

Relationship Estimator

This is a small program that is intended to make the DNA Prediction Chart Spreadsheet a bit easier to use. It is based entirely on the data in this spreadsheet plus some interpolation of missing values.

The algorithm to determine the most likely relationship(s) is very simple and based on summing the score of valid entries in the table for a given input. It is very much an experiment and test. It is likely to be less accurate with close relationships where there is missing data in the spreadsheet. You can save the match information that you generate. At the moment this is just stored but I plan to allow export in csv or other formats.

This works only for Windows. It requires Microsoft dotNET 4.0 which should ship with current versions of Windows. Also note that there is no internationalization of the code at this time so please enter any decimals using the point as a decimal separator (52.9). Using a comma separator will confuse the program and you will get wrong results.

Getting Started

Unzip the contents of the distribution file into a suitable folder and double click RelationshipEstimator.exe. You should see the small app as shown below

The screenshot shows the 'Relationship Estimator' application window. At the top, there is a text input field (1) for entering segment lengths. Below it, a section titled 'Based on Seg of 5 cM or greater' contains four numeric input fields: 'Total cM' (2) with value 1877.0, 'Longest Segment cM' (3) with value 125.3, '# of Segments' (4) with value 50, and '% Shared' (5) with value 25.360. A 'Source:' dropdown menu (6) is set to '23andMe/FTDNA/GEDMatch'. A 'Process' button (7) is located below these fields. To the right, an 'Estimated Relationship(s)' list (8) shows 'Uncle/Aunt/Niece/Nephew [11111]', 'Grandparent/Grandchild [11010]', and 'Half Sibling [11010]'. Below this is a 'Save Results' section with two text input fields for 'Person 1' (9) and 'Person 2' (10), a 'Known Rel' dropdown (11) set to 'Uncle/Aunt/Niece/Nephew', and a 'Save' button (12). A 'Find' button (13) is also present. At the bottom, there are 'Clear' (14), 'Delete' (15), 'Report' (16), and 'Close' (17) buttons. The version number 'v01.03.00' is displayed in the bottom left corner.

This is Windows 10 so it may look a little different on your computer. All the functions are available on one screen:

1. Raw segment lengths can be entered in this box. You can enter all the segments for a match or just those greater than 5cM. Any less than 5cM will be discarded when the raw data is summarized.
2. The summary of raw segment lengths entered in (1) or directly entered values
3. The source is where the data came from. At the moment we have a choice of Ancestry or 23andMe/FTDNA/GEDMatch. Ancestry provides different information so to get valid results you will need to select the correct source.
4. The Process button will execute the estimator and results will be shown in
5. The results box contains the best estimates for the relationship
6. These boxes let you enter the names of the two people the match is about. The program remembers names and will offer previously entered names when you start typing
7. If you have a known relationship then this can be entered and will be stored with the match data
8. Click Save to save the match information
9. If you want to go back to some previously saved match data then you can do this by entering some name information into the first person field and get a list of matches for that name. This will also update the relationship estimates in case the base data has been updated
10. Clear all the data

11. Delete a saved record
12. If you find that the program does not correctly report a known relationship then you can use the Report button to tell us about it so we can update the information that the program uses.
13. Close the application

Enter Raw Data

Enter Segment lengths in cM separated by forward slash '/'

12.8/ 52.9/ 13.1/ 22.2/ 19/ 7.2

Enter the segment lengths separated by the forward slash '/'. You can include spaces to make it easier to read if you wish. Whole numbers do not need any decimals. There is some rudimentary checking that the entries are all numeric, Clicking Process will generate the summary data and the relationship estimates

Calculate the Summary Data

Based on Seg of 5 cM or greater

Total cM	1877.0
Longest Segment cM	125.3
# of Segments	50
% Shared	25.360

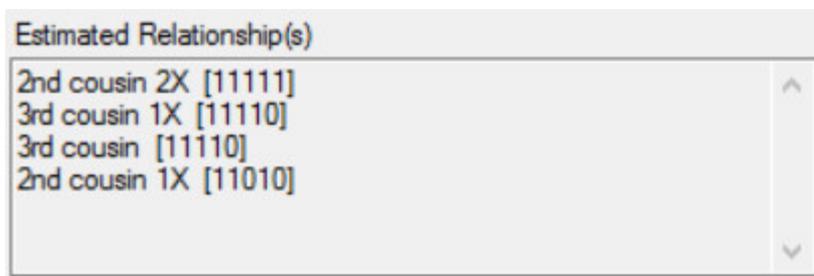
Source:

23andMe/FTDNA/GEDMatch

Process

You can either enter the raw data as described above or enter the summary data directly. If you enter your own summary data then please remember that the estimating algorithms are designed to use only segments of 5cM or larger. If you use a different cut-off value either higher or lower than 5cM then the relationship estimates will probably be unreliable and inaccurate. Click Process to proceed. Please be sure to enter the correct source for the data you are using.

Display the Estimated Relationship(s) for 23andMe, FTDNA, or GEDMatch Data

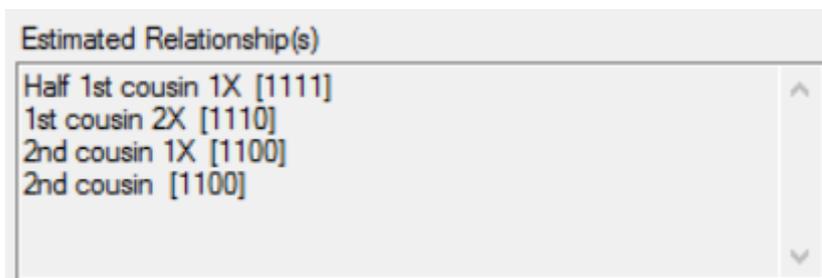


The program uses the data from the DNA prediction chart to identify likely relationships. Each relationship is compared to the summary data and a score is accumulated based on how many bits of data in the summary match the relationships values. The maximum score is Five. The five criteria checked are

- Total cM range
- Largest segment cM range
- # segments range
- Percent DNA shared range
- Closest to average shared DNA. This is applied to only one of the estimated relationships and can be a tie breaker

the set of 1s and 0s in brackets after the estimated relationship show the five criteria and there is a 1 if the match meets that criteria and a 0 if not. So in the example about 2C2R has a 1 for each criteria so it fits all the criteria. 2C1R at the bottom meets Total cM range, Largest segment range and percent shared DNA range. It does not meet the number of segments range nor was it closest to the average percent shared. All the above relationships are probably possible for the match data provided.

Display the Estimated Relationship(s) for Ancestry Data



Ancestry results are displayed the same way however only four criteria are used (Ancestry does not provide largest segment information).

Enter Summary Data Directly

Based on Seg of 5 cM or greater		Estimated Relationship(s)	
Total cM	<input type="text" value="1877.0"/>	Uncle/Aunt/Niece/Nephew [11111] Grandparent/Grandchild [11010] Half Sibling [11010]	
Longest Segment cM	<input type="text" value="125.3"/>		
# of Segments	<input type="text" value="50"/>		
% Shared	<input type="text" value="25.360"/>		
Source:	<input type="text" value="23andMe/FTDNA/GEDMatch"/>	Save Results	
		Person 1	<input type="text" value="Michael V..."/>
		Person 2	<input type="text" value="Maureen..."/>
		Known Rel	<input type="text" value="Uncle/Aunt/Niece/Nephew"/>

The summary above is for a known Aunt/Nephew relationship and shows that relationship as the most likely. If you have the summary data from somewhere else like GEDMatch then just enter it and press Process to get the results.

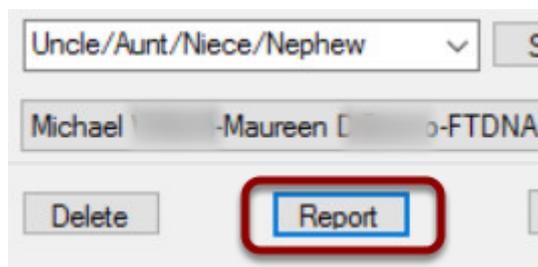
NOTE that the current DNA Prediction spreadsheet has quite a lot of gaps in the data for close relationships so this program is more likely to get these wrong. In this case it seems to have worked OK.

Saving the Match Data

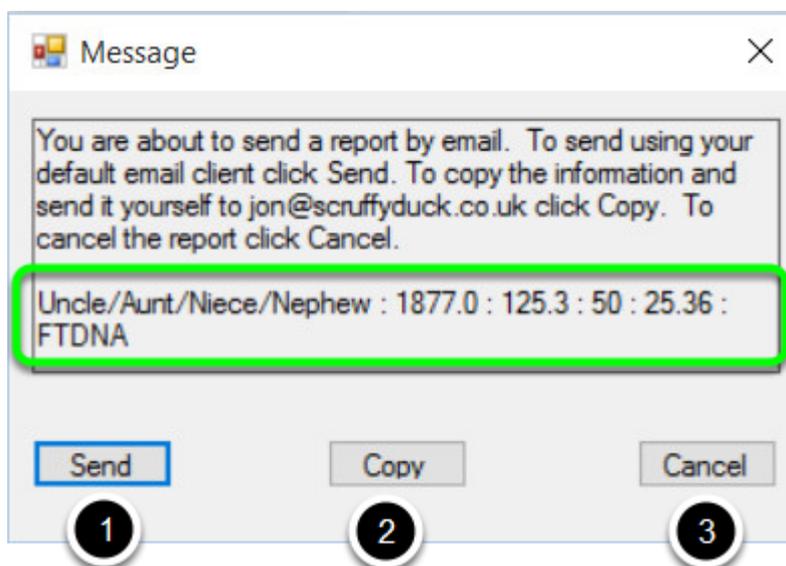
Save Results	
Person 1	<input type="text" value="Michael..."/>
Person 2	<input type="text" value="Maureen..."/>
Known Rel	<input type="text" value="Uncle/Aunt/Niece/Nephew"/>
	<input type="button" value="Save"/>

Enter the names of the people sharing the match. If the name is already in the database then you will get a list of possibles after you type a couple of letters and if the person is listed you can just click to add. Enter the known relationship if you have one and click Save

Reporting a Wrong Relationship Estimate



This program uses information based on the ranges of shared DNA for different relationships. If you have a known relationship but the program does not report it properly then you can let us know so that we can improve the accuracy of the estimates in the future. Note that the Report button will only be active if there is a known relationship selected. If this box is empty then you cannot send a report until you select a relationship to report. To send a report then click the Report button



When you click the button you will get a small message as shown above. The message shows the information that is going to be sent. You have several choices:

1. Click Send to open your default email program with a new message complete to send. All you need to do is send the pre-prepared message. Of course you can add some comments before you send it but please do not change the Subject or the body text. If you get an error when trying to open your default email program then you can try the Copy method below.
2. If you use only web mail and do not have a default email program set then you can copy the important information using the Copy button. The information shown in the green box will be copied to your Windows Clipboard. From there you can paste it into your own email message. Please give the email the Subject **Relationship Estimator** and send it to
3. If you decide not to send the report then just click Cancel

Finding a Match

The screenshot shows a web interface for finding matches. It includes input fields for 'Person 1' (containing 'Michael') and 'Person 2' (containing 'Maureen'). A dropdown menu for 'Known Rel' is set to 'Uncle/Aunt/Niece/Nephew'. A 'Save' button is visible. Below these are 'Find' and 'Clear' buttons. A dropdown menu is open, displaying a list of matches with columns for Person 1, Person 2, and the match type. The first two items are highlighted in blue.

Person 1	Person 2	Match Type
Michael	Maureen	-FTDNA
Michael	Maureen	-FTDNA
Michael	Maureen	-GEDMatch
Maureen	o-Rose	s-Other
Maureen	o-Lyn N	Other
Maureen	o-Carol	Other

Use the Clear button to clear any existing data and enter the name of the person to find matches for in Person 1. Click Find and get a list of matches to that person. Select one from the list to view.