

Welcome to the manual for VCFtoTree (Version 3.0.0). We appreciate you using our software, and hope to make it as straightforward and user-friendly as possible. We understand that many of our users may have little to no prior experience in computer programming, so this manual aims to address and troubleshoot any issues that may arise specifically related to the installation of the program, which can be programming-heavy. Should you have questions not addressed in the manual, please feel free to contact:

Omer Gokcumen, omergokc@buffalo.edu

Duo Xu, duoxu@buffalo.edu

Izzy Starr, istarr@buffalo.edu

Installing the VCFtoTree v3.0.0 software

VCFtoTree was designed for Mac OS users and requires a computer running El Capitan or higher.

Before installing VCFtoTree v3.0.0, you will need to install the following tools:

- Mac Command Line Tools
- samtools
- tabix
- wget
- gcc
- Dendroscope

If these tools are already installed on your computer, please ignore the following directions and proceed to page 3 for a step-by-step instruction for using VCFtoTree v3.0.0.

To install Mac Command Line Tools, you will need to download them from the Apple website (<https://developer.apple.com/opensource/>). On the right side of the screen you will see the option to 'view downloads', and log in using your Apple ID on the following page. After logging in, you will be brought to a page where you can download the correct version for your operating system.

For example, if you are running Mac OS El Capitan (10.11.x), download the appropriate version.

Description	Release Date
+ Command Line Tools (macOS 10.13) for Xcode 9	Sep 17, 2017
+ Command Line Tools (macOS 10.12) for Xcode 9	Sep 17, 2017
+ Command Line Tools for Xcode 8.3.2	Apr 17, 2017
+ Command Line Tools (macOS 10.12) for Xcode 8.3	Mar 26, 2017
+ Command Line Tools (macOS 10.12) for Xcode 8.2	Dec 12, 2016
+ Command Line Tools (macOS 10.11) for Xcode 8.2	Dec 11, 2016
+ Command Line Tools (macOS 10.12) for Xcode 8.1	Oct 26, 2016
+ Command Line Tools (macOS 10.12) for Xcode 8	Sep 12, 2016

Download and install the package.

To install samtools, tabix, and gcc, you will first need to install Homebrew. Homebrew installs development tools missed by Mac Command Line Tools. To do this, you will need to pull up Terminal (which can easily be found using the Spotlight Search function), and paste the following prompt and press 'enter':

```
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

It will begin installing Homebrew and require your password to continue the installation process. When it has finished installing it will prompt you for new input. You will be able to install samtools, tabix, and gcc using the following command. Input each command individually when prompted by the '\$'.

```
brew install samtools
brew install tabix
brew install wget
brew install gcc
```

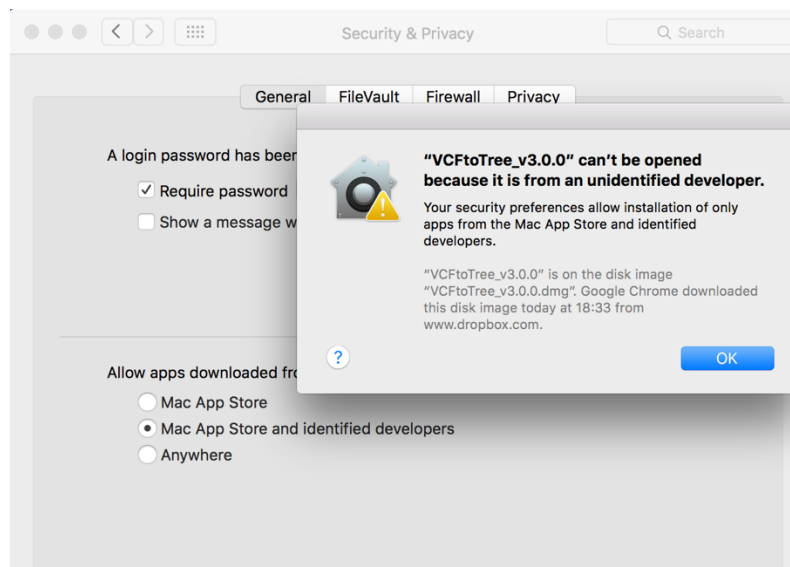
In order to confirm that these three tools are installed correctly, run each of the following commands individually when prompted by the '\$'.

```
find /usr/local/bin/samtools
find /usr/local/bin/tabix
find /usr/local/bin/wget
find /usr/bin/gcc    ← note that the command for gcc is different from the other commands
```

Download and install Dendroscope, in order to read your tree file. (<http://dendroscope.org/>)

Download VCFtoTree v3.0.0. Open the "dist" folder and drag the dog icon to copy it to the desktop. Make sure that the application has permission to run from Security & Privacy under System Preferences.

https://www.dropbox.com/sh/sewjxmh9207fp89/AAASfUGqMBtd2_HO3CR7XA3ha?dl=0

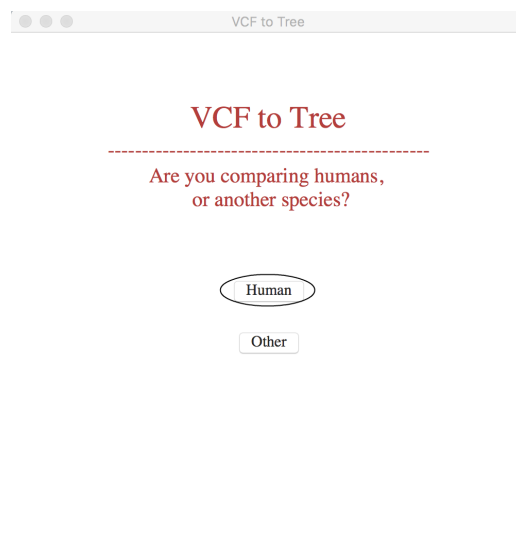


Double click on the dog icon on the desktop to open. The app will open, but does not open automatically. You will need to pull it up from the icon bar on your screen.

Using the VCFtoTree v3.0.0 software

As an example, we will walk through the process of creating the tree with VCFtoTree that was used in our paper *The psoriasis-associated deletion of late cornified envelope genes LCE3B and LCE3C has been maintained under balancing selection since Human Denisovan divergence* (Pajic et. al.). This paper discusses the LCE3BC deletion (Hg19, chr1:152,555,540–152,587,750). To create this tree, we will use the 5.6kb “target” sequence (chr1: 152,587,904–152,593,549).

The first screen asks if you are comparing humans or another species. LC3EBC is a human gene, so we’ll choose ‘Human’.



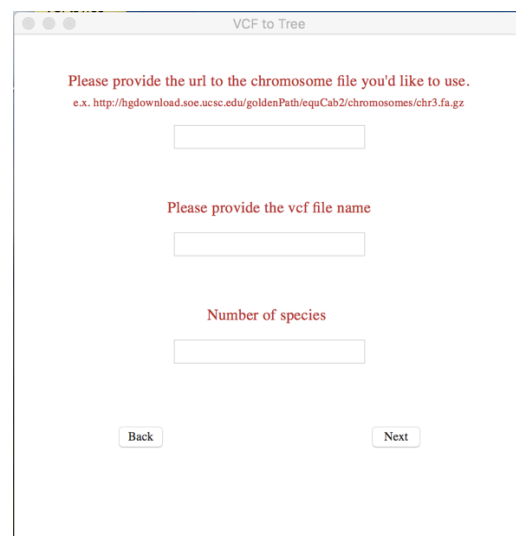
VCF to Tree

Are you comparing humans,
or another species?

Human

Other

If you would like to compare with another species, the program requests the URL to the chromosome file you’d like to use, the VCF file name, and the number of species.



VCF to Tree

Please provide the url to the chromosome file you'd like to use.
e.x. <http://hgdownload.soe.ucsc.edu/goldenPath/equCab2/chromosomes/chr3.fa.gz>

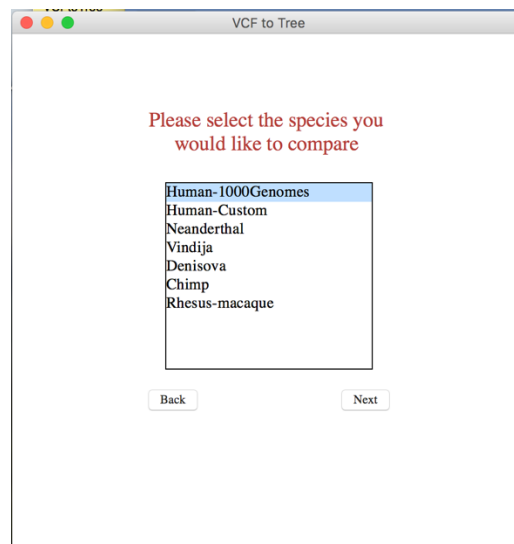
Please provide the vcf file name

Number of species

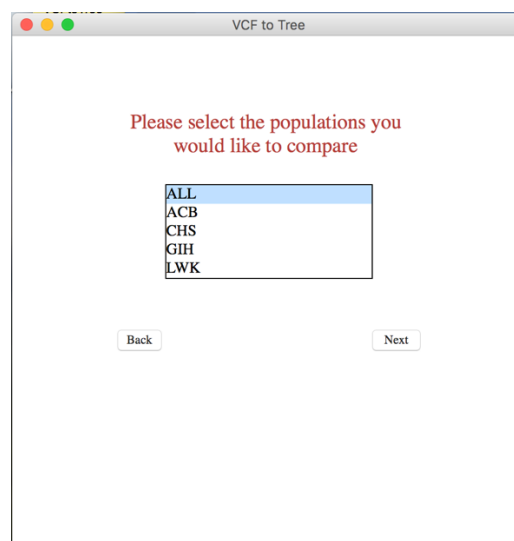
Back Next

The next screen asks you to select the species you would like to compare to. If you have chosen “Other Species” or choose to use the “Human-Custom” to compare with and your customized VCF file or reference are from a local address you will need to enter the full address of the file. Please use a compressed version of the file ending in “*.fa.gz” or “*.vcf.gz”. To enter the URL of a file, please include the “http://” part of the URL. To continue with our example, we will compare our LCE3BC target sequence to the Human Genome Project (Human-1000Genomes).

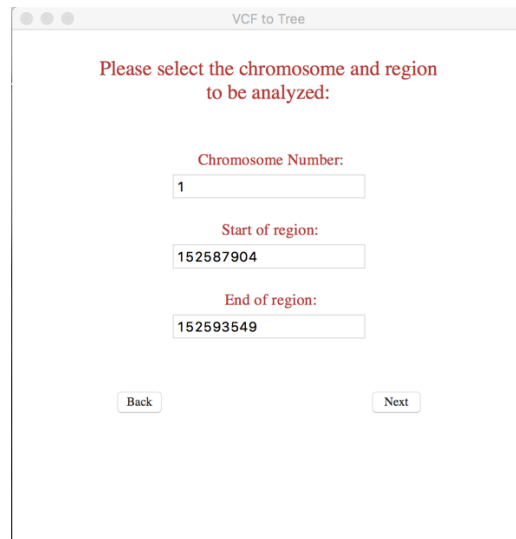
At this time the Vindija data does not work.



On the next screen you can select which populations you'd like to create a phylogeny from. We will choose 'ALL' to compare the VCF files from all populations.

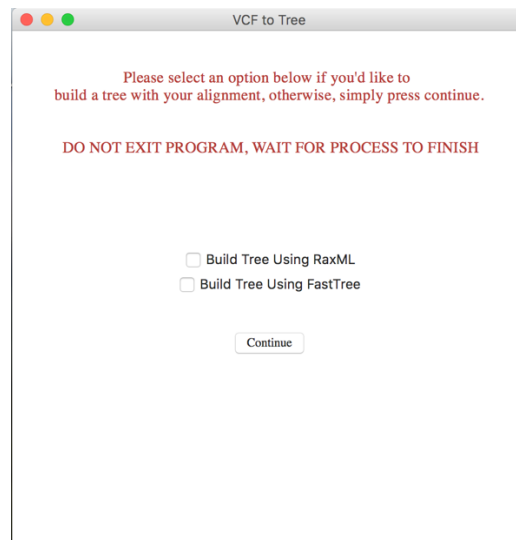


The next screen is where you will input your chromosome and region number. The LCE3BC deletion is located on chromosome 1, and the targeted sequence is 152,587,904–152,593,549. Do not include commas when inputting coordinates, as the tree will not be created.



The screenshot shows a web browser window titled "VCF to Tree". The main heading in red text says "Please select the chromosome and region to be analyzed:". Below this, there are three input fields. The first is labeled "Chromosome Number:" and contains the value "1". The second is labeled "Start of region:" and contains the value "152587904". The third is labeled "End of region:" and contains the value "152593549". At the bottom of the form, there are two buttons: "Back" and "Next".

On the next screen you are given the option of selecting between RAXML and FastTree for building your tree. Both RAXML and FastTree will create the same tree. RAXML is the traditional method of creating trees, while FastTree (<http://www.microbesonline.org/fasttree/>) is the newer version. You are welcome to choose whichever you prefer. If you do not wish to build a tree and simply need the alignment, click "continue" without choosing.



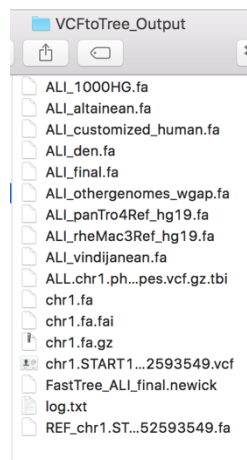
The screenshot shows a web browser window titled "VCF to Tree". The main heading in red text says "Please select an option below if you'd like to build a tree with your alignment, otherwise, simply press continue.". Below this, there is a red instruction: "DO NOT EXIT PROGRAM, WAIT FOR PROCESS TO FINISH". Further down, there are two radio button options: "Build Tree Using RaxML" and "Build Tree Using FastTree". At the bottom, there is a "Continue" button.

The next screen will show that the process has been completed. A folder on your desktop will be created called "VCFtoTree_Output." It may take some time for the folder to populate, especially if you use RAxML to build the tree. When the folder has populated, it will open and the process is complete. Once it has populated, you will want to rename the folder something for this specific tree as VCFtoTree will create the same output folder each time you run it.



Process Complete!

When your output folder populates it should list the following files:



Using Dendroscope, open the “***.newick” file. This is your tree.

