

Subject: evoldir question
Date: Thursday, 13 March 2014 1:13:26 am Australian Eastern Standard Time
From: Catherine Linnen
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Hi Jenny,

I just saw your email on evoldir and coincidentally, I just gave a lab on the STRUCTURE>Structure Harvester>CLUMPP>DISTRUCT pipeline in my ecological genetics class. It turns out that Structure Harvester produces input files for CLUMPP and this is very easy to run in command line once you have all the needed files (parameter file, input files) plus the executable all in the same folder. I have attached the lab exercise and am happy to provide additional input files if it would be helpful.

For Distruct on mac, I had the same problem and the easiest thing to do is to write Noah Rosenberg to get source code and compile on your own mac. This can be done as follows:

If you haven't already done so, you will need to install Xcode + command line tools for this (should be freely available from apple). Once that is installed, all you do is open up the terminal application, navigate to the directory that contains the source code. Then, type: gcc distruct1.1.c -o **distruct**

(note, for the -o flag, you can name it anything you want, but this is how you will call the program).

Hope that helps!
Catherine Linnen

--

Catherine R. Linnen, Ph.D.
Assistant Professor
Department of Biology
University of Kentucky
200A Thomas Hunt Morgan Building
Lexington, KY 40506
website: http://www.uky.edu/~cli242/Linnen_Lab/Home.html

Phone (lab): 859-323-3160
Phone (cell): 617-970-7717

Subject: Fwd: Distruct
Date: Thursday, 13 March 2014 1:46:04 am Australian Eastern Standard Time
From: Shannon Corrigan
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Hi Jenny,

I saw your post regarding Distruct on EvolDir. I recently had the same issue and queried it on the structure user forum. The curator sent me a recompiled version of distruct for mac and it works just fine now. I've forwarded it to you here.

Best of luck,
Shannon

----- Forwarded message -----

From: **Vikram Chhatre** <crypticlineage@gmail.com>
Date: Fri, Feb 7, 2014 at 4:47 PM
Subject: Distruct
To: Shannon Corrigan <shancorrigan1@gmail.com>

--
Shannon Corrigan, PhD
Postdoctoral Researcher
Hollings Marine Laboratory
331 Fort Johnson
Charleston, SC 29412
Phone: +1 843-725-4886
Fax: +1 843-762-8737
Cell: +1 415-359-7970
<http://prosper.cofc.edu/~sharkevolution/>

From : Catherine R. Linnen, Ph.D.

BIO 621 Lab 4: Population assignment analysis (3/6/14)

Intro:

In today's lab, we will be replicating some of the analyses from Konečný et al. 2012 ("Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa"). In particular, we will use data from 14 microsatellite markers to investigate genetic structure within our sample of 257 individuals using STRUCTURE, and then we will process these data using STRUCTURE HARVESTER, CLUMPP, and DISTRUCT.

Assignment:

As you go, save relevant output into a text file entitled "yourname_lab4results.txt" with a descriptive header for each set of results. Answer any questions throughout the exercises in a second file entitled "yourname_lab4questions.txt". In addition, for this lab, you will send a third file that is in ps (postscript) format that contains your STRUCTURE plot created in DISTRUCT (entitled "yourname_lab4plot.ps"). All 3 files should be emailed to me **by 5pm Friday 3/7**.

Note: for your results and question files, please include your name on the first line.

Setup:

We will be running a series of 4 analyses...with all the input and output files, this can quickly get confusing, so it's a good idea to organize folders by analysis. To facilitate this, I have created a Lab4_analyses folder in Dropbox. This contains 4 folders corresponding to the 4 modules below (some of which contain some required parameter files that you will use). These are named "Structure", "Harvester", "Clumpp", and "Distruct" (very creative, I know). I have also included the Structure input file for the rat data ("Ratgeno.txt"; alternatively, you are welcome to use your own file). Prior to starting this lab, copy and paste the Lab4_analyses folder into a convenient place.

Module 1: Running STRUCTURE

After generating basic summary stats (see Lab 3), the next step in a demographic analysis is to determine what our units of analysis should be. Although one approach would be to treat every single sampling location as an independent evolutionary unit, closely located populations could be very similar due to gene flow (alternatively, what looks like a single population to us may be comprised of multiple semi-isolated subpopulations). For this reason, many investigators use programs such as STRUCTURE to investigate genetic structure within their sample and to assign individuals to clusters.

1. Open STRUCTURE and choose the *New Project* option in the *File* menu.
2. In the first window, name your project whatever you want, choose the empty "Structure" folder for today's lab, and select the Ratgeno.txt file (or your Rat STRUCTURE data file that you created last week). Hit "Next".
3. Fill in the number of individuals (257), ploidy (2), number of loci (14), and missing data value (-9). Hit "Next".
4. Indicate that the data file contains a row of marker names and stores data for individuals in a single line (no other boxes should be checked). Hit "Next".

5. Indicate that data file contains individual ID and putative population origin. Hit “Finish”.
6. After you confirm that the details match your data, hit “Proceed”.
7. Now that your data are loaded, you will want to define a set of analysis parameters. Note that we are going for a speedy analysis today and that the settings we choose do not reflect recommended practices. To create a new parameter set, select the option *New* in the *Parameter Set* menu. You should get a window with four tabs, each with a couple of different options.
8. For today’s purposes, choose the following settings (we will talk more about what these mean once everyone gets things up and running):
 - a. Run length
 - i. Length of burnin period: 10,000 (generally 10,000-100,000 is sufficient; the rat paper used a burnin of 200,000)
 - ii. Number of MCMC reps after burnin: 10,000 (generally you will want to do much more than this; 100,000 is a good start...the rat paper used 600,000)
 - b. Ancestry model
 - i. Select “Use Admixture Model” and nothing else (this is what was chosen in the rat paper)
 - c. Allele frequency model
 - i. Select “Allele frequencies correlated” (in practice, this doesn’t usually make much of a difference)
 - d. Advanced
 - i. Use default options (i.e., compute probability of the data selected and nothing else).
 - e. Hit “OK” and name the parameter set whatever you want (and hit “OK” again).
9. Now that you have data and model settings loaded, go to the *Project* menu and select *Start a Job*. When the Structure Scheduler window comes up, you will have another series of options. Choose the following settings:
 - a. Select your parameter set that you just created
 - b. Set K from 1 to 5 (generally, the recommendation is to do 1 to the suspected true # of populations plus 3; rat paper used 1 to 10)
 - c. Choose 10 iterations per K (generally you want to do at least 20-30 to get good estimates of ancestry memberships; rat paper did 100).
 - d. Hit Start
10. On my laptop, this took ~15 min to run. We will use that time to discuss some general concepts related to today’s lab. Once everyone’s STRUCTURE analysis is completed, we will look at some output together.
11. **Copy results from a SINGLE K=4 replicate and paste this to your results file (include an informative label above these results).**
12. Examine the output for your 10 K=4 replicates. First, look at the “data plot” for your alpha values. **QUESTION 1: What do these plots suggest with respect to your burnin of 10000 steps (i.e., was this a sufficient burnin)? Explain your answer.**
13. Next, look at the “bar plot” for some of your K=4 replicates. **QUESTION 2: Based on the bar plots and mean alpha values reported in your results files, what can you infer with respect to admixture (i.e., is it high or low)? Explain your answer.**

Module 2: Processing data and choosing an optimal K in structure harvester

Now that you've run your analysis under several different Ks, we need to decide which is the best K. Although there has been some disagreement as to how to best choose K, perhaps the most widely used method for estimating K was suggested by Evanno et al. (2005) (available on dropbox). The basic idea behind this method, which is referred to as the ΔK method, is that when K is below the appropriate # of populations, $\text{Pr}(X|K)$ will be low. But $\text{Pr}(X|K)$ values will increase as K increases and then plateau as K exceeds the # of populations. This method therefore looks for that plateau—when several K values have similar $\text{Ln}(\text{Pr}X|K)$ values, it is assumed that the smallest K is correct. The ΔK method is implemented in Structure Harvester, so we will use this to select a K. Note that Evanno et al. recommend using ΔK together with $\ln P(K)$, alpha estimates, and individual assignment patterns when choosing K.

1. Look at your Results folder from structure. Zip all of your results files into one zip archive. (Note: if you are using Winzip, use “legacy compression” to ensure harvester can read your data).
2. Go to <http://taylor0.biology.ucla.edu/structureHarvester/>; click on “Choose file” and navigate to the zip archive you just made.
3. Hit “Harvest”
4. When the analysis is completed, you will see html results. **Copy and paste these into your results file with an informative label.**
5. You will see a line at the beginning of these results that says: “Single file archive including this page, all images, all clumpp files: [download](#). [.tar.gz]”. Click on the “Download” link and save your Harvester results to your “Harvester” folder. This will contain all of the data that was printed on the webpage, plus data files that are formatted for CLUMPP.
6. **QUESTION 3: What was your optimal K? How did this compare to everyone else's in the class and to the paper? What might explain this variation?**

Module 3: Summarizing data from replicate STRUCTURE runs in CLUMPP

As STRUCTURE produces multiple output files for a given K, we'd like to be able to summarize our results across all of these. CLUMPP provides a convenient way to summarize these results. And conveniently, Structure Harvester not only tells us which K to use, but also formats our data for CLUMPP. Here, we will assume that the optimal K is 3 (note that the paper analyzed results for both 3 and 4).

1. Running CLUMPP requires that a data file (either a *popfile* or an *indfile*), a parameter file (*paramfile*), and the program are all located in the same folder. Prior to starting CLUMPP, copy and paste your K3.indfile (includes ancestry of each individual, this is usually what people report) and K3.popfile (averages ancestry across individuals in each predefined population) from your Structure Harvester results into your “Clumpp” folder. Also copy and paste the executable Clumpp program that you downloaded into the same folder (all input files must be in the same folder as the program). You should already have two parameter files in the folder

- (Rats_paramfile_ind and Rats_paramfile_pop) corresponding to analyses for individual and population analyses.
2. Now that you have the program, indfile, and paramfiles in the same folder, you will start CLUMPP in one of two ways:
 - a. From the command line prompt in MAC/UNIX/LINUX (in macs, you can access this by opening the terminal application), you will first navigate to the folder containing the program and your input files (cd /pathtodirectory; tip: one easy way to do this is to open the terminal application, type cd followed by a space, then drag and drop your desired folder into the terminal window; this should give you the path and you can hit enter), then type `./CLUMPP Rats_paramfile_ind`
 - b. For a windows, open the command prompt (which can be accessed by going to the START menu, clicking on Run, and typing cmd), then navigate to where CLUMPP is located (by typing cd c:wherever your clump directory is). Then execute the program by typing `CLUMPP Rats_paramfile_ind`.
 3. Repeat step 2 for the population results. Specifically, type `./CLUMPP Rats_paramfile_pop` (or `CLUMPP Rats_paramfile_pop` if you are using windows).
 4. **Copy and paste the contents of the .miscfile for both the population and individual analyses in your Results file (use an informative label).**
 5. **QUESTION 4: What do your results indicate about the consistency of results across multiple independent runs?**

Module 4: Visualizing results in *distruct*

Distruct produces publication quality figures from STRUCTURE or CLUMPP (or any program that creates a Q-matrix) analyses (e.g., see figure 2e in the rat paper). This program requires a *drawparams* file, which contains all of the information needed to produce figures, including data file names, number of clusters (K), number of individuals (*NUMINDS*), and number of populations (*NUMPOPS*). This file also contains all other parameters that define graphical characteristics (see *Distruct* manual for details). Like CLUMPP, the program and all data files must be in the same folder and this can be run using the command line. For this analysis, we will create a nice plot for the Rat data that includes 3 clusters, sampling location names, and vertical lines separating sampling locations.

1. Again, we will start by getting all the data and your program in the right place. Your CLUMPP analysis produced two data files that you will need for *Distruct*. In the “Distruct” folder, you should see 3 files: *drawparams* (settings for making figures), *Rats.names* (contains the names of each of the 24 sampling locations), and *rats.perm* (contains color scheme for your 3 clusters). Copy and paste the “Rats_K3_ind.outfile” and “Rats_K3_pop.outfile” from the “Clumpp” folder into your “Distruct” folder. Then, copy the *Distruct* executable that you either downloaded or compiled and paste it into this folder.
2. You will run *distruct* in the command line very similarly to the way you did Clumpp. Specifically, open up terminal/command line, navigate to the Distruct folder, then type `./distruct` (or `distruct` for windows) [*distruct* assumes that your parameter file is called *drawparams*, which ours is.]
3. This will produce a ps file that can be opened by Preview, adobe illustrator, or free postscript viewers such as ghostview (http://pages.cs.wisc.edu/_ghost).
4. Just for fun, alter the colors in your *rats.perm* file and/or any other settings in your *drawparams* file (see *Distruct* manual for figure appearance options, including color scheme), rerun *distruct* and **send me the final ps file you created**. You can make this as pretty (or ugly) as you want!

Subject: Re: Other: CLUMPP and DISTRUCT
Date: Wednesday, 12 March 2014 11:24:38 pm Australian Eastern Standard Time
From: Vikram Chhatre
To: Jennifer Ovenden
CC: Dent Earl
Category: CLUMPP/DISTRUCT

Hi Jenny,

Note that the Distruct Mac version is not broken. It merely needs to be recompiled on newer versions of OSX. The source code is freely available from the author.

Also, while both CLUMPP and DISTRUCT are commandline utilities, they are really easy to set up. Harvester will give you input files for CLUMPP. No special formatting is needed for CLUMPP output files before they are fed to DISTRUCT except for a simple name change.

Hope that helps.
Vikram

Vikram Chhatre, PhD
Keller Lab - Appalachian Laboratory
University of Maryland Center for Environmental Sciences
301 Braddock Rd, Frostburg, MD 21532

Phone: (979)204-2294
Email: vchhatre@al.umces.edu

On Wed, Mar 12, 2014 at 2:42 AM, <evoldir@evol.biology.mcmaster.ca> wrote:

--_000_CF45CEF2D430jovendenuqueduau_
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?

CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.

The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).

Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.

Best, Jenny Ovenden

** New UQ email address**

j.ovenden@uq.edu.au

Jennifer Ovenden

<http://molecularfisherieslaboratory.com.au>

<http://www.uq.edu.au/sbms/staff/jennifer-ovenden>

--_000_CF45CEF2D430jovendenuqeduau_

Content-Type: text/html; charset="us-ascii"

Content-ID: <FDF2D8FD07780544BCA1979A863BEAB5@exchange.uq.edu.au>

Content-Transfer-Encoding: quoted-printable

<html>

<head>

<meta http-equiv="Content-Type" content="text/html; charset=us-ascii">

</head>

<body style="word-wrap: break-word; -webkit-nspace-mode: space; -webkit-line-break: after-white-space; color: rgb(0, 0, 0); font-size: 14px; font-family: Calibri, sans-serif;">

<div>

<div>Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?</div>

<div>

</div>

<div>CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.</div>

<div>

</div>

<div>The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).</div>

</div>

<div>

</div>

<div>Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.</div>

<div>

</div>

<div>

<p style="margin: 0px 0px 0px 36px; text-indent: -36px; font-size: 12px; font-family: Helvetica;">

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.</p>

</div>

<div>

</div>

<div>Best, Jenny Ovenden</div>

<div>

</div>

<div>

<div>

<div style="font-size: 16px;">*****</div>

<div>* New UQ email address*</div>

<div>j.ovenden@uq.edu.au</div>

<div>

</div>

<div>Jennifer Ovenden</div>

</div>

<div><http://molecularfisherieslaboratory.com.au></div>

<div><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden" style="font-size:

```
12px;">http://www.uq.edu.au/sbms/staff/jennifer-ovenden</a></div>  
</div>  
</body>  
</html>
```

```
--_000_CF45CEF2D430jovendenuqeduau_--
```

Subject: Other: CLUMPP and DISTRUCT
Date: Wednesday, 12 March 2014 7:28:43 pm Australian Eastern Standard Time
From: Roy Francis
To: Jennifer Ovenden, evoldir@evol.biology.mcmaster.ca
Category: CLUMPP/DISTRUCT

Hi Jenny,

If you have some basic knowledge of R, then the package pophelper can do the job. It does pretty much what Structure Harvester does. It can also plot figures.

The developer version is available here with basic instructions:

<https://github.com/royfrancis/pophelper>

This is a new package and available only on github at the moment. Not available on CRAN yet. The functionalities are expected to be added/improved.

Feel free to contact me if you have any issues/questions/comments/suggestions.

Best Regards,
Roy

From: evoldir@evol.biology.mcmaster.ca [evoldir@evol.biology.mcmaster.ca]
Sent: Wednesday, March 12, 2014 7:47 AM
To: Per Kryger
Subject: Other: CLUMPP and DISTRUCT

--_000_CF45CEF2D430jovendenuqeduau_
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?

CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.

The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).

Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.

Best, Jenny Ovenden

** New UQ email address**

j.ovenden@uq.edu.au

Jennifer Ovenden
<http://molecularfisherieslaboratory.com.au>
<http://www.uq.edu.au/sbms/staff/jennifer-ovenden>

--_000_CF45CEF2D430jovendenuqueduau_
Content-Type: text/html; charset="us-ascii"
Content-ID: <FDf2D8FD07780544BCA1979A863BEAB5@exchange.uq.edu.au>
Content-Transfer-Encoding: quoted-printable

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=us-ascii">
</head>
<body style="word-wrap: break-word; -webkit-nspace-mode: space; -webkit-line-break: after-white-space; color:
rgb(0, 0, 0); font-size: 14px; font-family: Calibri, sans-serif;">
<div>
<div>Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?</div>
<div><br>
</div>
<div>CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces
the STRUCTURE figure.</div>
<div><br>
</div>
<div>The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however,
I assume PC version is OK).</div>
</div>
<div><br>
</div>
<div>Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar
solution for CLUMPP and DISTRUCT would be most welcome.</div>
<div><br>
</div>
<div>
<p style="margin: 0px 0px 0px 36px; text-indent: -36px; font-size: 12px; font-family: Helvetica;">
Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE
output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-
011-9548-7.</p>
</div>
<div><br>
</div>
<div>Best, Jenny Ovenden</div>
<div><br>
</div>
<div>
<div style="font-size: 16px;"><span class="Apple-style-span" style="font-size:
12px;">*****</span></div>
<div>*<span style="font-weight: bold;"> New UQ email address</span>*</div>
<div><a href="mailto:j.ovenden@uq.edu.au">j.ovenden@uq.edu.au</a></div>
<div><span style="font-size: 12px;"><br>
</span></div>
<div><span style="font-size: 12px;">Jennifer Ovenden</span></div>
</div>
<div><a href="http://molecularfisherieslaboratory.com.au" style="font-size:
12px;"><a href="http://molecularfisherieslaboratory.com.au">http://molecularfisherieslaboratory.com.au</a></div>
<div><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden" style="font-size:
12px;"><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden">http://www.uq.edu.au/sbms/staff/jennifer-ovenden</a></div>
</div>
</body>
</html>
```

--_000_CF45CEF2D430jovendenuqeduau_--

Subject: Re: Other: CLUMPP and DISTRUCT
Date: Wednesday, 12 March 2014 7:23:08 pm Australian Eastern Standard Time
From: Violeta Muñoz-Fuentes
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Hi Jennifer,

I am running CLUMPP and DISTRUCT from the command line, in a Mac.

In reference to the problem you mention with DISTRUCT, I experienced the same issue and found a solution. On the DISTRUCT webpage there are three distributions. One works in one of my Macs, and a different one in the other! I managed to make them work by typing in the Terminal window, where you have placed the programmes:

```
chmod u+x NameOfProgramme
```

I did this for the three distributions, and then tested which one produced the postscript file with the example files.

I hope this helps!

Violeta

--

Dr Violeta Munoz-Fuentes

Senckenberg Research Institute and Natural History Museum Frankfurt
Conservation Genetics Group
Clamecystrasse 12, D-63571 Gelnhausen
Tel.: +49-6051-61954-3129

www.senckenberg.de/
www.senckenberg.de/naturschutzgenetik
www.senckenberg.de/conservationgenetics

Senckenberg Gesellschaft fr Naturforschung
Rechtsfhiger Verein gem§ 22 BGB
Senckenberganlage 25
60325 Frankfurt

On 12/03/2014 07:49, evoldir@evol.biology.mcmaster.ca wrote:

```
--_000_CF45CEF2D430jovendenuqeduau_  
Content-Type: text/plain; charset="us-ascii"  
Content-Transfer-Encoding: quoted-printable
```

Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?

CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.

The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).

Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.

Best, Jenny Ovenden

** New UQ email address**

j.ovenden@uq.edu.au

Jennifer Ovenden

<http://molecularfisherieslaboratory.com.au>

<http://www.uq.edu.au/sbms/staff/jennifer-ovenden>

--_000_CF45CEF2D430jovendenuqeduau_

Content-Type: text/html; charset="us-ascii"

Content-ID: <PDF2D8FD07780544BCA1979A863BEAB5@exchange.uq.edu.au>

Content-Transfer-Encoding: quoted-printable

<html>

<head>

<meta http-equiv="Content-Type" content="text/html; charset=us-ascii">

</head>

<body style="word-wrap: break-word; -webkit-nspace-mode: space; -webkit-line-break: after-white-space; color: rgb(0, 0, 0); font-size: 14px; font-family: Calibri, sans-serif;">

<div>

<div>Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?</div>

<div>

</div>

<div>CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.</div>

<div>

</div>

<div>The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).</div>

</div>

<div>

</div>

<div>Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.</div>

<div>

</div>

<div>

<p style="margin: 0px 0px 0px 36px; text-indent: -36px; font-size: 12px; font-family: Helvetica;">

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.</p>

</div>

<div>

</div>

<div>Best, Jenny Ovenden</div>

```
<div><br>
</div>
<div>
<div>
<div style="font-size: 16px;"><span class="Apple-style-span" style="font-size:
12px;">*****</span></div>
<div>*<span style="font-weight: bold;"> New UQ email address</span>*</div>
<div><a href="mailto:j.ovenden@uq.edu.au">j.ovenden@uq.edu.au</a></div>
<div><span style="font-size: 12px;"><br>
</span></div>
<div><span style="font-size: 12px;">Jennifer Ovenden</span></div>
</div>
<div><a href="http://molecularfisherieslaboratory.com.au" style="font-size:
12px;">http://molecularfisherieslaboratory.com.au</a></div>
<div><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden" style="font-size:
12px;">http://www.uq.edu.au/sbms/staff/jennifer-ovenden</a></div>
</div>
</body>
</html>
```

--_000_CF45CEF2D430jovendenuqeduau_--

Subject: AW: Other: CLUMPP and DISTRUCT
Date: Wednesday, 12 March 2014 6:20:39 pm Australian Eastern Standard Time
From: Wielgoss Sebastien
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Dear Jenny.

Well, my experience with DISTRUCT is also mixed, but it runs on Windows, and has a great manual.

I only want to share a couple of comments from my experience with it, which are no real solutions, but might help you around some obstacles. I have recently downloaded DISTRUCT, and it did not run well with my already prepared input files I had created years ago. The Mac version does not seem to support some or all of the new Intel-based architectures (at least it wouldn't run on a 64-bit MacOSX 10.8). So, I tried it with Windows 7.

I looked at the provided example files ("casia"-dataset), and carefully adapted my own input files.

However, nothing worked, and my produced output postscript file was empty (0kb). This is when I started copy-pasting my data into the "casia"-files, and all of a sudden it was okay. There must be something about these files that makes them compatible with DISTRUCT.

So, I basically agree with you that it would be nice to get versions of both programs that are less complicated to work with.

Cheers

Sébastien

Sébastien Wielgoss

Postdoctoral Fellow

ETH Zürich
Institute of Integrative Biology
Chair for Evolutionary Biology
CHN H74

Postal:
Universitätsstrasse 16
8092 Zürich
Switzerland

fon: (+41) 4463-27623

ResearchGate: https://www.researchgate.net/profile/Sebastien_Wielgoss

Website: <https://sites.google.com/site/sebsurinternet>

Von: evoldir@evol.biology.mcmaster.ca [evoldir@evol.biology.mcmaster.ca]

Gesendet: Mittwoch, 12. März 2014 07:55

An: Wielgoss Sebastien

Betreff: Other: CLUMPP and DISTRUCT

--_000_CF45CEF2D430jovendenuqeduau_
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?

CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.

The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).

Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.

Best, Jenny Ovenden

** New UQ email address**
j.ovenden@uq.edu.au

Jennifer Ovenden
<http://molecularfisherieslaboratory.com.au>
<http://www.uq.edu.au/sbms/staff/jennifer-ovenden>

--_000_CF45CEF2D430jovendenuqeduau_
Content-Type: text/html; charset="us-ascii"
Content-ID: <DFD2D8FD07780544BCA1979A863BEAB5@exchange.uq.edu.au>
Content-Transfer-Encoding: quoted-printable

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=us-ascii">
</head>
<body style="word-wrap: break-word; -webkit-nspace-mode: space; -webkit-line-break: after-white-space; color:
rgb(0, 0, 0); font-size: 14px; font-family: Calibri, sans-serif;">
<div>
<div>Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?</div>
<div><br>
</div>
<div>CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces
the STRUCTURE figure.</div>
<div><br>
</div>
<div>The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however,
I assume PC version is OK).</div>
</div>
<div><br>
</div>
<div>Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar
solution for CLUMPP and DISTRUCT would be most welcome.</div>
```

```
<div><br>
</div>
<div>
<p style="margin: 0px 0px 0px 36px; text-indent: -36px; font-size: 12px; font-family: Helvetica;">
Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE
output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-
011-9548-7.</p>
</div>
<div><br>
</div>
<div>Best, Jenny Ovenden</div>
<div><br>
</div>
<div>
<div>
<div style="font-size: 16px;"><span class="Apple-style-span" style="font-size:
12px;">*****</span></div>
<div>*<span style="font-weight: bold;"> New UQ email address</span>*</div>
<div><a href="mailto:j.ovenden@uq.edu.au">j.ovenden@uq.edu.au</a></div>
<div><span style="font-size: 12px;"><br>
</span></div>
<div><span style="font-size: 12px;">Jennifer Ovenden</span></div>
</div>
<div><a href="http://molecularfisherieslaboratory.com.au" style="font-size:
12px;">http://molecularfisherieslaboratory.com.au</a></div>
<div><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden" style="font-size:
12px;">http://www.uq.edu.au/sbms/staff/jennifer-ovenden</a></div>
</div>
</body>
</html>
```

--_000_CF45CEF2D430jovendenuqeduau_--

Subject: CLUMMP and DISTRUCT

Date: Wednesday, 12 March 2014 5:34:03 pm Australian Eastern Standard Time

From: Mark Ravinet

To: Jennifer Ovenden

Category: CLUMPP/DISTRUCT

Hi Jenny,

I recently used CLUMPP and DISTRUCT on Mac OS X without any issues. There are updated 2012 and 2013 versions of DISTRUCT on the program website:

<http://www.stanford.edu/group/rosenberglab/distructDownload.html>

I used the 2013 version.

Probably the easiest thing to do is set up a paramfile for individuals, a param file for populations and then run a bash script using command flags to alter those. Attached are two basic scripts I used to do this - you could quite easily automate them a bit more - i.e. get them to read the number of individuals and so on. I didn't bother because I only needed to run STRUCTURE once for an MS. Let me know if you'd like any more info on how to use them.

All the best

Mark

Subject: Re: Other: CLUMPP and DISTRUCT
Date: Wednesday, 12 March 2014 5:25:06 pm Australian Eastern Standard Time
From: Josef Janoušek
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Dear Jenny,

do you have troubles with programs only on Mac or also Windows?

For CLUMPP - you can use input files from Structure Harvester, there you first need to evaluate G' statistics or another stats (indiv.misc file), select similar runs; output files are produced directly by CLUMPP - and those files you can use for distruct.

Distruct - rename the files from CLUMPP "xxxx.indivq" / "xxxx.popq". Output file from Distruct is "[xxx.ps](#)" - this you can open as pdf file.

I hope it helps at least a bit.

Best,

Josef Janousek, Czech Republic

2014-03-12 7:46 GMT+01:00 <evoldir@evol.biology.mcmaster.ca>:

--_000_CF45CEF2D430jovendenuqeduau_
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?

CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT produces the STRUCTURE figure.

The challenge is that both are command-line programs, and DISTRUCT mac version seems broken (however, I assume PC version is OK).

Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar solution for CLUMPP and DISTRUCT would be most welcome.

Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361, doi:10.1007/s12686-011-9548-7.

Best, Jenny Ovenden

** New UQ email address**

j.ovenden@uq.edu.au

Jennifer Ovenden

<http://molecularfisherieslaboratory.com.au>

<http://www.uq.edu.au/sbms/staff/jennifer-ovenden>

--_000_CF45CEF2D430jovendenuqeduau_

Content-Type: text/html; charset="us-ascii"

Content-ID: <FD2D8FD07780544BCA1979A863BEAB5@exchange.uq.edu.au>

Content-Transfer-Encoding: quoted-printable

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=us-ascii">
</head>
<body style="word-wrap: break-word; -webkit-nspace-mode: space; -webkit-line-break: after-white-space; color:
rgb(0, 0, 0); font-size: 14px; font-family: Calibri, sans-serif;">
<div>
<div>Does anyone have a user-friendly way of inputting data into CLUMPP and DISTRUCT?</div>
<div><br>
</div>
<div>CLUMPP averages results across STRUCTURE simulations once K has been selected, and DISTRUCT
produces the STRUCTURE figure.</div>
<div><br>
</div>
<div>The challenge is that both are command-line programs, and DISTRUCT mac version seems broken
(however, I assume PC version is OK).</div>
</div>
<div><br>
</div>
<div>Dent Earl and Bridgett vonHoldt produced an great solution for processing STRUCTURE output. A similar
solution for CLUMPP and DISTRUCT would be most welcome.</div>
<div><br>
</div>
<div>
<p style="margin: 0px 0px 0px 36px; text-indent: -36px; font-size: 12px; font-family: Helvetica;">
Earl, D. A., vonHoldt, B. M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE
output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361,
doi:10.1007/s12686-011-9548-7.</p>
</div>
<div><br>
</div>
<div>Best, Jenny Ovenden</div>
<div><br>
</div>
<div>
<div style="font-size: 16px;"><span class="Apple-style-span" style="font-size:
12px;">*****</span></div>
<div>*<span style="font-weight: bold;"> New UQ email address</span>*</div>
<div><a href="mailto:j.ovenden@uq.edu.au">j.ovenden@uq.edu.au</a></div>
<div><span style="font-size: 12px;"><br>
</span></div>
<div><span style="font-size: 12px;">Jennifer Ovenden</span></div>
</div>
<div><a href="http://molecularfisherieslaboratory.com.au" style="font-size:
12px;">http://molecularfisherieslaboratory.com.au</a></div>
<div><a href="http://www.uq.edu.au/sbms/staff/jennifer-ovenden" style="font-size:
12px;">http://www.uq.edu.au/sbms/staff/jennifer-ovenden</a></div>
</div>
</body>
</html>
```

--_000_CF45CEF2D430jovendenuqeduau_--

Subject: Fwd: Other: Software.pophelper_R_package
Date: Friday, 14 March 2014 1:34:26 am Australian Eastern Standard Time
From: Vikram Chhatre
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

You may have already noticed this. Seems like this is kind of what you were asking for.

V

----- Forwarded message -----

From: <evoldir@evol.biology.mcmaster.ca>
Date: Thu, Mar 13, 2014 at 1:20 AM
Subject: Other: Software.pophelper_R_package
To: crypticlineage@gmail.com

I work with population genetics and population structure of honey bees. Therefore I had been writing some R scripts for my personal use and someone recently suggested that I make it into a package that can be shared. Therefore I have recently put together some useful stuff into an R package called pophelper.

Basically, this R package does what STRUCTURE HARVESTER does. pophelper works with structure and TESS output files. I think it may be useful for people working with STRUCTURE, TESS and R. This package requires some basic R skills to use. pophelper can be used to tabulate runs, summarise them, perform Evanno method, export to CLUMPP format, generate barplot figures etc.

The package can be installed from github for the moment. Once a stable version is established, it will be uploaded to CRAN.

<https://github.com/royfrancis/pophelper>

I would very much appreciate if you could share this with the community. I am also hoping people are willing to give this a try and provide some feedback/comments/suggestions/criticisms etc. Feel free to contact me at roy.m.francis@outlook.com if you run into any issues or have any questions etc.

Sincerely,Roy Francis

Roy Francis <roy.m.francis@outlook.com>

Subject: Following your question on CLUMPP and DISTRUCT
Date: Tuesday, 18 March 2014 12:11:35 am Australian Eastern Standard Time
From: Naama Kopelman
To: Jennifer Ovenden
Category: CLUMPP/DISTRUCT

Dear Jennifer,

We are about to submit a paper describing a new web-server, CLUMPAK, which provides a full pipeline for analyzing the output of global-ancestry programs, such as STRUCTURE. CLUMPAK is already available online, at <http://clumpak.tau.ac.il>.

It was developed by myself and colleagues, including Noah Rosenberg and Mattias Jakobsson, who developed DISTRUCT and CLUMPP.

CLUMPAK takes as input the results produced by STRUCTURE for a range of K values, and with any number of replicates for each K. The main pipeline will detect distinct solutions and separate runs which fall into these solutions, and produce high quality images for distinct solutions across the entire K range. In addition, a separate feature of CLUMPAK implements the Evanno method for detecting the optimal K value.

Please feel free to test CLUMPAK, and don't hesitate to consult me if you have any questions. We will appreciate your feedback :-)

Best,

Naama Kopelman

Tel Aviv University