|  |  |
| --- | --- |
| http://icons.iconarchive.com/icons/visualpharm/icons8-metro-style/256/Sciences-Classes-Dna-helix-icon.png  Genetic Genealogy Kit  Free and Open Source Software | Tutorial and Documentation  A kit management, analysis and matching tool for Autosomal DNA, X-DNA, Y-DNA and Mitochondrial DNA.  Felix Chandrakumar <i@fc.id.au> |

Contents

[Document Revision History 3](#_Toc398706002)

[Getting Started 3](#_Toc398706003)

[Version Differences 3](#_Toc398706004)

[Launching the Kit 3](#_Toc398706005)

[Kits Management 4](#_Toc398706006)

[New Kit 4](#_Toc398706007)

[Name and Kit Number 4](#_Toc398706008)

[Autosomal 4](#_Toc398706009)

[Y DNA 5](#_Toc398706010)

[Mt-DNA 8](#_Toc398706011)

[Save Kit 9](#_Toc398706012)

[Enable Kit 9](#_Toc398706013)

[Disable Kit 9](#_Toc398706014)

[Delete Kit 10](#_Toc398706015)

[Edit Kit 10](#_Toc398706016)

[Quick Edit 11](#_Toc398706017)

[Export Kit 13](#_Toc398706018)

[Import Kit 13](#_Toc398706019)

[Autosomal Analysis 14](#_Toc398706020)

[One to One comparison 14](#_Toc398706021)

[One to Many comparison 15](#_Toc398706022)

[Admixture 17](#_Toc398706023)

[Runs of Homozygosity 20](#_Toc398706024)

[Phasing Utility 21](#_Toc398706025)

[Process Kits 22](#_Toc398706026)

[Y-DNA Analysis 24](#_Toc398706027)

[ISOGG Y-Tree 24](#_Toc398706028)

[Mitochondrial Analysis 25](#_Toc398706029)

[Mt-DNA Phylogeny 25](#_Toc398706030)

[Mito Map 25](#_Toc398706031)

[Tools 27](#_Toc398706032)

[Settings 27](#_Toc398706033)

[Conclusion 28](#_Toc398706034)

[Roadmap 28](#_Toc398706035)

[Known Bugs 28](#_Toc398706036)

[ISOGG Y-Tree 28](#_Toc398706037)

[Phased Segment Visualizer 28](#_Toc398706038)

[Known Flaws 29](#_Toc398706039)

[Acknowledgements 29](#_Toc398706040)

[Data and Information 29](#_Toc398706041)

[Icons 29](#_Toc398706042)

[License 30](#_Toc398706043)

[Developer 30](#_Toc398706044)

[Appendix 31](#_Toc398706045)

[Reference Populations 31](#_Toc398706046)

[Source Code 32](#_Toc398706047)

[Database Structure (ggk.db) 32](#_Toc398706048)

[Table – kit\_master 32](#_Toc398706049)

[Table - cmp\_status 33](#_Toc398706050)

[Table – cmp\_autosomal 33](#_Toc398706051)

[Table – cmp\_mrca 34](#_Toc398706052)

[Table –cmp\_phased 34](#_Toc398706053)

[Table – ggk\_settings 34](#_Toc398706054)

[Table – kit\_autosomal 35](#_Toc398706055)

[Table – kit\_mtdna 35](#_Toc398706056)

[Table – kit\_phased 35](#_Toc398706057)

[Table – kit\_roh 35](#_Toc398706058)

[Table – kit\_ysnps 36](#_Toc398706059)

[Table – kit\_ystr 36](#_Toc398706060)

[File Format (.ggk) 36](#_Toc398706061)

[Sample Structure 36](#_Toc398706062)

# Document Revision History

|  |  |  |
| --- | --- | --- |
| Date | Author | Description |
| 03-Jul-2014 | Felix Chandrakumar | Version 1.1 Release documentation |
| 17-Sep-2014 | Felix Chandrakumar | Note for Comparisons and Admixture Analysis. |

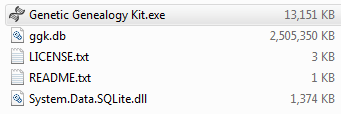
# Getting Started

## Version Differences

There are two versions available to download. A large version (GGK.zip ~ 400 MB) and a reduced version (GGK Reduced.zip ~ 11 MB). Both versions are basically the same except the database (ggk.db) of larger version will have reference samples for the calculation of admixture. For those who require admixture results can opt for the reduced version. Please refer Appendix to know the list of reference populations included in the larger version.

## Launching the Kit

The download, irrespective of larger or reduced version, it contains 3 important files that are required to function normally.



Double click on ‘Genetic Genealogy Kit.exe’ to open the tool. For the larger version, the tool may look like getting stuck at ‘Checking Integrity of DB …’. It will take a while to start the first time, so please wait until it finishes and the status changed to ‘Done’.

# Kits Management

## New Kit

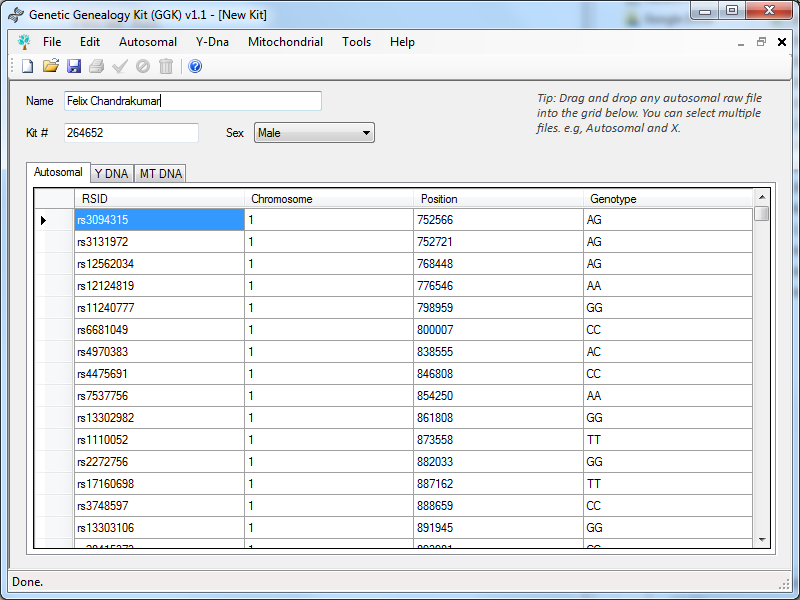
A new kit window can be open either by using File (menu) and New (sub-menu) or by using the New kit icon () on the toolbar.

### Name and Kit Number

All kits require a Name and a Kit number which are mandatory. The kit number can match the DNA testing company’s kit number. However, please make sure it is unique in your list of kits managed by you in the database.

### Autosomal

To enter autosomal data, you can drag and drop a FTDNA or 23andMe data files. For FTDNA you may need to select both autosomal and X-chromosome and drag both files into the data grid under the tab which says autosomal.



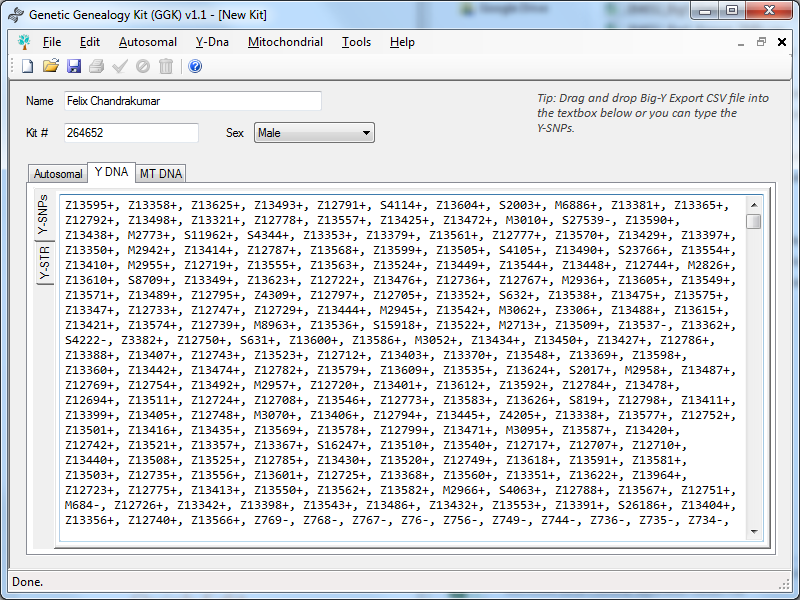
You should be able to edit and change any value.

### Y DNA

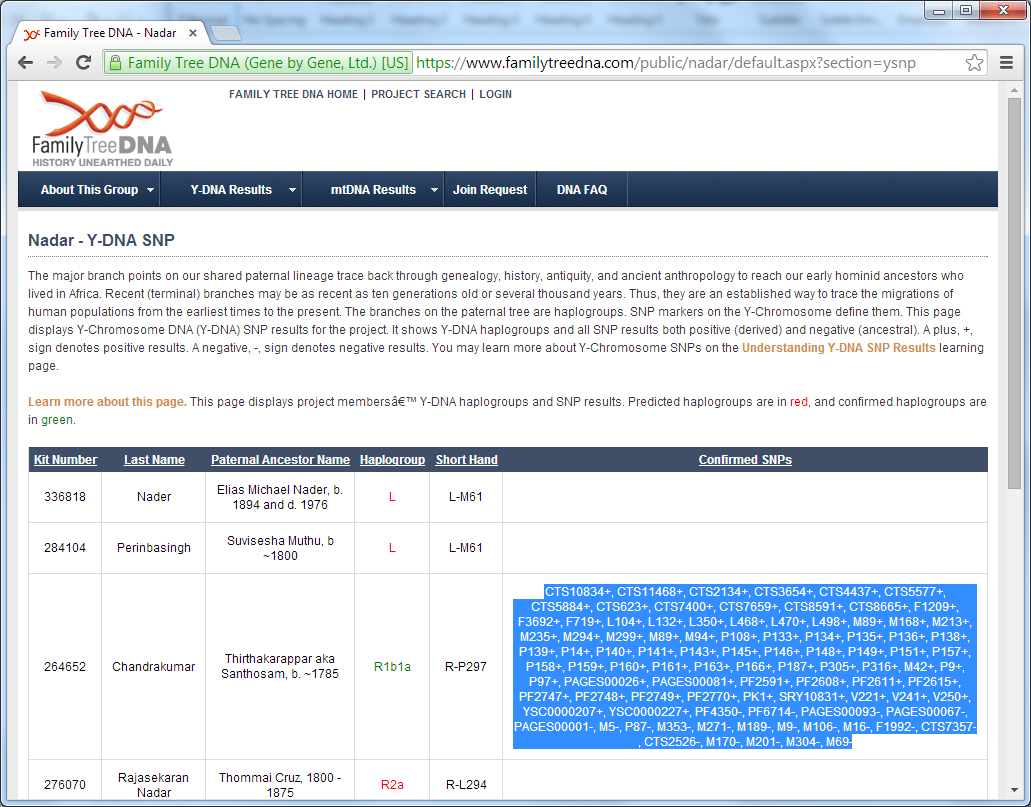
You can enter Y-DNA details from the Y DNA tab.

#### Y-SNPs

You can enter all the Y-SNPs separated by comma. If you had done Big-Y, The easiest way to populate Y-SNPs is by downloading the CSV export (e.g., <kit-no>\_BigY\_Data\_yyyymmdd.csv) from FTDNA Big-Y page and drag drop into the text area. The software will automatically populate the Y-SNPs for you.



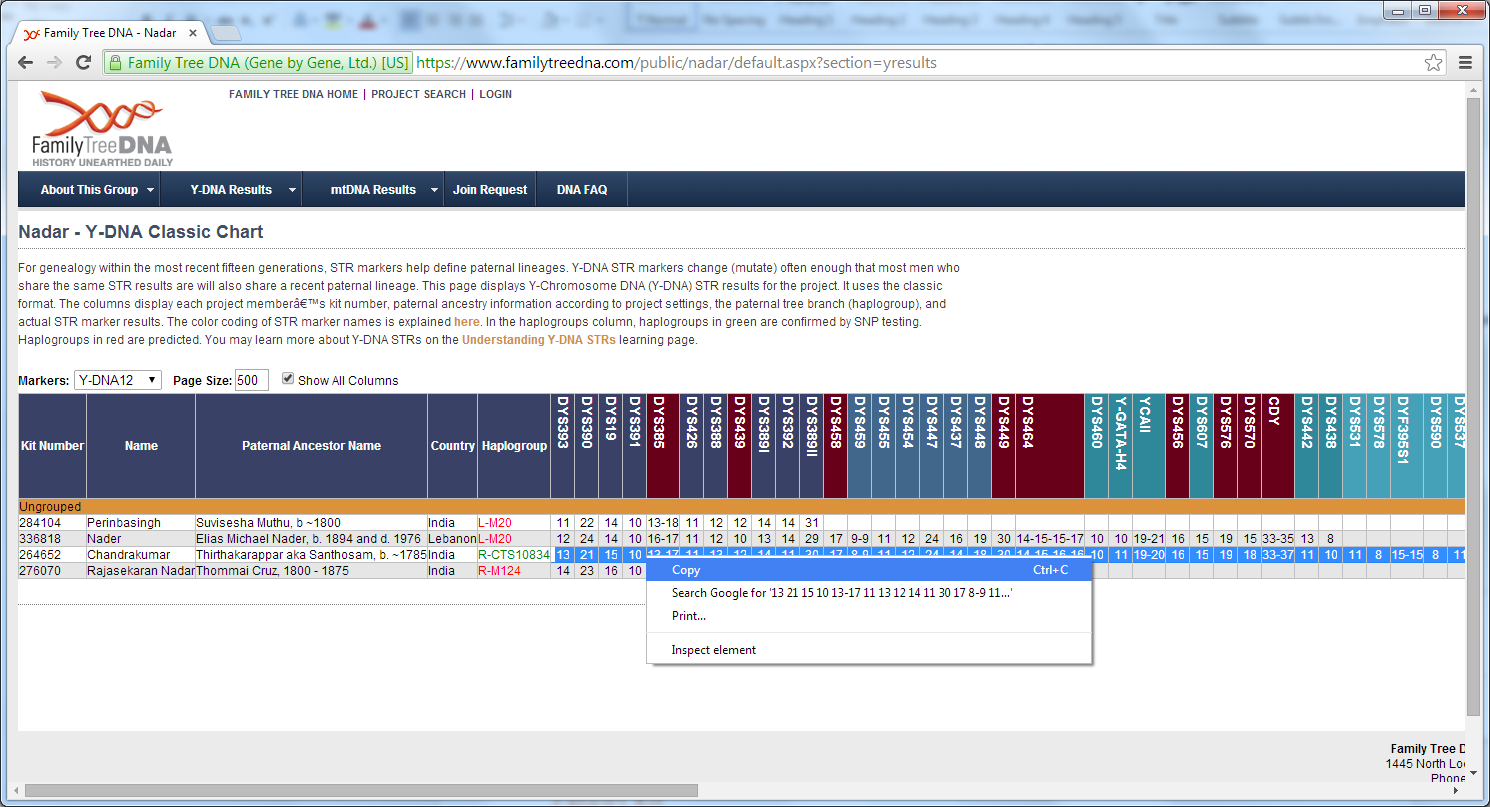
You can also manually add additional SNPs which you know that are positive or negative.



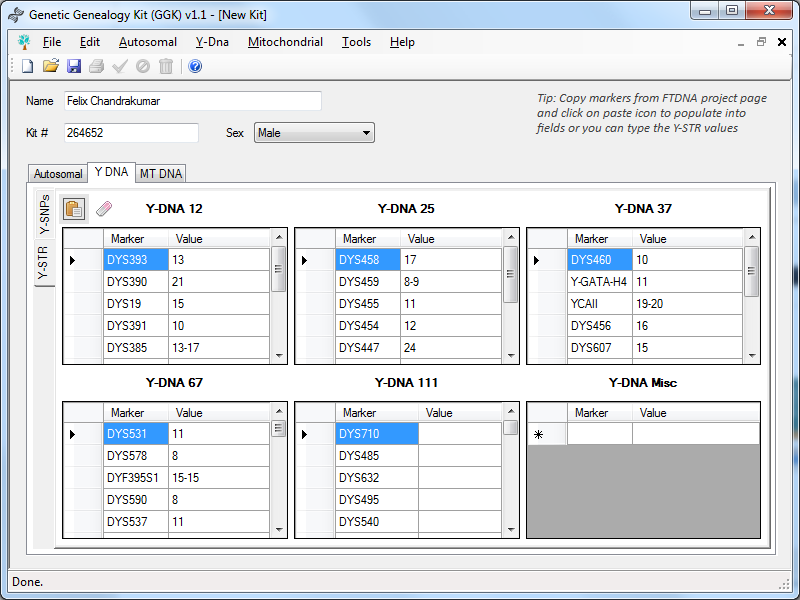
You can also copy the SNPs from FTDNA project page and paste them into the Y-SNPs text area.

#### Y-STR

Entering Y-STR values is one of the boring stuff as it has lots of values and prone to errors. To make things easier, you can copy the values from FTDNA project page and paste them into the Y-STR sections using the paste icon ().



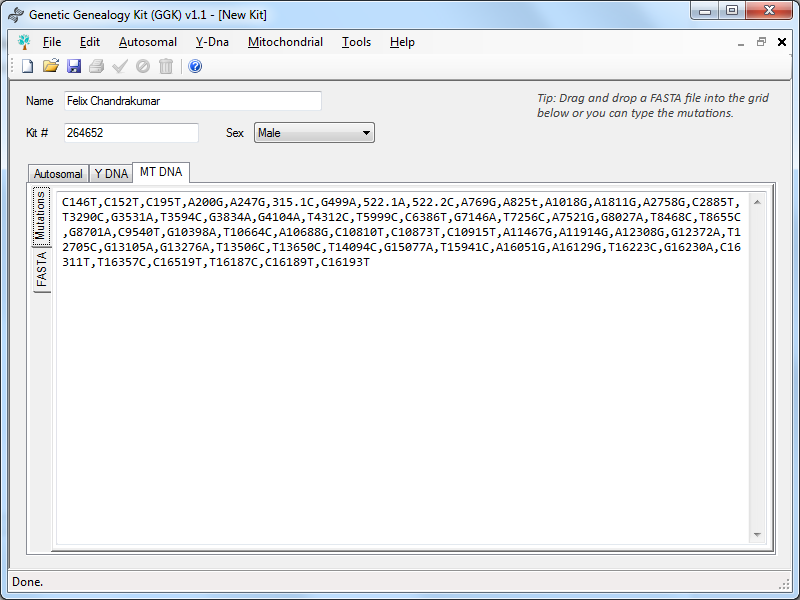
Go to a FTDNA project page, select all the SNPs, right click and copy.



Once copied, you can paste it directly into the Y-STR section. For other Y-STR values which are not available, you can manually enter them in Y-DNA Misc.

### Mt-DNA

Drag and drop a FASTA file into the text area. This will automatically convert to mutations and paste the FASTA file into the FASTA tab section.



You can also edit them or manually enter them as comma separated values.

### Save Kit

Once all the available values are entered, you can simply click the save icon () on the toolbar or the Edit (menu) and Save (sub-menu).

The saving may take a while as it takes nearly half a million SNP values and bulk inserts them into the database. So, please be patient when saving.

### Enable Kit

If the kit is disabled, you can enable it by either using the enable icon () from toolbar or the Edit (menu) and Enable (sub-menu).

### Disable Kit

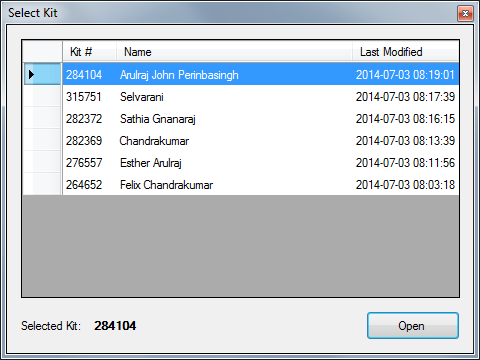
You can disable a kit by either using the disable icon () from toolbar or the Edit (menu) and Disable (sub-menu).

### Delete Kit

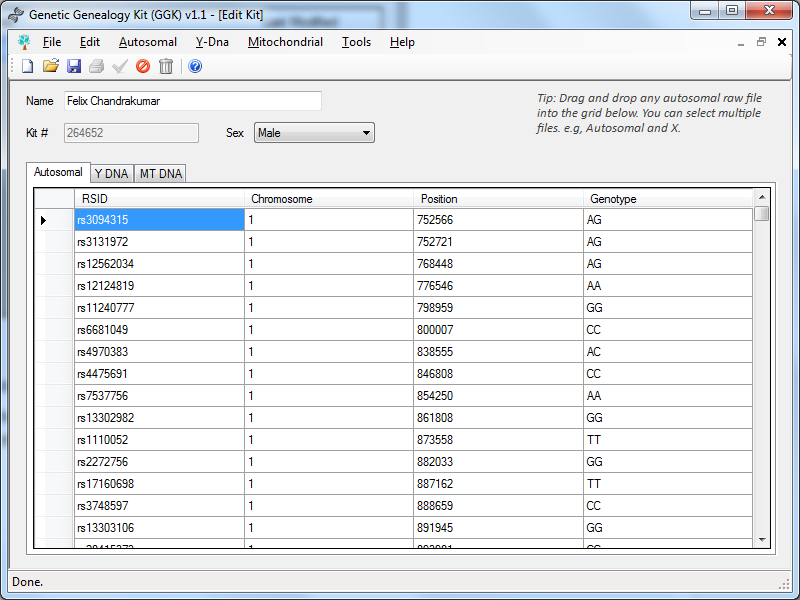
You can delete a kit by either using the disable icon () from toolbar or the Edit (menu) and Delete (sub-menu). It is important to note that, delete can sometimes take longer to finish. If you have many kits added and a Kit Processing is done, then deletion can take even longer to complete. This is because, when deleting a kit, it actually deletes everything associated with it, including the comparison data with every other kit and reference populations. Hence, it is always better to disable if there is no real requirement to delete it completely.

## Edit Kit

To open a kit to edit, you can either use the open icon () from toolbar or the File (menu) and Open (sub-menu).



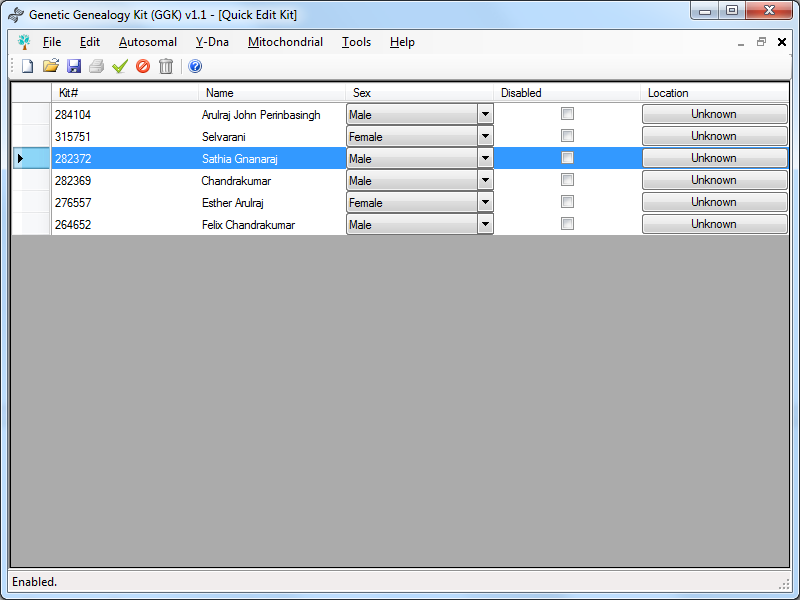
The select kit will list all the kits you had saved. Click ‘Open’ button to open the kit. Once open, you can change all values except the Kit number.



Once the changes are done, you can save the kit by using the save icon () from toolbar or File (menu) and Save (sub-menu).

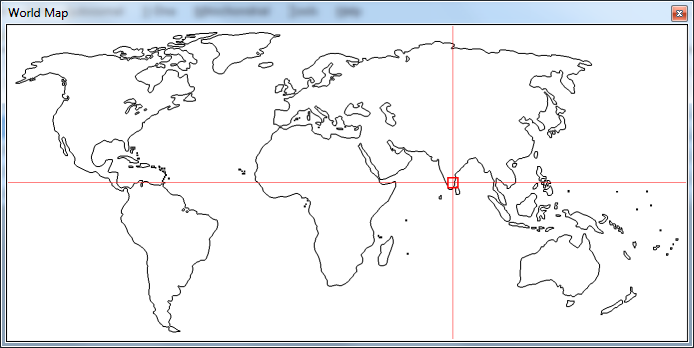
## Quick Edit

The purpose of Quick Edit is to quickly edit the kits which doesn’t require any changes to the actual genetic data values. Quick Edit can be accessed from Edit (menu) and Quick Edit (sub-menu).



You can modify a kit or multiple kits here. By pressing control key you should be able to multi-select kits. You can select the kit(s) and save (), enable (), disable () and delete () from the toolbar and/or from the corresponding menus.

You can also select a location by clicking the location button.



The location data is saved in an internal format.

## Export Kit

All kits that are added can also be exported. A kit containing Autosomal data, X-DNA, Y-DNA and mt-DNA can be exported as a .ggk file (Genetic Genealogy Kit file). This file has the sole purpose of exporting all data with the ability to reimport back without having to reenter any of the DNA details. This helps to share a DNA kit file with all details which can be reimported. A kit can be exported from the File (menu) and Export (sub-menu).

Export Kit supports the following formats:

* Genetic Genealogy Kit format – *contains Autosomal data, X-DNA, Y-DNA and mt-DNA*
* FTDNA format – *Autosomal and X DNA*
* 23andMe format - *Autosomal and X DNA*

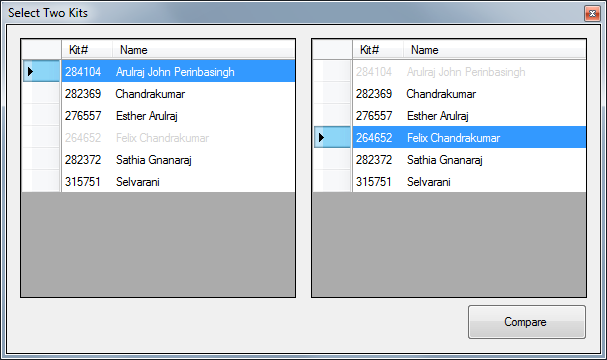
## Import Kit

All kits exported into .ggk file (Genetic Genealogy Kit file) can be reimported into GGK. Import supports only Genetic Genealogy Kit file format. A kit can be imported from the File (menu) and Import (sub-menu).

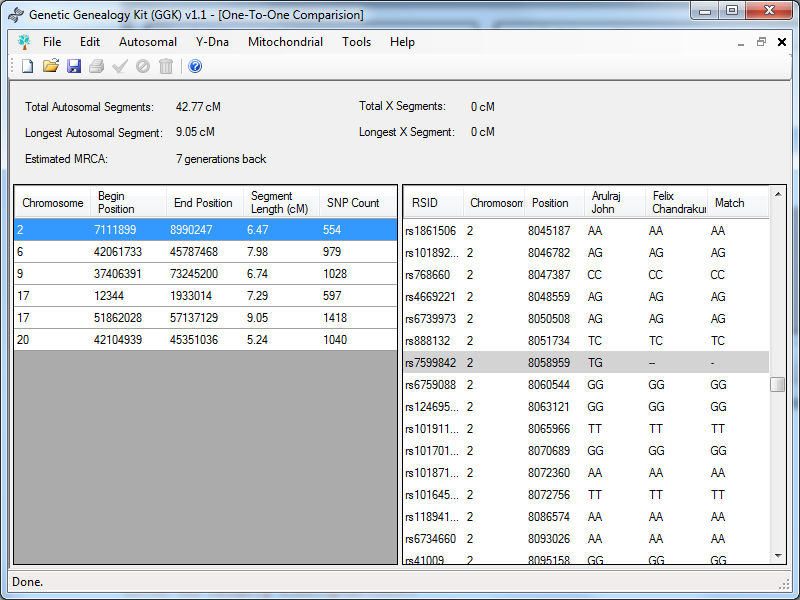
# Autosomal Analysis

## One to One comparison

One-to-one comparison can be done by Autosomal (menu) and One to One (sub-menu). Two kits must be selected and click compare.



The comparison can take a while to finish if ‘Process Kits’ wasn’t done.



The left side data grid lists the matching segments and upon selection of each segment, the right side lists the SNPs of both kits and how they match. A no call match will be highlighted by a dull gray background and a non-matching SNP included as a match in the matching segment will have a bright red background.

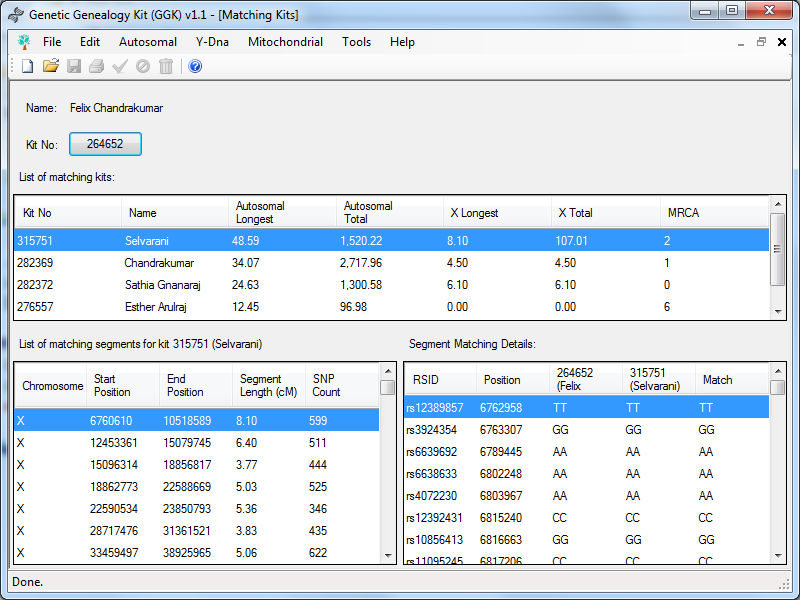
If any of the two kit is phased, upon double click on the segment, you can view how the Phased Segment Visualizer to see how the kit’s paternal and maternal match the segment.

You can click the Save icon () in toolbar (or) File (menu) and Save (sub-menu) to get the SNPs for common ancestor.

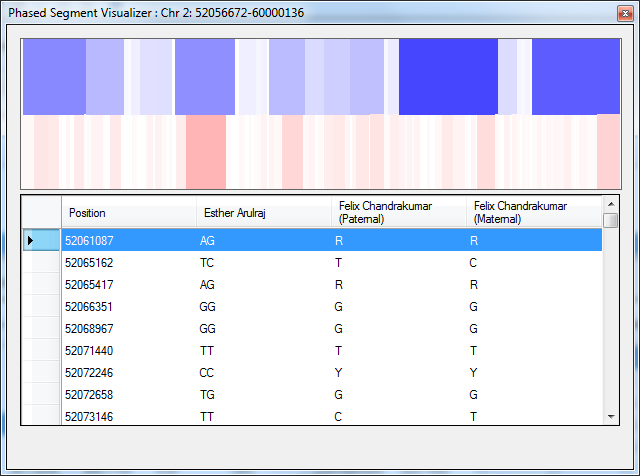
**Note:** *If you had changed parameters that affect comparison from Tools (menu) -> Settings, you must use Autosomal Analysis -> Process kits and check ‘Delete Existing Comparisons’ and process the kits again. This is required because, comparisons are cached to improve speed and performance.*

## One to Many comparison

One-to-Many will not work until the ‘Process Kits’ is completed. For more information on kits processing, refer to ‘Process Kits’ section. Once-to-Many comparison can be done by Autosomal (menu) and One to Many (sub-menu).



Upon selecting a kit, it lists the matching kits in the top data grid. When a matching kit is selected, the bottom left data grid displays the matching segment. Upon selecting a matching segment, the right side data grid displays the segment details SNP by SNP for further investigation.   
If any of the kit is phased, then upon double clicking the matching segment, the phased segment visualizer opens up.

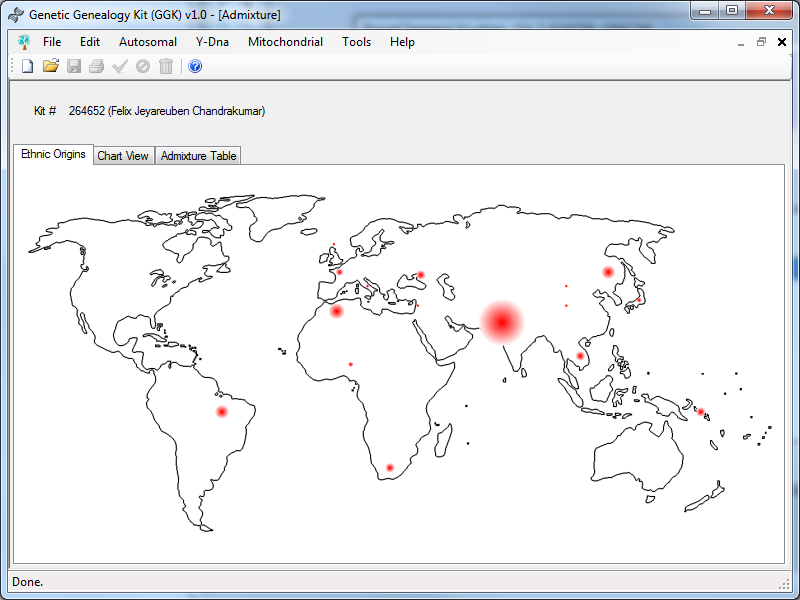


The blue represents the paternal and the red represents the maternal. The intensity is directly related to the sequence length of match SNPs before a mismatch is found. It gives a visualization of how the compound segment is getting matched and from where the segment came from i.e., either paternal or maternal.

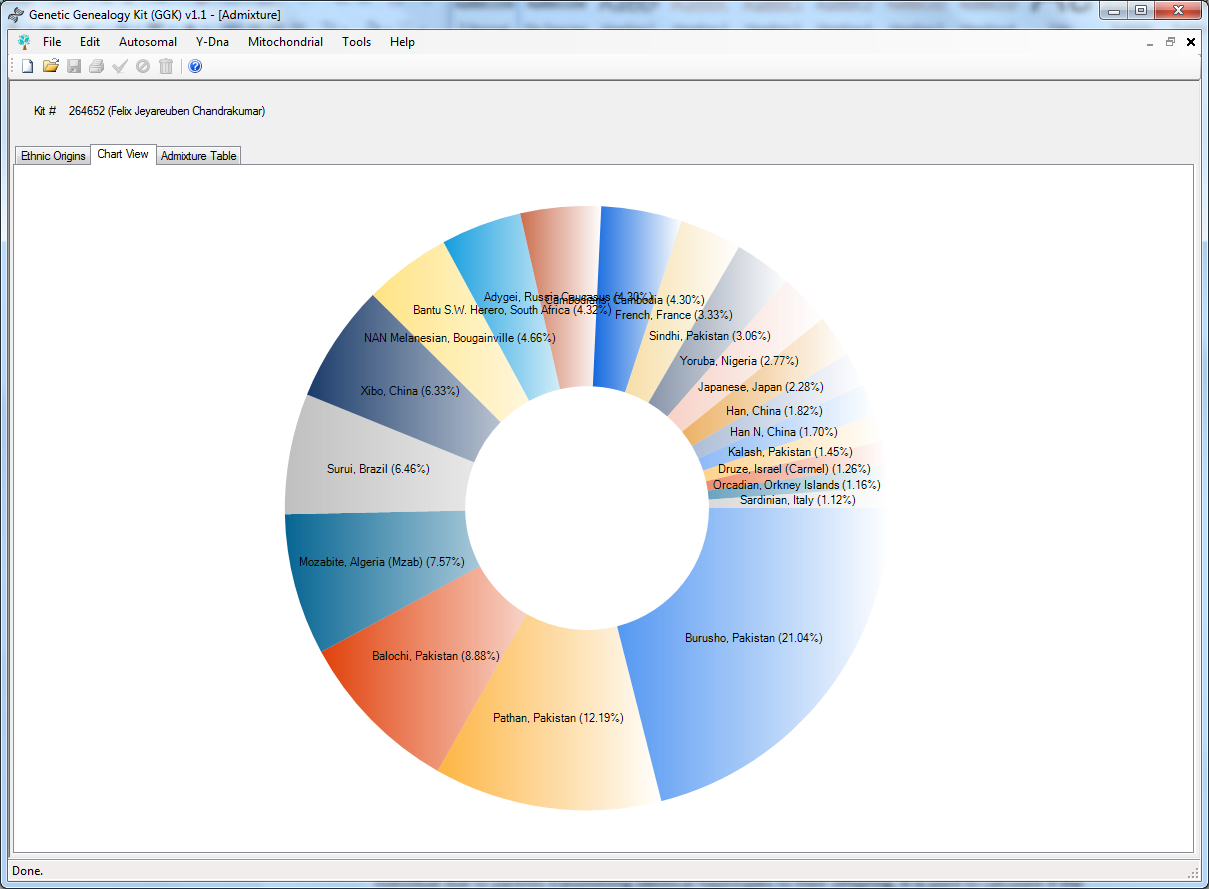
**Note:** *If you had changed parameters that affect comparison from Tools (menu) -> Settings, you must use Autosomal Analysis -> Process kits and check ‘Delete Existing Comparisons’ and process the kits again. This is required because, comparisons are cached to improve speed and performance.*

## Admixture

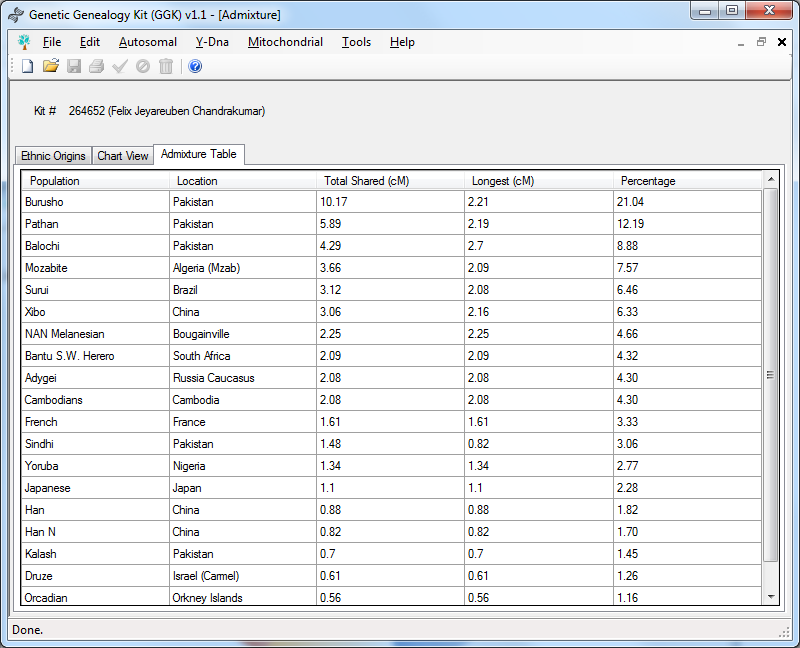
Admixture can be calculated by Autosomal (menu) and Admixture (sub-menu). The admixture is calculated based on compound segments among the reference populations. 56 reference population samples are included in the large version which assists in the admixture calculation. In order for admixture to work, the ‘Process Kit’ must be executed and completed with ‘No Admixture Calculation’ unchecked.



Ethnic origins will be plotted on the world based on compound segment matches.



The chart view will provide a pie-chart and the admixture table will provide the details of the compound segment matches.

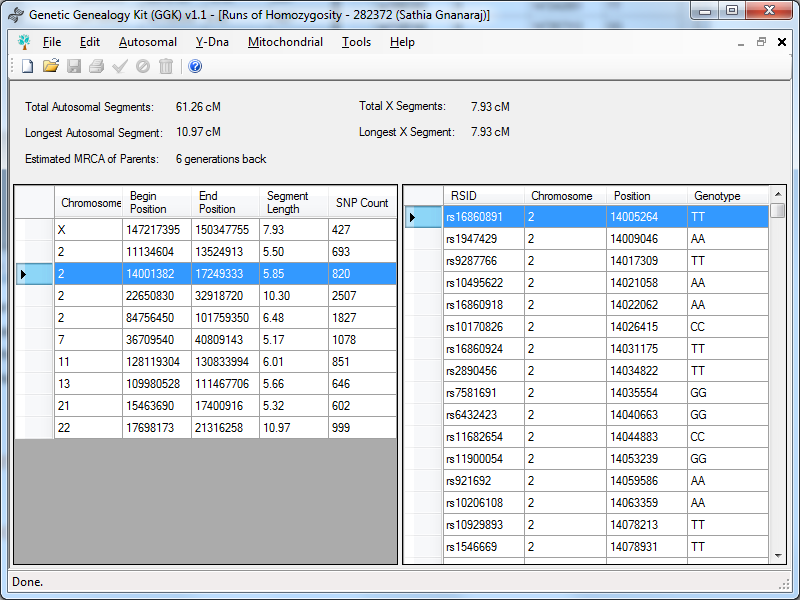


The admixture table provides the details of the compound segment matches.

**Note:** *Any segment length greater than 3 cM in length will be ignored and the lower limit is based on the settings value (default value is 0.5 cM). This is because, any segment length greater than 3 cM represents you belong to a population and you already know that you belong to them. It is based on the fact that people from same/similar geographical locations have segment matches at 3 cM. However, any compound segments less than 3 cM is will be of interest as it details the admixture from other populations and unknown geographical locations. This autosomal admixture is not about finding which population you belong to, but rather, which populations your ancestors were from beyond genealogical timeframe. For example, my ancestors and I were from South India. I wouldn’t be interested in knowing that my ancestral ethnicity is South India which I already know, but rather what other populations I match beyond paper trail and genealogical timeframe.*

## Runs of Homozygosity

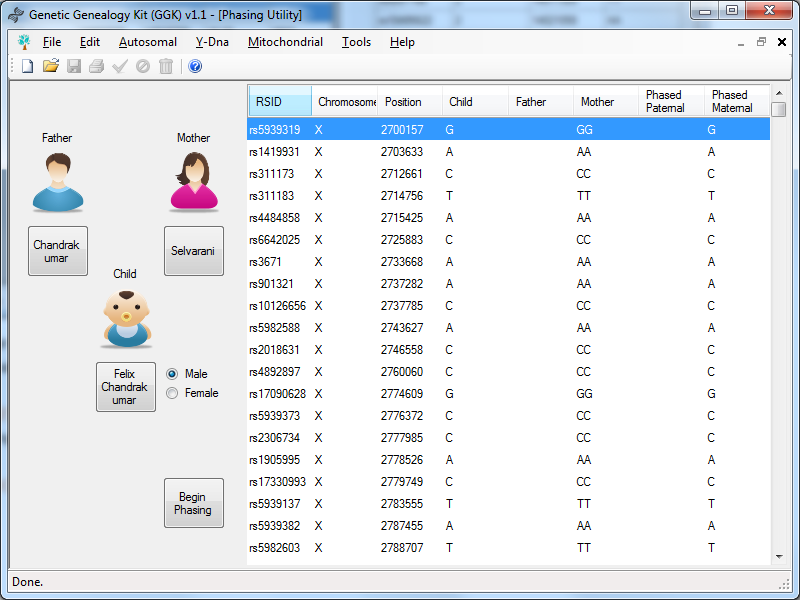
Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes that are present in an individual due to parents transmitting identical haplotypes to their offspring. It is used to calculate if the parents are related/cousins. Runs of Homozygosity can be done by Autosomal (menu) and Runs of Homozygosity (sub-menu).



Similar to the one to one comparison, Runs of Homozygosity also has the left side data grid populated with the matching segments and upon selection of each segment, the right side lists the SNPs.

## Phasing Utility

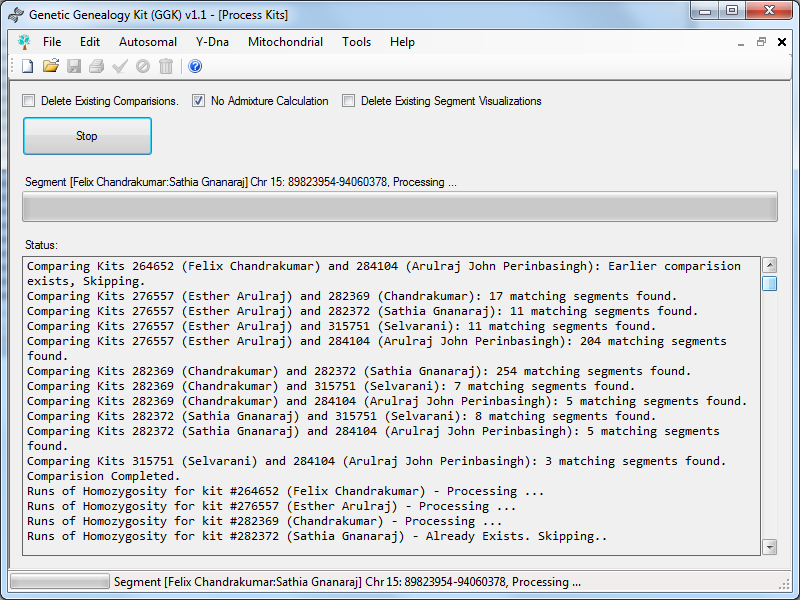
Phasing is the process of determining which DNA came from the mother, and which came from the father. Phasing can be done by Autosomal (menu) and Phasing Utility (sub-menu).



Click on the button ‘Select Father’ to select the father’s kit. Similarly select the mother’s and the child’s kit. Choose the sex of the child and click ‘Begin Phasing’. You must select at-least 1 parent and the child for phasing to begin. Once phasing is done, you can view matching segments of the phased kit to determine how the segment is matching and where the segment came from. Always phase the kits before proceeding to ‘Process kits’.

## Process Kits

Process Kits allows you to do all the required comparisons, admixture calculations, runs of homozygosity, phasing the matching segments for all phased kits, etc as one batch process. Process Kits can take a long time to complete. Hence, it is advisable to begin the processing of kits overnight. Kits can be processed by Autosomal (menu) and Process Kits (sub-menu).



Process Kits comes with three options.

#### Delete Existing Comparisons

Process kits takes the default thresholds from the settings. However, if you wish to reprocess the kits, you can check this option to delete existing comparisons.

#### No Admixture Calculation

Considering there are 56 reference population samples, and if you only have a few kits and don’t require admixture calculation for those kits, then you can check this option.

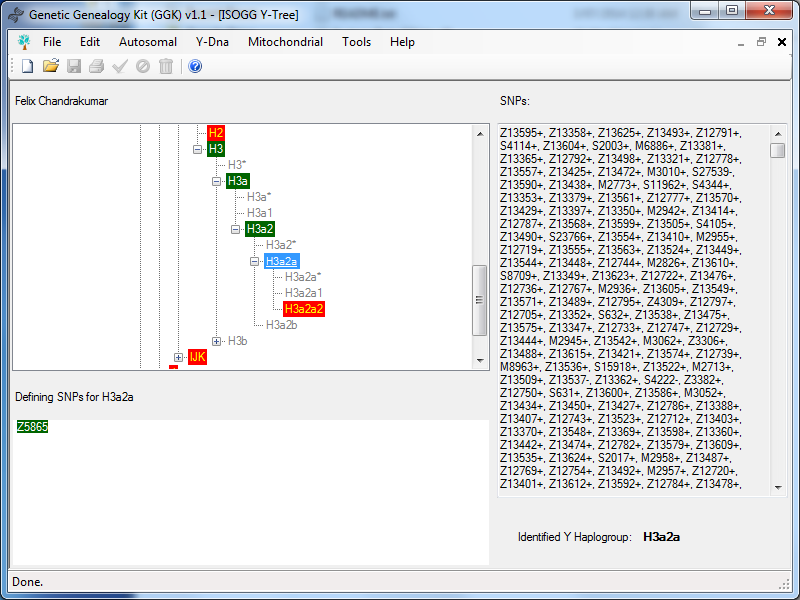
#### Delete Existing Segment Visualizations

If you wish to reprocess phased segment visualizations, you can do so by checking this option. This again is useful only when you change the thresholds.

# Y-DNA Analysis

## ISOGG Y-Tree

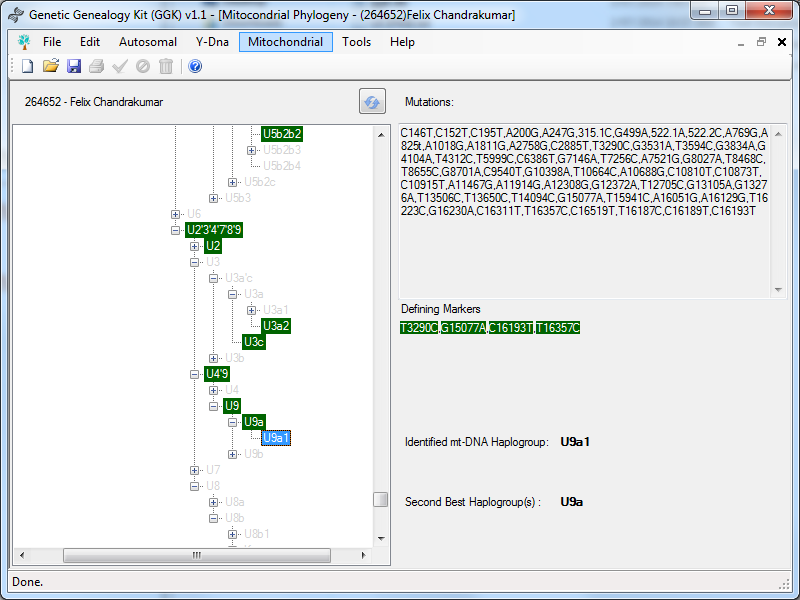
Since 2006 ISOGG has hosted the regularly updated online ISOGG Y-chromosome phylogenetic tree. The ISOGG Y-Tree which is included in GGK was retrieved on 10th March 2014. Kits having Y-SNP details can be plotted on ISOGG Y-Tree by Y-Dna (menu) and ISOGG Y-Tree (sub-menu). Y-SNPs that are positive are plotted as green and the negative SNPs are plotted as red. Identified Y-Haplogroup will also be displayed.



# Mitochondrial Analysis

## Mt-DNA Phylogeny

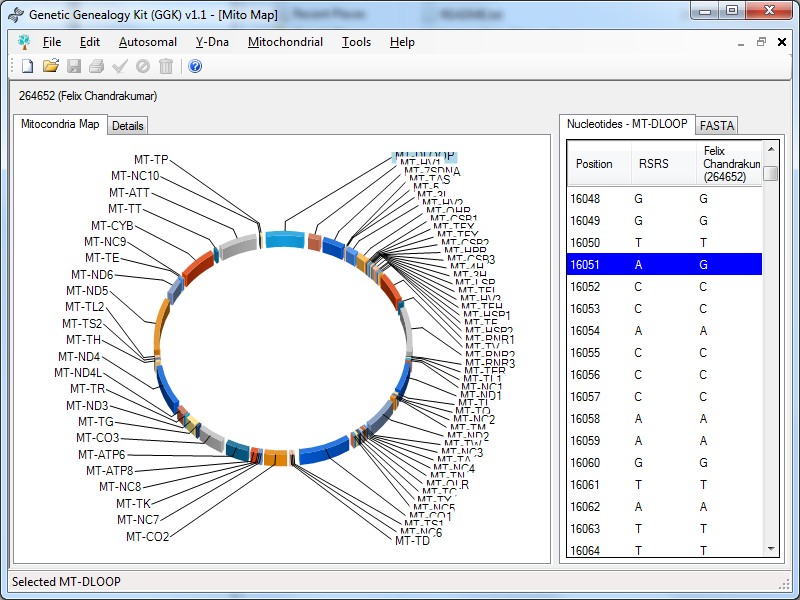
Kits having Mt-DNA details can be plotted on Human Mt-DNA Phylogeny by Mitochondrial (menu) and Mt-DNA Phylogeny (sub-menu).



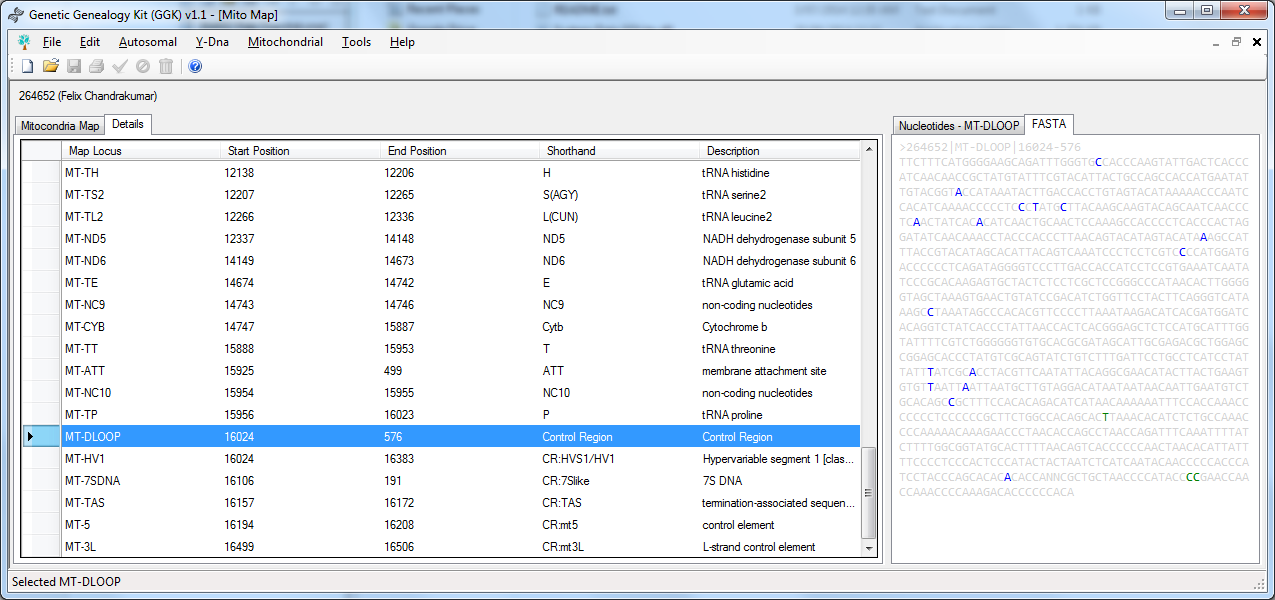
You can click the Save icon () in toolbar (or) File (menu) and Save (sub-menu) to get the mtDNA report for the kit.

## Mito Map

Kits having Mt-DNA details can be mapped, charted and visualized by Mitochondrial (menu) and Mito Map (sub-menu).



On the above visualization, you can left click and drag up/down to rotate vertically, and right click and drag left/right to rotate horizontally. You can click on each segment/ gene and view the comparison between RSRS and the selected kit.

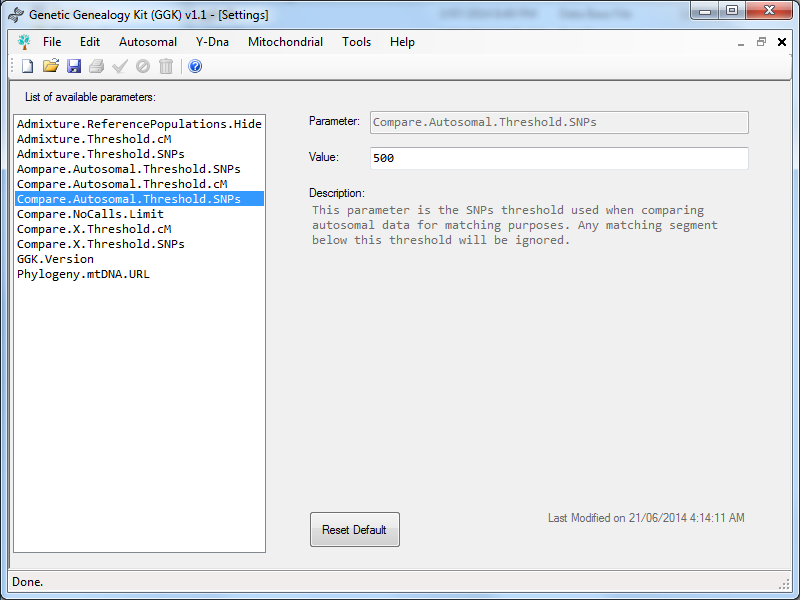


You can also view the details and get the FASTA for that section/gene alone with mutations plotted.

# Tools

## Settings

Some settings can be customized using Tools (menu) and Settings (sub-menu). You can select a setting which is listed on the left side, change the value and click on the Save icon () in toolbar (or File menu and select Save). You can always revert the values by clicking ‘Reset Default’ which automatically reverts and saves it.



# Conclusion

## Roadmap

The core idea for this is GGK software to have one central DNA kit management tool for genetic genealogy which is capable of doing everything related to genetic genealogy.

It is planned to have the following features in coming future versions apart from bug fixes.

* Y-DNA Comparisons
* Mt-DNA Comparisons
* Y-STR TMRCA calculations
* Custom reference populations for admixture calculations
* Phased Paternal/Maternal to kit.
* Segment wise search.
* Triangulations
* Encrypted export file.

If you have any ideas, feel free to contact me.

## Known Bugs

### ISOGG Y-Tree

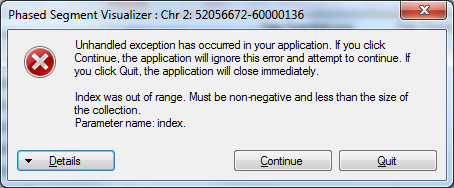
When selecting a kit for Y-DNA ISOGG Y-Tree when Y-SNPs where not entered for the Kit, it will throw an error similar to the one below.



The workaround is not to select a kit for ISOGG Y-Tree which doesn’t have Y-SNPs entered, or just click continue and ignore the error.

### Phased Segment Visualizer

In the Phased Segment Visualizer, when you use the mouse to click at the left corner of the image, you get the below error.



The workaround is just to click continue and ignore the error.

The mentioned above bugs will be fixed in the next upcoming version. If you face any bugs, feel free to contact me and I will do my best to fix it.

## Known Flaws

* The help icon in toolbar doesn’t work.
* The print icon in toolbar never gets enabled.

Both functionalities are never implemented. It will be implemented in future versions.

## Acknowledgements

### Data and Information

* HGDP Autosomal Reference Population: *http://www.hagsc.org/hgdp/files.html*
* ISOGG Y-Tree: *http://isogg.org/tree/*
* mtDNA Phylogeny: *http://www.mtdnacommunity.org/downloads/mtDNAPhylogeny.xml*
* Mitocondria Locations: *http://www.mitomap.org/bin/view.pl/MITOMAP/GenomeLoci*

### Icons

* Application Icon: *https://www.iconfinder.com/icons/171268/tree\_icon#size=128*
* Eraser Icon: *http://www.iconarchive.com/show/farm-fresh-icons-by-fatcow/draw-eraser-icon.html*
* Paste Icon: *http://findicons.com/icon/67168/gnome\_edit\_paste?id=67897*
* Delete Icon: *https://www.iconfinder.com/icons/100980/delete\_trash\_icon#size=16*
* Disable Icon: *http://findicons.com/icon/116499/150?id=293646*
* Enable Icon: *http://findicons.com/icon/2982/tick*
* Girl Icon: *http://www.iconarchive.com/show/pretty-office-9-icons-by-custom-icon-design/baby-girl-icon.html*
* Boy Icon: *http://www.iconarchive.com/show/pretty-office-9-icons-by-custom-icon-design/baby-boy-icon.html*
* Woman Icon: *http://www.iconarchive.com/show/pretty-office-4-icons-by-custom-icon-design/woman-icon.html*
* Man: *http://www.iconarchive.com/show/pretty-office-2-icons-by-custom-icon-design/man-icon.html*
* Refresh Icon: *http://www.iconarchive.com/show/basic-2-icons-by-pixelmixer/reload-icon.html*

## 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 Adelaide, South Australia. He is married and have two daughters. He finished his Bachelor of Engineering (Computer Science) in 2003 and Master of Science (Cyber Security and Forensic Computing) from University of South Australia in 2014. 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) and U9 mtDNA haplogroup FTDNA projects.

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

# Appendix

## Reference Populations

There are 56 sample reference populations included in the large version. The samples are chosen in such a way that 1 sample is taken for each population to reduce the size and time required to compare. Below are the list of reference population samples.

|  |  |
| --- | --- |
| **HGDP Sample** | **Population** |
| HGDP01385 | Adygei, Russia Caucasus |
| HGDP00058 | Balochi, Pakistan |
| HGDP01408 | Bantu N.E., Kenya |
| HGDP00993 | Bantu S.E. Pedi, South Africa |
| HGDP00994 | Bantu S.E. S.Sotho, South Africa |
| HGDP01034 | Bantu S.E. Tswana, South Africa |
| HGDP01033 | Bantu S.E. Zulu, South Africa |
| HGDP01035 | Bantu S.W. Herero, South Africa |
| HGDP01031 | Bantu S.W. Ovambo, South Africa |
| HGDP00611 | Bedouin, Israel (Negev) |
| HGDP00448 | Biaka Pygmies, Central African Republic |
| HGDP00011 | Brahui, Pakistan |
| HGDP00356 | Burusho, Pakistan |
| HGDP00712 | Cambodians, Cambodia |
| HGDP01308 | Dai, China |
| HGDP00557 | Druze, Israel (Carmel) |
| HGDP00521 | French, France |
| HGDP01361 | French Basque, France |
| HGDP00973 | Han, China |
| HGDP01291 | Han N, China |
| HGDP00103 | Hazara, Pakistan |
| HGDP01237 | Hezhen, China |
| HGDP00747 | Japanese, Japan |
| HGDP00766 | Japanese, Japan |
| HGDP00290 | Kalash, Pakistan |
| HGDP00998 | Karitiana, Brazil |
| HGDP01323 | Lahu, China |
| HGDP00131 | Makrani, Pakistan |
| HGDP00911 | Mandenka, Senegal |
| HGDP00864 | Maya, Mexico |
| HGDP00982 | Mbuti Pygmies, Democratic Republic of Congo |
| HGDP01198 | Miaozu, China |
| HGDP01263 | Mozabite, Algeria (Mzab) |
| HGDP00662 | NAN Melanesian, Bougainville |
| HGDP01337 | Naxi, China |
| HGDP01172 | North Italian, Italy (Bergamo) |
| HGDP00794 | Orcadian, Orkney Islands |
| HGDP01212 | Oroqen, China |
| HGDP00675 | Palestinian, Israel (Central) |
| HGDP00542 | Papuan, New Guinea |
| HGDP00234 | Pathan, Pakistan |
| HGDP01048 | Pima, Mexico |
| HGDP00880 | Russian, Russia |
| HGDP00991 | San, Namibia |
| HGDP00667 | Sardinian, Italy |
| HGDP01335 | She, China |
| HGDP00163 | Sindhi, Pakistan |
| HGDP00833 | Surui, Brazil |
| HGDP01347 | Tu, China |
| HGDP01101 | Tujia, China |
| HGDP01167 | Tuscan, Italy |
| HGDP01305 | Uygur, China |
| HGDP01249 | Xibo, China |
| HGDP00946 | Yakut, Siberia |
| HGDP01186 | Yizu, China |
| HGDP00927 | Yoruba, Nigeria |

## Source Code

Genetic Genealogy Kit (GGK) is built in C# and requires Microsoft .Net 4.0. The source code is uploaded to GitHub at <https://github.com/fcidau/Genetic-Genealogy-Kit>

## Database Structure (ggk.db)

The database used by Genetic Genealogy Kit (ggk.db) is a SQLite Database.

### Table – kit\_master

#### DDL

CREATE TABLE [kit\_master] (

[kit\_no] TEXT NOT NULL,

[name] TEXT NOT NULL,

[sex] CHAR NOT NULL DEFAULT U,

[disabled] INTEGER NOT NULL DEFAULT 0,

[reference] INTEGER NOT NULL DEFAULT 0,

[x] INTEGER DEFAULT 0,

[y] INTEGER DEFAULT 0,

[roh\_status] INTEGER NOT NULL DEFAULT 0,

[last\_modified] DATETIME DEFAULT CURRENT\_TIMESTAMP,

CONSTRAINT [sqlite\_autoindex\_kit\_master\_1] PRIMARY KEY ([kit\_no]));

#### Description

The table contains the kit details.

### Table - cmp\_status

#### DDL

CREATE TABLE [cmp\_status] (

[cmp\_id] INTEGER NOT NULL PRIMARY KEY AUTOINCREMENT,

[kit1] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[kit2] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[status\_autosomal] INTEGER NOT NULL DEFAULT 0,

[status\_ysnp] INTEGER NOT NULL DEFAULT 0,

[status\_ystr] INTEGER NOT NULL DEFAULT 0,

[status\_mtdna] INTEGER NOT NULL DEFAULT 0,

[at\_longest] DOUBLE NOT NULL DEFAULT (0.0),

[at\_total] DOUBLE NOT NULL DEFAULT (0.0),

[x\_longest] DOUBLE NOT NULL DEFAULT (0.0),

[x\_total] DOUBLE NOT NULL DEFAULT (0.0),

[mrca] INTEGER NOT NULL DEFAULT 0,

[mt\_haplogroup] TEXT,

[y\_haplogroup] TEXT,

[last\_processed] DATETIME NOT NULL DEFAULT CURRENT\_TIMESTAMP);

#### Description

The table contains status of comparison.

### Table – cmp\_autosomal

#### DDL

CREATE TABLE [cmp\_autosomal] (

[segment\_id] INTEGER NOT NULL PRIMARY KEY AUTOINCREMENT,

[cmp\_id] INTEGER NOT NULL REFERENCES [cmp\_status]([cmp\_id]) ON DELETE CASCADE ON UPDATE CASCADE,

[kit1] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[kit2] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[chromosome] TEXT NOT NULL,

[start\_position] INT NOT NULL,

[end\_position] INT NOT NULL,

[segment\_length\_cm] DOUBLE NOT NULL,

[snp\_count] INT NOT NULL,

[segment\_type] CHAR NOT NULL DEFAULT U,

[last\_modified] DATETIME NOT NULL DEFAULT CURRENT\_TIMESTAMP);

#### Description

The table contains the segment details for one to one comparison.

### Table – cmp\_mrca

#### DDL

CREATE TABLE [cmp\_mrca] (

[rsid] TEXT NOT NULL,

[chromosome] TEXT NOT NULL,

[position] INTEGER NOT NULL,

[kit1\_genotype] TEXT,

[kit2\_genotype] TEXT,

[match] TEXT,

[segment\_id] INTEGER NOT NULL REFERENCES [cmp\_autosomal]([segment\_id]) ON DELETE CASCADE ON UPDATE CASCADE);

#### Description

The table contains the SNP details for one to one comparison.

### Table –cmp\_phased

#### DDL

CREATE TABLE [cmp\_phased] (

[phased\_kit] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[match\_kit] TEXT NOT NULL,

[chromosome] TEXT NOT NULL,

[start\_position] INTEGER NOT NULL,

[end\_position] INTEGER NOT NULL,

[segment\_image] BLOB,

[segment\_xml] TEXT);

#### Description

The table holds the phased segment details along with the phased segment image and the segment details as XML.

### Table – ggk\_settings

#### DDL

CREATE TABLE [ggk\_settings] (

[key] TEXT NOT NULL,

[value] TEXT NOT NULL,

[description] TEXT NOT NULL,

[readonly] INT NOT NULL DEFAULT 0,

[last\_modified] DATETIME NOT NULL DEFAULT CURRENT\_TIMESTAMP,

CONSTRAINT [sqlite\_autoindex\_ggk\_settings\_1] PRIMARY KEY ([key]));

#### Description

The table holds the settings.

### Table – kit\_autosomal

#### DDL

CREATE TABLE [kit\_autosomal] ([kit\_no] TEXT REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[rsid] TEXT NOT NULL,

[chromosome] TEXT NOT NULL,

[position] INTEGER NOT NULL,

[genotype] TEXT NOT NULL,

CONSTRAINT [sqlite\_autoindex\_kit\_autosomal\_1] PRIMARY KEY ([kit\_no], [rsid]));

#### Description

The table holds the autosomal SNP values for each kit including reference samples.

### Table – kit\_mtdna

#### DDL

CREATE TABLE [kit\_mtdna] (

[kit\_no] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[mutations] TEXT NOT NULL,

[fasta] TEXT,

CONSTRAINT [sqlite\_autoindex\_kit\_mtdna\_1] PRIMARY KEY ([kit\_no]));

#### Description

The table holds the mtDNA values as comma separated mutations.

### Table – kit\_phased

#### DDL

CREATE TABLE [kit\_phased] (

[kit\_no] TEXT REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[rsid] TEXT NOT NULL,

[chromosome] TEXT NOT NULL,

[position] INTEGER NOT NULL,

[paternal\_genotype] TEXT NOT NULL,

[maternal\_genotype] TEXT NOT NULL,

[paternal\_kit\_no] TEXT,

[maternal\_kit\_no] TEXT,

CONSTRAINT [sqlite\_autoindex\_kit\_phased\_1] PRIMARY KEY ([kit\_no], [rsid]));

#### Description

The table holds the phasing SNP results.

### Table – kit\_roh

#### DDL

CREATE TABLE [kit\_roh] ([kit\_no] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[chromosome] TEXT NOT NULL,

[start\_position] INTEGER NOT NULL,

[end\_position] INTEGER NOT NULL,

[segment\_length\_cm] DOUBLE NOT NULL DEFAULT (0.0),

[snp\_count] INTEGER NOT NULL);

#### Description

The table holds the ROH segment values.

### Table – kit\_ysnps

#### DDL

CREATE TABLE [kit\_ysnps] (

[kit\_no] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[ysnps] TEXT NOT NULL,

CONSTRAINT [] PRIMARY KEY ([kit\_no]));

#### Description

The table holds the Y-SNP values. The Y-SNP values are comma separated.

### Table – kit\_ystr

#### DDL

CREATE TABLE [kit\_ystr] ([kit\_no] TEXT NOT NULL REFERENCES [kit\_master]([kit\_no]) ON DELETE CASCADE ON UPDATE CASCADE,

[marker] TEXT NOT NULL,

[value] TEXT NOT NULL,

CONSTRAINT [sqlite\_autoindex\_kit\_ystr\_1] PRIMARY KEY ([kit\_no], [marker]));

#### Description

The table holds the Y-STR values.

## File Format (.ggk)

Genetic Genealogy Kit file is a GZip compressed text file. The text file has the following structure.

### Sample Structure

**@KIT@**

264652

**@NAME@**

Felix Chandrakumar

**@AUTOSOMAL@**

<DocumentElement>

<kit\_autosomal>

<kit\_no>264652</kit\_no>

<rsid>rs3094315</rsid>

<chromosome>1</chromosome>

<position>752566</position>

<genotype>AG</genotype>

</kit\_autosomal>

<kit\_autosomal>

<kit\_no>264652</kit\_no>

<rsid>rs3131972</rsid>

<chromosome>1</chromosome>

<position>752721</position>

<genotype>AG</genotype>

</kit\_autosomal>

:

:

<kit\_autosomal>

<kit\_no>264652</kit\_no>

<rsid>rs669237</rsid>

<chromosome>X</chromosome>

<position>154916845</position>

<genotype>GG</genotype>

</kit\_autosomal>

</DocumentElement>

**@AUTOSOMAL-SCHEMA@**

<?xml version="1.0"?><xs:schema id="NewDataSet" ...

**@YSNPS@**

Z13595+, Z13358+, Z13625+, ...

**@YSTR@**

CDY=33-37,DYF395S1=15-15,DYF406S1=11,DYS19=15,DYS385=13-17, ...

**@MTDNA@**

C146T,C152T,C195T,A200G,A247G,315.1C,G499A,522.1A,522.2C,A769G, ...

**@FASTA@**

PjI2NDY1MixIVlIyLENSLEhWUjENCkdBVENBQ0FHR1RDVEFBVENBQ0dBV… EcNCg==

Each section is separated by a tag in a new line prefixed and suffixed by @ symbol. The allowed values are @KIT@, @NAME@, @AUTOSOMAL@, @AUTOSOMAL-SCHEMA@, @YSNPS@, @YSTR@, @MTDNA@, @FASTA@. All values after the tag line must be followed by a single line value except @AUTOSOMAL@.

#### @KIT@

It is a single line value. The next line will have the Kit Number.

#### @NAME@

It is a single line value. The next line will have the name.

#### @AUTOSOMAL@

It is a multiline value. The value is an XML.

#### @AUTOSOMAL-SCHEMA@

It is a single line value. The new lines are removed to linearize the XML schema.

#### @YSNPS@

It is a single line comma separated value.

#### @YSTR@

It is a single line comma separated value. Each STR name and value is again separated by ‘=’.

#### @MTDNA@

It is a single line comma separated value.

#### @FASTA@

It is a single line value. The fasta file value is encoded with Base64