

User's Guide for Program CLAM (Classification Method)

by

Anne Chao, National Tsing Hua University, Taiwan 30043

Shang-Yi Lin, National Tsing Hua University, Taiwan 30043

Table of Contents:	Page
Software Requirements	1
Operating System Requirements	1
Overview	1
Download	2
Program Settings	2
Data Formats	3
Running Procedure and Output	4
Reference	8

Software Requirements

1. *R software* (R is free. You can download R from <http://www.r-project.org/>).
CLAM uses R is to produce a two-dimensional plot showing classification results for all species
2. *Adobe Acrobat: Acrobat* is used to generate a pdf file for the two-dimensional plot.

Operating System Requirements

Windows XP, Vista, or Windows 7.

Overview

The program **CLAM** (Classification Method) is written in both C and R languages; C is used for efficient computing whereas R is used for plotting. CLAM classifies two groups of species (say, A and B) based on frequency data into four categories: Generalist, Group A Specialist, Group B Specialist, and Too Rare to Classify. In addition to the classification results for each species, CLAM also outputs a two-dimensional plot showing classification results for all species. The methodology with examples is presented in the following paper:

Chazdon, R. L., Chao, A., Colwell, R. K., Lin, S.-Y., Norden, N., Letcher, S. G., Clark, D. B., Finegan, B. and Arroyo J. P. (2011). A novel statistical method for classifying habitat generalists and specialists. *Ecology*, 92(6), 2011, pp. 1332-1343.

If you use CLAM to prepare data for publication, you should cite the above paper along with the following reference for CLAM:

Chao, A. and Lin, S. Y. (2011) Program CLAM (Classification Method). Program and User's Guide at <http://purl.oclc.org/clam>

Download

The program CLAM can be downloaded free of charge from <http://purl.oclc.org/clam>. In order to keep a record and contact you for future updated versions/information, you are asked to register before downloading. After registration, you are directed to download the program "CLAM.zip". Double-click this zip file to unzip the program and store the program in a designated directory (automatically created) that will be named CLAM. Then check that there are five files in the directory CLAM: "CLAM.bat," "data.txt," "graph.r," "graph(black-white).r," and "classpro.exe." In the file "data.txt," we store the data used in Chazdon et al. (2011) as a demonstration example. We suggest that you first run this example and check the output with the output given in this guide in order to gain familiarity with the program and to check your installation.

Program Settings

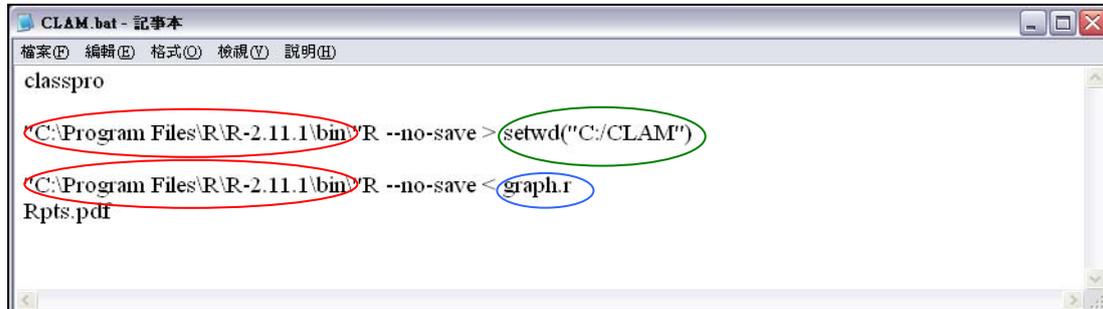
Right click the file "CLAM.bat", and select "edit". Then a Notepad window will appear as shown in Figure 1. Four modifications are necessary:

- (1) Modify the source path for R (as indicated in Figure 1; red). For example, if your copy of R is installed in C:\Program Files\R\R-2.11.1\ (a default directory name when downloading R, where R-2.11.1 is the version of R), then modify the path to read: C:\Program Files\R\R-2.11.1\bin\. (Remember to add \bin\.) There are two instances of this path to be modified.
- (2). Specify the source path (as indicated in Figure 1; green) for CLAM. For example, if the your CLAM directory is "C:\CLAM", then it should be modified to "setwd("C:/CLAM")". (*Please note the difference between forward slash and backslash.*)
- (3). Specify the graph type (as indicated in Figure 1; blue). If you want a color graph, then keep the default, "graph.r". If you prefer a black-and-white graph, then

modify it to “graph(black-white).r”.

Remember to save and close this file after modifying it.

Figure 1. Setting up for CLAM by editing the text file CLAM.bat



Data Formats

All species frequency data must be stored in the tab-delimited plain text file called “data.txt”. When you download the program, data for the example used in Chazdon et al. (2011) are stored in “data.txt”. You must replace these with your own data in this file or replace the file with new file named “data.txt”. Data must be arranged in three columns, separated by tab characters: (1) species names, (2) frequency (or abundance) for Group A, and (3) frequency (or abundance) for Group B. An example in the downloaded “data.txt” is shown in Figure 2. Space characters are not allowed in the column of species names. For example, the second species, *Aegiphila falcata*, should be modified to *Aegiphila_falcata*, *Aegiphila.falcata* or some other code for this species that contains no space characters. You can use Excel to edit the data first, and then save the file as “data.txt”, using the “Save As → Tab-Delimited-Text” command from the Excel File menu.

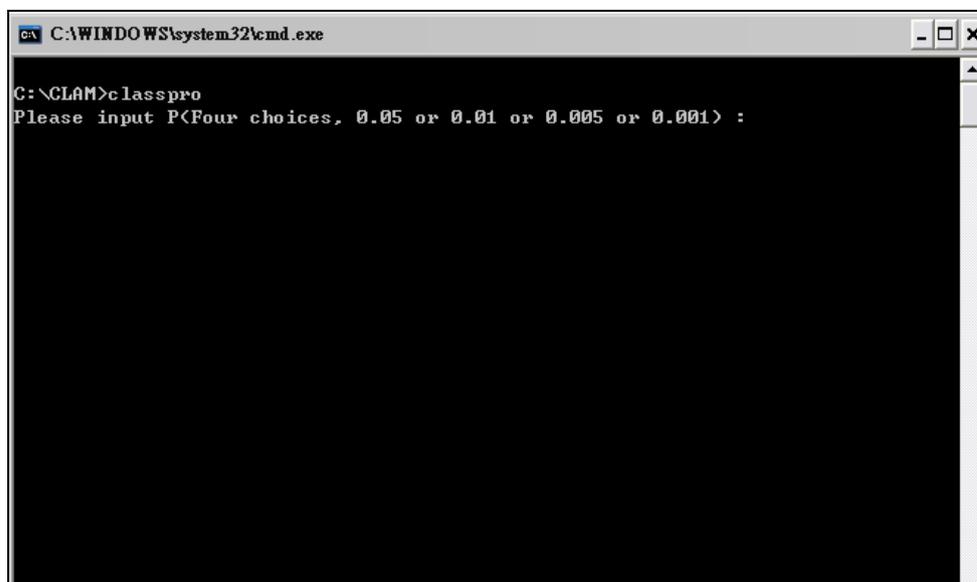
Figure 2. Data structure

Abarema_adenophora	3	8
Aegiphila_falcata	1	1
Alchomea_costaricensis	1	32
Alchomea_latifolia	3	22
Alchomeopsis_floribunda	18	16
Alibertia_atlantica	3	0
Allophylus_psilospermus	1	0
Ampelocera_macrocarpa	26	1
Anacardium_occidentale	0	1
Anaxagorea_crassipetala	56	8

Running Procedure and Output

Step 1. After setting up the CLAM.bat file and the data input file, click the file “CLAM.bat” in the directory CLAM. Then the window for entry of significance probability (P) in Figure 3 will appear. Input the value of P you require (there are four choices: 0.05, 0.01, 0.005 and 0.001). If your goal is to classify only one target species, $P = 0.05$ is suggested. If your target is to classify all species, $P = 0.005$ is suggested. (See Chazdon et al. 2011 for explanation of this recommendation.) After entry, click enter to save.

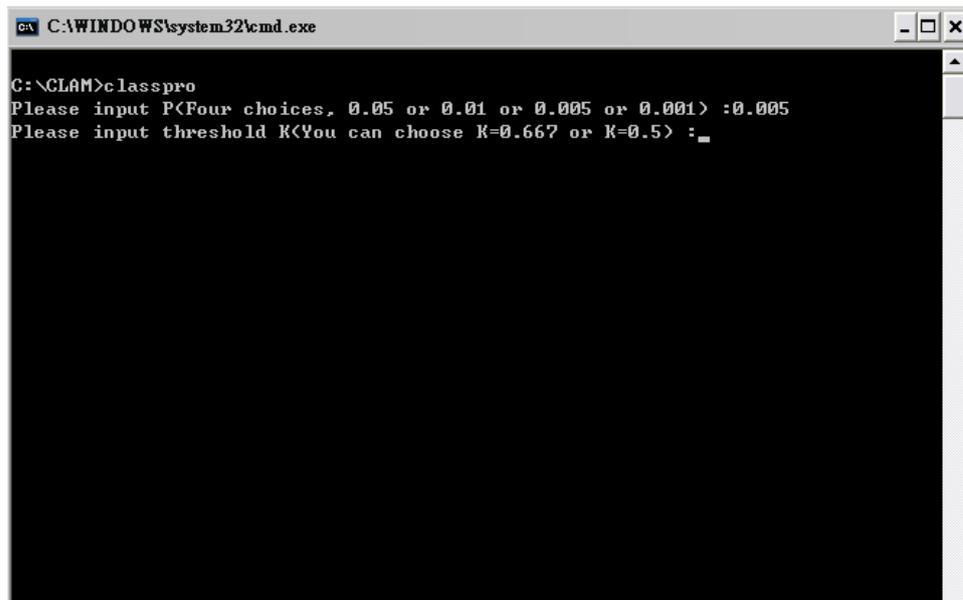
Figure 3. Interface window for P entry



```
ca C:\WINDOWS\system32\cmd.exe
C:\CLAM>classpro
Please input P<Four choices, 0.05 or 0.01 or 0.005 or 0.001> :
```

Step 2. After Step 1, the interface window for entry of threshold K is shown as in Figure 4. The user should choose a threshold ($K = 0.667$ for super-majority threshold or $K = 0.5$ for simple-majority threshold). After entry, click enter to save. The program starts after the entry of K .

Figure 4. Interface window for K entry



When the program finishes computation, the CLAM.bat window will close, and there will be six new files in the directory CLAM: “List Table.xls,” “Count Table.xls,” “Rplots.pdf,” “pri_line.txt,” “sec_line.txt,” and “rare_line.txt.” The three files “pri_line.txt,” “sec_line.txt,” and “rare_line.txt,” are just used for producing the graph, so these three files are not of any further use and can be ignored. The three useful output files are:

- (1) “Rplots.pdf”: the classification plot of all species produced by R. An example is shown in Figure 5. In Program Settings, if you select “graph.r”, then a color plot will be obtained as in Figure 5a. If you select “graph(black-white).r”, then a black and white plot will be obtained as in Figure 5b.
- (2) “Count Table.xls” contains some basic information and the classification results. It contains three parts: *Basic data information*; *Classification summary* (number of species and percentages in each of the four categories: *Generalist*, *First group specialist*, *Second group specialist* and *Too rare to classify*); and *Species classification results* (frequency data and classification category for each species), as shown in Figure 6.
- (3) “List table.xls” contains species lists for each of the four classification categories (*Generalist*, *First group specialist*, *Second group specialist* and *Too rare to classify*). For each category, the species name, its abundance in the first group, and its abundance in the second group are provided, as shown in Figure 7.

If you want to run any other data set, just repeat the above steps 1, 2, and 3 in the section “**Running Procedure and Output**.” However, *before running additional*

datasets, all six files must be closed, and the three useful files must be saved under different file names, because the six files will be replaced by the results obtained from running a new dataset. Each additional dataset must be in a file named “data.txt”.

Figure 5a. Classification Plot (color)

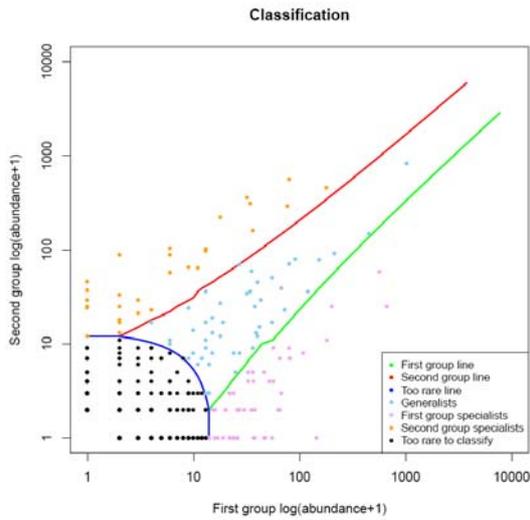


Figure 5b. Classification Plot (black and white)

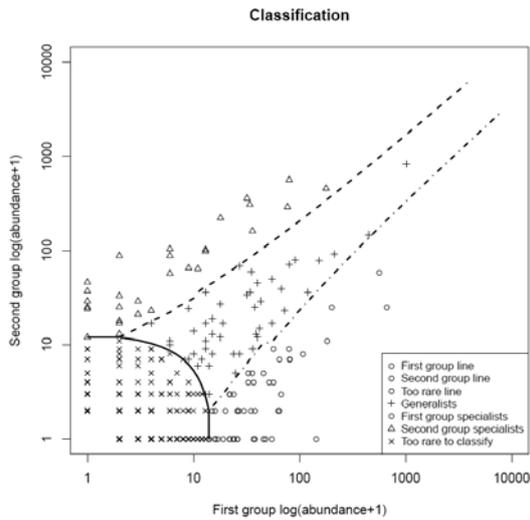


Figure 6. Output in the file “Count Table.xls”

(1). BASIC DATA INFORMATION:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	(1). BASIC DATA INFORMATION:													
2	Number of observed individuals in group 1	7768												
3	Number of observed individuals in group 2	5921												
4	Number of observed species in group 1	299												
5	Number of observed species in group 2	238												
6	Number of observed shared species	178												
7	Number of observed total species	359												
8	Estimated sample coverage(C1) in group 1	0.992533												
9	Estimated sample coverage(C2) in group 2	0.987671												
10	Estimated sample coverage(C1*) in group 1	0.909516												
11	Estimated sample coverage(C2*) in group 2	0.844681												
12	f(min)	13												
13	f*(min)	11												
14	(2). CLASSIFICATION:													
15	Class	Species#	Percentage()											
16	Too rare to classify	231	64.35											
17	First group specialists	52	14.48											
18	Second group specialists	29	8.08											
19	Generalists	47	13.09											
20	Total	359	100											
21	(3). Species classification result													
22	Name	First group	Second grc	Group										
23	Abarema_adenophora	3	8	4										
24	Aegiphila_falcata	1	1	4										
25	Alchornea_costaricensis	1	32	3										
26	Alchornea_latifolia	3	22	3										
27	Alchorneopsis_floribunda	18	16	1										
28	Alibertia_atlantica	3	0	4										
29	Allophylus_psilospermus	1	0	4										
30	Ampelocera_macrocarpa	26	1	2										
31	Anacardium_occidentale	0	1	4										

Figure 7. Output in the file “List table.xls”

	Generalist			First group specialist			Second group specialists			Too rare to classify									
	Name	FG	SG	Name	FG	SG	Name	FG	SG	Name	FG	SG							
1	Alchornea	18	16	Alchornea	26	1	Alchornea	1	32	Alchornea	3	22	Alchornea	3	8				
2	Apeiba	54	49	Anaxagore	56	8	Alchornea	3	22	Alchornea	3	22	Alchornea	3	8				
3	Balizia	17	11	Andira	21	0	Byrsonima	2	20	Alchornea	3	22	Alchornea	3	8				
4	Byrsonima	12	35	Ardisia	29	1	Casearia	177	454	Alchornea	3	22	Alchornea	3	8				
5	Carapa	71	22	Borojoa	18	0	Clethra	1	17	Alchornea	3	22	Alchornea	3	8				
6	Casearia	8	6	Brosimum	21	2	Cordia	28	28	Alchornea	3	22	Alchornea	3	8				
7	Cecropia	8	23	Brosimum	80	6	Cordia	17	220	Alchornea	3	22	Alchornea	3	8				
8	Cecropia	12	16	Calophyllum	23	0	Drosera	8	64	Alchornea	3	22	Alchornea	3	8				
9	Cespedesia	33	35	Capparis	64	6	Goethalia	76	289	Alchornea	3	22	Alchornea	3	8				
10	Conceveiba	17	26	Cassipouira	36	0	Hampea	10	63	Alchornea	3	22	Alchornea	3	8				
11	Cordia	14	18	Chrysophyllum	13	0	Handroanthus	0	24	Alchornea	3	22	Alchornea	3	8				
12	Cupania	12	2	Colubrina	45	4	Hasseltia	0	36	Alchornea	3	22	Alchornea	3	8				
13	Dendropanax	152	77	Coussarea	23	0	Inga	12	97	Alchornea	3	22	Alchornea	3	8				
14	Euterpe	211	91	Drosera	32	4	Inga	0	28	Alchornea	3	22	Alchornea	3	8				
15	Guatteria	39	11	Dussia	36	3	Inga	35	160	Alchornea	3	22	Alchornea	3	8				
16	Hernandia	39	44	Dystovomium	54	0	Acaranda	5	56	Alchornea	3	22	Alchornea	3	8				
17	Hieronyma	5	10	Schweileri	25	1	Aetia	79	561	Alchornea	3	22	Alchornea	3	8				
18	Hieronyma	12	10	Aramea	13	0	Miconia	1	87	Alchornea	3	22	Alchornea	3	8				
19	Ilex	13	5	Aramea	143	0	Miconia	5	103	Alchornea	3	22	Alchornea	3	8				
20	Inga	12	8	Berlinandua	37	0	Ocotea	0	11	Alchornea	3	22	Alchornea	3	8				
21	Inga	42	28	Parcinia	19	1	Persea	1	12	Alchornea	3	22	Alchornea	3	8				
22	Inga	31	33	Suarea	64	1	Rollinia	5	87	Alchornea	3	22	Alchornea	3	8				
23	Inga	37	24	Suarea	62	1	Simarouba	33	306	Alchornea	3	22	Alchornea	3	8				
24	Inga	5	9	Suarea	16	0	Spondias	1	16	Alchornea	3	22	Alchornea	3	8				
25	Inga	9	13	Suarea	51	3	Vismia	0	23	Alchornea	3	22	Alchornea	3	8				
26	Lacmellea	26	7	Viartea	564	57	Vochysia	31	356	Alchornea	3	22	Alchornea	3	8				
27	Miconia	3	16	Acunaria	23	0	Vochysia	1	24	Alchornea	3	22	Alchornea	3	8				
28	Miconia	34	58	Ecythis	14	1	Xylopia	12	101	Alchornea	3	22	Alchornea	3	8				
29	Miconia	41	14	Onchocarpus	25	0	Zanthoxylum	0	45	Alchornea	3	22	Alchornea	3	8				
30	Ocotea	13	2	Marantochloa	13	0				Alchornea	3	22	Alchornea	3	8				
31	Ocotea	23	7	Finimartia	80	6				Alchornea	3	22	Alchornea	3	8				

Reference

Chazdon, R. L., Chao, A., Colwell, R. K., Lin, S.-Y., Norden, N., Letcher, S. G., Clark, D. B., Finegan, B. and Arroyo J. P. (2011). A novel statistical method for classifying habitat generalists and specialists. *Ecology*, 92(6), 2011, pp. 1332-1343.