

# AutoKinship tutorial

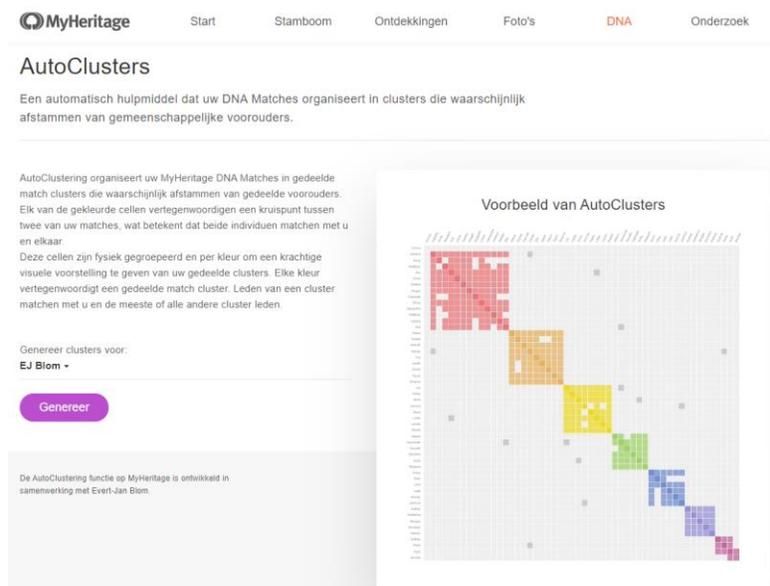
## Run AutoKinship analysis for a MyHeritage profile

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>



The screenshot shows the MyHeritage website interface. At the top, there is a navigation bar with the MyHeritage logo and menu items: Start, Stamboom, Ontdekkingen, Foto's, DNA, and Onderzoek. The main heading is 'AutoClusters'. Below the heading, there is a brief description: 'Een automatisch hulpmiddel dat uw DNA Matches organiseert in clusters die waarschijnlijk afstammen van gemeenschappelijke voorouders.' The main content area contains a paragraph explaining the clustering process: 'AutoClustering organiseert uw MyHeritage DNA Matches in gedeelde match clusters die waarschijnlijk afstammen van gedeelde voorouders. Elk van de gekleurde cellen vertegenwoordigen een kruispunt tussen twee van uw matches, wat betekent dat beide individuen matchen met u en elkaar. Deze cellen zijn fysiek gegroepeerd en per kleur om een krachtige visuele voorstelling te geven van uw gedeelde clusters. Elke kleur vertegenwoordigt een gedeelde match cluster. Leden van een cluster matchen met u en de meeste of alle andere cluster leden.' Below this text, there is a section 'Genereer clusters voor:' with the name 'EJ Blom' and a purple 'Genereer' button. At the bottom left, there is a small note: 'De AutoClustering functie op MyHeritage is ontwikkeld in samenwerking met Ever-Jan Blom.' On the right side, there is a diagram titled 'Voorbeeld van AutoClusters' showing a grid of colored squares (red, orange, yellow, green, blue, purple) representing clusters of DNA matches.

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/>)

### Other AutoCluster analyses

**Run AutoCluster using CSV files**  
Run an AutoCluster analysis using spreadsheet files containing matches and shared matches

**Run online AutoFastCluster**  
Run an AutoCluster analysis using an online spreadsheet and get results in seconds

**Recluster MyHeritage AutoClusters**  
Recluster your old AutoCluster or MyHeritage analysis

**Transform AutoCluster HTML**  
Transforms your old AutoCluster HTML files to an Excel file, enabling you to manually add matches and run the CSV clustering

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

Hi, eijblom,

Transform your old AutoCluster analyses to an Excel file. It will extract the matches as well as the shared matches from the AutoClusters chart and place them into an Excel file. For instance a MyHeritage analysis that is just missing your 405 cM cousin, or an old Ancestry analysis that you wish to extend with new matches.



37 Daniel Hedberg	40.7	36 Matt Lombardi	40.7
38 Daniel Hedberg	40.6	37 Richard Abrahamson	40.6
39 Dan Hedberg	40.6	38 Richard Abrahamson	40.6
40 Dan Hedberg	40.6	39 Richard Abrahamson	40.6
41 Dan Hedberg	40.6	40 Richard Abrahamson	40.6
42 Dan Hedberg	40.6	41 Richard Abrahamson	40.6
43 Dan Hedberg	40.6	42 Richard Abrahamson	40.6
44 Dan Hedberg	40.6	43 Richard Abrahamson	40.6
45 Dan Hedberg	40.6	44 Richard Abrahamson	40.6
46 Dan Hedberg	40.6	45 Richard Abrahamson	40.6
47 Dan Hedberg	40.6	46 Richard Abrahamson	40.6
48 Dan Hedberg	40.6	47 Richard Abrahamson	40.6
49 Dan Hedberg	40.6	48 Richard Abrahamson	40.6
50 Dan Hedberg	40.6	49 Richard Abrahamson	40.6
51 Dan Hedberg	40.6	50 Richard Abrahamson	40.6
52 Dan Hedberg	40.6	51 Richard Abrahamson	40.6
53 Dan Hedberg	40.6	52 Richard Abrahamson	40.6
54 Dan Hedberg	40.6	53 Richard Abrahamson	40.6
55 Dan Hedberg	40.6	54 Richard Abrahamson	40.6
56 Dan Hedberg	40.6	55 Richard Abrahamson	40.6
57 Dan Hedberg	40.6	56 Richard Abrahamson	40.6
58 Dan Hedberg	40.6	57 Richard Abrahamson	40.6
59 Dan Hedberg	40.6	58 Richard Abrahamson	40.6
60 Dan Hedberg	40.6	59 Richard Abrahamson	40.6
61 Dan Hedberg	40.6	60 Richard Abrahamson	40.6
62 Dan Hedberg	40.6	61 Richard Abrahamson	40.6
63 Dan Hedberg	40.6	62 Richard Abrahamson	40.6
64 Dan Hedberg	40.6	63 Richard Abrahamson	40.6
65 Dan Hedberg	40.6	64 Richard Abrahamson	40.6
66 Dan Hedberg	40.6	65 Richard Abrahamson	40.6
67 Dan Hedberg	40.6	66 Richard Abrahamson	40.6
68 Dan Hedberg	40.6	67 Richard Abrahamson	40.6
69 Dan Hedberg	40.6	68 Richard Abrahamson	40.6
70 Dan Hedberg	40.6	69 Richard Abrahamson	40.6
71 Dan Hedberg	40.6	70 Richard Abrahamson	40.6
72 Dan Hedberg	40.6	71 Richard Abrahamson	40.6
73 Dan Hedberg	40.6	72 Richard Abrahamson	40.6
74 Dan Hedberg	40.6	73 Richard Abrahamson	40.6
75 Dan Hedberg	40.6	74 Richard Abrahamson	40.6
76 Dan Hedberg	40.6	75 Richard Abrahamson	40.6
77 Dan Hedberg	40.6	76 Richard Abrahamson	40.6
78 Dan Hedberg	40.6	77 Richard Abrahamson	40.6
79 Dan Hedberg	40.6	78 Richard Abrahamson	40.6
80 Dan Hedberg	40.6	79 Richard Abrahamson	40.6
81 Dan Hedberg	40.6	80 Richard Abrahamson	40.6
82 Dan Hedberg	40.6	81 Richard Abrahamson	40.6
83 Dan Hedberg	40.6	82 Richard Abrahamson	40.6
84 Dan Hedberg	40.6	83 Richard Abrahamson	40.6
85 Dan Hedberg	40.6	84 Richard Abrahamson	40.6
86 Dan Hedberg	40.6	85 Richard Abrahamson	40.6
87 Dan Hedberg	40.6	86 Richard Abrahamson	40.6
88 Dan Hedberg	40.6	87 Richard Abrahamson	40.6
89 Dan Hedberg	40.6	88 Richard Abrahamson	40.6
90 Dan Hedberg	40.6	89 Richard Abrahamson	40.6
91 Dan Hedberg	40.6	90 Richard Abrahamson	40.6
92 Dan Hedberg	40.6	91 Richard Abrahamson	40.6
93 Dan Hedberg	40.6	92 Richard Abrahamson	40.6
94 Dan Hedberg	40.6	93 Richard Abrahamson	40.6
95 Dan Hedberg	40.6	94 Richard Abrahamson	40.6
96 Dan Hedberg	40.6	95 Richard Abrahamson	40.6
97 Dan Hedberg	40.6	96 Richard Abrahamson	40.6
98 Dan Hedberg	40.6	97 Richard Abrahamson	40.6
99 Dan Hedberg	40.6	98 Richard Abrahamson	40.6
100 Dan Hedberg	40.6	99 Richard Abrahamson	40.6

Bestand kiezen Geen bestand gekozen

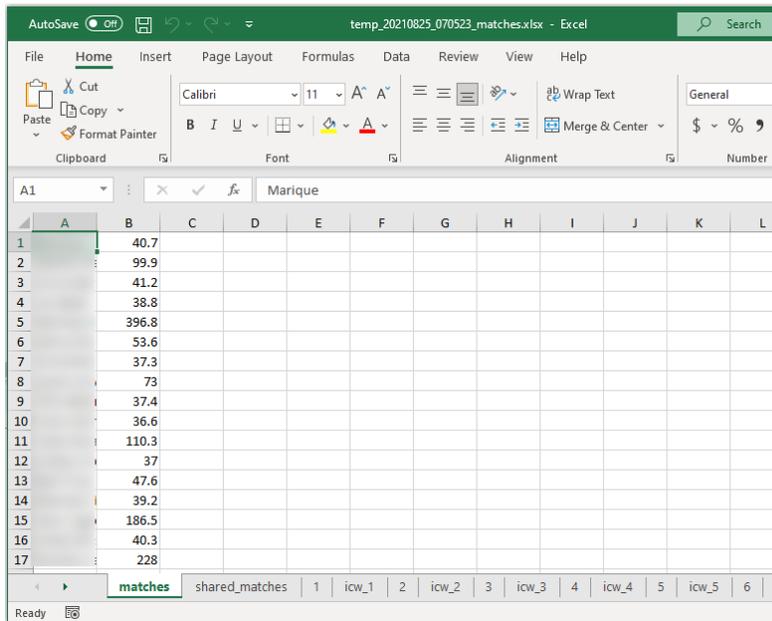
Transform HTML to Excel analysis

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.

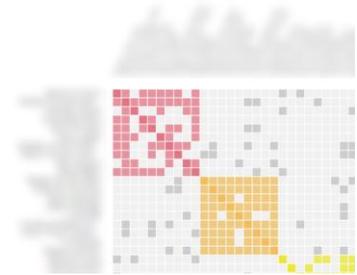


	A	B	C	D	E	F	G	H	I	J	K	L
1		40.7										
2		99.9										
3		41.2										
4		38.8										
5		396.8										
6		53.6										
7		37.3										
8		73										
9		37.4										
10		36.6										
11		110.3										
12		37										
13		47.6										
14		39.2										
15		186.5										
16		40.3										
17		228										

## AutoClusters

For: [redacted] · Kit: FT-100931 · August 13 2021

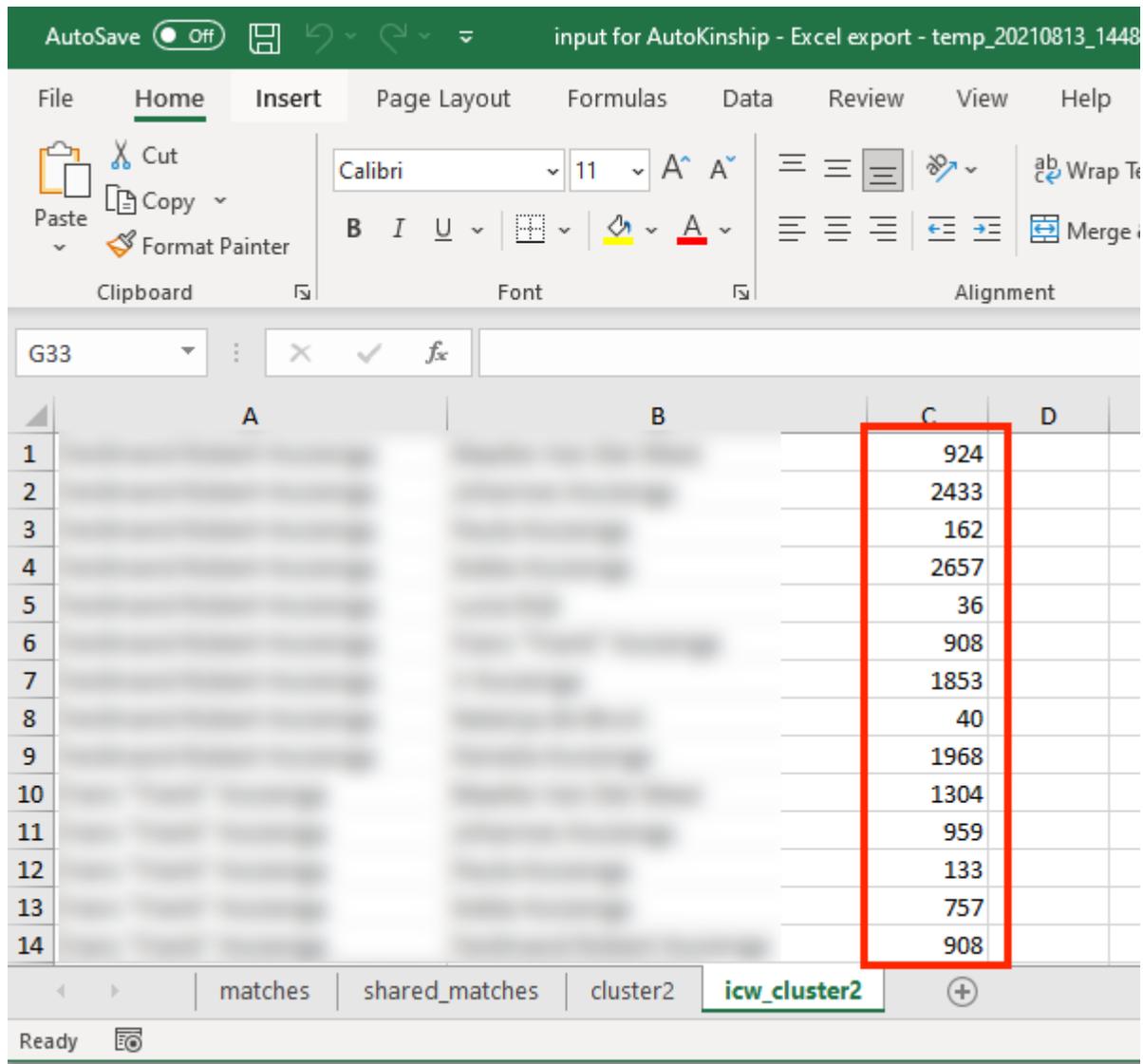
Order DNA Matches by: Cluster



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 3<sup>rd</sup> 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.

The screenshot shows the MyHeritage interface. At the top, there are navigation tabs: Start, Stamboom, Ontdekkingen, Foto's, DNA, and Onderzoek. Below the navigation, there are two main sections: 'Tested person on MH' and 'DNA match A'. The 'DNA match A' section shows 'Nicht 1 generatie verwijderd - 2e graad nicht' and 'Gedeelde DNA: 6,1% (435,1 cM)'. Below this is a section for 'Gedeelde DNA Matches' with a sub-header 'delen de volgende 376 DNA Matches'. The main content is a table of shared matches.

Geschatte relatie tot <b>tested person on MH</b>	Gedeelde Match	Geschatte relatie tot <b>DNA match</b>
Halfzuster, nicht	24,4% (1.730,4 cM)  <b>B</b>	2e graad nicht - 2e graad nicht 1 generatie verwijderd
Neef, halfbroer	21,8% (1.548,5 cM)  <b>C</b>	Neef 1 generatie verwijderd - 2e graad neef
Nicht 1 generatie verwijderd - 2e graad nicht	5,6% (396,8 cM)  <b>D</b>	Nicht, achternicht
Achterneef, neef	11,9% (841,4 cM)  <b>E</b> EJ Blom	Neef 2 generaties verwijderd - 4e graad neef
1e graad nicht 2 generaties verwijderd - 2e graad nicht 2 generaties verwijderd	1,6% (110,4 cM)  <b>F</b>	Achernicht, nicht 1 generatie verwijderd

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.

	A	B	C
1	DNA match A	shared match B	176.7
2		shared match C	249.1
3		shared match D	1083.6
4		shared match E	86.1
10	DNA match B	shared match A	
11		shared match C	
12		shared match D	
19	DNA match C	shared match A	
20		shared match B	
21		shared match D	

# 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>

Name of tested person  
John Smit

Max trees to report 10  
Max difference in generations 2 generations  
Set generational level tested person Do not set generational level

Select cM probabilities MyHeritage

Load existing WATO tree Bestand kiezen | Geen bestand gekozen

Perform AutoKinship analysis

Bulk import

Import DNA matches data | Import shared matches data

Click for a MyHeritage tutorial

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

Perform AutoKinship analysis

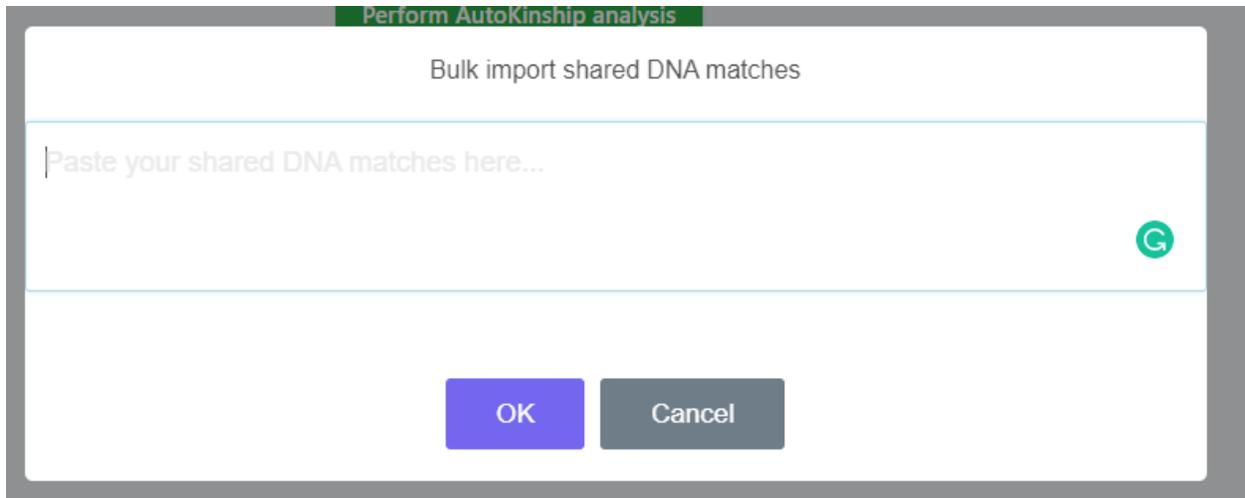
Bulk import DNA matches

Paste your DNA matches here...

OK Cancel

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.

DNA Match name	cM	Ge...
1730	-1	
1548	-1	
435		
397		
442		
100		
110		
101		
186		
37		
228		

Name of tested person: John Smit

Max trees to report: 10

Max difference in generations: 2 generations

Set generational level tested person: generation 0

Select cM probabilities: MyHeritage

Perform AutoKinship analysis

Bulk import

Import DNA matches data

Import shared matches data

Click for a MyHeritage tutorial

DNA Match name	Shared match	cM
		2905
		177
		245
		237
		81
		85
		96
		155
		68
		62
		2905
		0
		249
		262
		268
		126
		76
		157
		82
		48
		68
		177
		249

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 2<sup>nd</sup> and 3<sup>rd</sup> 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.

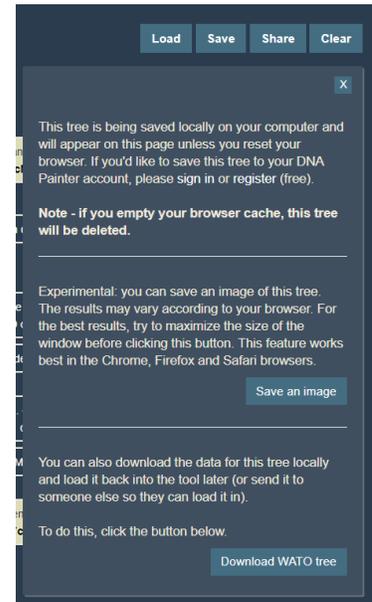
Load existing WATO tree ⓘ

Bestand kiezen Geen bestand gekozen

The existing tree can be generated using the WATO website:

<https://dnainter.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.



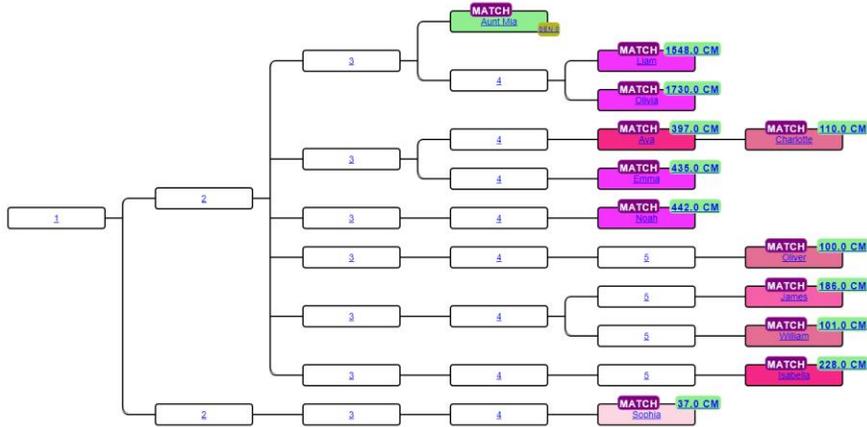
## Analyzing the AutoKinship results

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

 autokinshipTrees	8/14/2021 7:45 PM	File folder	
 autokinship.html	8/14/2021 7:44 PM	Chrome HTML Do...	49 KB
 autokinship.xlsx	8/14/2021 7:44 PM	Microsoft Excel W...	7 KB

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 [Britt 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	<a href="#">tree1</a>	12	7.871E-30	16	2
2	<a href="#">tree2</a>	12	3.877E-30	7.9	1
3	<a href="#">tree3</a>	12	3.877E-30	7.9	1.1
4	<a href="#">tree4</a>	12	3.500E-30	7.1	1
5	<a href="#">tree5</a>	12	3.500E-30	7.1	1.2
6	<a href="#">tree6</a>	12	2.991E-30	6.1	1.8
7	<a href="#">tree7</a>	12	1.666E-30	3.4	1
8	<a href="#">tree8</a>	12	1.666E-30	3.4	1.1

### DNA matrix

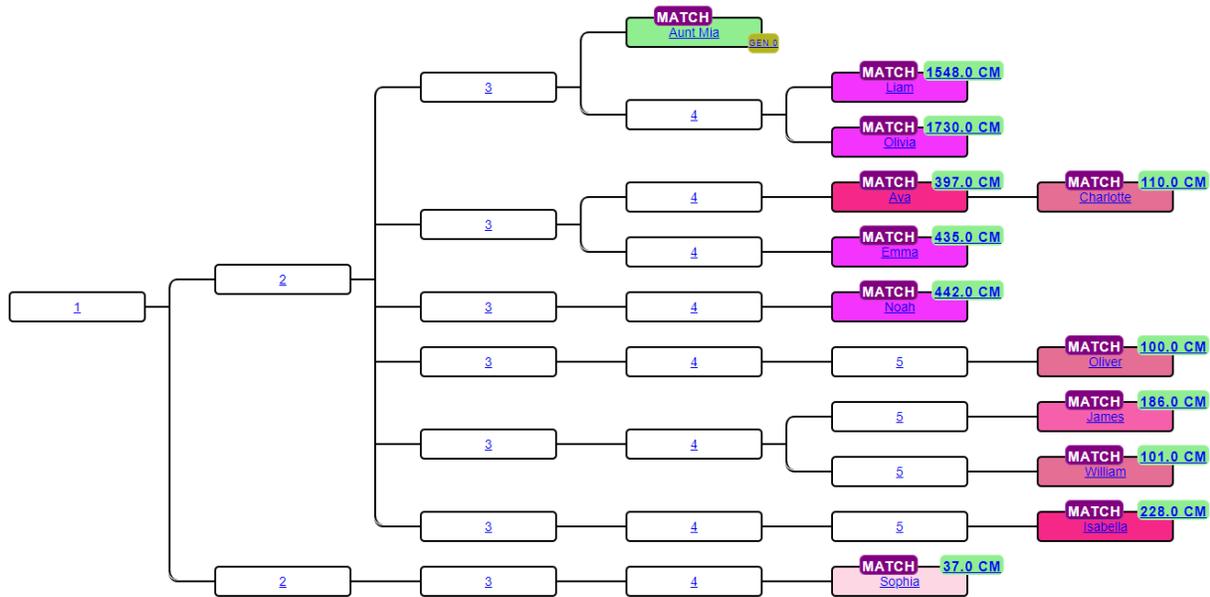
This relationship matrix shows the shared centimorgans between DNA matches.

[Download relationship matrix](#)

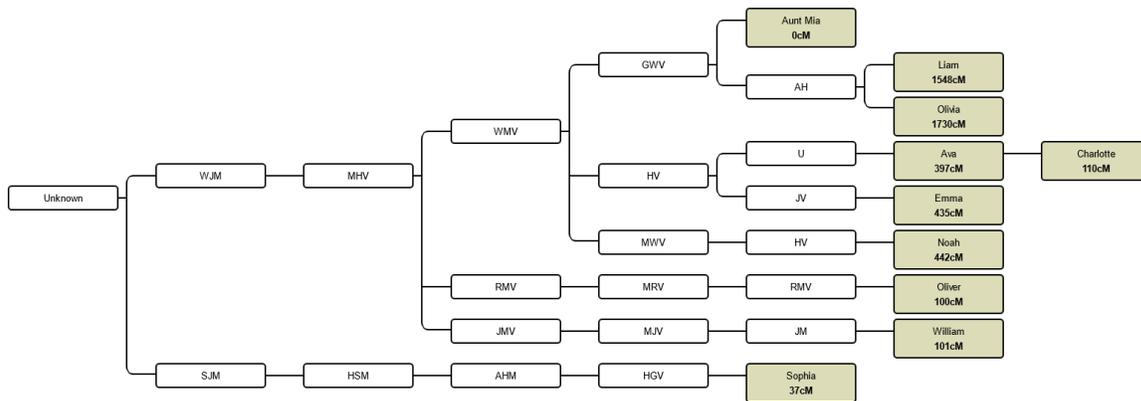
match	Ava	Olivia	Liam	Aunt Mia	Charlotte	Emma	Noah	James	William	Oliver	Isabella	Sophia
Ava	0	245	262	397	3452	1084	89	81	47	141	44	48
Olivia	245	0	2905	1730	85	177	237	155	96	81	62	68
Liam	262	2905	0	1548	76	249	268	82	157	126	68	48
Aunt Mia	397	1730	1548	0	110	435	442	186	101	100	228	37
Charlotte	3452	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.

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

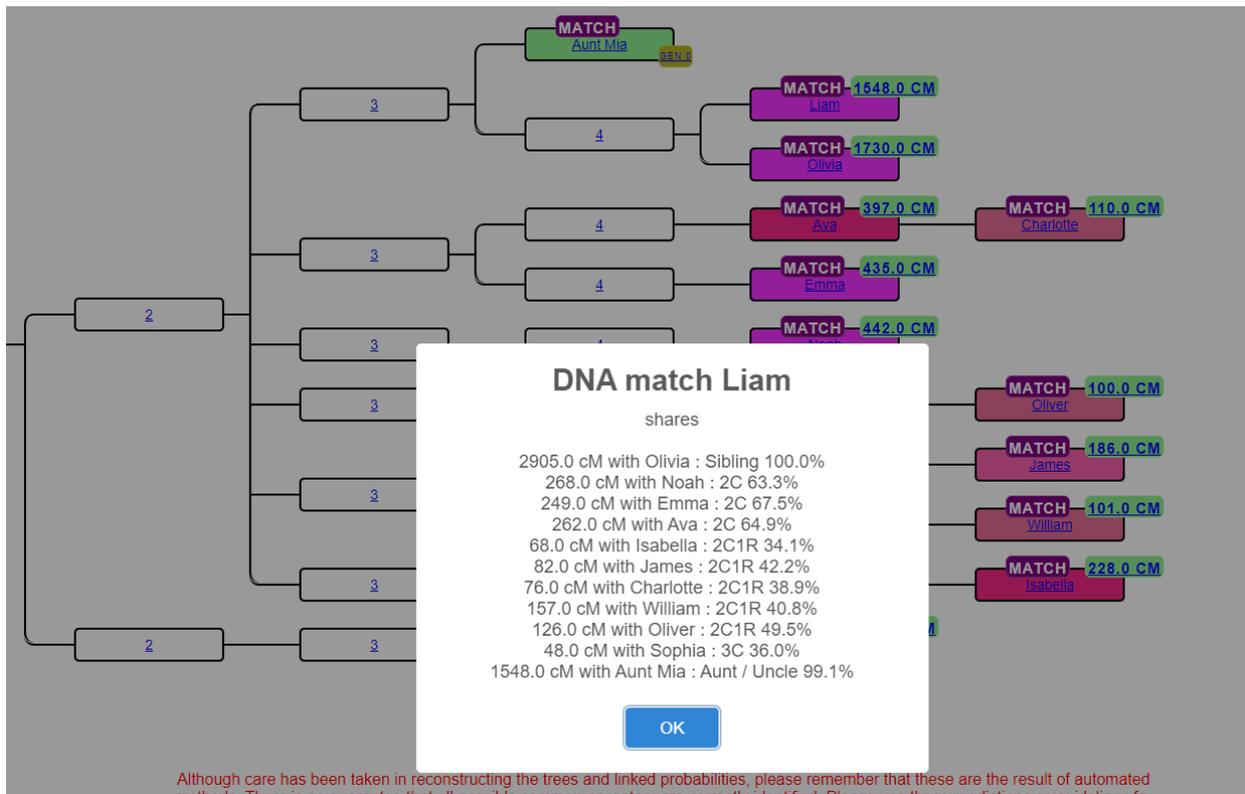


And this is the known tree



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.