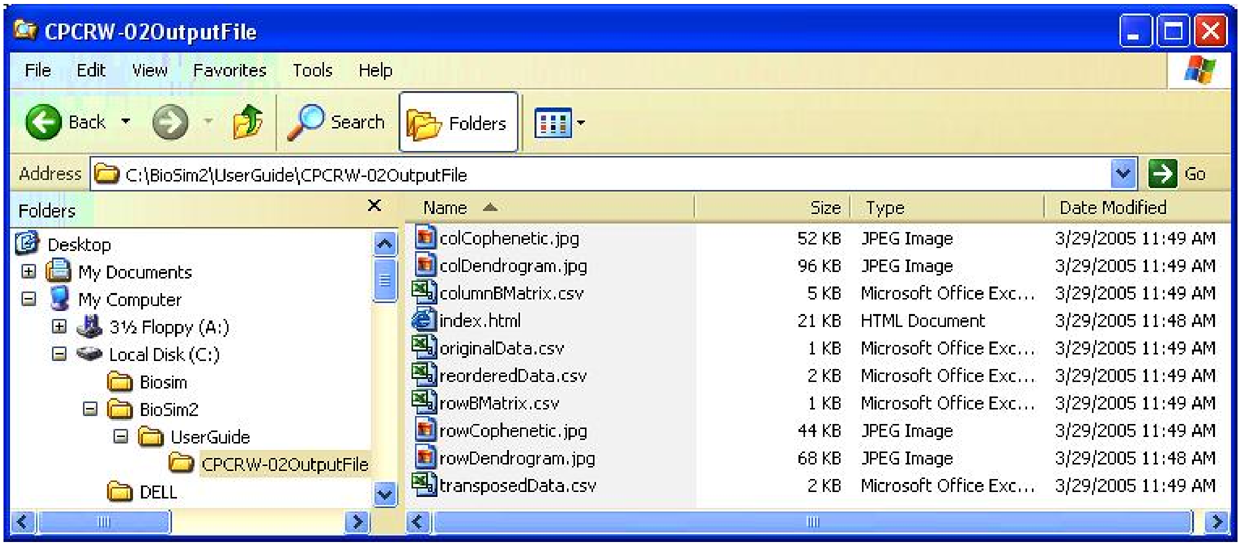
	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki	DCO-Lapp	DCO-Syno	DE-Cheli	DS-Meta	DS
CPC720R1	17	1	4	71	3	31	0	14	13	10
CPC720R2	16	3	3	53	27	37	0	5	16	22
CPC720R3	33	2	8	87	15	38	6	0	6	18
CC725R1	96	54	0	0	30	0	0	12	186	48
CC725R2	88	24	0	12	20	0	0	8	164	37
CC725R3	152	76	0	8	80	0	4	20	184	48
PC725R1	73	0	7	90	10	26	0	13	0	0
PC725R2	30	0	27	80	6	47	0	39	0	9
PC725R3	27	0	24	326	14	120	0	10	7	10

Data successfully processed.

Figure 3. Opening Screen with "Actions" Drop-down Menu Opened.

Save Report: Selecting the "Save Report" option from the Actions menu gives the user the ability to identify or create a folder in which to save all of the program outputs using a standard Windows dialog box. This option is also available via the keyboard shortcut Ctrl-R. Figure 4 shows the files that are automatically created and saved to a folder when this option is selected. The dendrograms and the cophenetic correlation plots for both rows and columns are saved as .jpg files. All of the data matrices (original input data, matrices of matrix of similarity values (Bs), and the reordered and transposed data files) are saved as CSV files. An html file, index.html, is also created which allows the user to view the output from the program using any of the common Internet browsers. Depending on the size of the data matrix and the speed of the computer, this may take some time. During this time, the standard hourglass symbol will appear. Once the output files have been saved, you will get a green message in the message window that reads "Report saved."

Expand and Compress Row and Column Dendrograms: Selecting either the "Expand Row or Column Dendrogram" option from the Actions menu increases the size of the row or column dendrogram by 50 percent. This is useful when the matrix size is greater than 50 rows or columns. These options are also available via the keyboard shortcuts Ctrl-2 and Ctrl-4 (Figure 3). Once a dendrogram has been expanded, it can be compressed to its original size by selecting either the “Compress Row or Column Dendrogram” option from the Actions menu. Again, these options are also available via the keyboard shortcuts Ctrl-1 and Ctrl-3 (Figure 3).



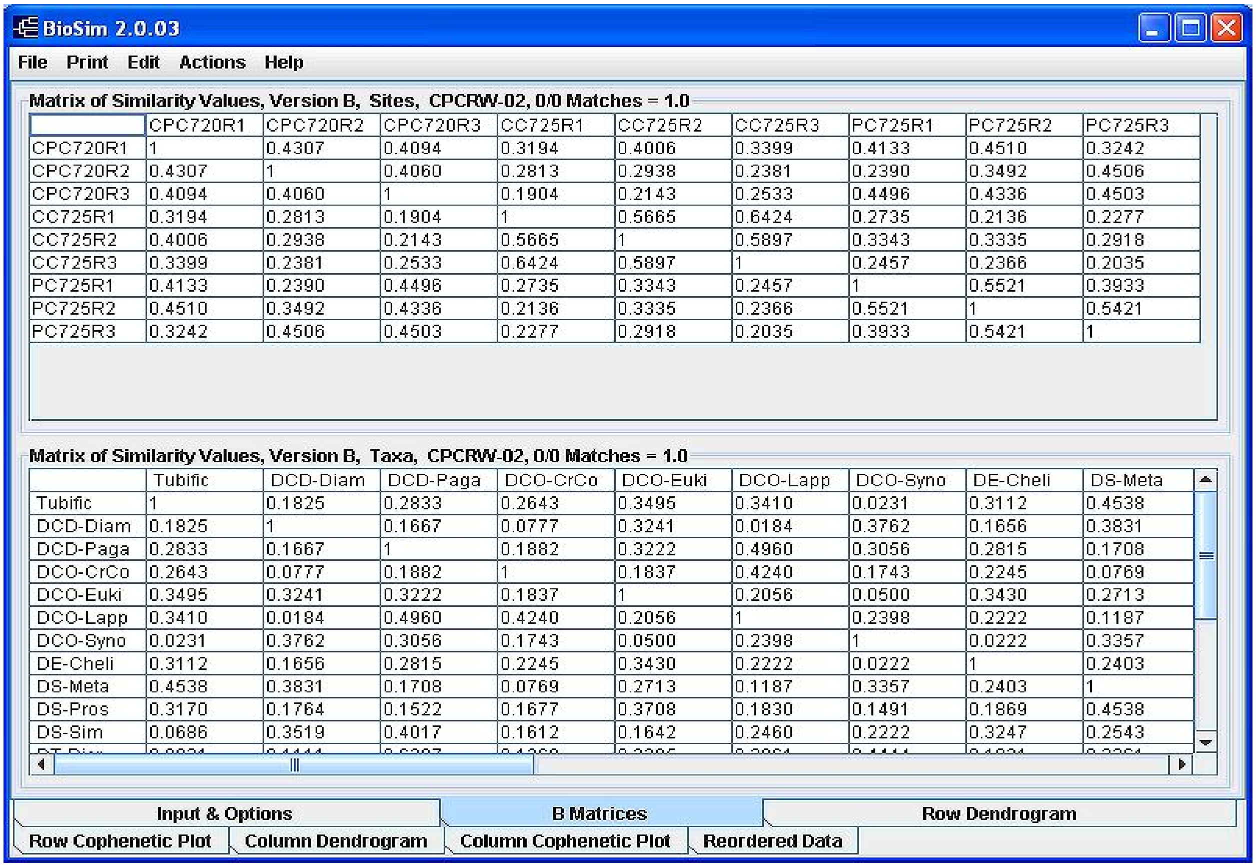
Address C:\BioSim2\UserGuide\CPCRW-02OutputFile

Name	Size	Type	Date Modified
colCophenetic.jpg	52 KB	JPEG Image	3/29/2005 11:49 AM
colDendrogram.jpg	96 KB	JPEG Image	3/29/2005 11:49 AM
columnBMatrix.csv	5 KB	Microsoft Office Exc...	3/29/2005 11:49 AM
index.html	21 KB	HTML Document	3/29/2005 11:48 AM
originalData.csv	1 KB	Microsoft Office Exc...	3/29/2005 11:49 AM
reorderedData.csv	2 KB	Microsoft Office Exc...	3/29/2005 11:49 AM
rowBMatrix.csv	1 KB	Microsoft Office Exc...	3/29/2005 11:49 AM
rowCophenetic.jpg	44 KB	JPEG Image	3/29/2005 11:49 AM
rowDendrogram.jpg	68 KB	JPEG Image	3/29/2005 11:48 AM
transposedData.csv	2 KB	Microsoft Office Exc...	3/29/2005 11:49 AM

Figure 4. Files Created and Saved Using the “Save Report” Option.

Once you have successfully processed the data, you can view the results by selecting the various tabs at the very bottom of the screen. These are shown in Figures 5-10.

Examples of saved graphics files are shown in Appendix A.



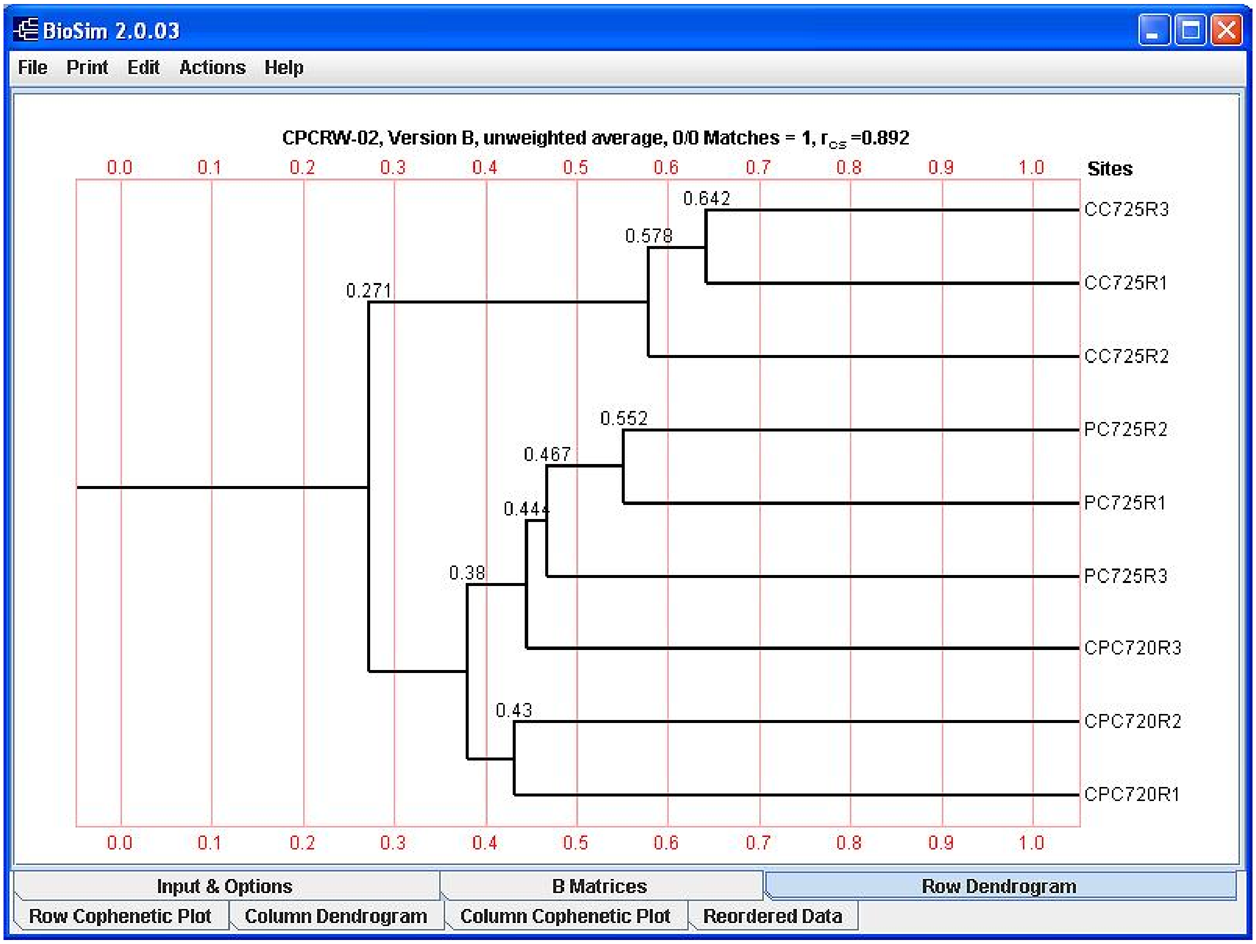
Matrix of Similarity Values, Version B, Sites, CPCRW-02, 0/0 Matches = 1.0

	CPC720R1	CPC720R2	CPC720R3	CC725R1	CC725R2	CC725R3	PC725R1	PC725R2	PC725R3
CPC720R1	1	0.4307	0.4094	0.3194	0.4006	0.3399	0.4133	0.4510	0.3242
CPC720R2	0.4307	1	0.4060	0.2813	0.2938	0.2381	0.2390	0.3492	0.4506
CPC720R3	0.4094	0.4060	1	0.1904	0.2143	0.2533	0.4496	0.4336	0.4503
CC725R1	0.3194	0.2813	0.1904	1	0.5665	0.6424	0.2735	0.2136	0.2277
CC725R2	0.4006	0.2938	0.2143	0.5665	1	0.5897	0.3343	0.3335	0.2918
CC725R3	0.3399	0.2381	0.2533	0.6424	0.5897	1	0.2457	0.2366	0.2035
PC725R1	0.4133	0.2390	0.4496	0.2735	0.3343	0.2457	1	0.5521	0.3933
PC725R2	0.4510	0.3492	0.4336	0.2136	0.3335	0.2366	0.5521	1	0.5421
PC725R3	0.3242	0.4506	0.4503	0.2277	0.2918	0.2035	0.3933	0.5421	1

Matrix of Similarity Values, Version B, Taxa, CPCRW-02, 0/0 Matches = 1.0

	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki	DCO-Lapp	DCO-Syno	DE-Cheli	DS-Meta
Tubific	1	0.1825	0.2833	0.2643	0.3495	0.3410	0.0231	0.3112	0.4538
DCD-Diam	0.1825	1	0.1667	0.0777	0.3241	0.0184	0.3762	0.1656	0.3831
DCD-Paga	0.2833	0.1667	1	0.1882	0.3222	0.4960	0.3056	0.2815	0.1708
DCO-CrCo	0.2643	0.0777	0.1882	1	0.1837	0.4240	0.1743	0.2245	0.0769
DCO-Euki	0.3495	0.3241	0.3222	0.1837	1	0.2056	0.0500	0.3430	0.2713
DCO-Lapp	0.3410	0.0184	0.4960	0.4240	0.2056	1	0.2398	0.2222	0.1187
DCO-Syno	0.0231	0.3762	0.3056	0.1743	0.0500	0.2398	1	0.0222	0.3357
DE-Cheli	0.3112	0.1656	0.2815	0.2245	0.3430	0.2222	0.0222	1	0.2403
DS-Meta	0.4538	0.3831	0.1708	0.0769	0.2713	0.1187	0.3357	0.2403	1
DS-Pros	0.3170	0.1764	0.1522	0.1677	0.3708	0.1830	0.1491	0.1869	0.4538
DS-Sim	0.0686	0.3519	0.4017	0.1612	0.1642	0.2460	0.2222	0.3247	0.2543

Figure 5. Tab “B Matrices” Showing the Similarity Values (B) for Rows (top) and Columns (bottom).



CPCRW-02, Version B, unweighted average, 0/0 Matches = 1, r_{cs} =0.892

Figure 6. Tab "Row Dendrogram" Showing the Dendrogram for Rows.

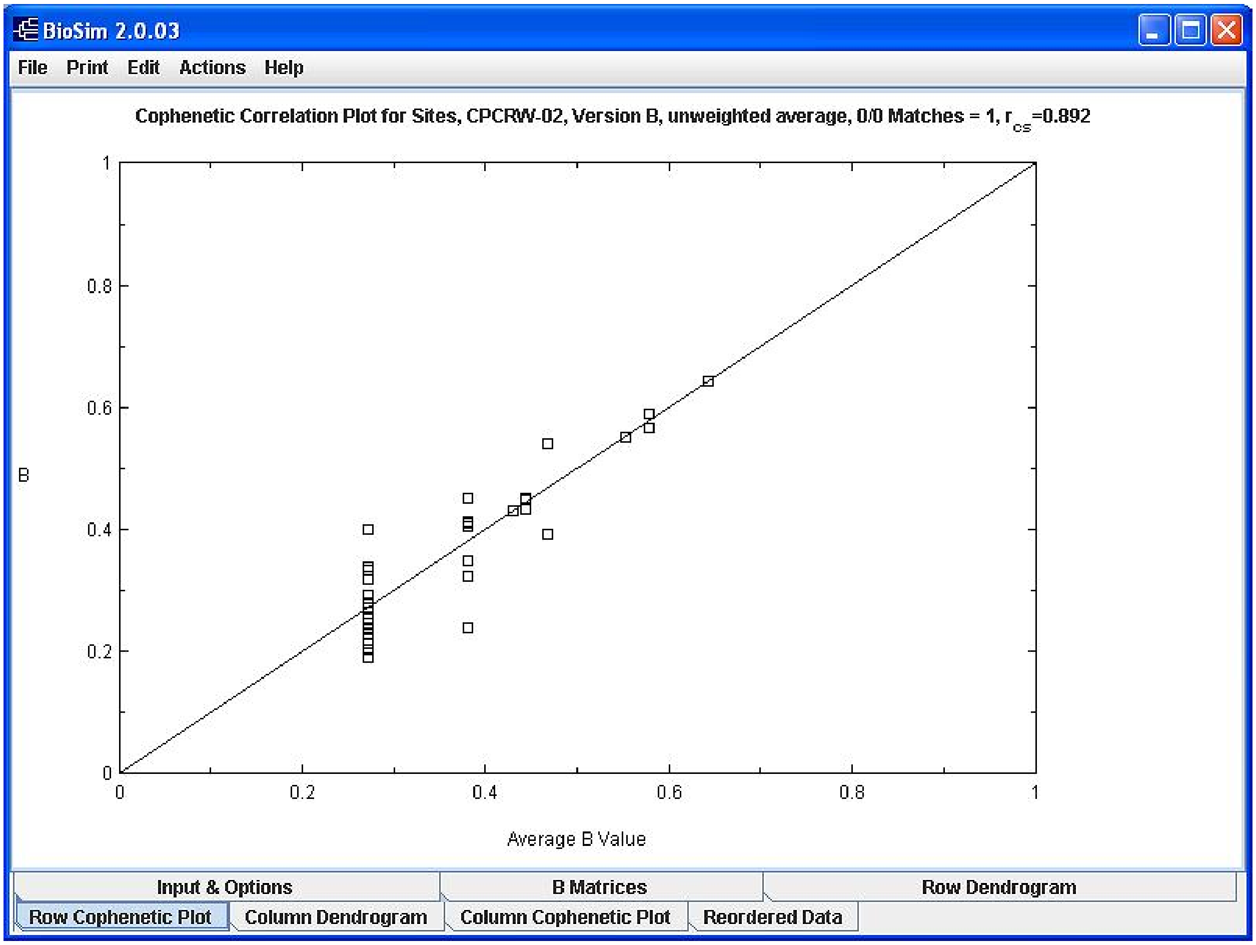


Figure 7. Tab “Row Cophenetic Plot” Showing the Cophenetic Correlation Plot for Rows.

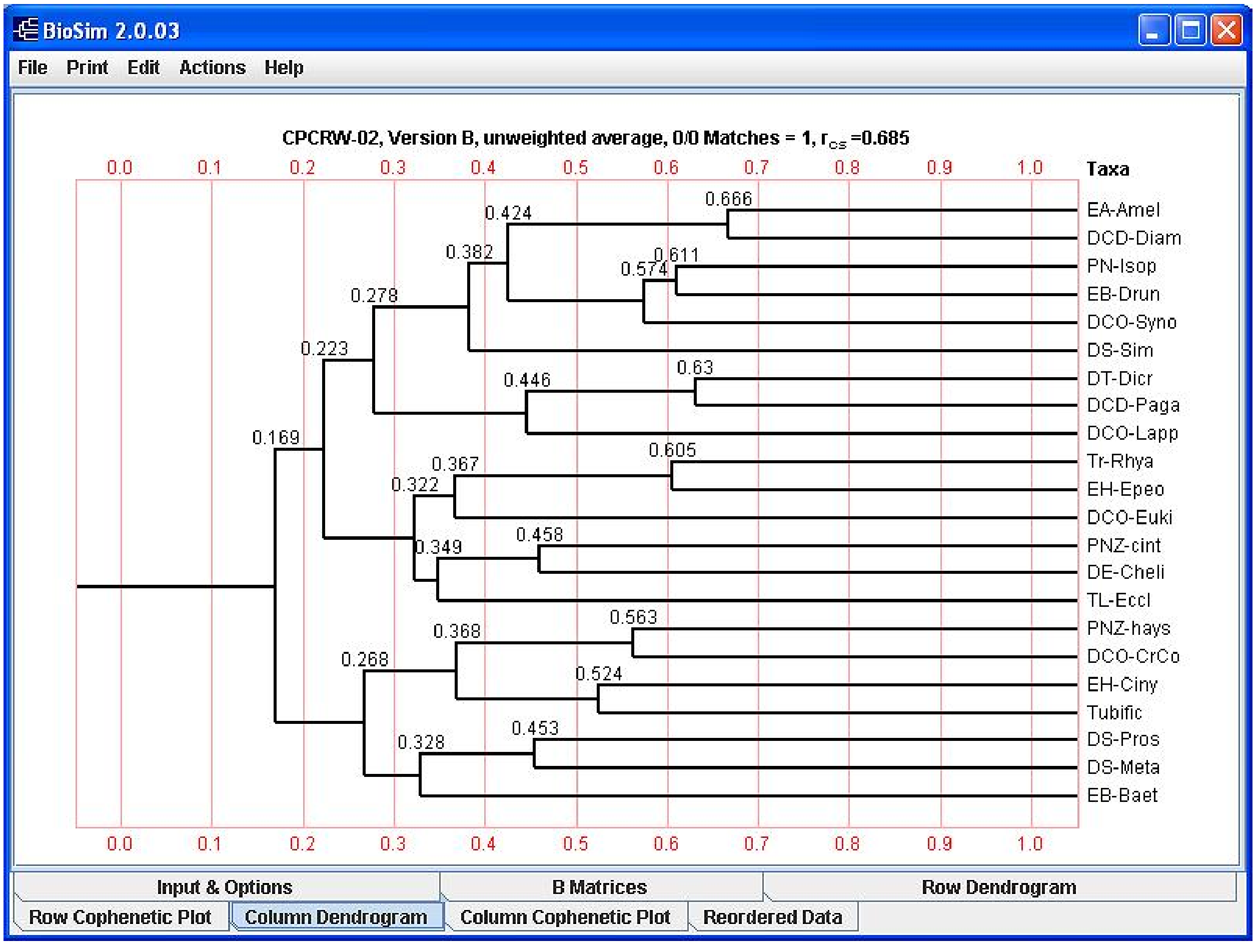


Figure 8. Tab “Column Dendrogram” Showing the Dendrogram for Columns.

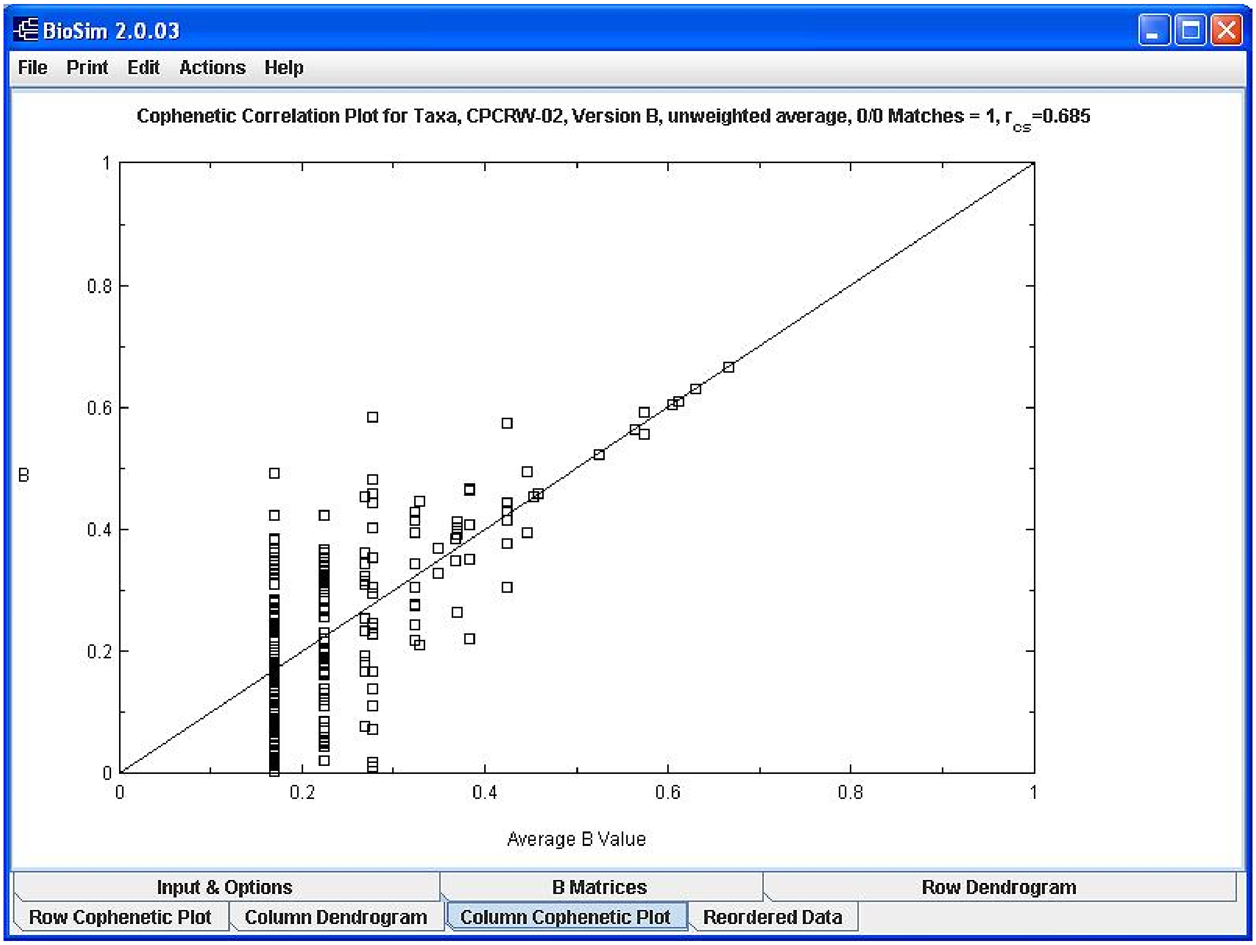
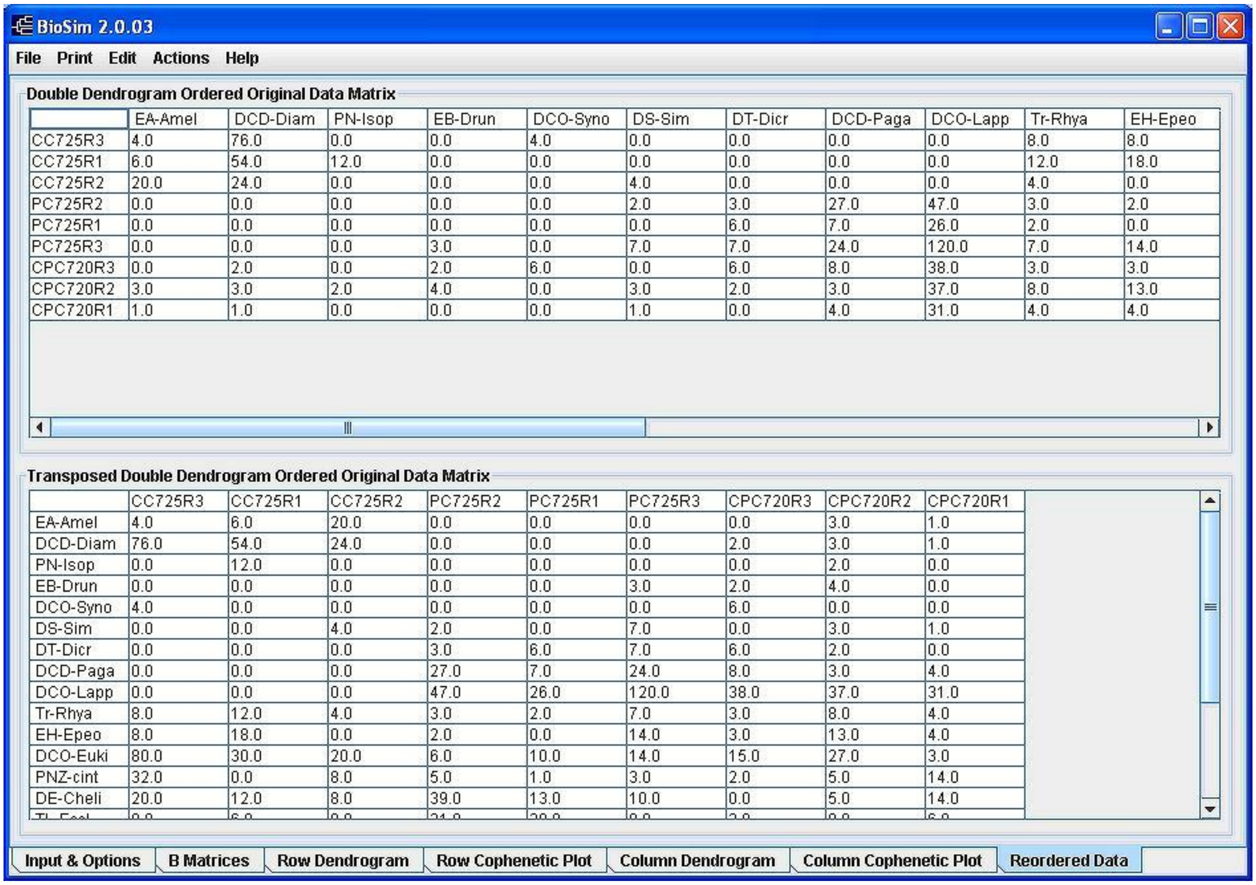


Figure 9. Tab "Column Cophenetic Plot" Showing the Cophenetic Correlation Plot for Columns.



Double Dendrogram Ordered Original Data Matrix

	EA-Amel	DCD-Diam	PN-Isop	EB-Drun	DCO-Syno	DS-Sim	DT-Dicr	DCD-Paga	DCO-Lapp	Tr-Rhya	EH-Epeo
CC725R3	4.0	76.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	8.0	8.0
CC725R1	6.0	54.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	18.0
CC725R2	20.0	24.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	4.0	0.0
PC725R2	0.0	0.0	0.0	0.0	0.0	2.0	3.0	27.0	47.0	3.0	2.0
PC725R1	0.0	0.0	0.0	0.0	0.0	0.0	6.0	7.0	26.0	2.0	0.0
PC725R3	0.0	0.0	0.0	3.0	0.0	7.0	7.0	24.0	120.0	7.0	14.0
CPC720R3	0.0	2.0	0.0	2.0	6.0	0.0	6.0	8.0	38.0	3.0	3.0
CPC720R2	3.0	3.0	2.0	4.0	0.0	3.0	2.0	3.0	37.0	8.0	13.0
CPC720R1	1.0	1.0	0.0	0.0	0.0	1.0	0.0	4.0	31.0	4.0	4.0

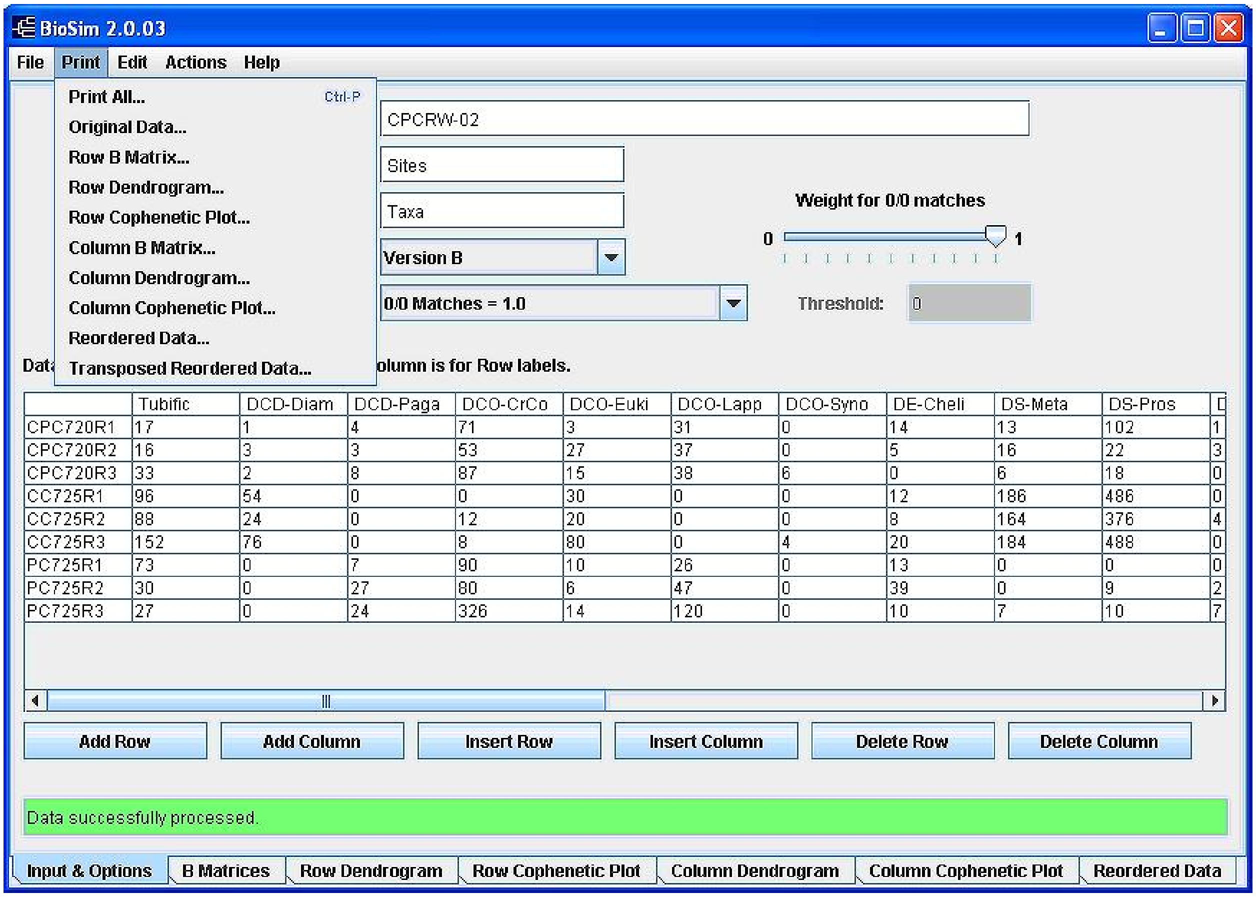
Transposed Double Dendrogram Ordered Original Data Matrix

	CC725R3	CC725R1	CC725R2	PC725R2	PC725R1	PC725R3	CPC720R3	CPC720R2	CPC720R1
EA-Amel	4.0	6.0	20.0	0.0	0.0	0.0	0.0	3.0	1.0
DCD-Diam	76.0	54.0	24.0	0.0	0.0	0.0	2.0	3.0	1.0
PN-Isop	0.0	12.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
EB-Drun	0.0	0.0	0.0	0.0	0.0	3.0	2.0	4.0	0.0
DCO-Syno	4.0	0.0	0.0	0.0	0.0	0.0	6.0	0.0	0.0
DS-Sim	0.0	0.0	4.0	2.0	0.0	7.0	0.0	3.0	1.0
DT-Dicr	0.0	0.0	0.0	3.0	6.0	7.0	6.0	2.0	0.0
DCD-Paga	0.0	0.0	0.0	27.0	7.0	24.0	8.0	3.0	4.0
DCO-Lapp	0.0	0.0	0.0	47.0	26.0	120.0	38.0	37.0	31.0
Tr-Rhya	8.0	12.0	4.0	3.0	2.0	7.0	3.0	8.0	4.0
EH-Epeo	8.0	18.0	0.0	2.0	0.0	14.0	3.0	13.0	4.0
DCO-Euki	80.0	30.0	20.0	6.0	10.0	14.0	15.0	27.0	3.0
PNZ-cint	32.0	0.0	8.0	5.0	1.0	3.0	2.0	5.0	14.0
DE-Cheli	20.0	12.0	8.0	39.0	13.0	10.0	0.0	5.0	14.0

Figure 10.Tab “Reordered Data” Showing the Data Reordered According to Both Row and Column Dendrograms.

4.1.3 "Print" Drop-down Menu

Figure 11 represents the opening screen with the "Print" drop-down menu opened. From this menu, you can select and print any of the following outputs: Row Dendrogram, Column Dendrogram, Reordered Data, Transposed Reordered Data, Row B Matrix, Column B Matrix, Row Cophenetic Plot, or the Column Cophenetic Plot. You can also print all of these outputs by selecting "Print All" or the keyboard shortcut Ctrl-P.



CPCRW-02

Sites

Taxa

Version B

0/0 Matches = 1.0

Weight for 0/0 matches

0 1

Threshold: 0

Dat olumn is for Row labels.

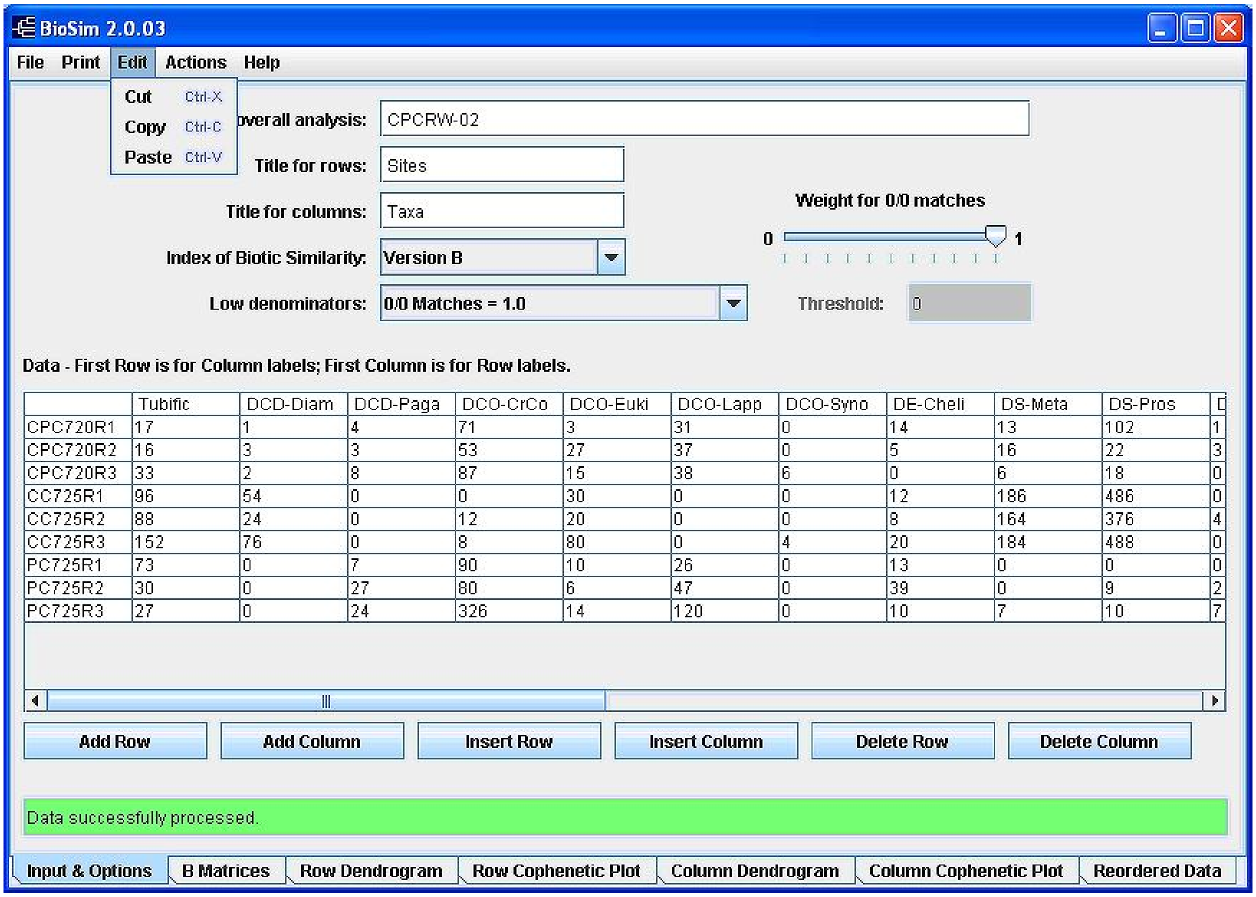
	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki	DCO-Lapp	DCO-Syno	DE-Cheli	DS-Meta	DS-Pros	D
CPC720R1	17	1	4	71	3	31	0	14	13	102	1
CPC720R2	16	3	3	53	27	37	0	5	16	22	3
CPC720R3	33	2	8	87	15	38	6	0	6	18	0
CC725R1	96	54	0	0	30	0	0	12	186	486	0
CC725R2	88	24	0	12	20	0	0	8	164	376	4
CC725R3	152	76	0	8	80	0	4	20	184	488	0
PC725R1	73	0	7	90	10	26	0	13	0	0	0
PC725R2	30	0	27	80	6	47	0	39	0	9	2
PC725R3	27	0	24	326	14	120	0	10	7	10	7

Data successfully processed.

Figure 11. Opening Screen with "Print" Drop-down Menu Opened.

4.1.4 “Edit” Drop-down Menu

Figure 12 represents the opening screen with the “Edit” drop-down menu opened. From this menu, you can highlight any of the input data fields and use the standard cut, copy, or paste functions. The standard keyboard shortcuts (Ctrl-X, Ctrl-C, and Ctrl-V) also work for these functions.



verall analysis: CPCRW-02

Title for rows: Sites

Title for columns: Taxa

Index of Biotic Similarity: Version B

Low denominators: 0/0 Matches = 1.0

Threshold: 0

Data - First Row is for Column labels; First Column is for Row labels.

	Tubific	DCD-Diam	DCD-Paga	DCO-CrCo	DCO-Euki	DCO-Lapp	DCO-Syno	DE-Cheli	DS-Meta	DS-Pros	D
CPC720R1	17	1	4	71	3	31	0	14	13	102	1
CPC720R2	16	3	3	53	27	37	0	5	16	22	3
CPC720R3	33	2	8	87	15	38	6	0	6	18	0
CC725R1	96	54	0	0	30	0	0	12	186	486	0
CC725R2	88	24	0	12	20	0	0	8	164	376	4
CC725R3	152	76	0	8	80	0	4	20	184	488	0
PC725R1	73	0	7	90	10	26	0	13	0	0	0
PC725R2	30	0	27	80	6	47	0	39	0	9	2
PC725R3	27	0	24	326	14	120	0	10	7	10	7

Data successfully processed.

Figure 12. Opening Screen with “Edit” Drop-down Menu Opened.

4.1.5 “Help” Drop-down Menu

The “Help” menu has two options. The “About BioSim2” option provides information on the authors and the version number of BioSim2. At present, the “Help for BioSim2” provides access to an html version of this User’s Manual.

References

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: metric redundancy and variability among reference stream sites, Environmental Toxicology and Chemistry, 11(4): 437-449.

Bonham-Carter, G.F. 1967. Fortran IV program for Q-mode cluster analysis of non-quantitative data using IBM 7090/7094 computers. Kans. Geol. Surv., Computer Contribution No. 17.

Gonzales, D.A., J.G. Pearson and C.F.A. Pinkham. 1993. User's manual for BioSim1, Beta Version 1.0. A Program that Applies the Coefficient of Biotic Similarity, B, to Complex Data Matrices. EPA 600/R-93/219. Environmental Monitoring Systems Laboratory, Las Vegas, NV.

Kaesler, R.L. 1970. The Cophenetic Correlation Coefficient in Paleoecology, Geological Society of America Bulletin. Lawrence, KS. pp 1261-1266.

Klemm, D.J., P.A. Lewis and J.M. Lazorchak. 1990. Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters. Aquatic Biology Branch, Quality Assurance Research Division, Environmental Monitoring Systems Laboratory. Cincinnati, OH. EPA-600/0-90-000, U.S. EPA, Washington, D.C.

Pankhurst, Richard J. 1991. Practical Taxonomic Computing. Cambridge University Press, Cambridge, 201 pp.

Pearson, J.G., and C.F.A. Pinkham. 1992. Strategy for data analysis in environmental surveys emphasizing the index of biotic similarity and BioSim1. Water Environ. Res., 64:901-909.

Pfalkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish. EPA-444/4-89-001, U.S. EPA, Washington, D.C.

Pinkham, C.F.A., and Pearson, J.G. 1976. Applications of a new coefficient of similarity to pollution surveys. J. Water Pollut. Control Fed., 48, 717.

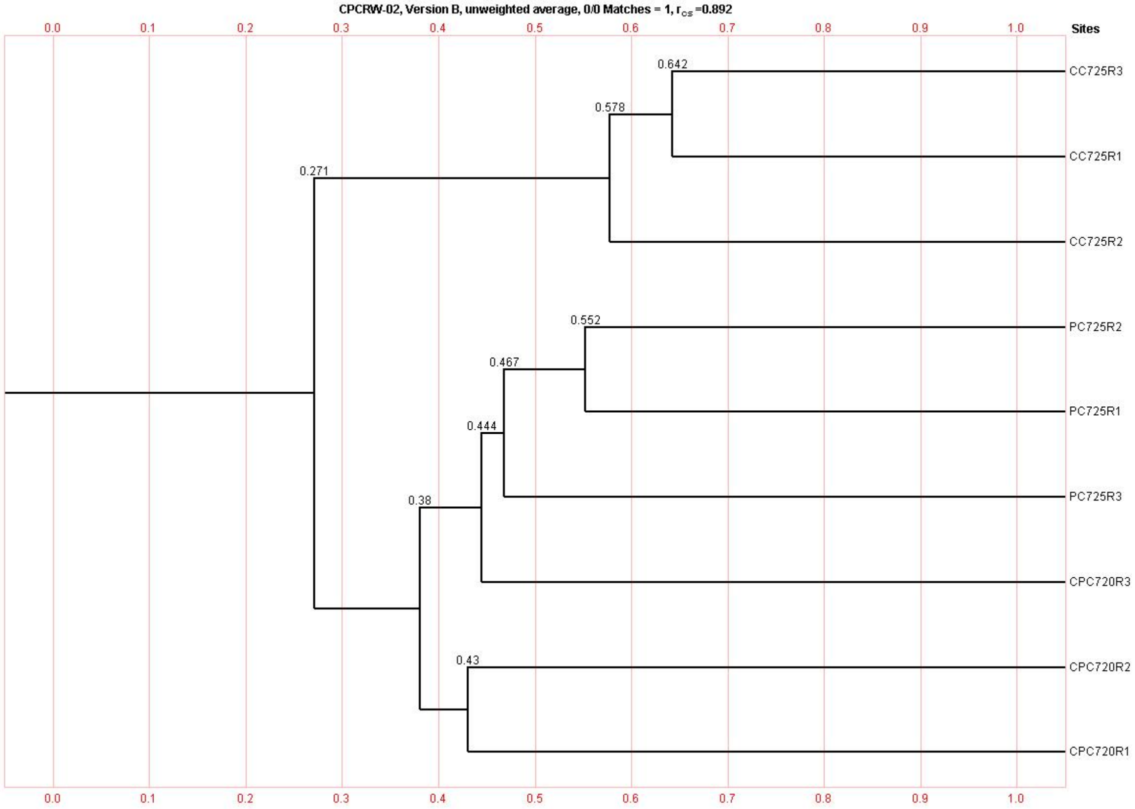
Pinkham, C.F.A., J.G. Pearson, W.L. Clontz and A.E. Asaki. 1975. A Computer Program for Calculations of Measures of Biotic Similarity Between Samples and the Plotting of the Relationship Between These Measures. EATR EB-TR-75013, Mar - Sep 74, 41pp.

Vermont Department of Environmental Conservation. 1990. Indirect Discharge Rule. Chap. 14: Environmental Protection Rules. Agency of Natural Resources, Waterbury, Vermont.

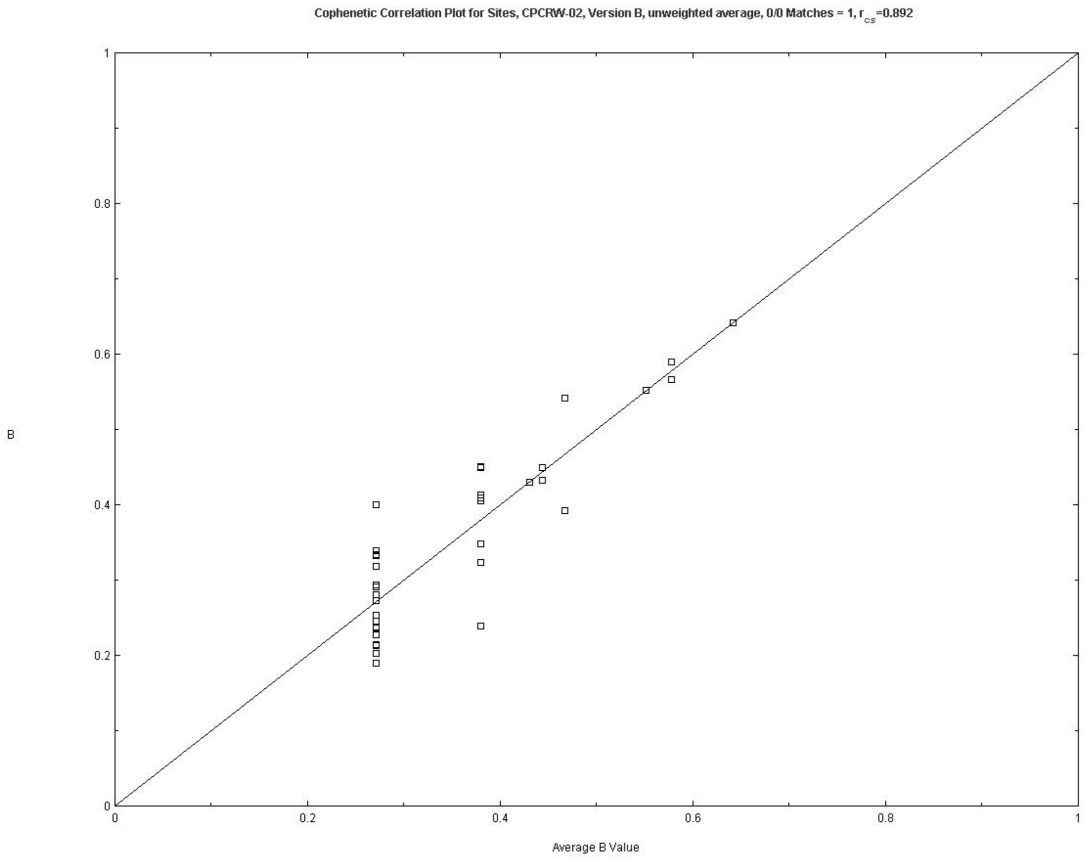
Appendix A

Samples of Saved Graphic Output Files

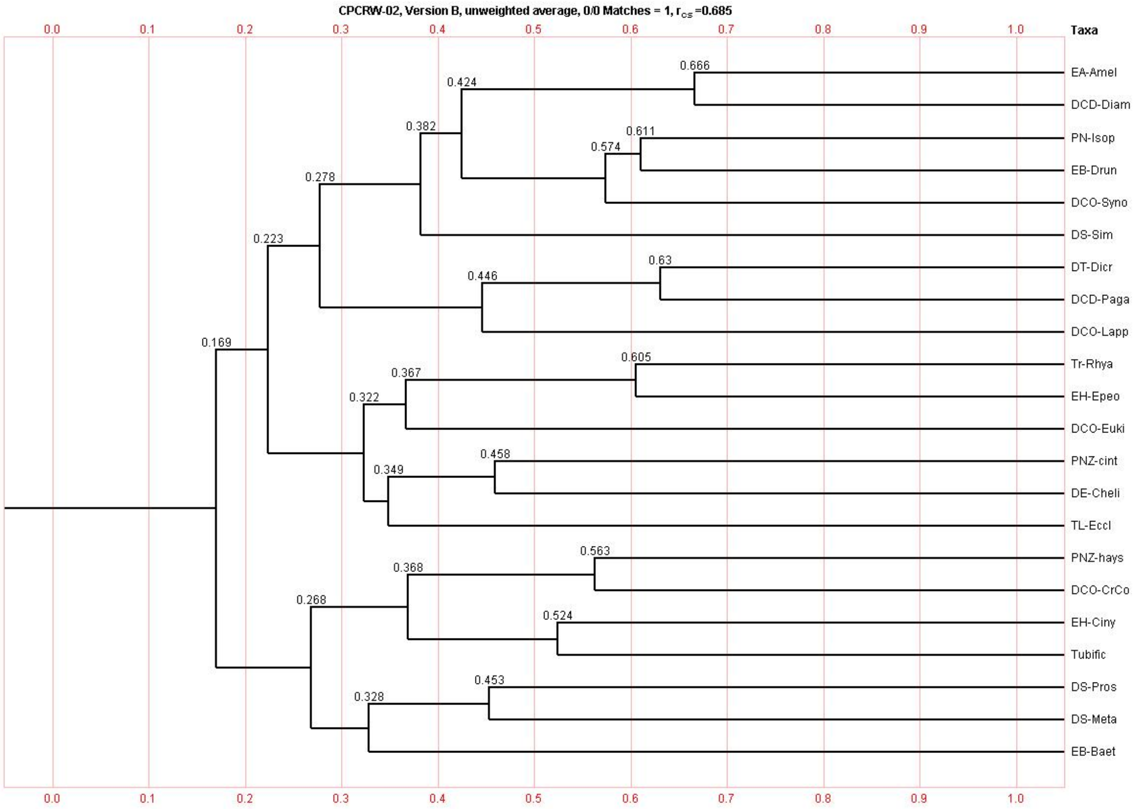
Row Dendrogram



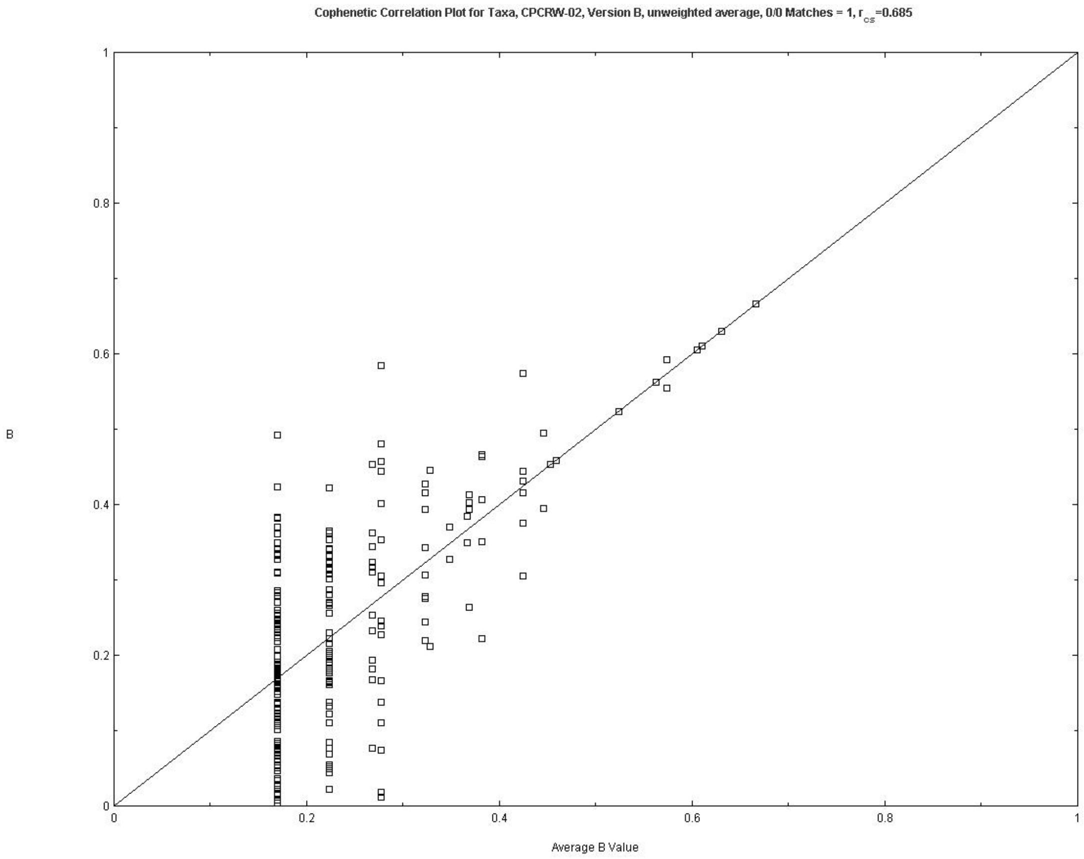
Row Cophenetic Correlation Plot



Column Dendrogram



Column Cophenetic Correlation Plot



EPA
United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

EPA/600/R-05/150
November 2005
www.epa.gov

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE □; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Recycled/Recyclable
Printed with vegetable-based ink on
paper that contains a minimum of
50% post-consumer fiber content
processed chlorine free