AutoKinship tutorial

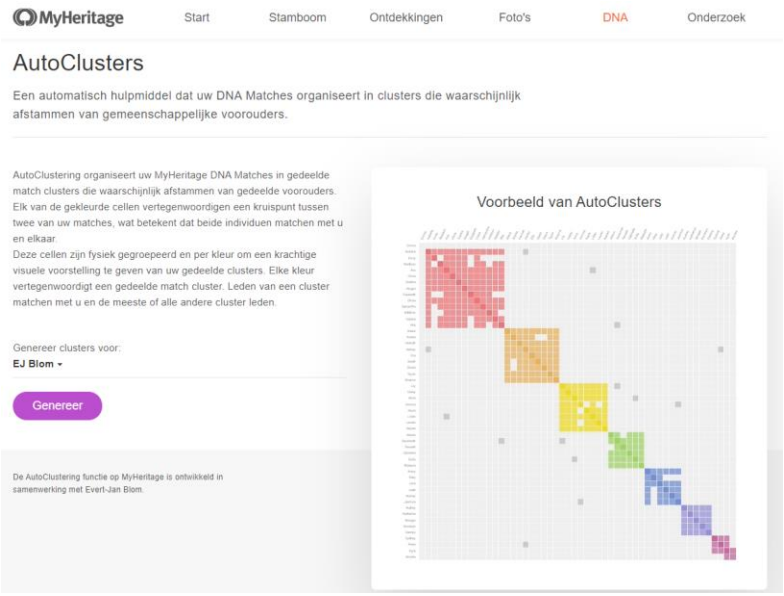
Run AutoKinship analysis for a MyHeritage profile

AutoKinship automatically predicts family trees based on the amount of DNA your DNA matches share with you and each other. Note that AutoKinship does not require any known genealogical trees from your DNA matches. Instead, AutoKinship looks at the predicted relationships between your DNA matches and calculates many different paths you could all be related to each other. The trees from our analysis are ranked and represent the most likely trees out of all the possibilities we calculated. An AutoKinship analysis can be started using this link: https://members.geneticaffairs.com/autokinship

In this tutorial, we will be using the results from MyHeritage AutoClusters to start the analysis. The same principle can be used for clustering results for 23andme and GEDmatch since they also provide the shared cM between shared matches.

Running MyHeritage AutoClusters

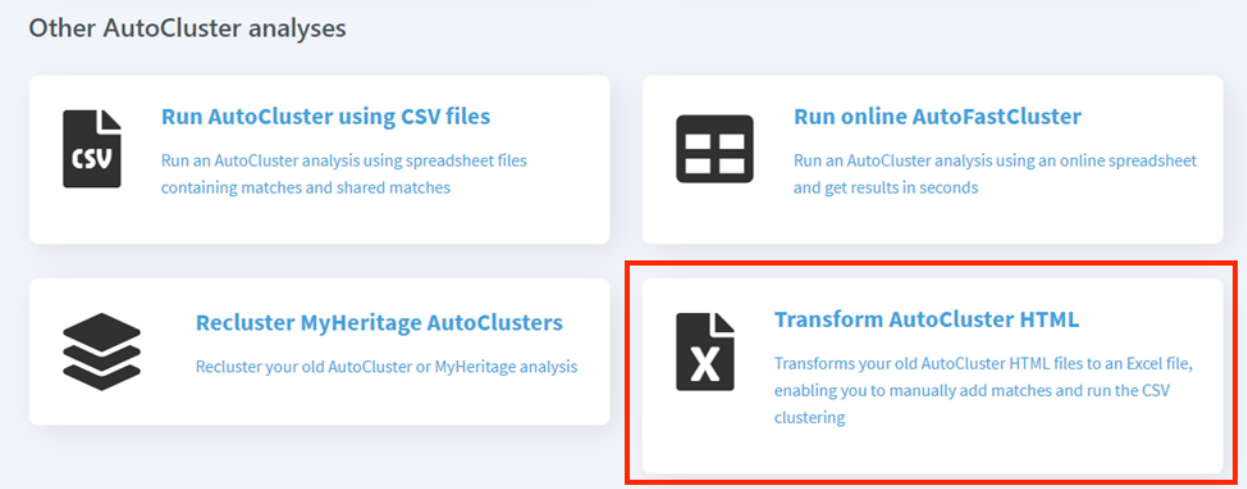
Open the AutoClusters page on MyHeritage: https://www.myheritage.com/dna/autoclusters



Select the profile of interest and generate the analysis by clicking on the generate button. Wait a while, save the attachments from the e-mails and go to the folder that holds the ZIP file. Extract the zip file, open HTML file, and find cluster of interest.

Transforming HTML to Excel

Next, we will transform our freshly generated AutoCluster HTML report to an Excel file. This feature will create an Excel with all the (shared) matches as well the (shared) matches per cluster. This blog post also discusses this feature: https://patriciacolemangenealogy.com/2021/02/23/819/)



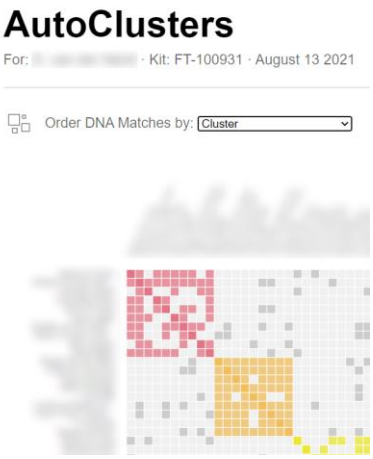
Visit this page: https://members.geneticaffairs.com/htmltoexcel

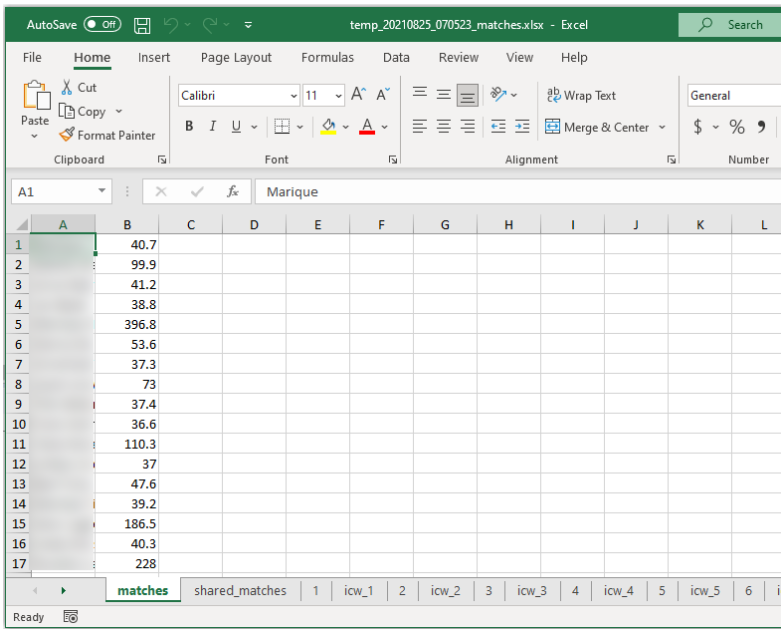
Select the MyHeritage AutoClusters HTML file and start the analysis by clicking on the green button. If all goes well, an Excel file is returned.

Locate cluster data in Excel file

Next, we need to extract the matches from the Excel that are in a certain cluster. The new version of the HTML to Excel tool creates separate sheets for each cluster.

Open the Excel file and locate the sheet that has the same name as the cluster. The shared matches are in a sheet with the same name but with the icw_ prefix.



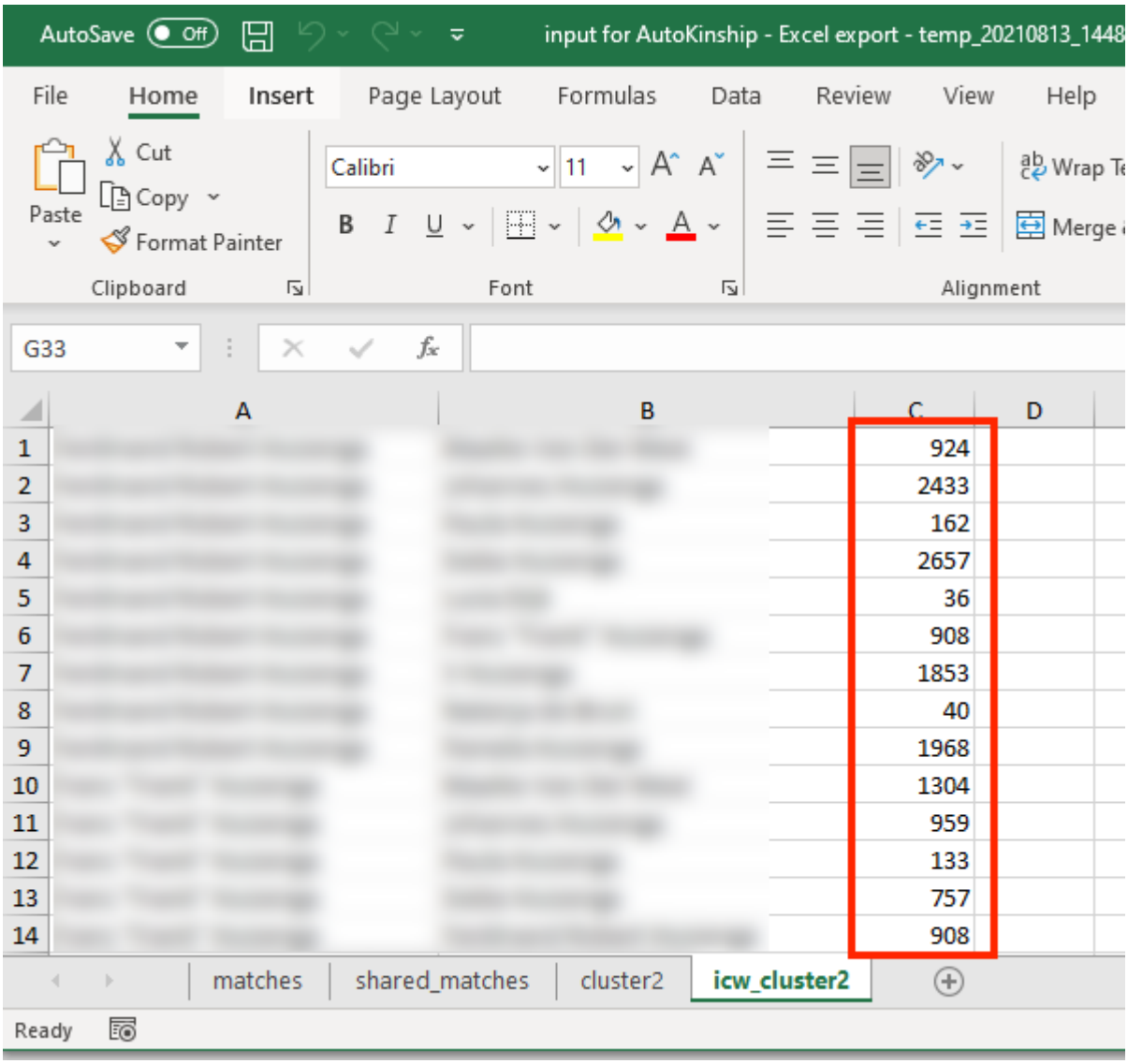


	A	B
1		40.7
2		99.9
3		41.2
4		38.8
5		396.8
6		53.6
7		37.3
8		73
9		37.4
10		36.6
11		110.3
12		37
13		47.6
14		39.2
15		186.5
16		40.3
17		228

Important: please round all cM values for the matches and shared matches to round numbers (e.g., 1.251,9 cM becomes 1252 cM)

Getting shared DNA for ICW matches

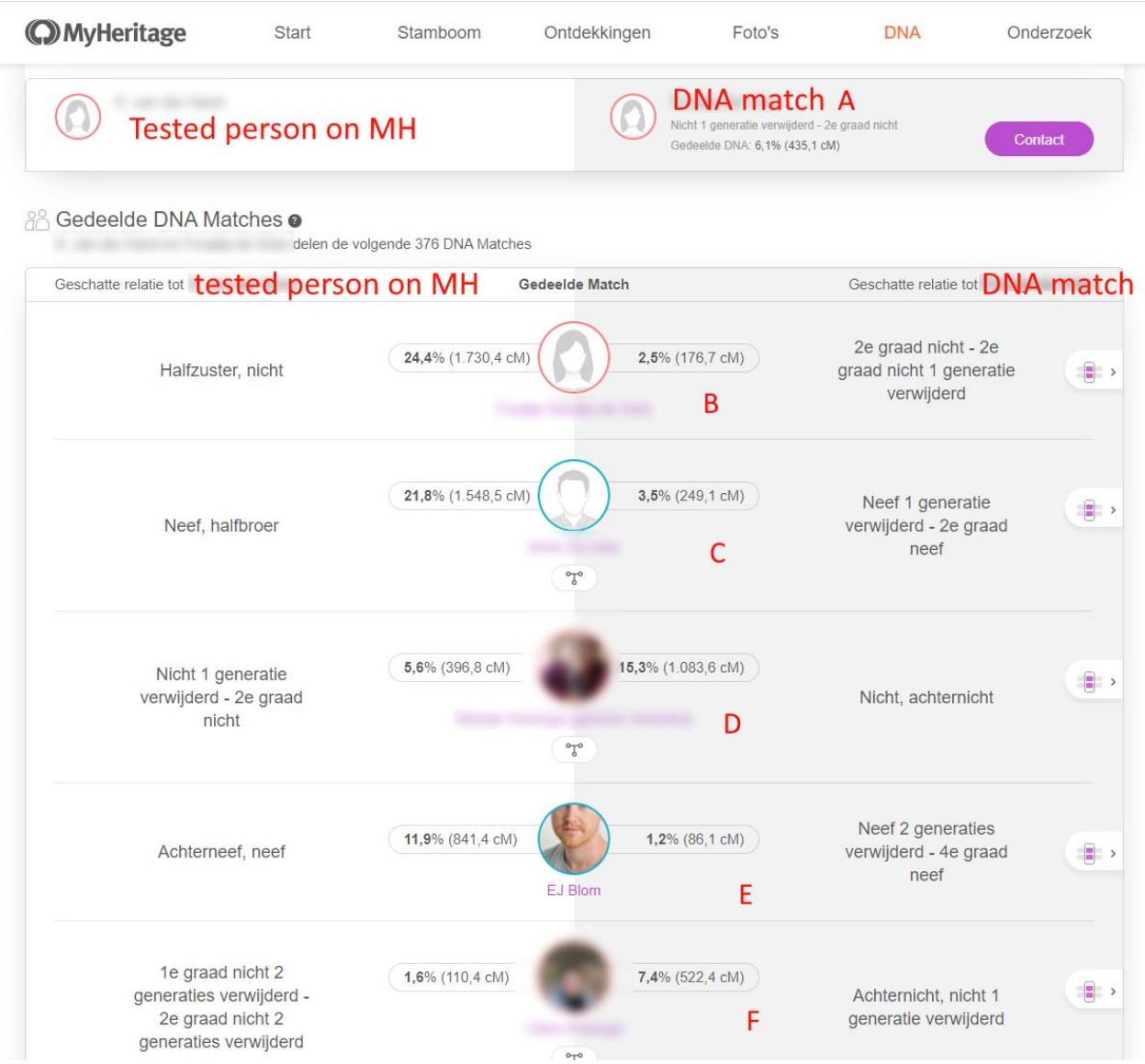
In the previous step, we located the DNA match information for a specific cluster as well as the shared match information. However, one crucial part of information is still missing. We need to include the shared DNA between the shared matches.



In my example Excel, I've already filled in the cM values between shared matches in the 3rd column, indicated in red.

To get this information, we need to find this information online on the MyHeritage page for each DNA match in the shared matches sheet.

In this case, we are examining DNA match A for my great aunt. If you scroll down, past the shared surnames/locations, you will see the shared DNA matches.

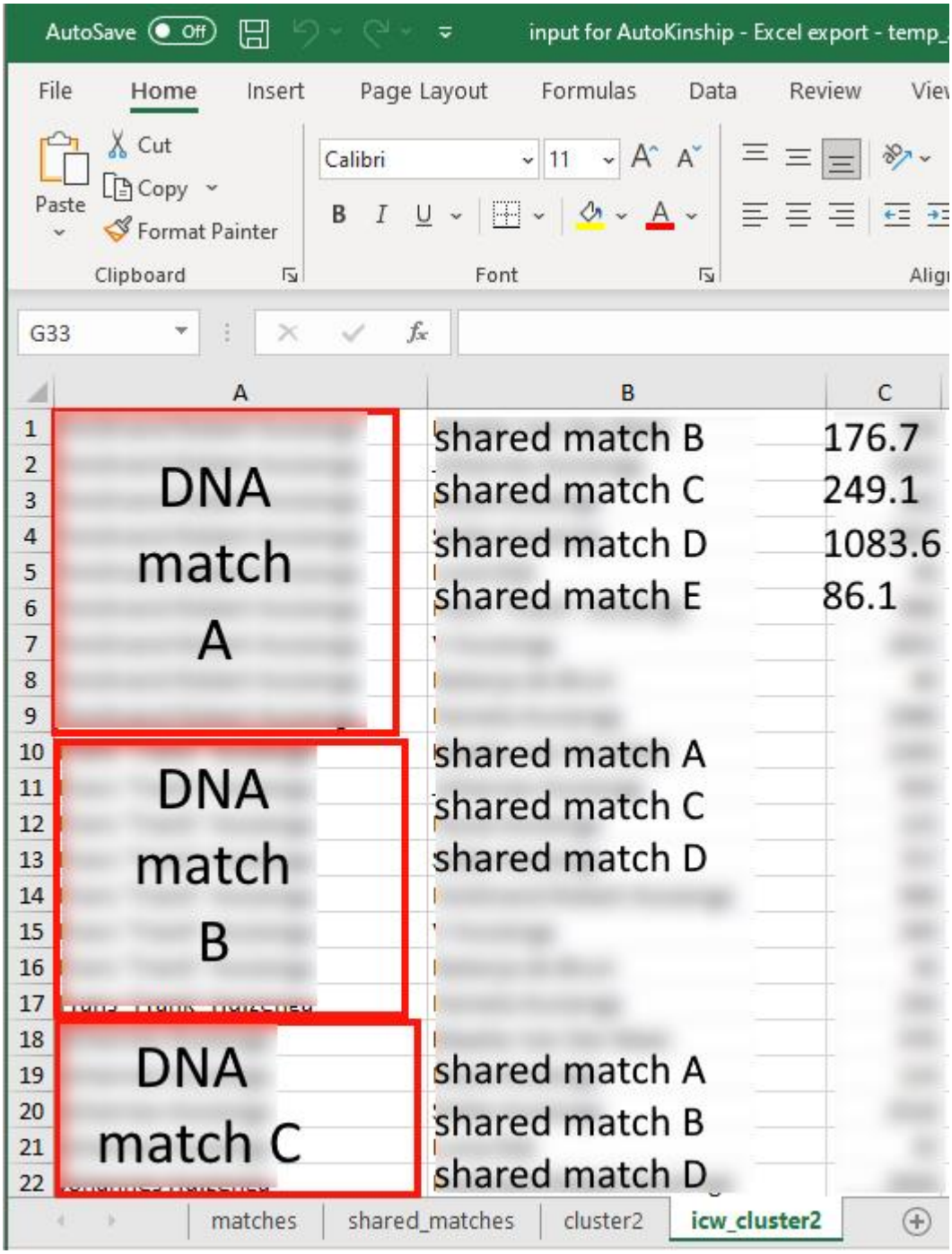


At the left side, you will see how much my great aunt shares with the shared matches. At the right side, is how much DNA match A shares with these shared matches.

In this example, my great aunt shares 435 cM with this DNA match A. This particular DNA match shares 176.7 cM with DNA match B, 249 cM with DNA match C, etc etc.

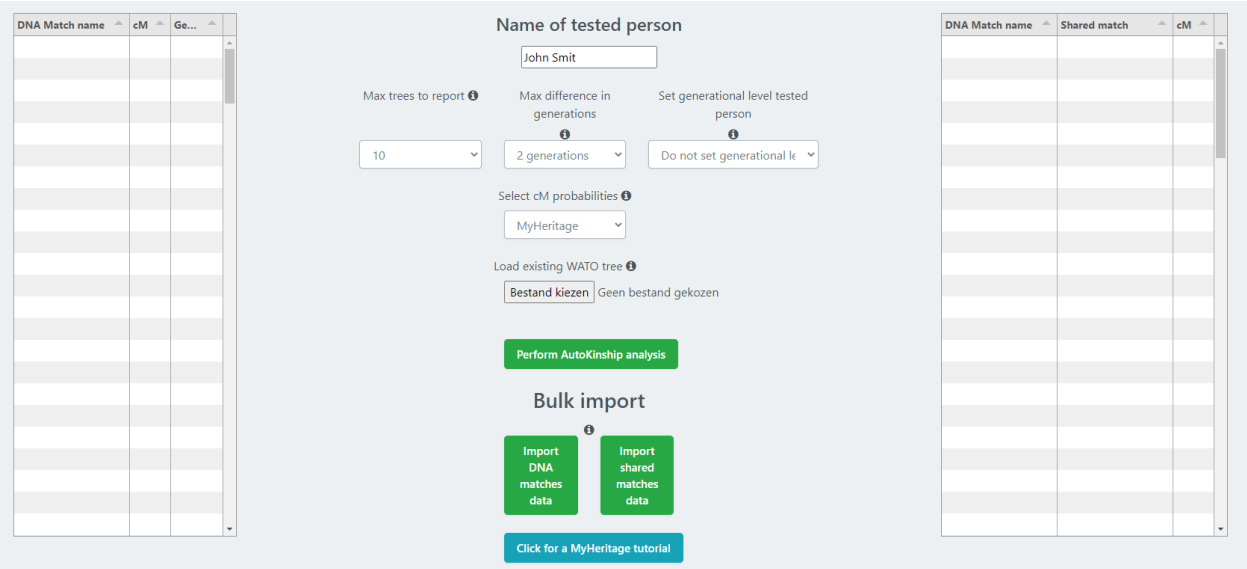
We now locate DNA match A in the list of shared matches. Now we locate each of those shared matches in the DNA match page that we discussed on the previous page and we fill in the missing cM information. For instance, DNA match A shares 176 cM with match B, and 249 cM with match C.

Always uses rounded cM for the shared cM values.

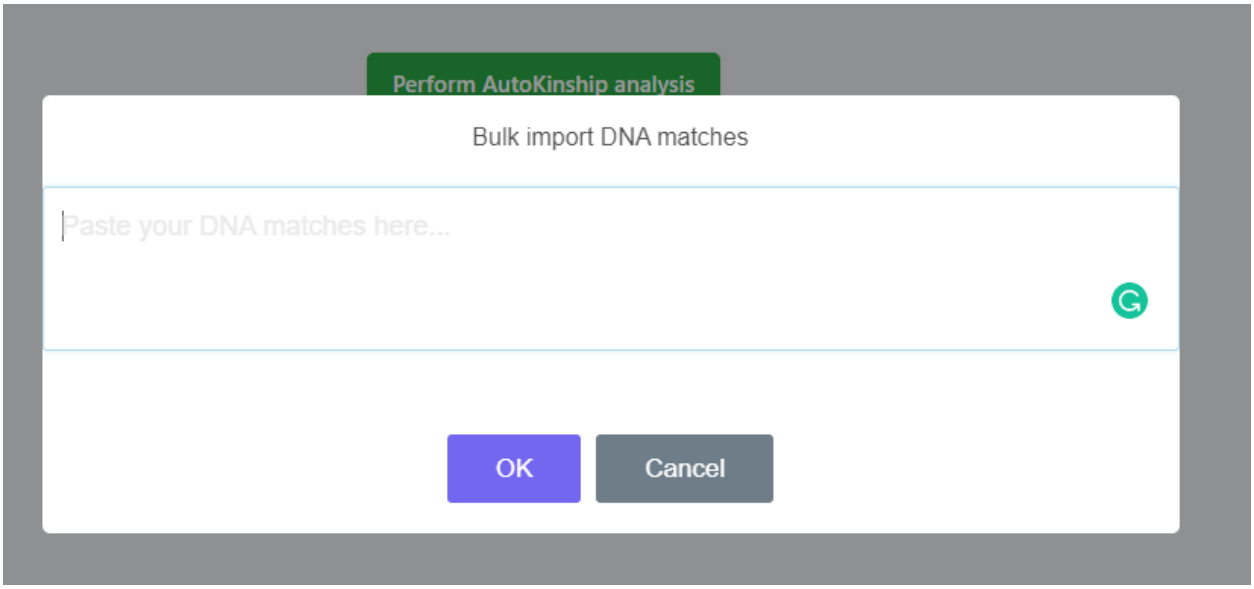


Filling in the AutoKinship spreadsheet

After finishing obtaining the shared DNA information, it's time to copy the information in our new spreadsheet. Open this page: https://members.geneticaffairs.com/autokinship

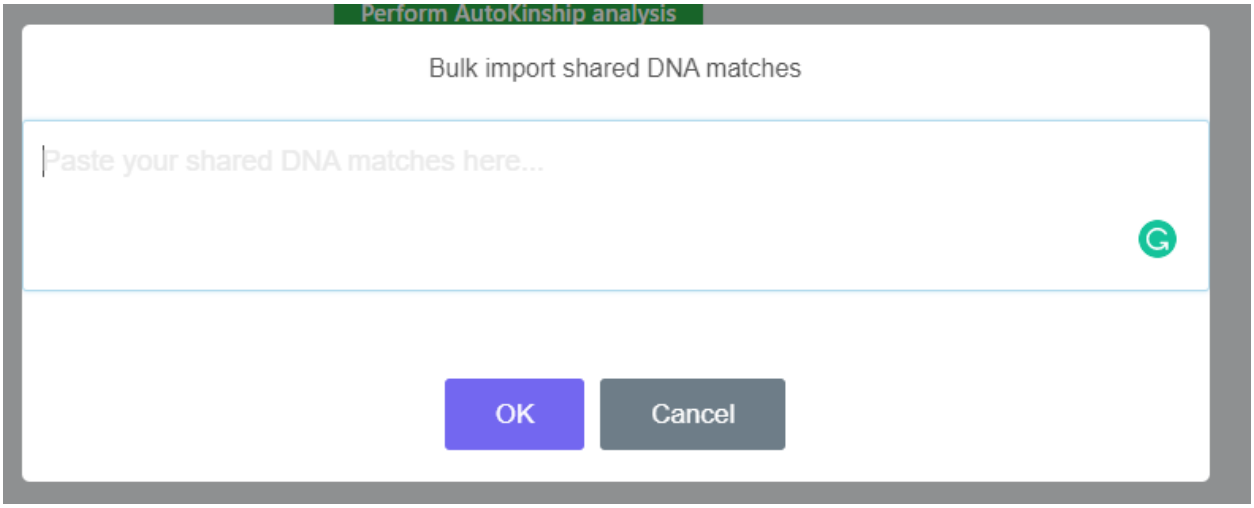


To import data from the spreadsheet, click on each of the buttons underneath "Bulk import". First, click on the "Import DNA matches data":

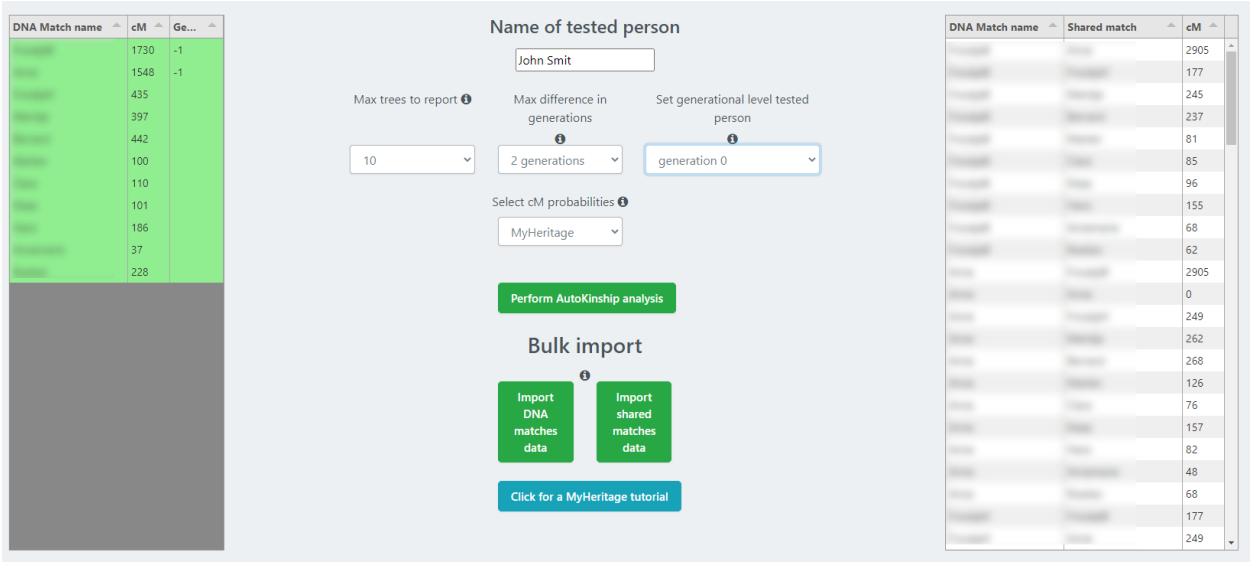


Select the DNA match information and cM values in the Excel sheet and copy it into the text field. If all goes well, the left panel will be filled with DNA matches.

Now repeat this process for the shared matches by clicking on the "Import shared matches data":



If all goes well, the matches and shared matches panel are populated with data.



There are now a couple of settings that can be used. For instance, if you know the generational level for some of the DNA matches. In this example, I've set the 2nd and 3rd match to generation -1. If no other match is linked to generation 0, the use of generation -1 is the same as using generation -2.

Using the same generation for matches will force the AutoKinship tool to only consider relationships that generate these matches on the same generational level.

Perhaps you also know that the tested person, for instance yourself or in my example my great aunt are one or more generations apart from some matches. For my data, I could set the generational level for my great aunt to 0 and set some of her matches (for instance my mother and her brother) to -1, because they are one generation closer.

Next, it is possible to specify the maximum difference in generations. This will remove trees from the output that contain DNA matches that are apart with a certain number of generations. Otherwise, the program might identify a possible but unlikely great-great-great-great-aunt. We probably want to identify DNA matches that are around our own generational level.

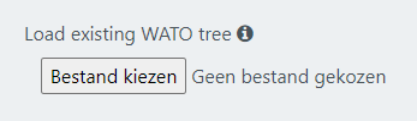
Last, we need to specify the cM probabilities. These probabilities are used to describe how likely a certain relationship is given a certain cM score.

The probabilities used by AutoKinship are based on simulated data, kindly provided by Brit Nicholson (methodology described here: https://dna-sci.com/2021/04/06/a-new-probability-calculator-for-genetic-genealogy). Brit has provided specific probabilities for MyHeritage, 23andme and GEDmatch data. In addition, it is also possible to use the probabilities from Ancestry as obtained from the WATO tool.

We are now ready to run the analysis! Click on the "Perform AutoKinship" analysis button and wait for the results to appear in your mailbox.

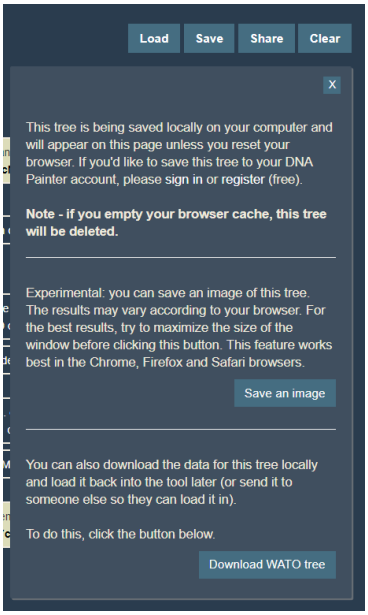
Using an existing tree with AutoKinship

Sometimes you already know the genealogical relationships of certain matches in a cluster. Ideally, one would like to use these relationships and let the AutoKinship figure out the remaining relationships. It is now possible to import an existing tree which will serve as a backbone for the AutoKinship analysis.



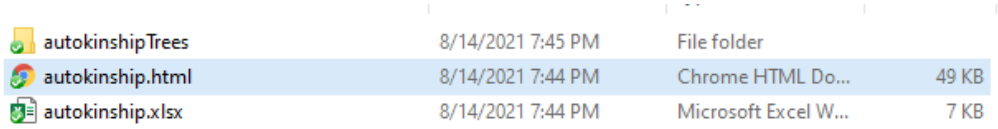
The existing tree can be generated using the WATO website: https://dnapainter.com/tools/probability

After generating the tree, first click on the save button and then select the "Download WATO tree" option. Select the tree from the AutoKinship interface.



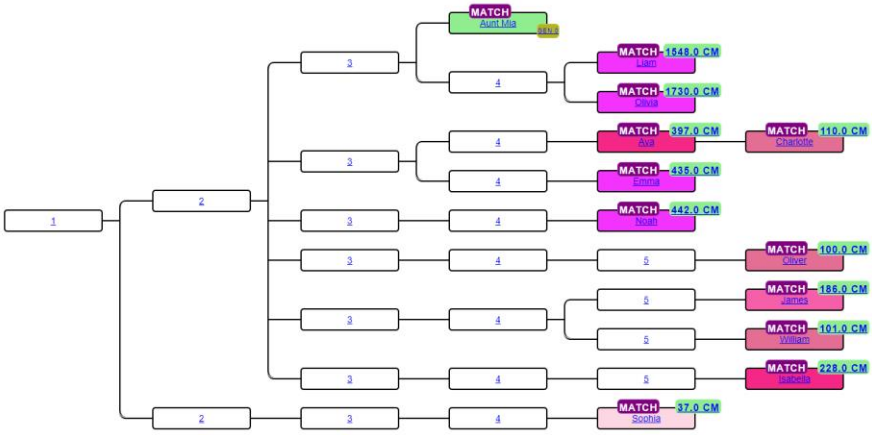
Analyzing the AutoKinship results

Save the ZIP file from the email, and unzip it to a folder. You will get a couple of files and a folder:



The AutoKinship.html landing page contains the best tree. Below is an explanation of the AutoKinship tool. Next, the ranked AutoKinship trees are provided in a table with some information concerning the combined odds ratio, which allows us to compare them with each other. Last, we provide a DNA matrix with all the shared cM information for all provided matches.

Reconstructed AutoKinship tree 8496 with probability 7.871931003146019E-30 and 12 persons



Explanation of the automated probability calculations

AutoKinship automatically predicts family trees based on the amount of DNA your DNA matches share with you and each other. Note that AutoKinship does not require any known genealogical trees from your DNA matches. Instead, AutoKinship looks at the predicted relationships between your DNA matches, and calculates many different paths you could all be related to each other. The trees from our analysis are ranked and represent the most likely trees out of all the possibilities we calculated.

The AutoKinship tool is available for 23andme profiles as well as a standalone tool. Using the standalone tool, users can fill in shared match information from DNA testing companies that provide the shared cM information between shared matches. Currently, only MyHeritage, 23andme and GEDmatch provide this information.

AutoCluster first organizes your DNA matches into shared match clusters that likely represent branches of your family. Everyone in a cluster will likely be on the same ancestral line, although the MRCA between any of the matches and between you and any match may vary. The generational level of the clusters may vary as well. One may be your paternal grandmother's branch, another may be your paternal grandfather's father's branch. Next, AutoKinship reconstructs and compares many different trees. AutoKinship works for persons with an unknown parentage to their birth families (for instance adoptees or donor conceived persons).

The probabilities used by this AutoKinship analysis are based on simulated data for MyHeritage matches and are kindly provided by Brit Nicholson (methodology described here). Based on the shared cM data between shared matches, we create different trees based on the putative relationships. We then use the probabilities to test every scenario which are then ranked.

Reconstructed AutoKinship trees

Reconstructed AutoKinship trees are sorted based on their combined probabilities. Next, we compare these trees based on the lowest ranking (lowest probability). The combined odds ratio reflects this comparison. In addition, the ratio as compared to the next AutoKinship tree is provided, allowing you to compare trees with each other.

Rank	Tree	#Matches	Prob	Combined odds ratio	Compared to previous
1	tree1	12	7.871E-30	16	2
2	tree2	12	3.877E-30	7.9	1
3	tree3	12	3.877E-30	7.9	1.1
4	tree4	12	3.500E-30	7.1	1
5	tree5	12	3.500E-30	7.1	1.2
6	tree6	12	2.991E-30	6.1	1.8
7	tree7	12	1.666E-30	3.4	1
8	tree8	12	1.666E-30	3.4	1.1

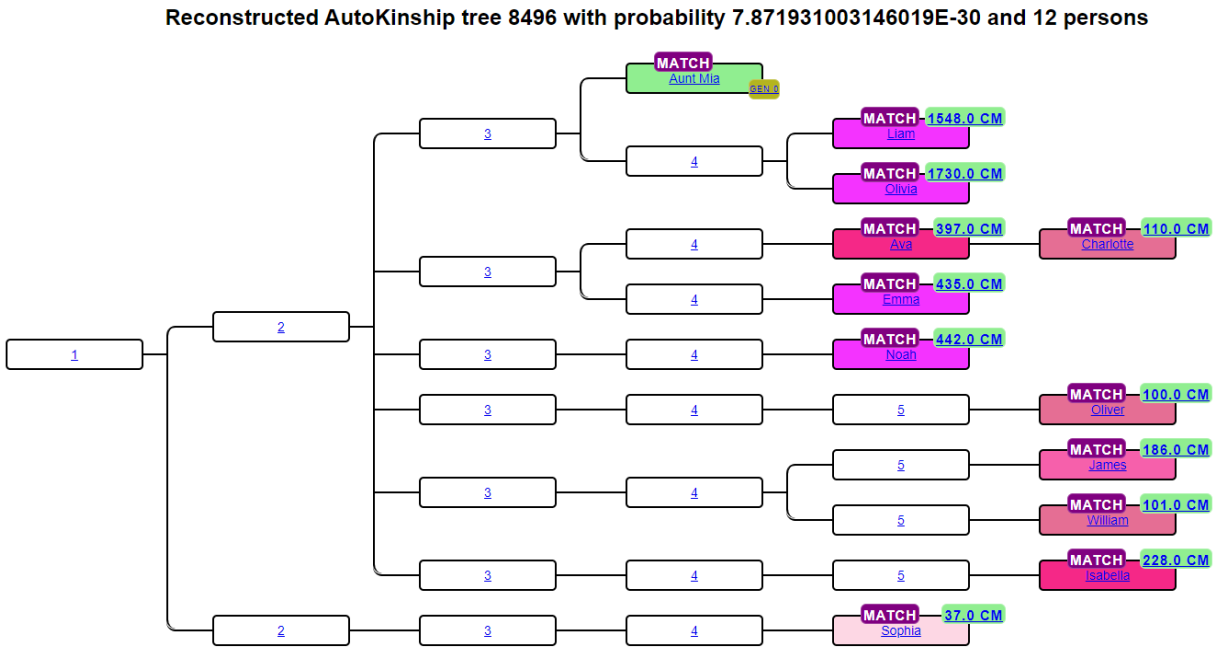
DNA matrix

This relationship matrix shows the shared centimorgans between DNA matches.

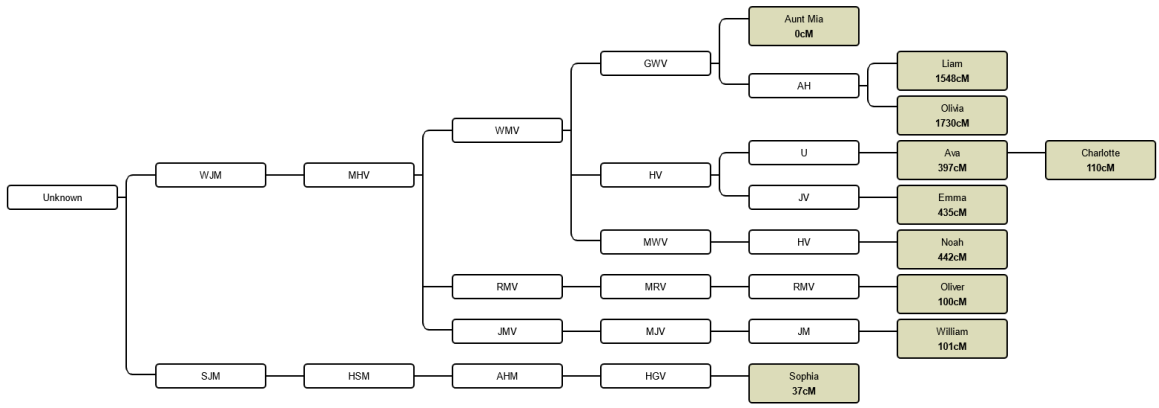
Download relationship matrix

match	Ava	Olivia	Liam	Aunt Mia	Charlotte	Emma	Noah	James	William	Oliver	Isabella	Sophia
Ava	0	245	262	397	3452	1084	89	81	47	141	44	46
Olivia	245	0	2905	1730	85	177	237	155	96	81	62	68
Liam	262	2905	0	1548	76	249	268	82	157	126	68	48
Aunt Mia	397	1730	1548	0	110	435	442	186	101	100	228	37
Charlotte	3542	85	76	110	0	522	43	0	28	34	18	27
Emma	1084	177	249	435	522	0	160	41	32	126	168	0
Noah	89	237	268	442	43	160	0	102	178	149	153	38
James	81	155	82	186	0	41	102	0	362	79	47	0
William	47	96	157	101	28	32	178	362	0	103	0	17
Oliver	141	81	126	100	34	126	149	79	103	0	60	24
Isabella	44	62	68	228	18	168	153	47	0	60	0	0

This is the best reconstructed tree.

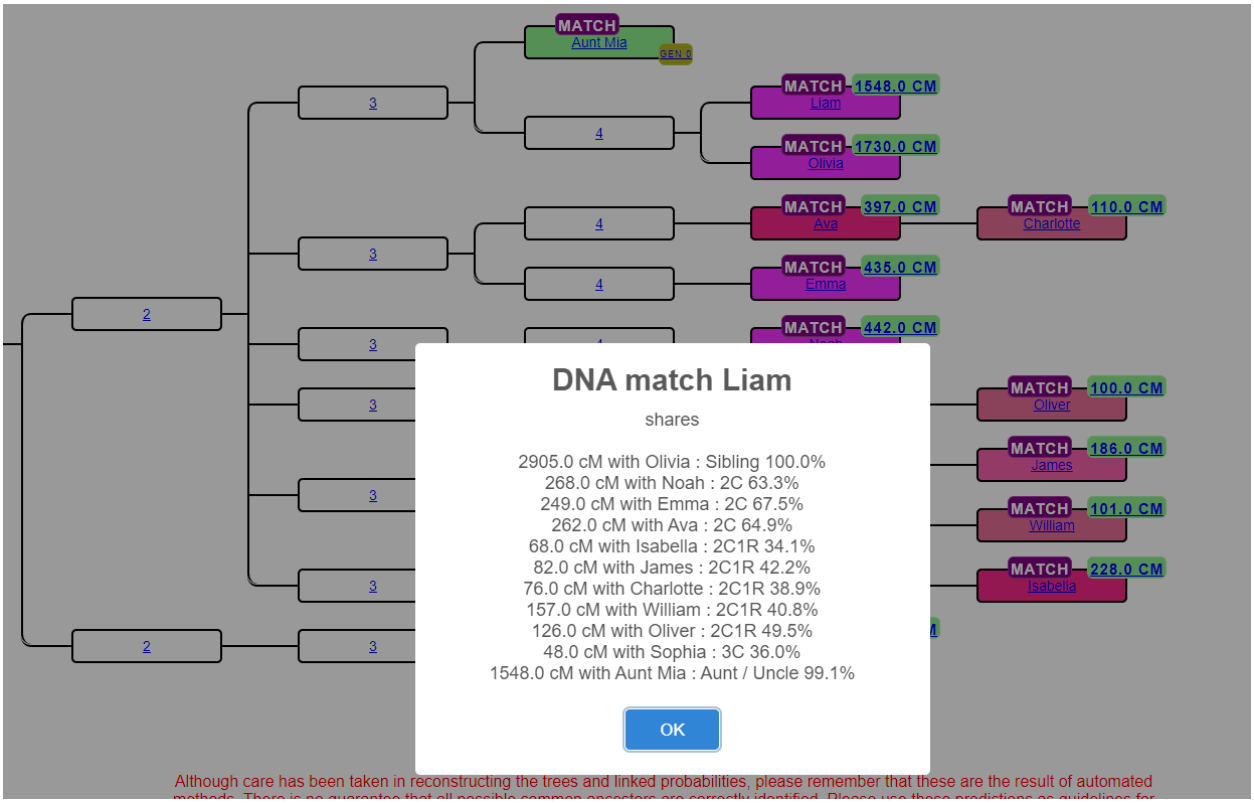


And this is the known tree



The predicted tree from AutoKinship is quite similar to the known tree, although some matches are off by one generation (in this case William & Sophia). Note that there are more matches in the AutoKinship results since I couldn't find trees for all the matches. But now that I have the AutoKinship results, I have some good clues where they might be positioned.

There are some pop-up badges that contain some relevant information. For instance, clicking on a DNA match will bring up the cM information from that match perspective:



Last, don't forget to examine the other predicted trees from the landing page. Quite often, matches are placed in the wrong generational level. There might be other trees that contain the same matches but with more accurate generational levels. You can also re-run the analysis and fixate some generational levels by supplying this information in the DNA matches panel.