|  |  |
| --- | --- |
| http://icons.iconarchive.com/icons/visualpharm/icons8-metro-style/256/Sciences-Classes-Dna-helix-icon.png  Autosomal Pedigree Creator  Free and Open Source Software | Tutorial and Documentation  A tool that automatically creates a pedigree tree based on segment matches from a set of autosomal files.  Felix Chandrakumar <i@fc.id.au> |

Contents

[Document Revision History 2](#_Toc404616726)

[Getting Started 2](#_Toc404616727)

[Folder Structure 2](#_Toc404616728)

[Kit Preparation 2](#_Toc404616729)

[User Interface 3](#_Toc404616730)

[Execution 3](#_Toc404616731)

[Pedigree Output 4](#_Toc404616732)

[Tracing the Connection 5](#_Toc404616733)

[XML Representation 6](#_Toc404616734)

[Matching Segments 8](#_Toc404616735)

[Output Interpretation 11](#_Toc404616736)

[Acknowledgements 12](#_Toc404616737)

[Data and Information 12](#_Toc404616738)

[Icons 12](#_Toc404616739)

[License 12](#_Toc404616740)

[Developer 13](#_Toc404616741)

[Appendix 14](#_Toc404616742)

[Source Code 14](#_Toc404616743)

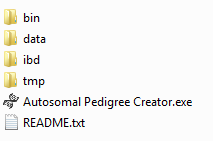
# Document Revision History

|  |  |  |
| --- | --- | --- |
| Date | Author | Description |
| 24-Nov-2014 | Felix Chandrakumar | Version 1.3 Release documentation |

# Getting Started

## Folder Structure

Download the Autosomal Pedigree Creator.zip file from the website which is usually less than 1 Mb and extracting it gives you the following files and folders.



* *bin* – contains bare minimum Graphviz binaries required to convert a .gv dot to PNG image file.
* *data* – intermediate folder
* *ibd* – intermediate folder
* *tmp* – intermediate folder
* *Autosomal Pedigree Creator.exe* – Executable
* *README.txt* – Readme file giving a quick overview of the software just in case you haven’t looked at the website.

## Kit Preparation

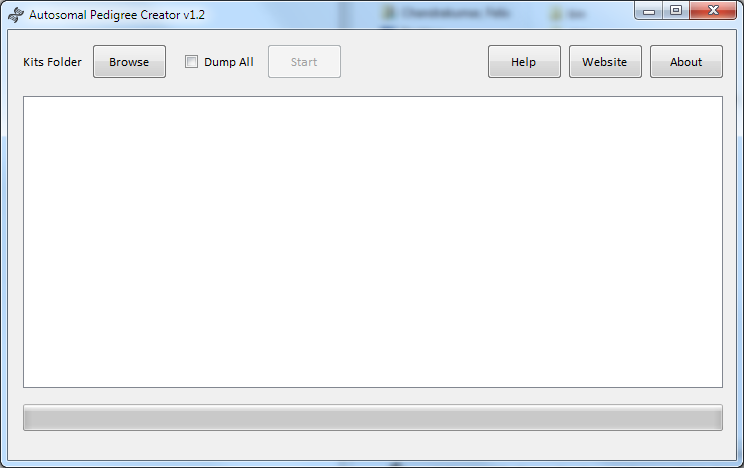
In order to use this tool, some basic preparation must be done. It is just renaming the files with humanly readable filenames. Please don’t change the file extensions. Please use only alphabets.

E.g.,

* 264652-autosomal-o37-results.csv.gz can be renamed to *Felix*.gz
* 264652-autosomal-o37-results.csv can be renamed to *Felix*
* genome\_v3\_Full\_20131006120000.zip can be renamed to *Felix*.zip

Once renamed, place all the renamed kits into a folder. This folder will be selected from the interface.

# User Interface

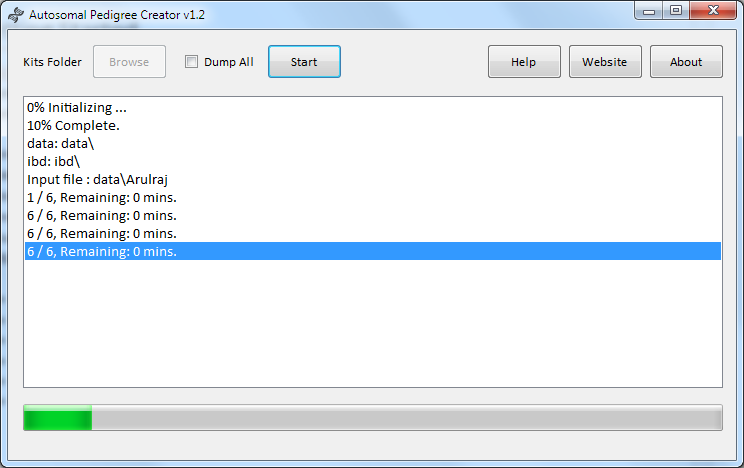


Usage of this tool is self-explanatory and below are the brief steps.

1. Click Browse and select folder where you had placed all the prepared kits.
2. Dump All –This option is only required when you have kits totally unrelated to each other and you want to dump every possible segment connection.
3. Click start and the process begins. The process can go on for a few minutes to several hours depending on the number of autosomal DNA files.

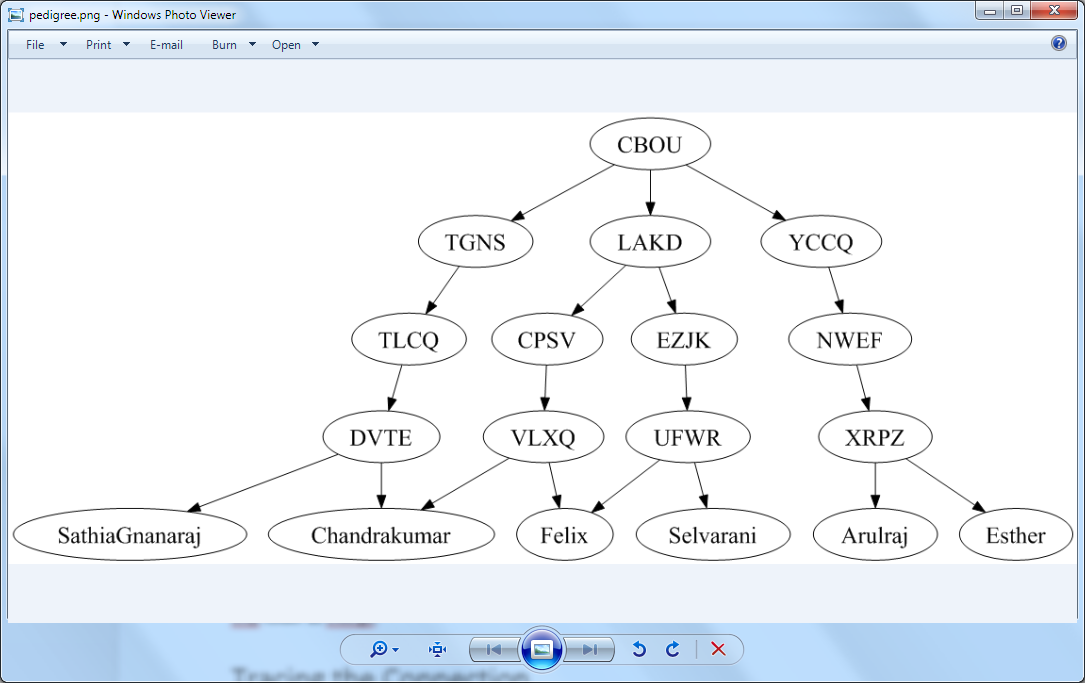
## Execution

The process executes sometimes for several hours. The progress seems to get stuck at 15% and then at 75%. This is not really stuck but it tries to extract as much information as possible in order to construct the tree and it does not know how far it has to go. Also, each comparison is done in parallel equal to the number of processors in your computer to accelerate the process.



# Pedigree Output

When the process finishes, a PNG file called pedigree.png will automatically open which contains the tree. For some reason if the PNG file didn’t open, then you can always find the file in the root folder of Autosomal Pedigree Creator.



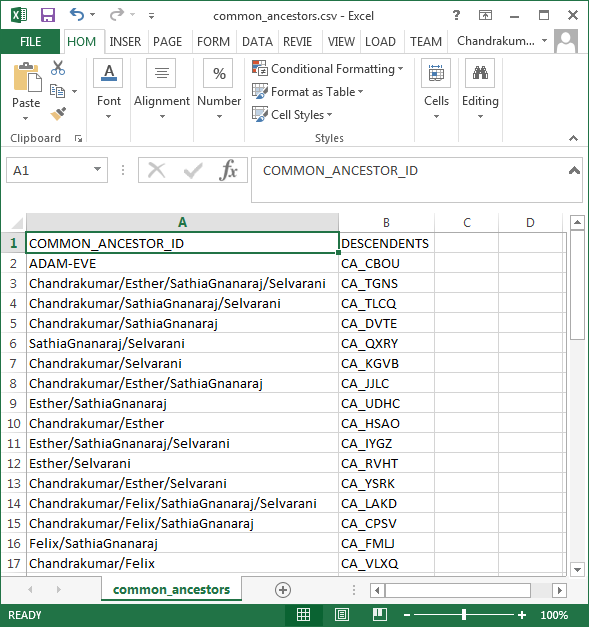
The tool uses Graphviz to generate the PNG file output from a .gv dot file. The .gv file can be found inside the tmp folder as tree.gv.

# Tracing the Connection

For some reason, if you want to check a connection between two common ancestors or two autosomal files, you can do so by following the below procedure.

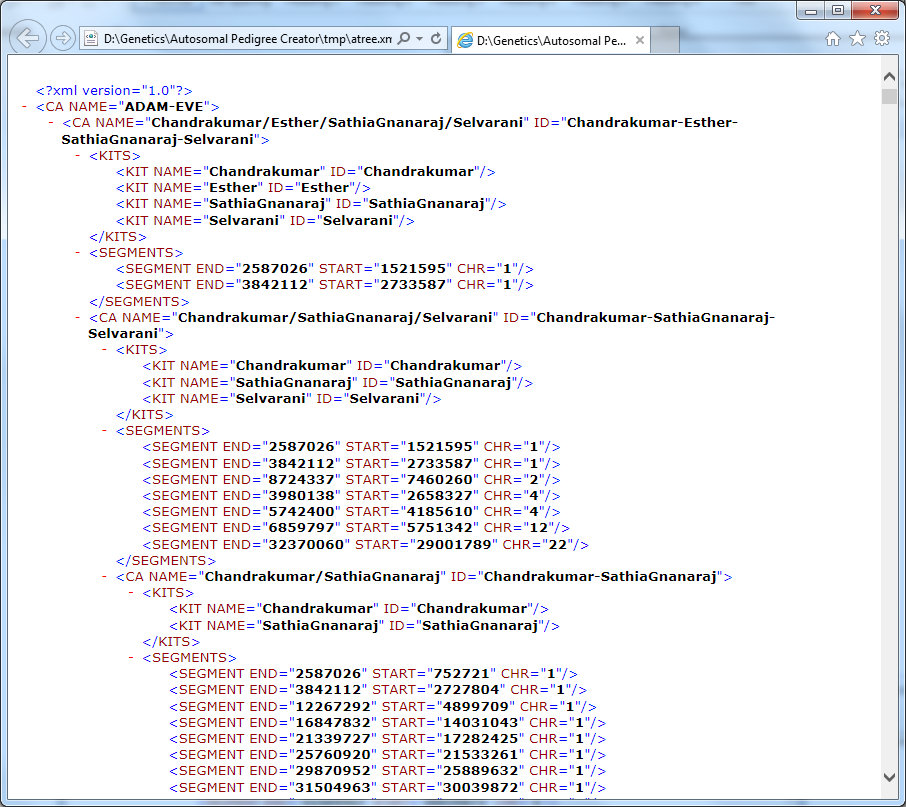
In the pedigree output, each line is a match, the terminals are autosomal files and the 4 letter ovals are common ancestors. The mapping between these 4 letters and what it means can be found inside tmp folder in the file common\_ancestors.csv which can be opened in excel.

As mentioned each arrow is a connection or a matching segment or a group of segments from a common ancestor.



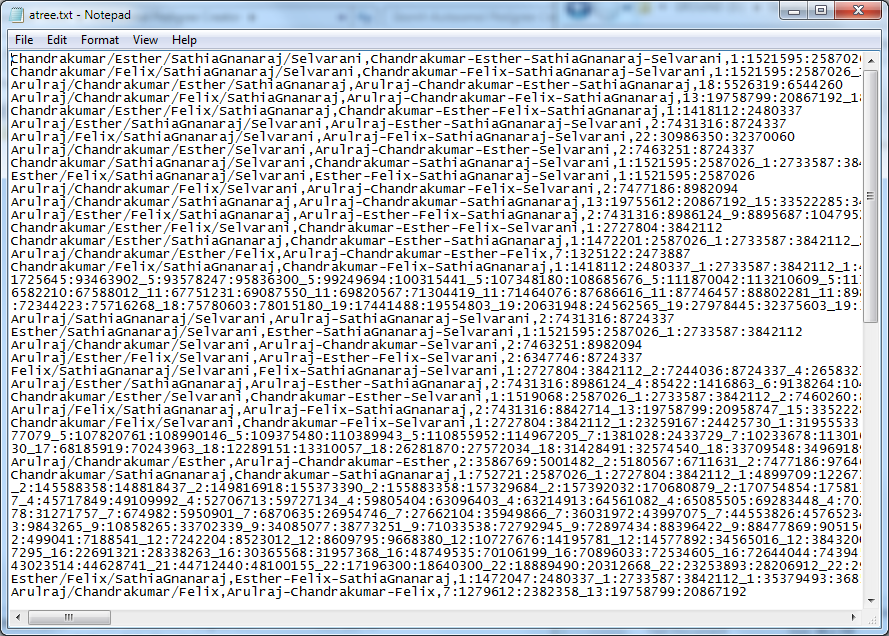
## XML Representation

The complete list of common ancestors and how each are related is present in the XML file atree.xml.



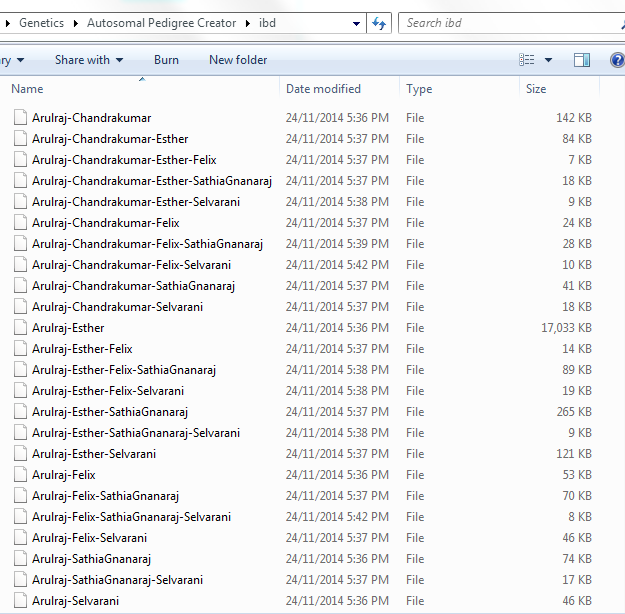
This file contains the common ancestor CA tag and the list of segments that match. Please note that all the sub nodes match all the segments at the parent level. Even though the root element is ADAM-EVE, its sub nodes are not automatically connected to the root. The root element is just for the sake of having a root element in XML and is not reproduced in the pedigree tree.

The XML is generated from a text file ‘atree.txt’. The XML file is simply a hierarchical representation of the text file.

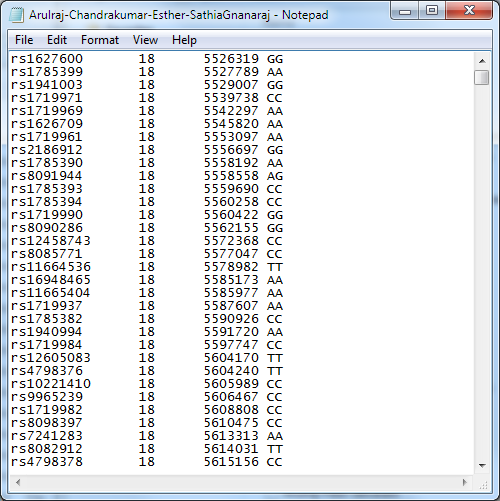


## Matching Segments

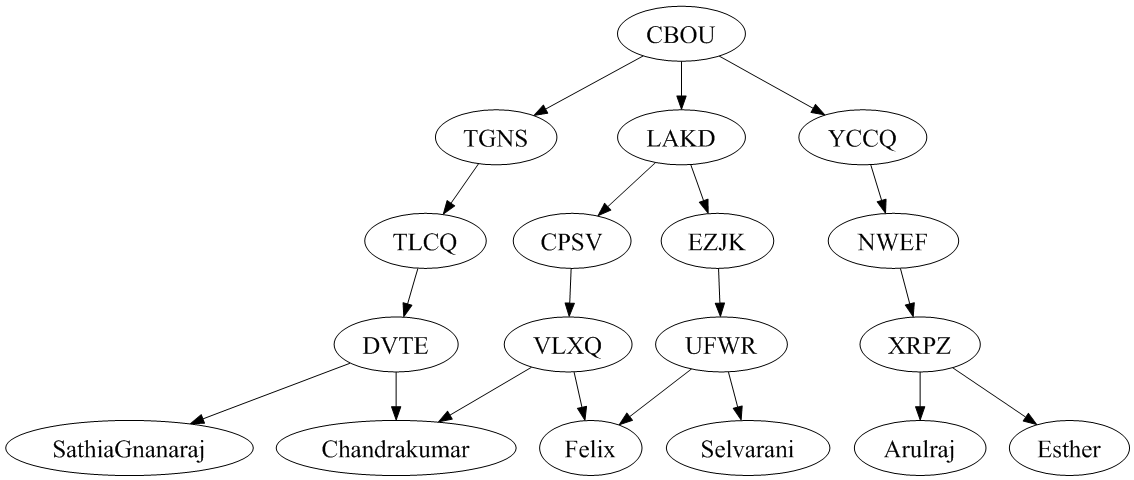
All matching segments can be found inside the ‘ibd’ folder. Please note ‘ibd’ is just a folder name and does not automatically mean they haven’t had recombination or Identity By Descent. However, all matching segments inside ‘ibd’ folder are compound segments.



Opening a file say, Arulraj-Chandrakumar-Esther-SathiaGnanaraj means, the segment is common across Arulraj, Chandrakumar, Esther and SathiaGnanaraj autosomal files and it represents the common ancestor for the kits involved.



# Output Interpretation



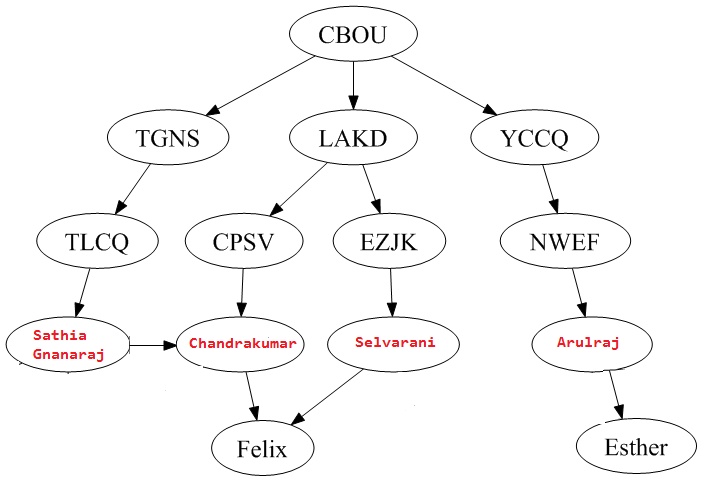
You might wonder why there are some common ancestors represented as 4 chars in ovals but has only one descendant common ancestor also represented as 4 chars in ovals. The reason is because, these intermediate common ancestors do have population data or segments matching the individuals but does not match the parents. If you want to include all such matching segments from population data, you can enable ‘Dump All’ option. However, be warned that ‘Dump All’ can create a clumsy pedigree because every individual may match every common ancestor depending on how close they are related.

The above output which is closely correct, but still requires some manual intervention and adjustments to get accurate pedigree.

For the above pedigree, below are the true relations.

* Felix (self)
* Chandrakumar (Father)
* Selvarani (Mother)
* Sathia Gnanaraj (Paternal grandfather)
* Esther (Wife)
* Arulraj (Father in law)

There is no common ancestors between Felix and Chandrakumar (because Chandrakumar is my father). So, VLXQ name represented as a common ancestor between myself and my father is none other than my father himself. Similarly for all parent/child relations. It is not possible to automate this using computer algorithms because, a computer can only say if a relation is parent/child but it cannot find who the parent is unless it has all the required data surrounding it which is not possible or feasible all the time. Changing the parent/child relations will lead to the below modified pedigree.



As you can see, I can infer the following from the autosomal pedigree tree.

* My wife’s tree is separate line.
* There are three individual common ancestors giving three lines.
* My parents are distant cousins.

# Acknowledgements

### Data and Information

* Graphviz: *http://www.graphviz.org*
* SharpZipLib: *http://icsharpcode.github.io/SharpZipLib/*

### Icons

* Application Icon: *https://www.iconfinder.com/icons/175409/dna\_helix\_icon#size=512*

## License

The MIT License (MIT)

Copyright © 2014 Felix Jeyareuben Chandrakumar

http://www.y-str.org

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software. THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

## Developer

Felix Jeyareuben Chandrakumar is a software professional working at Hewlett-Packard based out of Canberra, Australian Capital Territory. He is married and have two daughters. He finished his Bachelor of Engineering (Computer Science) a decade back and also done his Master of Science (Cyber Security and Forensic Computing) from University of South Australia. He develops free software for churches and for Christians in his part-time which are available at churchsw.org. He is an Australian by nationality and South Indian by ethnicity. Genealogy is his hobby and he develops several open source applications and maintains Genetic Genealogy Tools available at y-str.org website in his part-time. He is also the group administrator for Nadar (a South Indian community), U9 mtDNA haplogroup and Ancient DNA FTDNA projects.

He can be contacted anytime at i@fc.id.au

# Appendix

## Source Code

Autosomal Pedigree Creator is built in C# and requires Microsoft .Net 4.0. The source code is uploaded to GitHub at <https://github.com/fcidau/Autosomal-Pedigree-Creator>