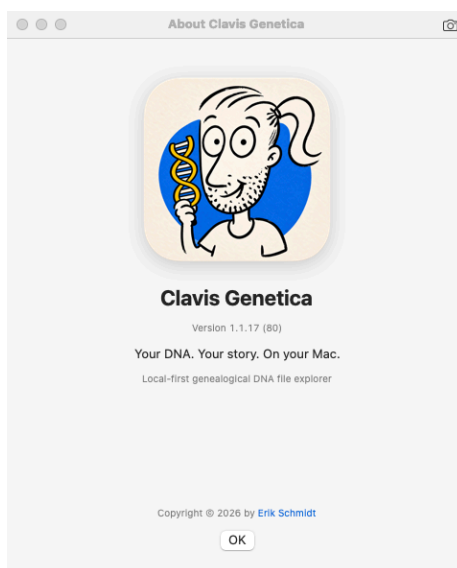


Clavis Genetica

User Manual

Your DNA. Your story. On your Mac.



App version 1.1.17

As of: October 10, 2026

Copyright © 2026 Erik Schmidt

Contents

1	Introduction	4
1.1	What Clavis Genetica is for	4
1.2	Principles	4
1.3	Requirements and installation	4
1.4	Updates	4
2	Getting started	6
2.1	The data folder	6
2.2	The sample family	7
2.3	The main window	8
2.4	Changing the language	9
2.5	The introduction	9
2.6	Screenshots from the app	10
3	Viewing files	11
3.1	Origin & creator	11
3.2	Chip raw data (autosomal)	11
3.3	Loading a file: summary, chromosome overview and data	14
3.4	Explained simply and What should I do next?	16
3.5	Local file and checksum	17
3.6	VCF files	17
3.7	BED files	18
3.8	BAM files and reads	18
3.9	FASTA files	19
3.10	Y-STR values	19
3.11	Match and segment lists	20
4	Ancestry and haplogroups	21
4.1	Maternal haplogroup (mtDNA)	21
4.2	Paternal haplogroup (Y-DNA)	23
4.3	Second opinion from FamilyTreeDNA	23
4.4	Ancient DNA on your line	24

5	Y-DNA and Big Y	25
5.1	Checking a Big Y data set	25
5.2	Y match list	27
5.3	The STR clock	28
6	Comparing files	29
6.1	Comparing two files	29
6.2	Multi-person comparison and triangulation	36
6.3	Separating parental sides (phasing)	37
7	Match lists and segments	39
7.1	The match list	39
7.2	The segment list and name matching	42
7.3	Overlapping segments	43
8	Family and family tree	48
8.1	People	48
8.2	Loading and linking a family tree	50
8.3	Family tree ↔ DNA	50
8.4	Analysing the family tree	51
8.5	The chromosome map	52
8.6	The family report	55
9	Research tools	56
9.1	The research log	56
9.2	Matches by hand	57
9.3	Looking up shared cM	58
9.4	Searching all files for a marker	59
10	Data, privacy and settings	61
10.1	Reference data	61
10.2	Privacy & security	62
10.3	Converting for upload	64
10.4	Medical data	64
10.5	Settings	65
11	Help in the app	67

A	Menus and keyboard shortcuts	69
B	Supported files	70
C	Troubleshooting	71

Introduction

1.1 What Clavis Genetica is for

If you have taken a DNA test for genealogy, you can usually download the “raw data” from the testing company: files with cryptic names and hundreds of thousands of lines. Clavis Genetica opens these files, recognises which company and which person they come from, and explains in plain language what they contain, what they are good for and where their limits lie.

The app also analyses the data: it determines haplogroups, compares files of relatives, checks Big Y data sets, works with match lists and segments, finds groups of matches that lie at the same place, paints a chromosome map and connects all of this with your own family tree.

1.2 Principles

- **Everything stays on your Mac.** Clavis does not upload any DNA data. The app only goes online to check for updates and to download public reference data when you explicitly ask for it (chapter 10.1).
- **Clavis does not change any DNA file.** Your own entries (notes, people, family tree, maps) are stored separately in the app’s own folder. The only exception: at the top level of the data folder Clavis places the small file `clavis-project.json`, by which it recognises the folder (section 2.1).
- **Pointers, not claims.** Wherever the app detects or estimates something, it states the basis and how certain it is. Relationship estimates, groups and time estimates are statements of probability, not proof.

1.3 Requirements and installation

Clavis Genetica runs on macOS 14 (Sonoma) or later. The app is delivered as a DMG file: open the DMG, drag Clavis Genetica into the Applications folder and start it. Next to the app, the DMG contains this manual as a PDF in the four program languages: `Handbuch (Deutsch).pdf`, `Manual (English).pdf`, `Manuel (Français).pdf` and `Handleiding (Nederlands).pdf`.

The app speaks German, English, French and Dutch. On first launch it uses the system language if it is one of these four, otherwise English. You can switch at any time with the flags at the top right of the toolbar (section 2.4).

1.4 Updates

Each time it starts, Clavis quietly checks once for a new version. If there is one, the **Check for updates** window opens. To check by hand, use Clavis Genetica › Check for updates () or the

settings.

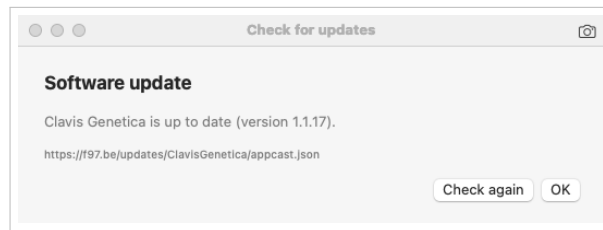


Figure 1.1: The **Check for updates** window when the app is up to date.

If an update is available, the window shows the release notes and the buttons **Later** and **Download, install & relaunch**. The download is verified before installation: it must come from the update server, and the checksum, signature, bundle identifier, version and developer ID must all match. Only then does a helper program replace the app and restart it. If anything fails, the old version is kept.

If the update server is on your local network, macOS asks for permission to access the local network. If the check fails, go to System Settings › Privacy & Security › Local Network and turn on **Clavis Genetica**.

Getting started

2.1 The data folder

Clavis works with a **data folder** that holds your DNA files. The **first level of subfolders is the person**: everything in `data/erik/...` belongs to “erik”. Deeper folders only serve to keep things tidy. The recommended layout is:

```
data/
  erik/
    atDNA/      chip raw data (MyHeritage, FTDNA, AncestryDNA, 23andMe ...)
    Y-DNA/      Big Y: CSV exports, variants.vcf, regions.bed, .bam + .bai
    mtDNA/      FASTA
  horst/
    atDNA/
```

2.1.1 Opening a folder

You choose a folder with

- the **Open** button in the toolbar,
- File > Open file or folder... () ,
- **Choose data folder** on the empty start screen or in the settings,
- or by dragging a folder into the window (“Drop a folder or DNA files here” appears).

A folder becomes the new data folder. Single files are added to the list; the person is then the name of the folder they are in. The data folder is remembered, even if you later move or rename it. **Refresh** () reads it in again.

If the chosen folder is one level too high (a single “person” that itself contains several person folders), Clavis asks: “This folder is probably one level too high”, with buttons to open the correct folder and **Leave as it is**.

2.1.2 What gets read

Clavis reads files with the extensions `.csv`, `.tsv`, `.txt`, `.vcf`, `.bed`, `.bam`, `.bai`, `.fasta`, `.fa`, `.fas` and `.fna`, and a gVCF also with the extension `.gvcf`. To be compared, however, a gVCF must end in `.vcf` (for example `sample.g.vcf`). When scanning the folder it only looks at the first 64 KB of each file (for BAM only at the header). It reads the actual content only when you click **Load file** in the detail view. This keeps even a folder with large BAM files fast.

ZIP and GZ archives are unpacked automatically, once, into a private cache; their files appear as “Archive.zip > File.csv”. Archives over 1 GB are not unpacked automatically, and unpacking

stops above 20 GB. If an archive could not be unpacked, it appears in the list as a single entry with an explanation.

Family trees (.ged, .gramps) do not appear in the file list; they are loaded in the **People** window (chapter 8). If there is exactly one GEDCOM file at the top level of a data folder for which no tree has been loaded yet, Clavis loads it automatically.

Your own work belongs to the data folder

Person profiles, the family tree, notes and groups for matches, name assignments, chromosome maps and the research log belong to the data folder in which you created them. When you switch folders, Clavis puts them aside and brings in those of the new folder. When you come back, everything is there again. If the new folder cannot be opened (for example because its saved work is damaged), Clavis reports that it could not be opened and gives the reason; the previous data folder then stays open with your work, and nothing is changed. So that Clavis can recognise a folder, it places the file `clavis-project.json` at its top level (an identifier, the folder name, the path and the app version; no DNA data). It does not appear in the file list. Please do not delete it: with it, Clavis finds your work even if you rename or move the folder. If you copy the folder in the Finder, the copy gets its own identifier and a copy of the current state of your work the first time it is opened; after that the two are independent. If the folder is write-protected, Clavis only remembers the link internally. If Clavis cannot save a change to people, the research log, matches entered by hand or notes for matches (for example because the drive is full), the window shows “Your changes could not be saved: ...” in orange, with the reason and the **Save again** button.

Back up your own work regularly with **Back up your own work** (section 10.2), especially before an app update.

Versions 1.1.13 to 1.1.16 sometimes did not load the work for a data folder that you had used before your first folder switch when you came back to it; the work was not lost. From version 1.1.17, Clavis brings this work back once by itself, at startup or when you switch to that folder: it only adds what is missing from the current state and overwrites nothing. The old storage stays in the app’s own folder as a safety copy.

2.2 The sample family

So that you can try things out without your own data, Clavis creates an **invented family**. You find the **Create sample family** button on the empty start screen and in the settings under **Data folder**. Clavis asks where the “Clavis-Beispielfamilie” folder should be created, creates it (an existing one is never overwritten) and offers to open it straight away: **Open sample family**.

The sample family contains

- for **Anna Beispiel** (born 1988) a match list with 42 matches and a segment list with 134 segments in FamilyTreeDNA format,
- for her father **Paul Beispiel** a match list with 26 matches and 109 segments,
- a family tree `Beispielfamilie.ged` with 32 people, which is loaded as soon as you open the folder,
- a file `LIESMICH.txt` with hints on what there is to discover: relatives on the father’s side also appear in Paul’s list, several matches overlap on chromosomes 7 and 11, there is a

pile-up on chromosome 15, and “Rike Entwurf” is not in the tree – where does she belong? The file is not a DNA file; Clavis does not show it in the file list.

The sample family contains no raw data, so it cannot show haplogroups or raw data comparisons. When the app language is Dutch, the folder is called “Clavis-Voorbeeldfamilie” and the people have Flemish names. To return to your own data, use **Choose data folder**; your own work is not touched.

Almost all figures of match lists, segments, groups and family trees in this manual show the sample family.

2.3 The main window

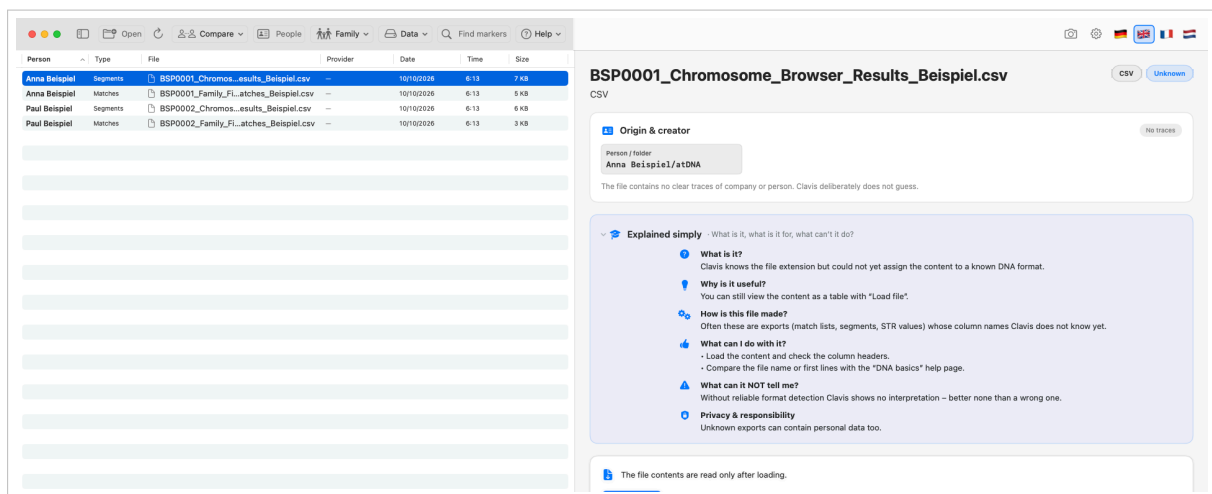


Figure 2.1: The main window with the sample family: the file list on the left, the detail view of the selected file on the right.

The main window has two columns: on the left the **file list**, on the right the **detail view** of the selected file (chapter 3).

2.3.1 The file list

Click a column header to sort by that column (a second click reverses the order); by default the list is sorted by person, then by file.

Person	Name of the person folder (first level).
Type	Short code of the file: atDNA, mtDNA, Big Y, Y-STR, Y-SNP, Y-DNA, BED, Seq., Segments, Matches. The tooltip names the kind of data.
File	Coloured icon for each kind of data (blue autosomal, pink mtDNA, orange Y, purple BED, turquoise sequence) and file name. An orange double document warns of a probable duplicate (same name and same size in another folder).
Provider	Detected company (FamilyTreeDNA as “FTDNA”); the tooltip names product and kit.
Date, Time, Size	Modification date and size of the file.

Below the list you see “Processed locally”, the note “Your files never leave this Mac.”, the path of the data folder and the **Quit Clavis Genetica** button.

2.3.2 The toolbar

From left to right:

Open	Open a data folder or single files.
Refresh	Read in the data folder again (circular arrow).
Compare	Menu with Compare the selected file with a second one , Multi-person comparison , Overlapping segments and Look up shared cM (chapter 6).
People	The People window (chapter 8).
Family	Menu with Analyse the tree , Chromosome map , Separate parental sides (phasing) , Research log and Matches by hand .
Data	Menu with Convert for upload , Family report (PDF) , Reference data , Privacy & security and, if enabled, Medical variants .
Find markers	Search in all files.

On the far right: the camera (**Picture of the window**), the gear (**Settings**) and four flags for the language.

2.4 Changing the language

Click one of the four flags (German, English, French, Dutch) at the top right to switch the language immediately; the active flag has a frame. Your choice is remembered. Alternatively: Settings › Language. The menus and buttons that macOS itself provides (such as **Services**, **Hide** and **Quit** in the app menu, the Edit and Window menus or **Hide Sidebar**) follow the chosen language only after Clavis is next started.

2.5 The introduction

On first launch the introduction (window **Show introduction**) opens with six steps: Welcome, Choose a folder, Select a file, Understand, Compare & discover, “Your data stays with you”. Use **Next** and **Back** to move through it, or **Skip** to leave it. You can open it again later with Help › Show introduction or in the settings.

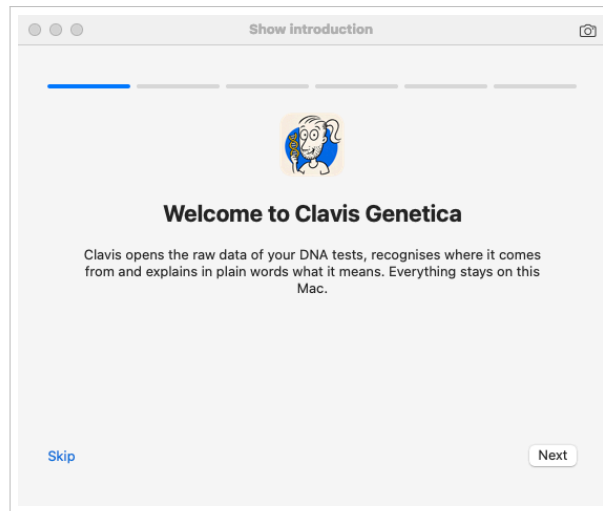


Figure 2.2: The first step of the introduction.

2.6 Screenshots from the app

Every Clavis window has a **camera icon** on the right of its title bar. Click it to save a picture of exactly this window as a PNG on the Desktop, with title bar, without shadow and without the mouse pointer, at double resolution on Retina displays. The file name is “Clavis-<window title>-<date>-<time>.png”; the Finder then shows the file. Data › Picture of the window () takes a picture of the frontmost window, including an open sheet (such as a quality report). Open menus and pop-overs do not appear in the picture.

Viewing files

When you select a file on the left, the detail view on the right shows everything Clavis knows about it. At the top are the file name, format and type and, for files that can be compared, the **Compare...** button. Below follow cards; which cards appear depends on the kind of file.

3.1 Origin & creator

This card appears for every file. As far as can be detected, it names **Person / folder**, **Company / lab**, **Test / product**, **Kit number**, **Tested person (per file)**, **Sample / sequencing ID**, **Created / exported**, **Reference genome** and **Programs / pipeline**.

At the top right it shows how certain the detection is: **Confidence: high** (a unique identifier in the file header or a typical file name pattern), **medium** (typical format) or **low** (a clue from the surroundings). Expanded, **How was this detected?** lists every piece of evidence, such as the file name, the header line, typical columns or markers. If Clavis finds no clear evidence, it deliberately does not guess.

Among the companies and formats it recognises are MyHeritage, FamilyTreeDNA (Family Finder, Big Y, Y-STR, SNP results, mtDNA), AncestryDNA, 23andMe and Living DNA, the reference genome (GRCh37, GRCh38, T2T) and analysis programs such as DRAGEN, GATK, Sentieon and DeepVariant. Under **Belongs together with** the card links files of the same kit or the same sample, such as BAM and BAI.

Detection is based only on the file name and the file header. File names may have been changed – the details are pointers, not proof.

3.2 Chip raw data (autosomal)

Chip raw data from MyHeritage, FamilyTreeDNA (Family Finder), AncestryDNA, 23andMe and Living DNA contain several hundred thousand measuring points (markers) with the person's genotype. For these files Clavis shows three analyses without the file having to be loaded first.

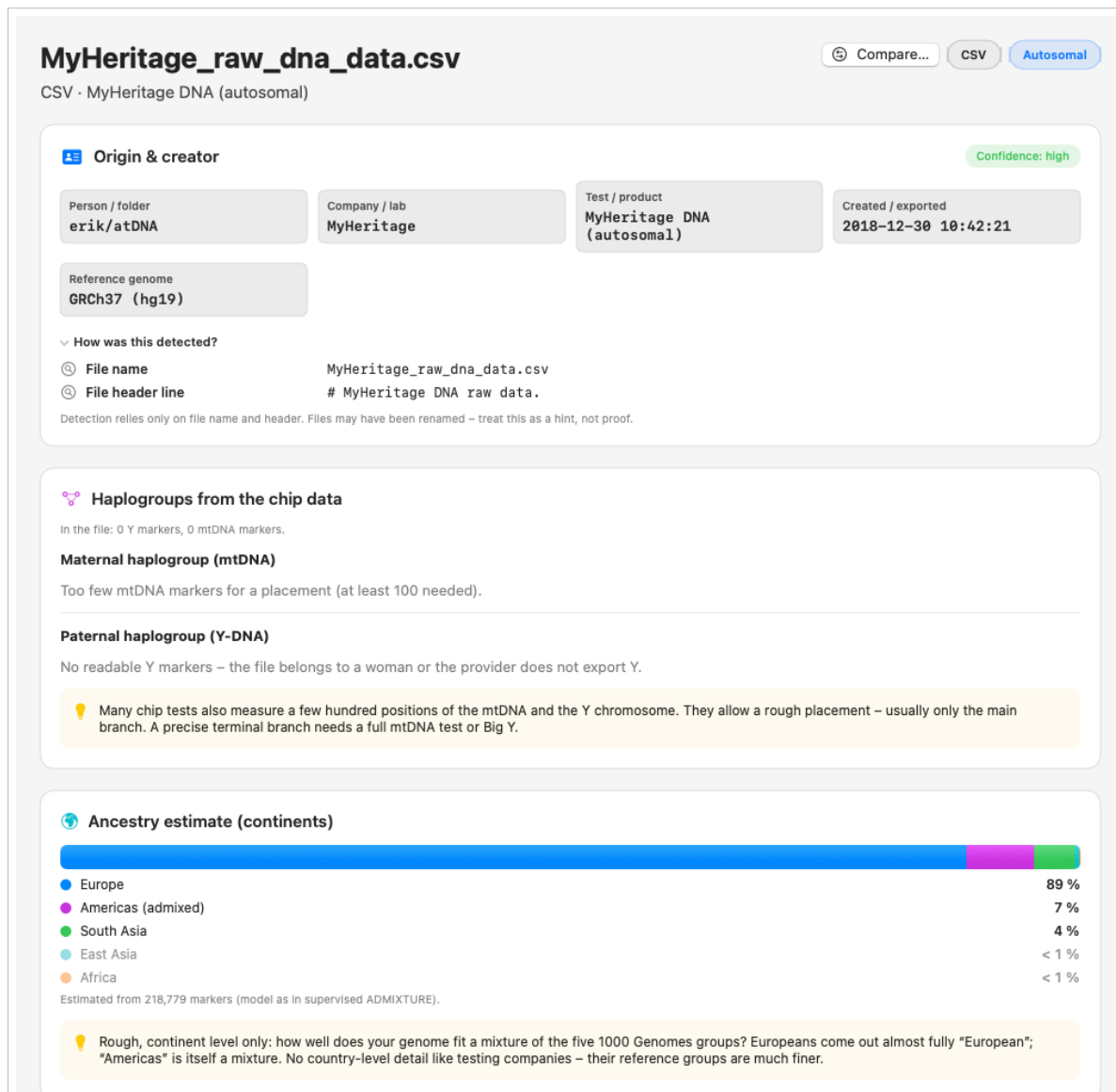


Figure 3.1: Chip raw data: Origin & creator, haplogroups from the chip data and the ancestry estimate.

3.2.1 Haplogroups from the chip data

Many chips also measure a few hundred positions of the mtDNA and the Y chromosome. Clavis states how many Y and mtDNA markers the file contains and, if there are enough, places them roughly in the tree – usually only down to the main branch.

- **mtDNA** needs at least 100 readable positions.
- **Y-DNA** needs the Y-SNP positions from YBrowse (available under **Reference data**, 3.8 MB) and a file in GRCh37. There is only a result if at least three derived SNPs lie on one path. Many chips (AncestryDNA, for example) measure hardly any named Y-SNPs.

For a precise terminal branch you need a full mtDNA test or Big Y (chapters 4 and 5).

3.2.2 Ancestry estimate (continents)

Click **Estimate** to calculate how well the data fit a mixture of five 1000 Genomes groups: Africa, Americas (admixed), East Asia, Europe and South Asia. This needs the **allele frequencies from 1000 Genomes** (under **Reference data**, 1.46 GB download) and a file in GRCh37. The calculation uses up to 200,000 markers and follows the “supervised” model of ADMIXTURE.

The estimate is rough and only valid at continent level. Europeans come out almost fully “European”; “Americas” is itself a mixture. There is no country-level detail as with the testing companies – their reference groups are much finer.

3.2.3 Are the parents related? (runs of homozygosity)

If at some place you inherit the same piece of a common ancestor from both your father and your mother, all markers there are homozygous. Long homozygous runs therefore show that the parents are related to each other: the longer and more numerous they are, the closer the relationship.

Click **Find runs** to search chromosomes 1 to 22. With **Runs from** you choose the minimum length (5, 7 or 10 cM). Clavis shows the number of runs, their total and the longest one, a graphic of the 22 chromosomes and a table with position, cM and number of markers. Expanded, **For comparison: expected total for children of related parents** gives reference values from siblings to third cousins.

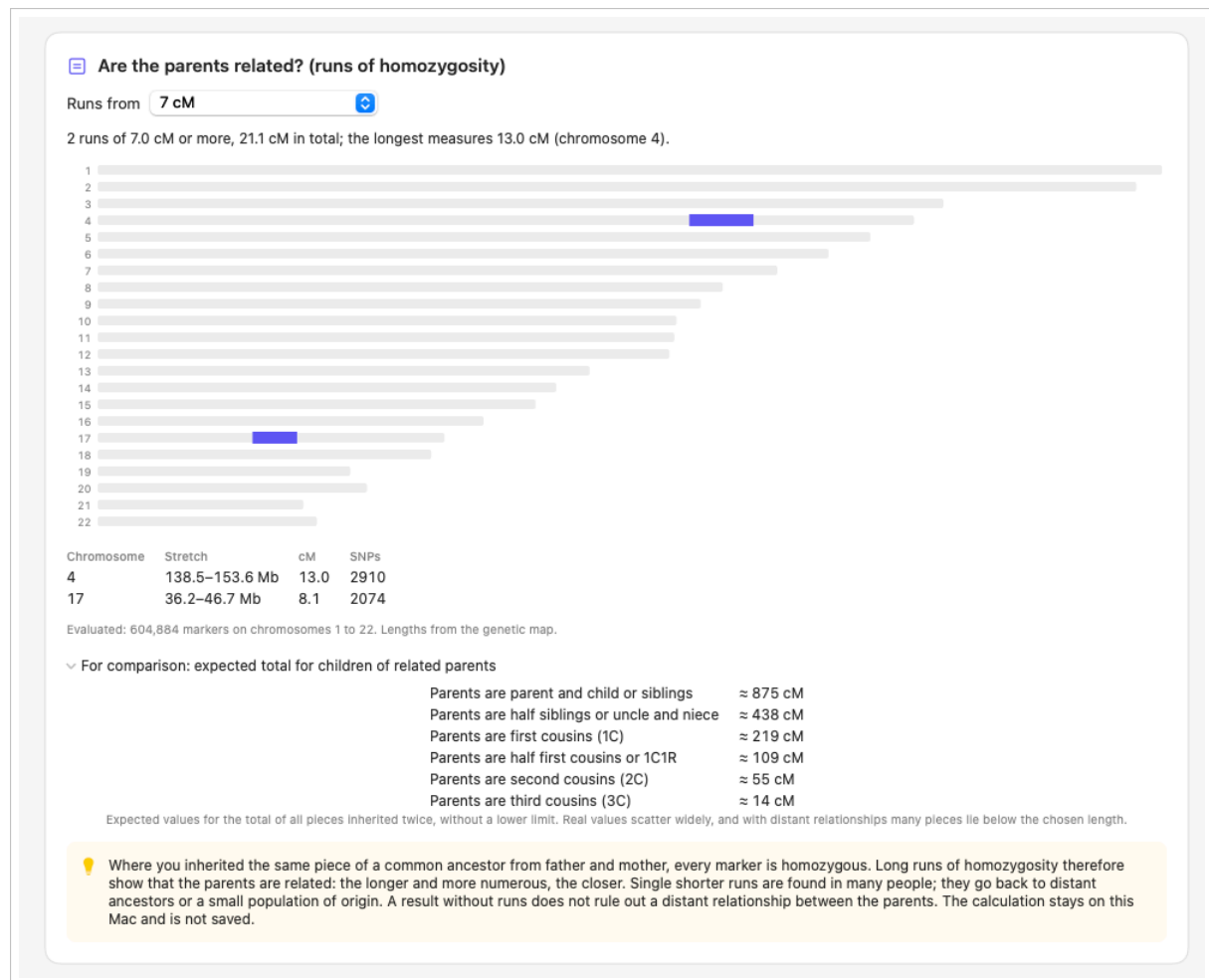


Figure 3.2: Runs of homozygosity: chromosome graphic, table and reference values.

- A run needs at least 300 markers and 50 markers per megabase; gaps of more than 2 Mb end it. A single heterozygous marker in the middle of a run is treated as a read error.
- With the **genetic map** loaded (reference data) the cM values are exact, otherwise they are estimated.
- The function needs chip raw data with at least 50,000 markers; a VCF only lists positions that differ from the reference and is not suitable.

Individual shorter runs are found in many people; they go back to distant ancestors or to a small population of origin. A result without runs does not rule out a distant relationship between the parents. The calculation stays on this Mac and is not saved.

3.3 Loading a file: summary, chromosome overview and data

The actual content of a file is only read when you click **Load file**. While it is loading, you see **Loading file...** with a progress bar and the **Cancel** button. If loading fails, a box appears with the cause, such as “File not found”, “File is empty” or “Invalid file format: ...”, and the **Try again** button. Once the file is loaded, you see:

Summary

Format, number of entries, missing values (no-calls), reference genome; below that the distribution over the chromosomes as clickable filters and, where applicable, warnings such as “Some rows have chromosome 0” (internal control markers of the chip, not errors). **Technical details from the file** shows all details from the file header.

Chromosome overview

Each row is one chromosome, drawn to scale. The darker a section, the more measuring points lie there; red places have many missing values. For BED files it shows the covered region.

Data viewer

The entries as a table with search and pages of 100 rows. Click a row to see its details: reference and variant, for VCF the sample values (FORMAT) and technical values (INFO) with an explanation when you hover over them, the position in the other reference genome (if the liftover data are loaded) and the **Original source line** with **Copy line**.

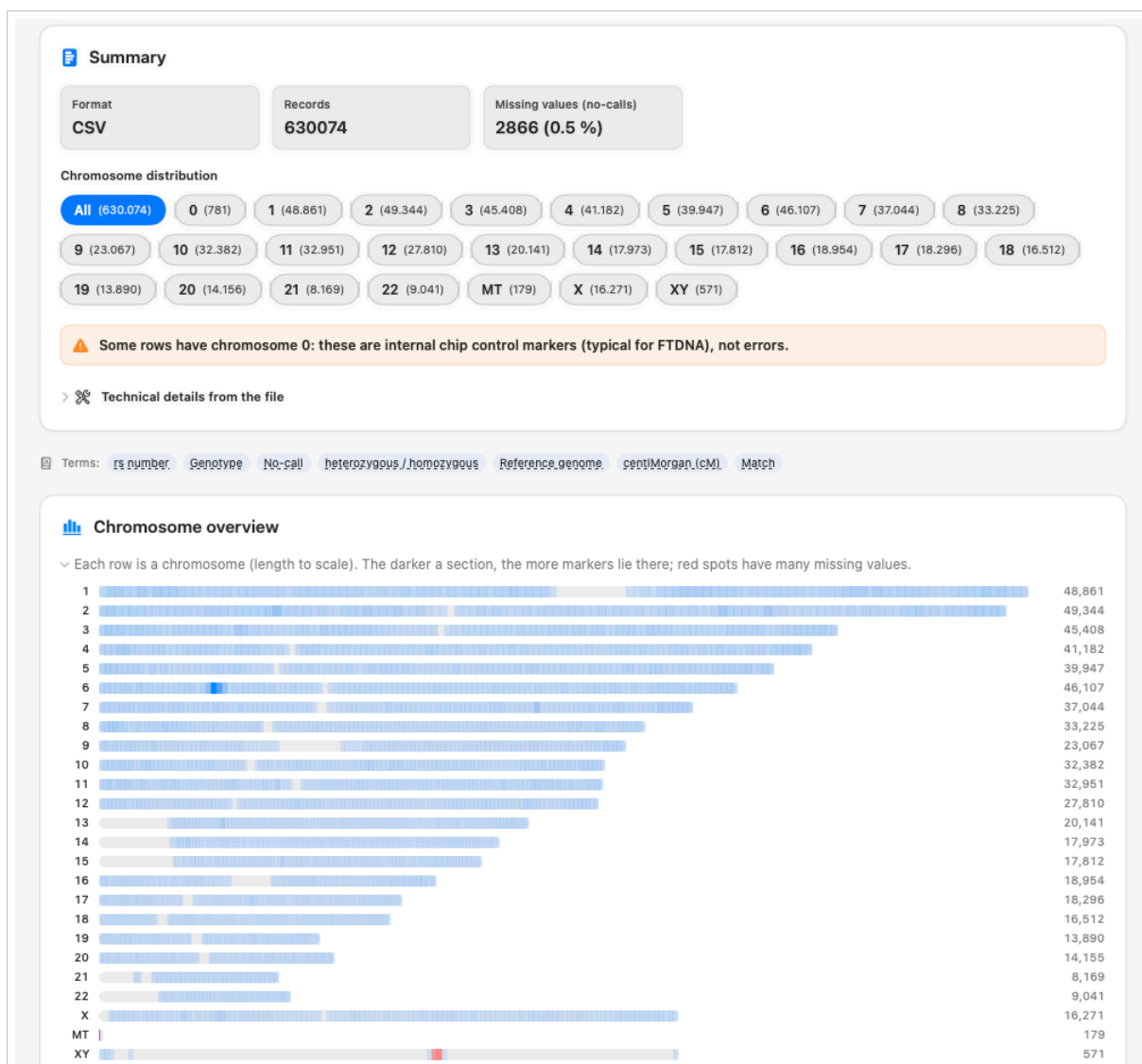


Figure 3.3: Summary and chromosome overview of a loaded chip file.

Data viewer 630,074 records

Search...

< 1 / 6.301 >

Variant	Chromosome	Position	Genotype	Quality (QUAL)	Filter
2010-08-Y-1221	0	0	AA	—	—
2010-08-Y-1364	0	0	GG	—	—
2010-08-Y-1995	0	0	TT	—	—
2010-08-Y-2045	0	0	CC	—	—
2010-08-Y-3042	0	0	TT	—	—
2010-08-Y-3189	0	0	AA	—	—
2010-08-Y-3314	0	0	TT	—	—
2010-08-Y-3348	0	0	TT	—	—
2010-08-Y-3576	0	0	TT	—	—

2010-08-Y-2045 CC

Chr 0 : 0

Original source line Copy line

2010-08-Y-2045, 0, 0, CC

Figure 3.4: The data viewer with a selected row and its original line.

Below the data, **Detailed explanation** explains the columns of the file type (for chip data: RSID, chromosome and position, genotype, no-calls). Clavis keeps at most three loaded files in memory; when one drops out, **Load file** appears again. The selected file and file 1 of a comparison never drop out.

3.4 Explained simply and What should I do next?

Explained simply answers six questions for every kind of file: What is it? Why is it useful? How is this file made? What can I do with it? What can it NOT tell me? Privacy & responsibility. Whether the card is expanded when you open a file is set in the settings. If you collapse or expand it in a file, this only applies to that view; the setting stays as it is.

✓ **Explained simply** · What is it, what is it for, what can't it do?

What is it?
This file is the result of a DNA chip test (genotyping) such as those offered by MyHeritage, FamilyTreeDNA (Family Finder), AncestryDNA or 23andMe. About 600,000–700,000 selected positions of your genome were measured from a saliva or cheek-swab sample. Each line is one position: marker ID (rs number), chromosome, position and your result, e.g. "AG" – one letter from your mother, one from your father.

Why is it useful?
You inherit autosomal DNA (chromosomes 1–22) half from each parent – and therefore a piece from every ancestral line, not just the direct paternal or maternal one. That makes it the key tool for finding relatives across all lines (reliably to about third cousins, sometimes further) and for ancestry estimates. The more distant a relationship, the less DNA is shared.

How is this file made?
The lab reads the chip and maps every signal to a position in the human reference genome (usually GRCh37). You then download the raw data from your account. Header lines (starting with #) often name the company, date and reference – Clavis uses exactly these traces to identify the creator.

What can I do with it?

- Upload the file to other companies or services (e.g. MyHeritage, FamilyTreeDNA, GEDmatch) to find more relatives – usually free.
- Compare two people (e.g. "erik" and "horst") to check a suspected relationship.
- Look up individual markers or check quality (share of missing values).
- Keep it as a backup – companies can close accounts.

What can it NOT tell me?
It is not a complete DNA sequence: only about 0.02 % of all positions were measured. It contains neither a finished ancestry breakdown nor a match list – both are computed by the company. Health conclusions are unreliable: chip data has an error rate and is not validated for medical use. Lines with "--" are gaps, not results.

Privacy & responsibility
DNA always reveals something about your parents, siblings, children and cousins. Upload raw data only where you trust the terms of service and only with the tested person's consent. Clavis processes everything on this Mac only – nothing is uploaded.

Figure 3.5: “Explained simply” expanded.

What should I do next? gives numbered suggestions for each kind of file and offers matching buttons: **Compare...**, **Family report (PDF)** and, for autosomal chip raw data, **Convert for upload**.

3.5 Local file and checksum

At the very bottom, **Local file** shows path, size and date with the **Reveal in Finder** button, and the **SHA-256 checksum** of the whole file for copying (automatically up to 50 MB, above that at the click of a button). The checksum is a “fingerprint”: if it matches for a copy, the copy is unchanged and complete – ideal for checking backups.

3.6 VCF files

A VCF (Variant Call Format) lists the positions where the sample differs from the reference, with genotype, quality and filter. Clavis reads the first sample column. In the data viewer, the details of a row also show **Sample values (FORMAT)** and **Technical values (INFO)**. The detailed explanation describes REF and ALT, genotype (0/1, 1/1), QUAL (Phred 20 = 99%, 30 = 99.9%), FILTER, INFO and FORMAT (AD, DP, GQ). A Y-VCF belongs to the Big Y data set (chapter 5).

If the file contains several samples, Clavis reads only the first; the summary then shows the note “The VCF file contains several samples. Clavis shows only one of them – which one is listed under “Technical details from the file”.”, and **Technical details from the file** names the sample that was read and all samples present under “Selected sample” and “Available samples”. You cannot choose a different sample.

A gVCF also contains **reference blocks**: stretches in which the sample was measured and matches the reference (rows with END= and ALT <NON_REF> or <*>). The data viewer shows

them as “chromosome:start-end” with the genotype “0/0 (reference block: n bp)”. Phased genotypes (with |) carry the addition “[phased]”, for example “0|1 (A|G [phased])”.

A genome VCF is useful for autosomal comparison as a gVCF or against chip raw data. Only positions measured in both files are compared, but a variants-only VCF lists only the differences from the reference. Two such files therefore usually share too few positions (section 6.1.2).

3.7 BED files

A BED file describes regions, such as the reliably read regions of a Big Y test. The coordinates are 0-based and half-open [start, end). The chromosome overview shows the covered share; the data viewer lists chromosome, start, end, length, name and strand. A BED file does not yet contain any genetic results.

3.8 BAM files and reads

A BAM file contains the raw reads of a sequencing run, the BAI file the index for it. Clavis reads the SAM header and, with an index, the number of reads per contig. After **Load file**, the section **BAM header and reference contigs** shows the contigs with length and mapped reads, the programs used in the analysis and the complete header.

BAM header and reference contigs ✔ BAI index found

The BAM file was verified locally and its SAM header was parsed. Full read-level alignment streaming is not enabled in this tier; no false analysis is simulated.

View reads at position ⓘ

Reference contigs

Contig / chromosome	Length (bp)	Mapped reads
chrY	57,227,415	7,734,492
chrM	16,569	0
chr1	248,956,422	0
chr2	242,193,529	0
chr3	198,295,559	0
chr4	190,214,555	0
chr5	181,538,259	0
chr6	170,805,979	0

Programs & pipelines

samtools v1.5 · samtools reheader /tmp/GRC183810426/0426-ebba33d9-c458-4c3a-8c70-4f4debe8569c_header_stripped.sam /tmp/GRC1838...

> Full SAM header text

Figure 3.6: BAM file: BAM header and reference contigs with the field “View reads at position”.

With an index you also get **View reads at position**: enter the chromosome (default chrY) and the position and click **Reads**. A window opens with the read depth, the distribution of bases (read forwards and backwards) and an excerpt of the reads of ± 30 bases, with the requested column marked in yellow. Reads with low mapping quality, bases with low base quality, duplicates and overlapping mates are not counted. A real variant shows the same base in almost all reads in both directions.

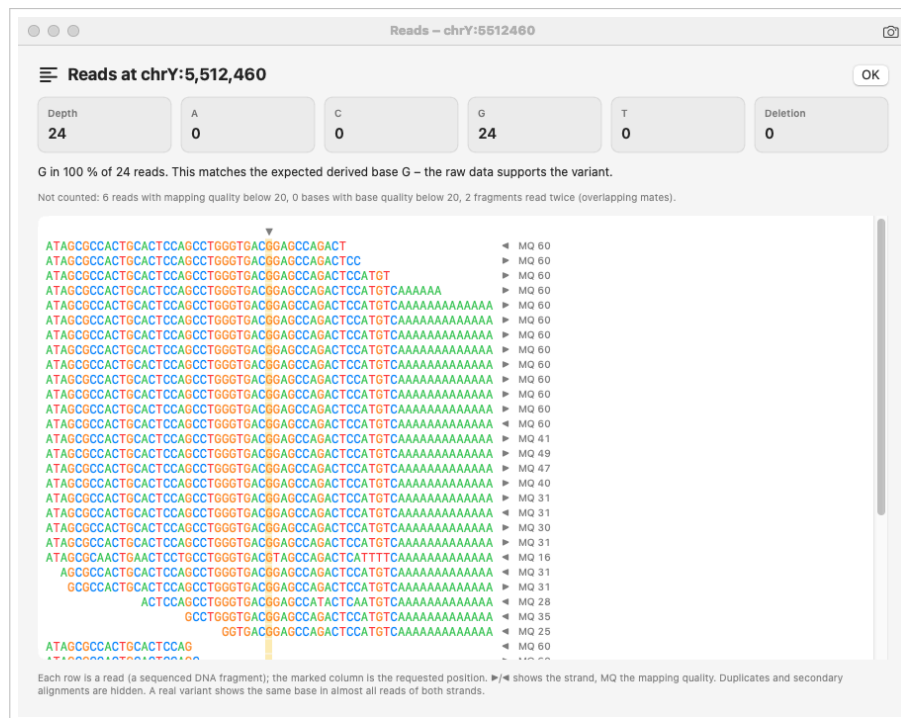


Figure 3.7: The “Reads at ...” window at a private variant: all reads show the expected base.

The BAM is the only file from which new analyses can be made later. Please keep a permanent backup of it.

3.9 FASTA files

FASTA files contain base sequences, usually the complete mtDNA. For mtDNA, Clavis determines the maternal haplogroup (chapter 4). After **Load file**, the **Sequence view** shows length, GC content and unknown bases (N) as well as the sequence page by page (1,200 bases per page, 60 per line). The field **Search base sequence / motif...** jumps to that position when you enter a number and finds the first occurrence of a motif when you enter letters. The first 100,000 bases of each sequence are shown.

3.10 Y-STR values

For a Y-STR file (such as Y-37, Y-111 or Big Y-700) the **Y-STR values** card shows the panel, the number of markers and a grid of all markers with their values for the first person in the file. With **Compare** Clavis calculates the genetic distance to another man (section 6.1.4).

Y-STR values

Big Y-700 (797 STR) 797 markers - 1 person(s) in the file

DYS393 12	DYS390 22	DYS19 14	DYS391 10	DYS385 13-18	DYS426 11	DYS388 15
DYS439 12	DYS389I 13	DYS392 11	DYS389II 32	DYS458 16	DYS459 8-9	DYS455 11
DYS454 11	DYS447 25	DYS437 14	DYS448 20	DYS449 27	DYS464 12-14-16-17	DYS460 10
Y-GATA-H4 10	YCAII 22-22	DYS456 16	DYS607 14	DYS576 15	DYS570 17	CDY 32-35
DYS442 12	DYS438 11	DYS531 11	DYS578 9	DYF395S1 15-16	DYS590 8	DYS537 11
DYS641 10	DYS472 8	DYF406S1 11	DYS511 9	DYS425 12	DYS413 21-23	DYS557 18
DYS594 10	DYS436 12	DYS490 12	DYS534 14	DYS450 8	DYS444 14	DYS481 25
DYS520 22	DYS446 13	DYS617 12	DYS568 11	DYS487 14	DYS572 11	DYS640 12
DYS492 12	DYS565 11	DYS710 37	DYS485 15	DYS632 8	DYS495 15	DYS540 12
DYS714 25	DYS716 27	DYS717 20	DYS505 12	DYS556 12	DYS549 12	DYS589 11
DYS522 13	DYS494 9	DYS533 12	DYS636 11	DYS575 10	DYS638 10	DYS462 10

Figure 3.8: The “Y-STR values” card of a Big Y-700.

3.11 Match and segment lists

Clavis recognises FamilyTreeDNA match lists and segment lists by their columns and shows them in views of their own; they are described in chapter 7. Y-DNA match lists are covered in chapter 5.

Ancestry and haplogroups

A **haplogroup** is a branch in the family tree of humankind. The maternal haplogroup follows the mtDNA (from the mother to all her children), the paternal one the Y chromosome (from father to son). A haplogroup is not a statement of origin in the sense of a nationality, and it is not a medical statement.

4.1 Maternal haplogroup (mtDNA)

For an mtDNA FASTA file Clavis determines the haplogroup immediately, without loading. The sequence is aligned to the rCRS reference and scored against PhyloTree Build 17 (5,435 branches) using the HaploGrep method; everything needed for this is built into the app.

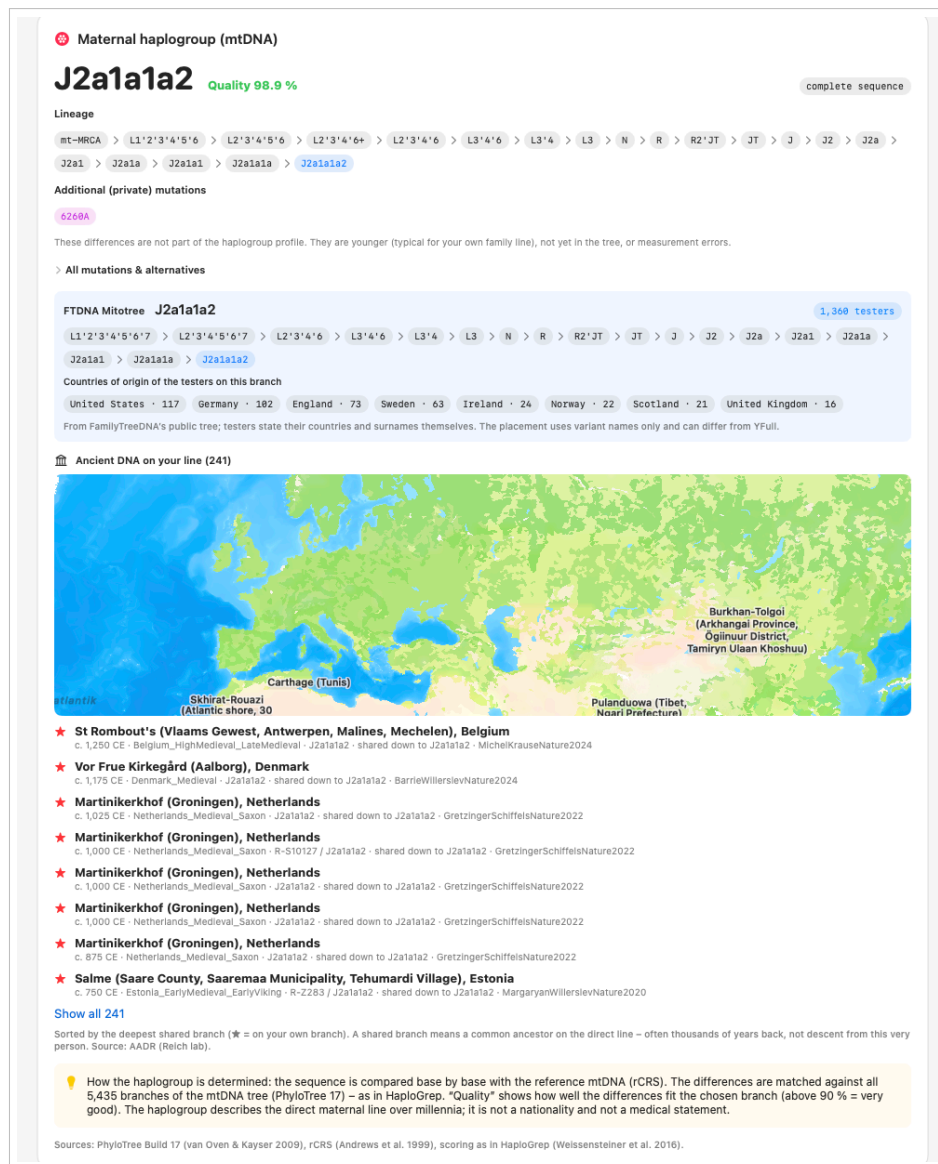


Figure 4.1: Maternal haplogroup with lineage, second opinion from the FTDNA Mitotree and ancient DNA on your own line.

The card shows

- the haplogroup with its **quality** (above 90% is very good) and whether the sequence is complete,
- the **Lineage** from the root to the haplogroup,
- **Additional (private) mutations** – younger, not yet in the tree, or measurement errors – and **Expected but not found**,
- expanded, **All mutations & alternatives**: all differences from the rCRS and the next best haplogroups.

4.2 Paternal haplogroup (Y-DNA)

For a Big Y “Named Variants” file or FamilyTreeDNA SNP results, Clavis looks for the deepest branch on whose path most derived SNPs lie. Without a download it uses a built-in basic tree with about 95 major branches. For the precise terminal branch, load the public **YFull tree** in the same card (**Load YFull tree (≈ 15 MB)** or **Import file...**); no DNA data are sent. With the YFull tree the card also states when the branch formed and when the most recent common ancestor of all its living carriers lived.

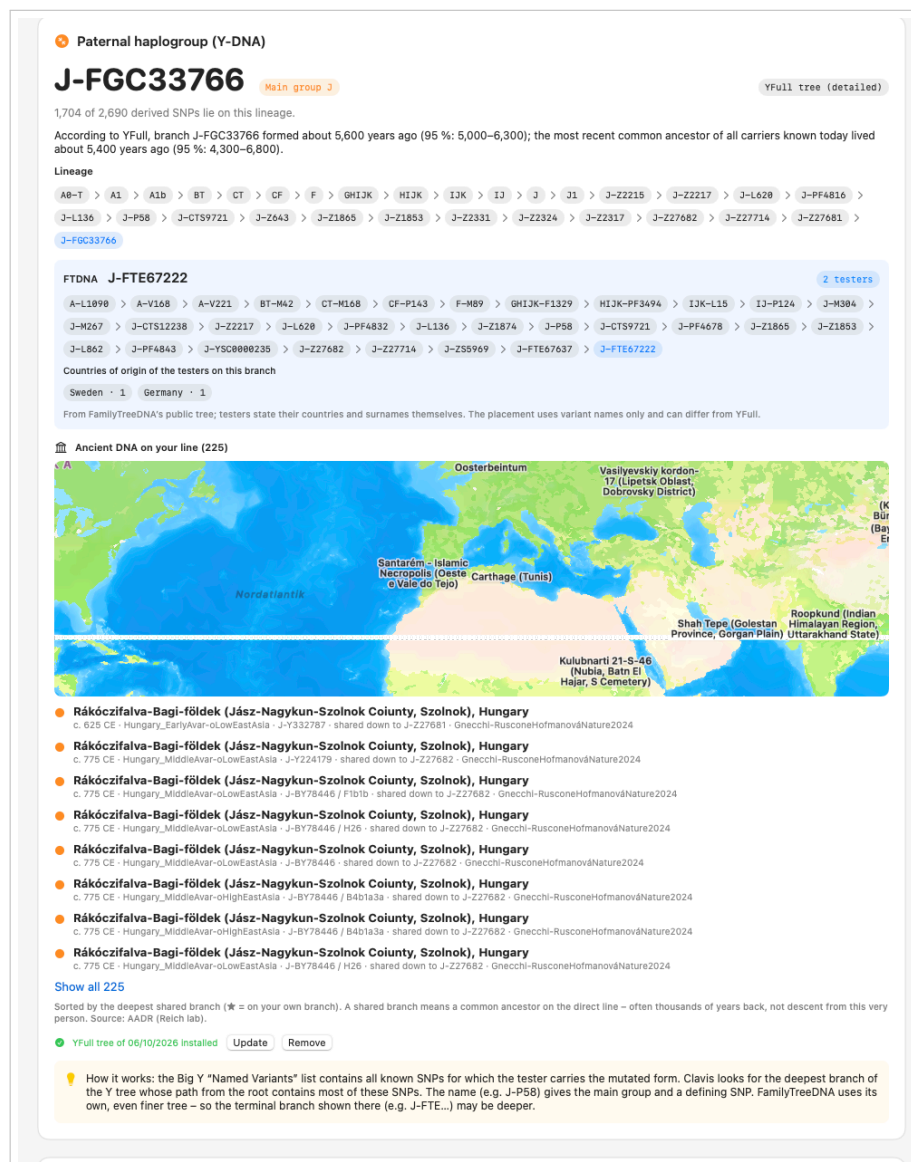


Figure 4.2: Paternal haplogroup from a Big Y file, with YFull tree, FTDNA second opinion and ancient DNA.

4.3 Second opinion from FamilyTreeDNA

If the FamilyTreeDNA haplotrees are loaded under **Reference data** (Y tree and the mtDNA tree “Mitotree”), a blue box appears in the haplogroup cards: the branch according to FamilyTreeDNA with its path, the number of testers, the testers’ countries of origin on this branch and, for Y, the common surnames. Countries and names are given by the testers themselves.

FamilyTreeDNA uses its own, finer tree; the terminal branch there may be deeper than at YFull.

4.4 Ancient DNA on your line

If the **AADR** data set (Allen Ancient DNA Resource, Reich Lab) is loaded under **Reference data**, the haplogroup cards show **Ancient DNA on your line**: published samples from excavations that share a branch with you, on a map (red: your own branch, orange: a shared parent branch) and as a list with site, date, group and publication. A shared branch means a shared ancestor in the direct line, often thousands of years back – not descent from that particular person.

Y-DNA and Big Y

5.1 Checking a Big Y data set

A Big Y test consists of several files that say little on their own: CSV lists (Named Variants, Private Variants, Matches), the VCF with the variants, the BED with the reliably read regions and the BAM with the reads. The **Check Big Y data set** card appears with every Y file of a person and brings together the files of the same kit. At the top it shows which parts were found (green tick) and which are missing.

Check data set now checks every private and every named variant: Does it lie in the BED region? Is it in the VCF as derived with PASS? How many reads support it? The result is:

confirmed	in the BED region, derived in the VCF, PASS, read depth at least 10
weakly supported	in the VCF, but low read depth, no PASS or a different base
outside BED	not in the reliably read region – treat with caution
not in VCF	no entry at this position (a different analysis version?)
contradicted	the VCF shows the original base
not checkable	the VCF is missing from the data set

5.1.1 The “Figures” view

The **Figures** view shows key figures: **Reliably read (BED)** (typically 14–16 Mb for Big Y-700), **Regions**, **VCF: PASS / total**, **Derived calls (PASS)**, **Heterozygous calls** (these should be rare in a man), **Typical read depth** (good from about 10×) and **Reads on chrY (BAI)**. Below follow the haplogroup, the haplogroups of the matches and tables of the private variants and of the named variants that are not clearly confirmed, with position, change, BED, genotype, read depth, quality, filter and assessment. If a BAM is present, **Reads** in each row opens the reads window at exactly this position (figure 3.7).

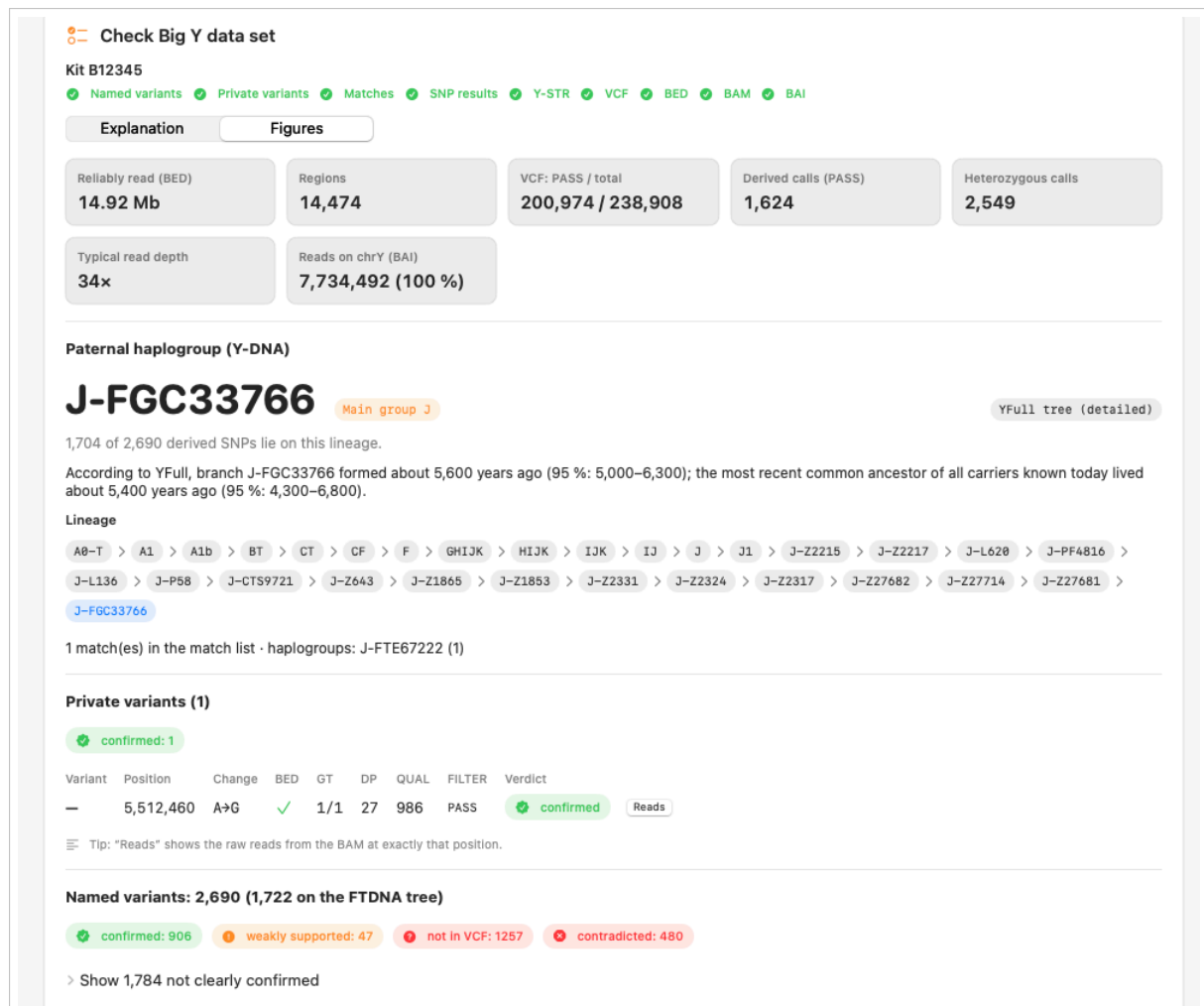


Figure 5.1: Check Big Y data set, “Figures” view.

5.1.2 The “Explanation” view

The **Explanation** view answers in everyday language: *Where are you on the tree? How well was your test read? Which variants are yours alone? What does your match mean? How old can the connection be?* Each section has an expandable **How does the app know this?**. If you enter your year of birth in the **Year of birth** field, the age estimate is converted into calendar years. It is based on the YFull mutation rate, scaled to the area of your BED file, with a 95% range.

Check Big Y data set

Kit B12345

Named variants Private variants Matches SNP results Y-STR VCF BED BAM BAI

Explanation Figures Year of birth

Where are you on the tree?

Your paternal line belongs to the branch J-FGC33766. A branch is a group of men who carry the same changes on the Y chromosome and therefore share an ancestor in the direct paternal line.

Your test found 2,690 known (named) variants. 1,722 of them have a fixed place on the tree; the remaining 968 are known but do not mark a branching point so far.

The branch says nothing about surnames, places or dates. It only tells you with whom you share the paternal line.

> How does the app know this?

How well was your test read?

The test reliably read 14.9 million positions of the Y chromosome. The Y chromosome has about 57 million positions; a large part consists of repeats that cannot be read unambiguously with short reads. That is normal for a Big Y.

A typical position was read 34 times. From about ten reads on, a result is considered solid.

> How does the app know this?

Which variants are yours alone?

Your test lists one private variant. Private variants are changes that have not been found in any other tester so far. They are the changes of your own line since the last common ancestor with your closest match.

Position	Change	Reads	Assessment
5512460	A→G	26 of 26	well supported

Well supported: 1. To be read with caution: 0.

> How does the app know this?

Figure 5.2: Check Big Y data set, “Explanation” view (excerpt).

The section “What does your match mean?” names your closest match. Whether the match really carries the original base at your positions can only be decided with his files.

5.1.3 PDF report

PDF report... saves “Clavis-BigY-<Kit>.pdf” with haplogroup, explanation, key figures and variant tables. Names of matches only appear in it if **Include match names in the PDF** is ticked.

5.2 Y match list

Clavis shows a FamilyTreeDNA Y-DNA match list (file name containing “match”) after **Load file** as a **Match list** with the columns name, haplogroup, Y-37, Y-67, Y-111, non-matching SNPs and earliest ancestor, with filter, sorting and an overview of the haplogroups. When you select a match, the detail view shows

- an **SNP time estimate**: from the number of non-matching known SNPs, the age of the common ancestor with a 95% range (more precise if the BED file is in the same folder; otherwise with 83 years per SNP),
- an **STR time estimate** from the genetic distance,
- with the YFull tree, the shared branch and its age,
- a **My notes** field, which is saved locally right away.

Unnamed private variants are missing from the company’s count, so the age tends to come out too low.

5.3 The STR clock

For Y-STR comparisons Clavis estimates from the genetic distance how many generations ago the common ancestor of the paternal line most likely lived (average rate 0.003 per marker and generation, 31 years per generation, from 12 markers). The estimate is very rough; back mutations are possible, and a distance of 0 does not rule out distant ancestors. Big Y gives more precise results.

Comparing files

6.1 Comparing two files

This is how you compare two files:

1. Select the first file in the list.
2. Click **Compare...** in the detail view, or use Compare › Compare the selected file with a second one (⌘D).
3. The file becomes “File 1”. In the list you can now only choose files that can meaningfully be compared with it; all others are greyed out.
4. Click the second file. The comparison starts immediately.

End comparison returns to the normal view.

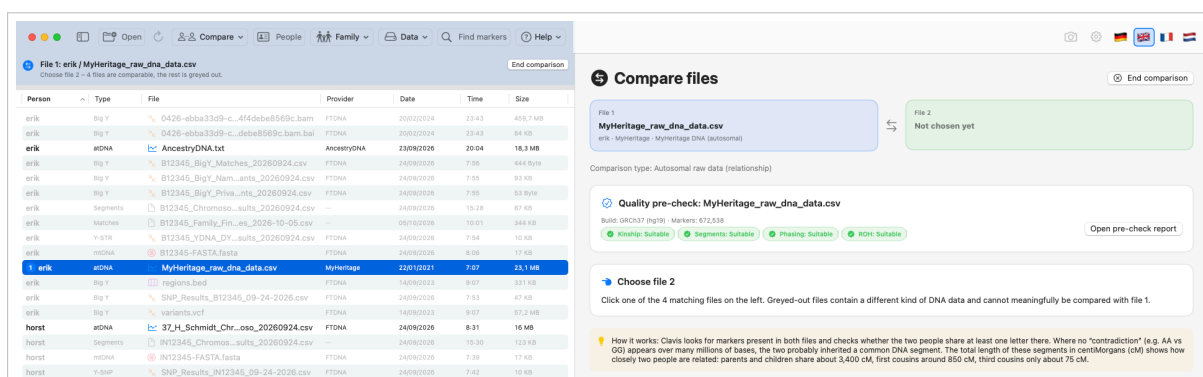


Figure 6.1: Comparison mode: file 1 is selected, only matching files can be clicked; on the right the “Quality pre-check” card.

You can compare:

Autosomal raw data	chip files with each other, plus genome VCFs, preferably as gVCF (relationship)
Y-DNA variants	Big Y CSV, SNP results and Y-VCF
Y-STR values	two Y-STR files, also a project table as file 2
Y match lists	two Y match lists (new, gone, shared)
mtDNA	two mtDNA FASTA files
BED regions	two BED files (coverage and overlap)

6.1.1 Quality pre-check

As soon as file 1 is chosen, the **Quality pre-check**: card with the name of the file appears below the **File 1** and **File 2** boxes – for every kind of comparison, also during and after the comparison. It names the reference genome (**Build**) and the number of markers, and coloured badges show what the file is suitable for: for chip raw data, VCF and BAM **Kinship**, **Segments**, **Phasing** and **ROH** (runs of homozygosity), for Y, mtDNA, BED and list files **Kinship**, **Segments**, **ClinVar** and **Haplogroup**. The pre-check is only a display; it does not prevent any comparison.

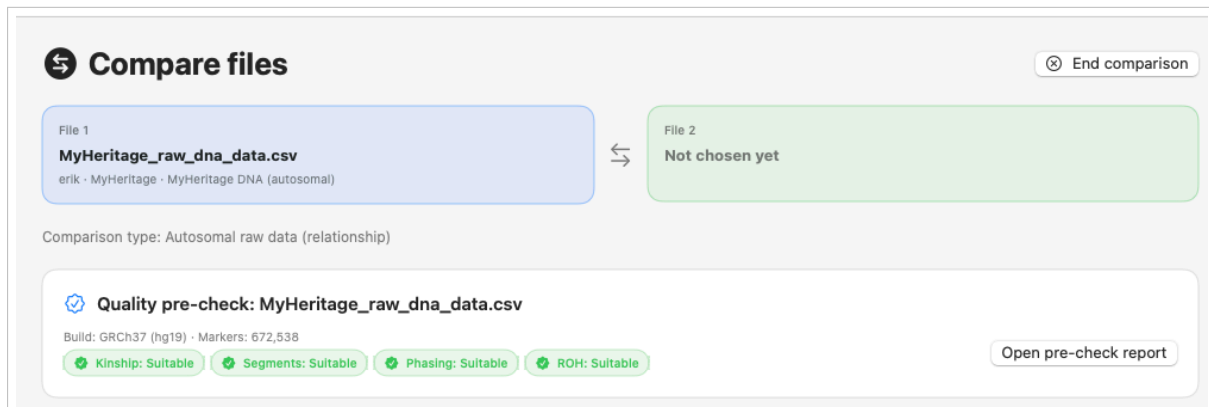


Figure 6.2: The “Quality pre-check” card for a loaded chip file: all four domains suitable.

Suitable (green)	The data are sufficient for this domain.
Limited (orange)	Usable, but with limitations; the report gives the reasons.
Unsuitable (red)	Too few or too uncertain data for this domain.
Not applicable (grey)	The domain does not fit the kind of file.

For chip raw data, VCF and BAM: with fewer than 5,000 markers or more than 20% missing values (no-calls) a file is **Unsuitable**; with fewer than 200,000 markers, more than 5% no-calls or an unconfirmed reference genome it is **Limited**. **ClinVar** is suitable from 5,000 markers, **Haplogroup** only for Y and mtDNA files. **ROH** is also **Limited** if more than 15 Mb without any measurement lie between two markers of a chromosome (**Measurement gap**). Regions in which no chip can measure do not count here: the centromeres, large heterochromatin blocks and the short arms of chromosomes 13, 14, 15, 21 and 22. The rules on the number of markers, no-calls and gaps do not apply to Y, mtDNA and BED files; by nature they contain few entries.

The pre-check works with the values of the loaded file: number of markers, no-call rate and (for chip raw data and VCF) measurement gap. If file 1 is chip raw data, a VCF or a BAM that has not been loaded yet, the card shows “Load the file so that Clavis can check its quality.” with the **Load file** button; the assessment appears after that.

Open pre-check report shows the **Quality report** sheet: file name, kind and reference genome, an overall verdict, the key figures **Total markers**, **Usable markers**, **No-call rate** and **Measurement gap (max)**, below that **Suitability by analysis** with reasons and recommendations for each domain and, if there are any, **Quality findings & risks**. **Copy report as Markdown** puts the report on the clipboard as text (in the app language, with the **Threshold version**); **Done** closes the sheet. With  **(Picture of the window)** you can save the window together with the sheet as a picture.

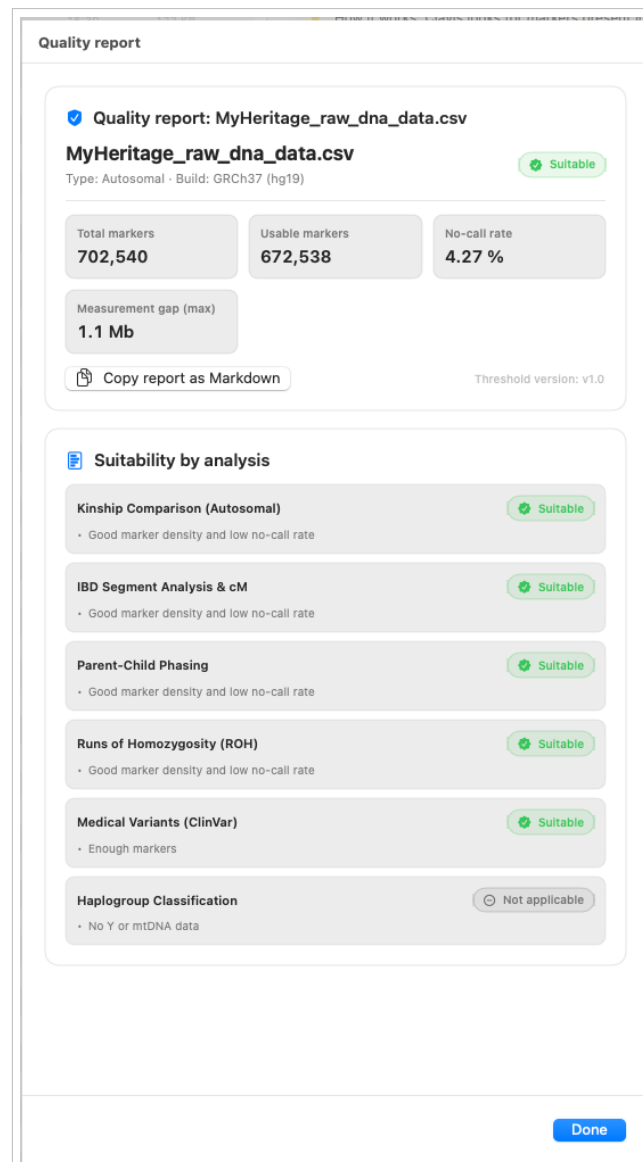


Figure 6.3: The quality report of a chip file, opened with “Open pre-check report”.

6.1.2 Autosomal comparison

Clavis looks for half-identical segments (at least 7 cM and 200 SNPs), fully identical regions and opposite homozygotes, and estimates the relationship from them.

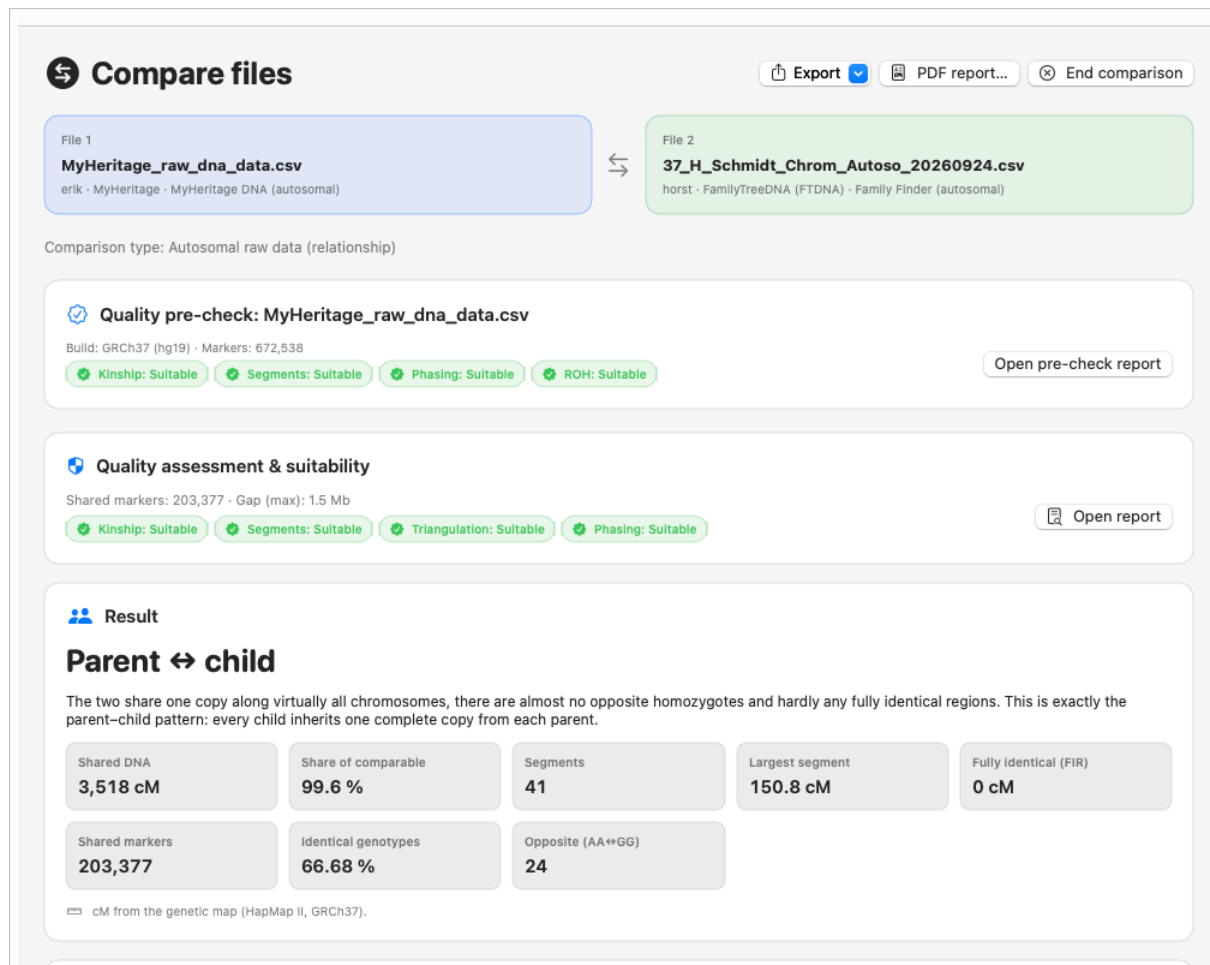


Figure 6.4: Autosomal comparison of two people: Quality pre-check, Quality assessment & suitability and the result “Parent ↔ child”.

Quality assessment & suitability

appears above the result and assesses the pair of files: **Shared markers**, the largest gap between two shared markers (**Gap (max)**) and the suitability for **Kinship**, **Segments**, **Triangulation** and **Phasing**, below that up to two findings (see below).

Result

the relationship estimate with an explanation and the key figures **Shared DNA**, **Share of comparable**, **Segments**, **Largest segment**, **Fully identical (FIR)** (around 25% for full siblings, almost zero for parent and child), **Shared markers**, **Identical genotypes** and **Opposite (AA↔GG)** (zero for parent and child, apart from measurement errors).

X chromosome

shared X cM, segments, the sex according to the data and which inheritance paths are possible. Some companies do not export the X chromosome.

Chromosome browser

blue: shared (one copy), green: fully identical (both copies).

Shared segments

table with chromosome, start, end, SNPs and cM; with allele frequencies loaded, the **Rare alleles** column, with chromosome bands loaded, the **Bands** column (a warning triangle marks segments at the centromere).

The relationship classes range from “Same person (or identical twins)” through “Parent ↔ child”, “Full siblings” and second- and third-degree relatives to “Third cousin (or similar)” and “Distant relationship possible”. If the files use different reference genomes, Clavis converts them with the liftover data or matches them by marker name. With the genetic map loaded, the cM values are more precise; below the result it says which method was used.

The values may differ by a few per cent from MyHeritage, FamilyTreeDNA or GEDmatch. The relationship is a statement of probability based on the Shared cM Project, not proof.

Quality of the pair of files. For **Segments**, **Triangulation** and **Phasing**: with fewer than 5,000 shared markers the pair is **Unsuitable**; with fewer than 100,000 shared markers, with an unconfirmed reference genome or with a gap of more than 15 Mb it is **Limited**. For **Kinship** the same limits apply, without the gap rule. As for a single file, the centromeres, heterochromatin and the short arms of the acrocentric chromosomes do not count towards the gap. With an unconfirmed reference genome, the report recommends checking the reference genome (GRCh37 or GRCh38) with the provider; in the multi-person comparison you can choose it for each file.

Open report shows the **Quality report: file comparison** sheet: the two file names and their reference genomes with the verdict **Compatible** or **Incompatible**, the key figures **Shared markers**, the share of each file, **Identical genotypes (IBS 2)**, **Opposite homozygotes (IBS 0)** and **Shared gap (max)**, below that the suitability for each domain and, if there are any, the **Findings & caveats**. A finding shows its severity as **Note**, **Warning** or **Critical**.

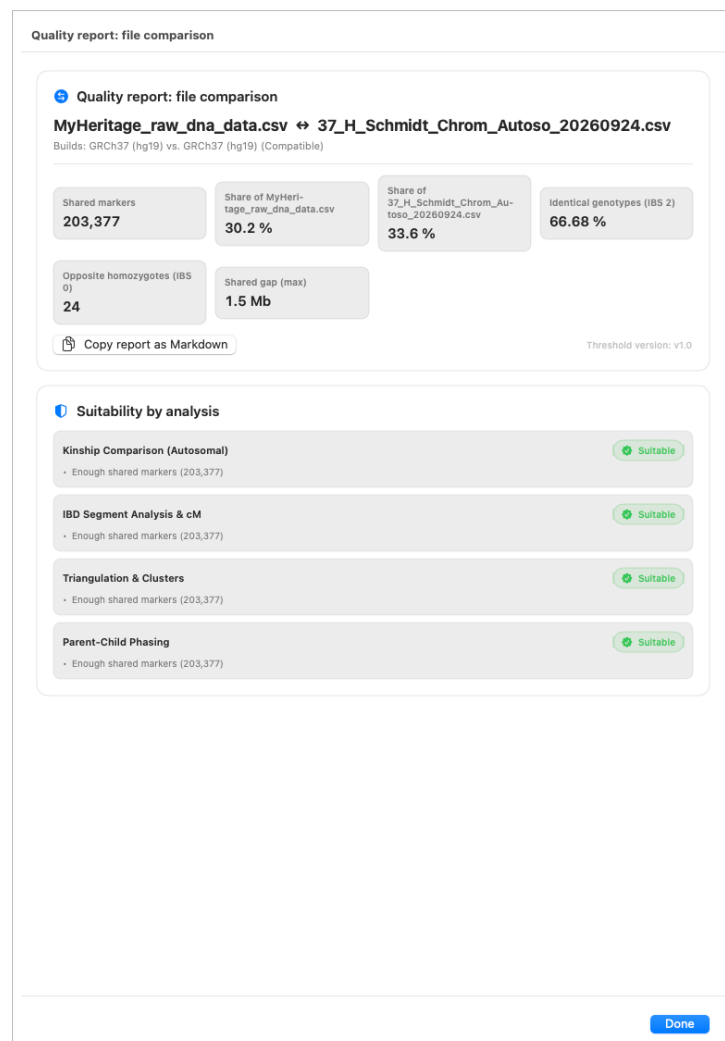


Figure 6.5: The “Quality report: file comparison” sheet, opened with “Open report”.

VCF files. From a VCF, only positions with FILTER “PASS” or “.” count and, where the file contains these values, QUAL of at least 20, read depth (DP) of at least 5 and genotype quality (GQ) of at least 15. Rows with multi-base or symbolic alleles are skipped. Only positions measured in both files are compared; a gVCF contributes its reference blocks. Two variants-only VCFs therefore usually share too few positions, and the result reads “Too few shared markers”. Below the result, a note says what the comparison is based on: for two variants-only VCFs only on the positions listed in both files (a rough guide), for a gVCF on its reference blocks, and, if only few positions were found at the same place, on matching by marker name (rs numbers).

6.1.3 mtDNA comparison

Both haplogroups, the **Genetic distance** (without hotspots and insertions, only in regions covered by both sequences), the shared mutations and the mutations that occur in only one file. A distance of 0 means an identical maternal line, 1–2 “very similar”.

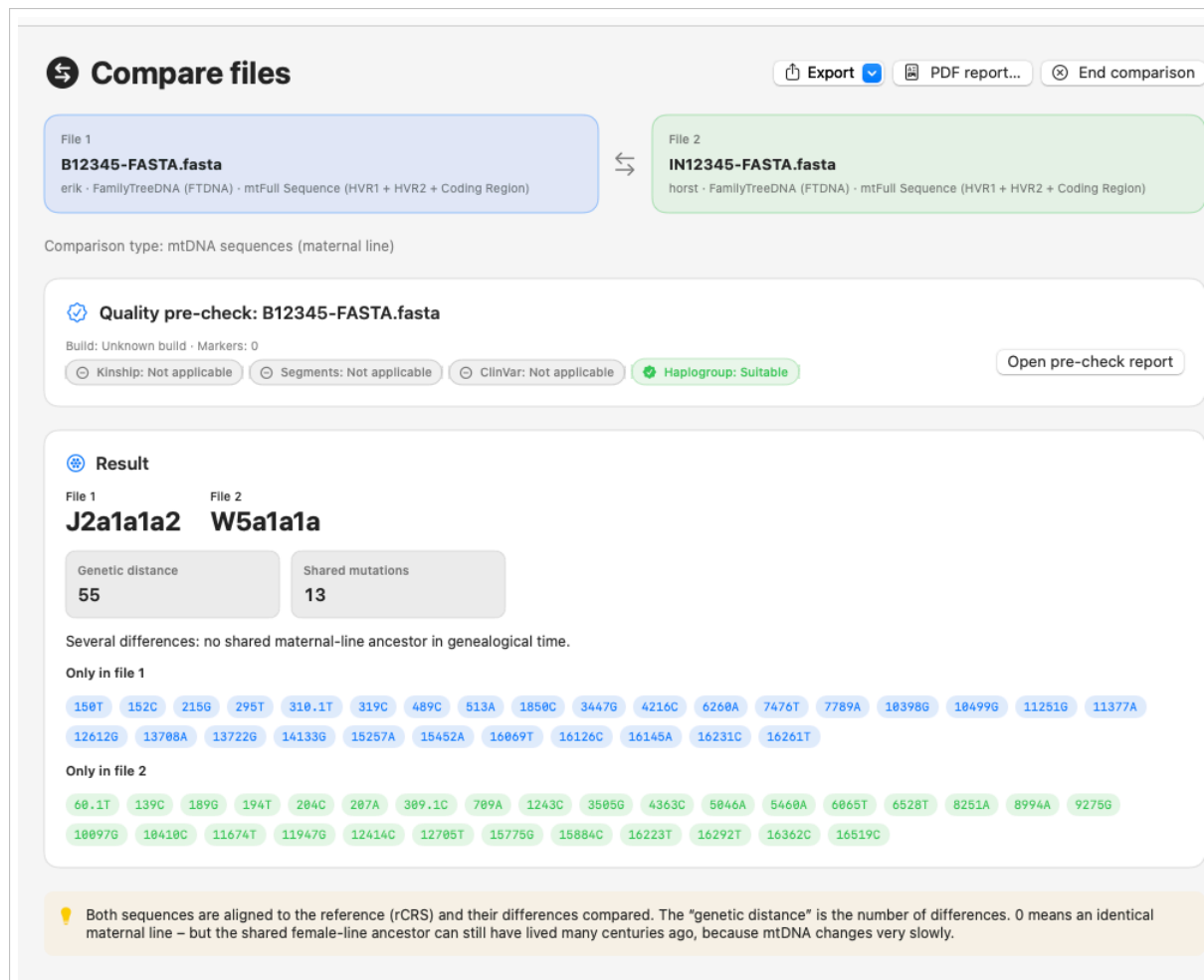


Figure 6.6: mtDNA comparison of two sequences.

6.1.4 Y-DNA comparisons

- **Y variants:** shared variants, variants in only one file and the haplogroups of both sides. Two lists of SNP names are matched by SNP name; “only in ...” then only counts SNPs that were explicitly tested negative.
- **Y-STR:** The first man in file 1 is compared with all men in file 2 (at most 50, sorted by distance). For each man: **Genetic distance**, **Markers compared**, the differences as “Marker: A → B” and the STR clock. Counting follows the FamilyTreeDNA rules (DYS389II without 389I; multi-copy markers by the infinite alleles model).
- **Y match lists:** matches new in file 2, no longer in file 2, and in both lists.

6.1.5 Export and PDF

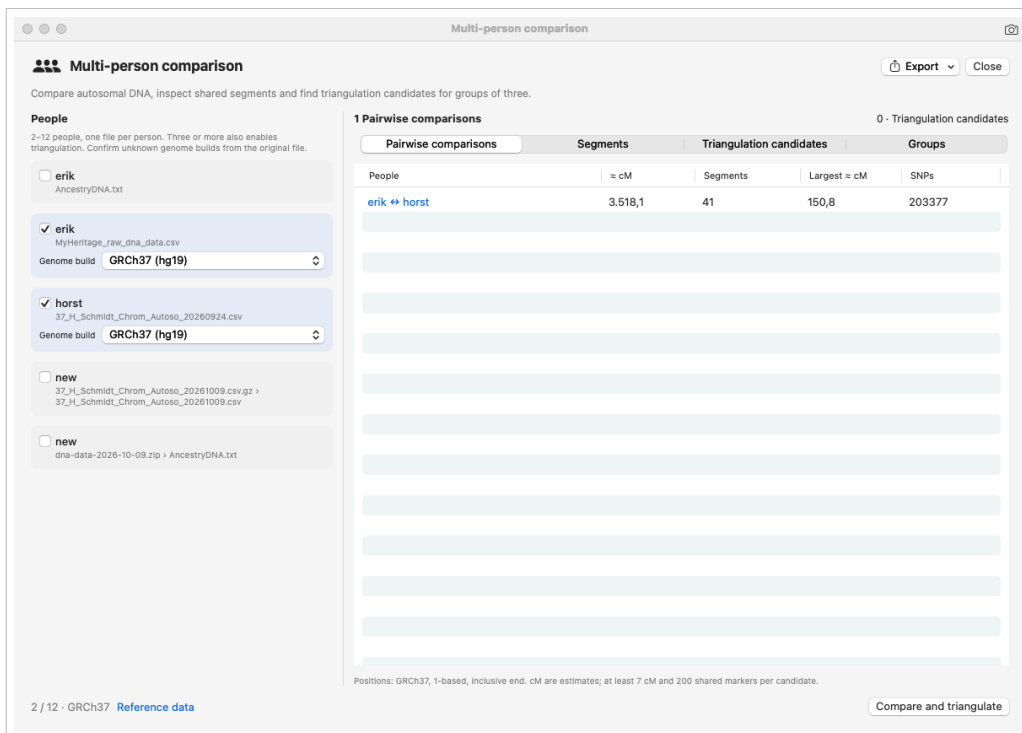
Once there is a result, **Export** saves the comparison results and the segments (or individual results) as CSV or TSV: UTF-8, English column names, decimal point, 1-based coordinates; cells beginning with = + - @ are protected against being read as formulas. **PDF report...** creates “Clavis-Vergleich.pdf” with the files, result, key figures and segment table.

For an autosomal comparison, both also contain the quality assessment: the CSV file has the rows `quality_thresholds_version`, `max_shared_gap_bp` and `suitability_...` for the four do-

mains (values suitable, limited, unsuitable, notApplicable). The PDF has a section “Quality assessment & suitability” with domain, status and reasoning, and a section on methods and reproducibility: an analysis ID (new for every report), a reproducibility hash (the same for unchanged files), the method, the time and, for each input file, its size, reference genome and SHA-256 checksum; with the genetic map, also the reference data used. For the checksums Clavis reads both files completely, so the PDF report takes a little while for large genome VCFs. These sections, too, are in the app language.

6.2 Multi-person comparison and triangulation

Compare › Multi-person comparison () compares 2 to 12 people at once, one autosomal file per person. On the left you choose the files and, for each file, the genome build (GRCh37 or GRCh38; unknown ones must be set). **Compare and triangulate** loads each file once, compares all pairs and, with three or more people, checks all groups of three.



The screenshot shows the 'Multi-person comparison' window. On the left, under 'People', there are four entries: 'erik' (AncestryDNA.txt), 'erik' (MyHeritage_raw_dna_data.csv) with genome build 'GRCh37 (hg19)', 'horst' (37_H_Schmidt_Chrom_Autoso_20260924.csv) with genome build 'GRCh37 (hg19)', and 'new' (37_H_Schmidt_Chrom_Autoso_20261009.csv.gz) with genome build 'GRCh37 (hg19)'. The 'Pairwise comparisons' tab is active, showing a table with columns: People, cM, Segments, Largest cM, and SNPs. The first row shows 'erik ↔ horst' with 3,518.1 cM, 41 segments, 150.8 largest cM, and 203,377 SNPs. Below the table, there are tabs for 'Segments' and 'Triangulation candidates'. At the bottom right, there is a 'Compare and triangulate' button.

Figure 6.7: Multi-person comparison with the “Pairwise comparisons” tab.

Pairwise comparisons

table of all pairs with cM, segments, largest segment and SNPs.

Segments

the full comparison view for a selected pair, with the **Quality assessment & suitability** card (section 6.1.2).

Triangulation candidates

regions shared by all three pairs of a group of three, where at every jointly tested marker one allele occurs in all three people (at least 7 cM, 200 markers, 20 markers per Mb). Unphased data give candidates, not proof.

Groups

groups from your own files: from the point of view of a selected person, with an ad-

justable upper limit (“Relatives up to”) and connection threshold (“Connected from”), as a coloured matrix.

Export saves pair results, segments and triangulation candidates as CSV or TSV. GRCh38 files are converted to GRCh37; this needs the liftover file. Chromosomes 1–22 are analysed.

6.3 Separating parental sides (phasing)

If one parent has tested, for many markers you can tell which allele the child got from which parent. Family > Separate parental sides (phasing) opens the window for this: choose the **Child** and the **Tested parent**, tick the other people under **Matches (other people)** and click **Determine sides**.

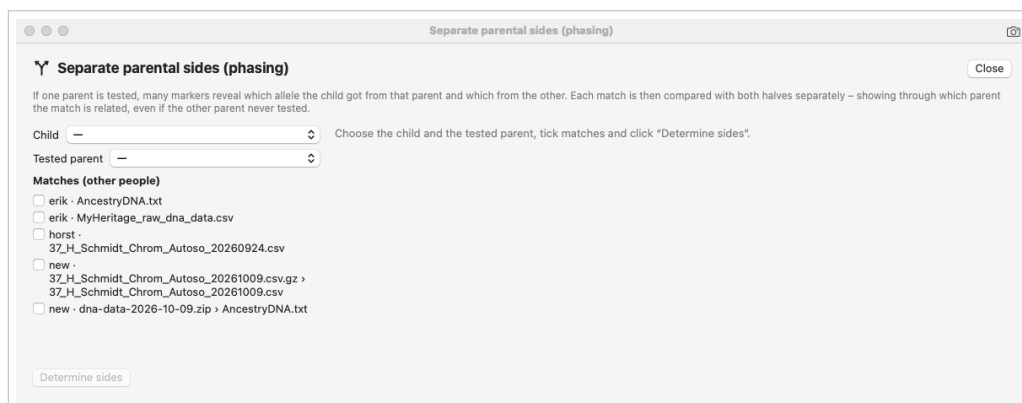


Figure 6.8: The “Separate parental sides (phasing)” window before making a selection.

Clavis first checks that child and parent really are a parent–child pair. The result shows the cM between the two, the assigned and unassignable markers and the rate of Mendelian conflicts (for real pairs usually well below 0.5%). Below that, a badge shows how suitable the pair is for phasing (**Suitable**, **Limited** or **Unsuitable**, following the rules in section 6.1.2); if the conflicts are above 1%, **Elevated error rate** follows in orange with the value; in that case, check the files and which one is the child and which the parent. The badge is based on the reference genomes of the two files. For each selected match, a table gives the cM on the tested parent’s side, on the other side and unphased, together with a graphic with two tracks.

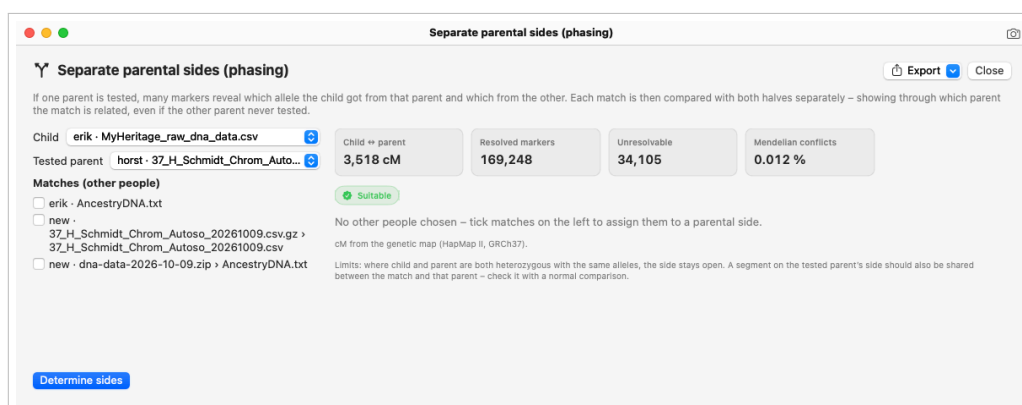


Figure 6.9: Separating parental sides: result for child and tested father, with no matches selected yet.

Before **Add to chromosome map** you state whether the tested parent is the paternal or the maternal one (... is the ...: **Paternal** or **Maternal**); the segments then go into the child's chromosome map with their side (section 8.5). **Export > Segments** saves the separated segments as CSV or TSV. Where child and parent are heterozygous in the same way, the side stays open.

Chapter 7

Match lists and segments

This chapter applies to the exports from **FamilyTreeDNA Family Finder**: the match list and the segment list from the chromosome browser. For companies that no longer export a match list (MyHeritage, AncestryDNA), there is **Matches by hand** (section 9.2).

Do not rename the files

Clavis links the match list, the segment list and the lists of relatives through the person folder and the file name: match lists contain “Family_Finder_Matches”, segment lists “Chromosome_Browser”. If you rename the exports, these links are lost. If there are several matching files, Clavis takes the newest one (by modification date).

7.1 The match list

When you select a person’s match list and load it, the **Autosomal match list** card appears.

Autosomal match list (42)

Over 200 cM: 5 · 40 to 200 cM: 4 · 20 to 40 cM: 12 · under 20 cM: 21
According to the provider paternal: 18 · maternal: 3 · both: 0 · not assigned: 21
Compared with Paul Beispiel: 17 matches are also in that list, and for 17 of them a segment is found at the same place.
Linked to the tree: 1 matches, 0 of them outside the expected range.
X matches: none – FamilyTreeDNA reports no shared X segment for any match of this list.
Segments from BSP0001_Chromosome_Browser_Results_Beispiel.csv

Filter (name, place, note...) Sort **Shared cM** Side **All sides**

Group **All groups** Overlapping segments

Compare with **Paul Beispiel** **all matches** Tree **all** Save as CSV

Name	Shared cM	Longest s...	Side	Segments	Compared person	Group	Tree
Paul Beispiel	3,373.0	265.6	paternal	22	not in list		—
Mira Muster	870.0	68.1	maternal	24	not in list		—
Clara Probe	450.0	47.4	paternal	15	15 of 15 segments	Segmentgruppe 1 (väterlic...	1st cousins on
Jonas Beispiel	230.0	40.2	paternal	9	9 of 9 segments	Segmentgruppe 1 (väterlic...	—
Theo Vorlage	210.0	37.8	maternal	8	not in list		—
Ida Vorlage	180.0	34.6	maternal	7	not in list		—
Rike Entwurf	120.0	30.5	—	5	not in list		—
Henrik Probe	75.0	27.9	paternal	3	3 of 3 segments	Segmentgruppe 1 (väterlic...	—
Malte Beispiel	60.0	23.1	paternal	3	3 of 3 segments	Segmentgruppe 1 (väterlic...	—
Enno Fernver...	38.1	21.0	—	2	not in list		—
Thea Fernver...	37.3	20.1	—	2	not in list		—

Figure 7.1: Anna’s match list from the sample family, compared with her father’s list.

7.1.1 The summary

Above the table you see

- the distribution by shared DNA: over 200 cM, 40 to 200 cM, 20 to 40 cM, under 20 cM,
- the sides as given by the company (paternal, maternal, both, not assigned),
- with a comparison person selected, how many matches also appear in that person's list and for how many a segment lies at the same place,
- with a family tree, how many matches are linked and how many lie outside the expected range,
- the X matches (also "none") and, with a family tree, how many of them fit X inheritance,
- with an older list of the same person, how many matches are new and how many have disappeared,
- which segment list the segments come from – and, in orange, whether names in the segment list are not assigned to any match.

7.1.2 Filtering and sorting

Search field

filters by name, the match's surnames, the company's estimate and your own note.

Sort by shared cM, longest segment, match date or name (forms of address such as "Mrs." are ignored).

Side all sides, paternal, maternal, both, not assigned or **X matches only**.

Group

only matches in a group you have assigned yourself.

Compare with

another person in the data folder with a segment list, plus a filter: all matches, "also in that person's list", "segment at the same place", "not in that person's list".

Tree all, linked, not linked, outside the range.

new only

only matches added since the older list.

Click a column header to sort the table; you can drag column widths. The columns are Name, Shared cM, Longest segment, Side, Segments, Compared person, Group, Tree, Provider's estimate and Match date. **Save as CSV** saves the list as currently shown, with English column headers.

7.1.3 A match in detail

Click a match to see its details below:

- key figures (shared cM, longest segment, side, company's estimate, X match, haplogroups),
- **Possible relationships according to the Shared cM Project**,
- **In the family tree** (see below),
- the surnames from the match's family tree,

- **Shared segments** with chromosome, start, end, cM and SNPs; with a comparison person selected, “found again” or “not found” for each segment,
- the fields **Contact** and **Group** (with **Apply**) and **My notes** (with **Save note**).

The screenshot displays the Clavis Genetica interface. At the top, there's a navigation bar with tabs for 'Compare with', 'Tested person', 'My matches', and 'Tree'. Below this is a table listing matches. The match 'Clara Probe' is highlighted in blue. Below the table, the detailed view for 'Clara Probe' is shown. It includes summary statistics: Shared cM (450.0 cM), Longest segment (47.4 cM), Side (paternal), and Provider's estimate (1st Cousin - 2nd Cousin). Below this, there's a section for 'Possible relationships according to the Shared cM Project' and 'In the family tree'. A link 'Clara Probe (1962-)' is shown. Below that, a paragraph explains the relationship fit. A table of 'Shared segments (15)' follows, listing chromosomes, start/end positions, cM, SNPs, and whether the segment was 'found again'. At the bottom, there's a 'Contact' field and an 'Apply' button.

Name	Shared cM	Longest s...	Side	Segments	Compared person	Group	Tree
Paul Beispiel	3,373.0	265.6	paternal	22	not in list		—
Mira Muster	870.0	68.1	maternal	24	not in list		—
Clara Probe	450.0	47.4	paternal	15	15 of 15 segments	Segmentgruppe 1 (väterlic... 1st cousins	
Jonas Beispiel	230.0	40.2	paternal	9	9 of 9 segments	Segmentgruppe 1 (väterlic...	—
Theo Vorlage	210.0	37.8	maternal	8	not in list		—
Ida Vorlage	180.0	34.6	maternal	7	not in list		—
Rike Entwurf	120.0	30.5	—	5	not in list		—
Henrik Probe	75.0	27.9	paternal	3	3 of 3 segments	Segmentgruppe 1 (väterlic...	—
Malte Beispiel	60.0	23.1	paternal	3	3 of 3 segments	Segmentgruppe 1 (väterlic...	—
Enno Fernver...	38.1	21.0	—	2	not in list		—
Thea Fernver...	37.3	20.1	—	2	not in list		—

Clara Probe

Shared cM: **450.0 cM** | Longest segment: **47.4 cM** | Side: **paternal** | Provider's estimate: **1st Cousin - 2nd Cousin** | mtDNA: **T2b**

Possible relationships according to the Shared cM Project
 Half 1st cousins · 1st cousins once removed · 2nd cousins · 1st cousins twice removed · 1x great-aunt/-uncle · 1st cousins

In the family tree
 🔗 **Clara Probe (1962-)** [Unlink](#)

According to the tree: 1st cousins once removed. Expected on average 433 cM, observed range 102 to 980 cM. Measured: 450 cM. This fits the relationship but does not prove it yet, because other relationships have this range too. Common ancestors: Hermann Probe (1901–1968), Bertha Platzhalter (1905–1990). Side according to the tree: paternal.

Surnames from the match's tree
 Probe / Platzhalter

Shared segments (15)

Chromosome	Start	End	cM	SNPs	Paul Beispiel
2	150,169,001	185,495,545	36.5	8422	found again
4	61,379,276	78,341,068	18.1	4171	found again
6	91,481,901	106,937,127	17.0	3925	found again
8	59,502,613	87,274,642	30.8	7110	found again
9	53,387,352	94,114,760	45.9	10599	found again
10	64,376,977	78,109,164	17.2	3970	found again
10	93,802,542	130,222,245	45.5	10520	found again
11	60,202,495	101,561,598	47.4	10953	found again
13	12,252,817	30,310,574	19.7	4558	found again
13	92,200,491	105,835,272	14.9	3452	found again
16	6,335,235	25,961,286	28.4	6563	found again
17	15,558,201	36,525,096	33.1	7647	found again
18	59,162,124	73,686,277	21.9	5062	found again
20	31,484,284	56,794,944	43.4	10034	found again
21	14,597,561	37,932,158	30.1	6958	found again

Three statements must be kept apart. A match also being in the compared person's list means the match may be related through that person. That person sharing a segment with the match at the same place supports your segment coming from that side. If a segment is not found, it may come from the other side, be false, or simply lie below the display threshold for the compared person. The comparison is made by the match's name.

Contact: [Apply](#)

Figure 7.2: The match “Clara Probe” in detail, linked to the family tree.

Notes, contacts and groups are only saved locally, not in the DNA file. A note is saved when you click **Save note** or switch to another match – not if you close the window before that.

7.1.4 Linking matches with the family tree

If a family tree is loaded and the tested person is linked to their entry in the **People** window (chapter 8), you can assign each match to a person in the tree: through a name suggestion (**Same name in the tree**) or through the search field **Search person in tree....** Clavis then calculates the relationship according to the tree, the expected cM with range, the common ancestors and the

side, and assesses whether the measured cM fit. If the company names a different parental side from the tree, Clavis warns you.

For **X matches** Clavis also checks whether X-DNA can reach both of you at all through the common ancestors – a man inherits his X only from his mother. The verdict is “can”, “cannot” or “unclear” (if the sex of people in the tree is missing). If an X match is not linked, **Show X ancestors** shows which ancestors lie on your X lines.

7.1.5 Comparing with a relative's list

If a second person has tested – ideally a parent –, **Compare with** shows whether your matches also appear in that person's list and whether a segment is found again at the same place (at least 1 Mb overlap).

Keep three statements apart: The match is also in the comparison person's list. The comparison person shares a segment with the match at the same place. Only from this does it follow that your segment comes from that side. The comparison works by the match's name.

7.2 The segment list and name matching

After loading, the segment list (export from the chromosome browser) shows the distribution of the segments by size (15 and over, 10 to 15, 7 to 10, under 7 cM), the **Name matching with the match list** box, search, sorting, the **Segments from** filter and the table with name, chromosome, start, end, cM and SNPs.

Segment list (134)

From 15 cM: 108 · 10 to 15 cM: 17 · 7 to 10 cM: 9 · under 7 cM: 0 · from 42 matches

Name matching with the match list

42 names in the segment list, 42 of them joined with a match of the match list.
 Compared with: BSP0001_Family_Finder_Matches_Beispiel.csv
☒ All names are joined.

Filter (name, place, note...) Sort **Position** Segments from — Overlapping segments

Name	Chromosome	Start	End	cM	SNPs
Paul Beispiel	1	746,266	248,429,320	265.6	61354
Fenja Fernverwandt	1	57,054,695	79,528,489	24.1	5568
Mira Muster	1	91,718,242	130,147,153	41.2	9516
Mira Muster	1	186,275,232	224,889,731	41.3	9543
Thea Fernverwandt	1	212,662,693	228,675,902	17.2	3965
Paul Beispiel	2	742,975	242,452,156	249.7	57670
Ida Vorlage	2	90,930,880	101,763,190	11.2	2581
Clara Probe	2	150,169,001	185,495,545	36.5	8422
Mira Muster	2	160,355,342	190,809,465	31.5	7277
Paul Beispiel	3	718,262	197,411,420	217.6	50256
Mira Muster	3	6,717,232	43,159,993	40.3	9305
Peer Fernverwandt	3	31,823,124	50,110,977	20.2	4665
Rieke Fernverwandt	3	46,240,590	60,817,403	16.2	3743
Quirin Fernverwandt	3	58,461,997	69,623,347	12.3	2851

Every row is a segment you share with a match. As a working rule: from 15 cM very probably genuine, 10 to 15 cM probably genuine, 7 to 10 cM uncertain and only usable together with further evidence, below that do not use. These are rules of thumb, not fixed limits.

Figure 7.3: Anna's segment list with the “Name matching with the match list” box.

Rule of thumb for size: 15 cM and over very likely genuine, 10–15 cM probably genuine, 7–10 cM uncertain, below that do not use.

7.2.1 Name matching

The FamilyTreeDNA segment list only gives the match's name. Clavis assigns each name to a match in the same person's match list. It uses first, middle and last names; forms of address ("Mr.", "Mrs."), titles and suffixes do not get in the way, and a comma in a name is treated as a hyphen, as FamilyTreeDNA writes it in the segment list.

The box states how many names are assigned. If names are left over, ... **names without an entry in the match list – assign** lists them; with **Assign** you pick the right one from similar matches. Names assigned by hand appear under ... **names assigned by hand** and can be reversed with **Undo**. Names that belong to several matches (same name) are listed separately by the box; their segments cannot be told apart. Assignments take effect immediately in the match list, in the comparison and in the overlapping segments.

7.3 Overlapping segments

The **Overlapping segments** button in the match list or segment list opens a window with two parts: **Places** and **Map and groups**.

7.3.1 Places

Clavis looks for places where the segments of several matches overlap (common stretch at least 1 Mb). You can set **Segments from** (default 12 cM), **Matches per place at least** (default 3), **Hide pile-ups** and **Count parents, children and siblings** (otherwise matches of 2,300 cM or more are left out). A **pile-up** (10 or more matches) rather points to an old, widespread segment.

Overlapping segments – Anna Beispiel

Places Map and groups

Segments from 12 cM Matches per place at least 3 ☐ Hide pile-ups ☐ Count parents, children and siblings

12 places where the segments of several matches overlap, 0 of them pile-ups. Evaluated: 93 segments.

Not counted: 1 match(es) of 2300 cM or more (parents, children, siblings), because they share DNA almost everywhere.

Chromosome	Common stretch	Matches	Longest segment	Side
4	30.5 – 52.3 Mb	3	51.2 cM	paternal: 2 - maternal: 1
7	117.8 – 121.7 Mb	6	68.1 cM	maternal: 3 - not assigned: 3
8	77.1 – 87.3 Mb	3	30.8 cM	paternal: 2 - not assigned: 1
11	54.3 – 58.7 Mb	3	30.0 cM	paternal: 2 - maternal: 1
11	60.2 – 69.0 Mb	3	47.4 cM	paternal: 2 - maternal: 1
11	68.3 – 71.0 Mb	3	47.4 cM	paternal: 2 - maternal: 1
11	72.4 – 88.1 Mb	3	47.4 cM	paternal: 3

Chromosome 7, 117.8 to 121.7 Mb: 6 matches

Name	cM	Start	End	Side
Mira Muster	68.1	80,457,862	140,242,779	maternal
Theo Vorlage	37.8	95,200,347	128,446,296	maternal
Ida Vorlage	34.6	100,647,626	131,044,845	maternal
Rike Entwurf	24.1	104,110,138	125,358,823	—
Quirin Fernverwandt	23.1	117,837,229	138,030,042	—

Next step: check at the provider whether these matches also match one another at this place. Only then is it a triangulation, and even that names no ancestor.

This list shows candidates, not evidence. The segment list only says that you share a segment with each of these matches at the same place. It does not say whether the matches match one another, and two of them can lie on different copies of the chromosome, one paternal and one maternal. The side comes from the provider. Short segments are often false, which is why the list only counts from the chosen length.

Figure 7.4: Overlapping segments: the place on chromosome 7 with six matches.

Click a place to see its matches with cM, position and side. If the company assigns them to different sides, there are at least two separate groups. **New check sheet** creates a pre-filled segment check sheet in the research log.

The list shows candidates, not evidence. Two matches at the same place may lie on different copies of your chromosome, one paternal and one maternal. Only if they also match each other at this place is it a triangulation.

7.3.2 Map and groups (segment groups)

This part uses the segment lists to form **groups of matches that lie at the same place on the same copy** – a complement to the companies' clusters, built from your own data. What matters is the **side** of each segment. Clavis determines it, strongest source first:

1. **A parent's data:** If you have chosen a tested person with a segment list under **Father** or **Mother**, and that person shares a segment with the match at the same place, your segment lies on that person's side.
2. **Exclusion:** If only one parent has tested and a match of 20 cM or more is completely missing from that parent's list, it counts as a match on the other side.
3. **Company:** the side that FamilyTreeDNA gives based on linked relatives (**Use the provider's side**).

Two matches are joined if their segments overlap by at least 1 Mb, the overlap does not lie in a pile-up, and both lie on the same copy. **Also join segments of unknown side** forms groups even without parental data, but these may then mix two lines. **Segments from** (default 10 cM) sets the minimum size.

The **Map** view shows all chromosomes: in grey, how many matches share a segment with you at a place; in orange the pile-ups; coloured bars the places of the groups. A click gives the number of matches at that place and highlights the group.

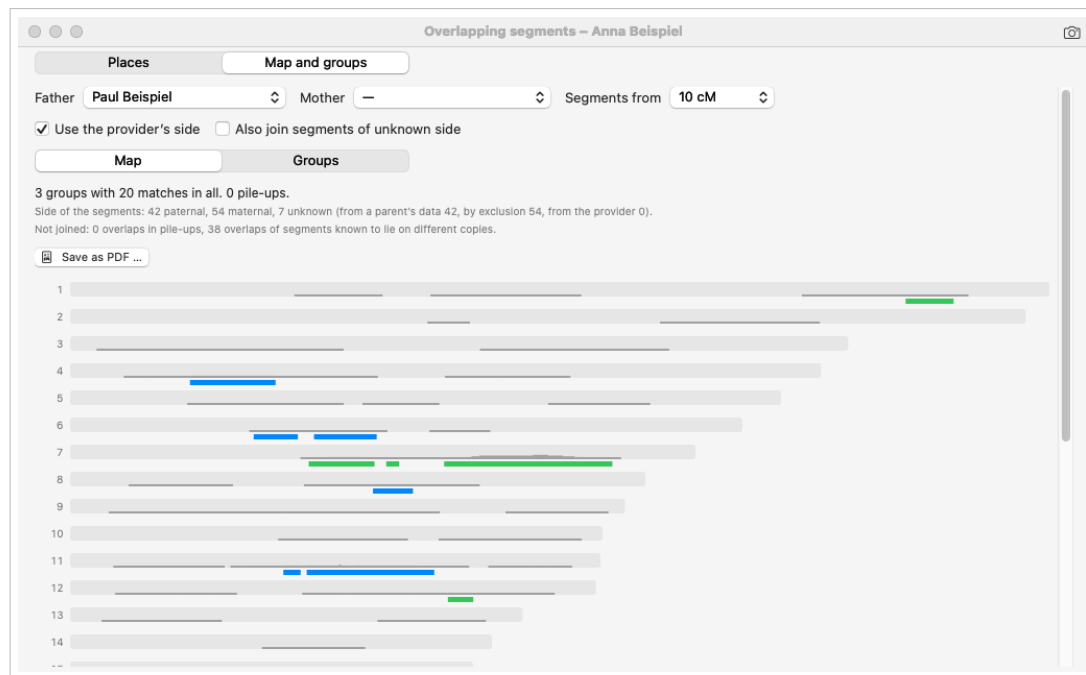


Figure 7.5: Map and groups, “Map” view, with Paul as the tested father.

The **Groups** view lists each group with its colour, copy (paternal, maternal, unknown), number of matches and its places. A click opens the group: its members with cM and side, including where the side comes from, and – with a family tree – how many members are linked, which common ancestors they share with you and through which **common line** the group’s segment would have come down. If the family tree and the segment data contradict each other, Clavis warns you.

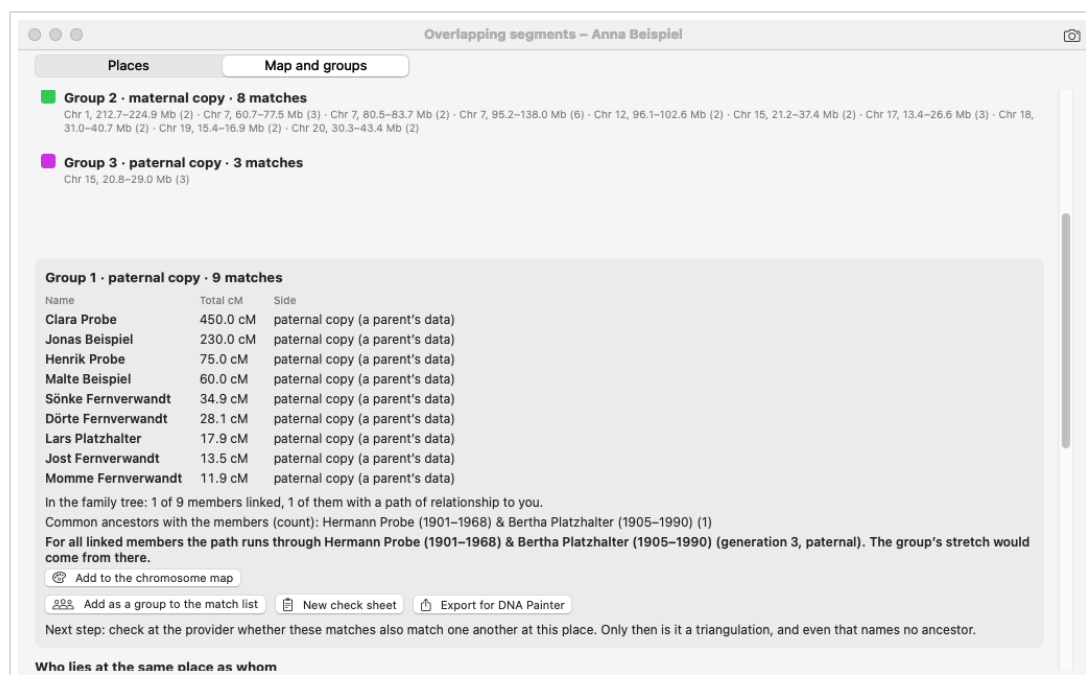


Figure 7.6: Group 1 in detail: members and the common line according to the family tree.

Each group has these buttons:

Add to the chromosome map

adds the group's segments to your chromosome map – under this ancestral couple if there is a common line, otherwise without an ancestor but with the group's side.

Add as a group to the match list

writes “Segment group n” as a group into the match list, but only for matches that do not have a group yet.

New check sheet

creates a pre-filled check sheet in the research log.

Export for DNA Painter

saves the group's segments as CSV in DNA Painter's custom format; **Export all groups for DNA Painter** does this for all groups.

Overlapping segments - Anna Beispiel

Places Map and groups

UNT 15, 20.8-29.0 MD (3)

Group 1 - paternal copy - 9 matches

Name	Total cM	Side
Clara Probe	450.0 cM	paternal copy (a parent's data)
Jonas Beispiel	230.0 cM	paternal copy (a parent's data)
Henrik Probe	75.0 cM	paternal copy (a parent's data)
Malte Beispiel	60.0 cM	paternal copy (a parent's data)
Sönke Fernverwandt	34.9 cM	paternal copy (a parent's data)
Dörte Fernverwandt	28.1 cM	paternal copy (a parent's data)
Lars Platzhalter	17.9 cM	paternal copy (a parent's data)
Jost Fernverwandt	13.5 cM	paternal copy (a parent's data)
Momme Fernverwandt	11.9 cM	paternal copy (a parent's data)

In the family tree: 1 of 9 members linked, 1 of them with a path of relationship to you.
 Common ancestors with the members (count): Hermann Probe (1901–1968) & Bertha Platzhalter (1905–1990) (1)
 For all linked members the path runs through Hermann Probe (1901–1968) & Bertha Platzhalter (1905–1990) (generation 3, paternal). The group's stretch would come from there.

Add to the chromosome map Add as a group to the match list New check sheet Export for DNA Painter

0 matches entered as "Segment group 1 (paternal copy)"; 9 already had a group and were left unchanged.
 0 new segments added to the chromosome map under "Hermann Probe (1901–1968) & Bertha Platzhalter (1905–1990)".
 Next step: check at the provider whether these matches also match one another at this place. Only then is it a triangulation, and even that names no ancestor.

Who lies at the same place as whom

	Clara Probe	Jonas Beispiel	Henrik Probe
Clara Probe	■	■	■
Jonas Beispiel	■	■	■
Henrik Probe	■	■	■

Figure 7.7: The buttons of a group and the message after adding it to the chromosome map.

Below the groups, the matrix **Who lies at the same place as whom** shows which matches are joined: coloured within a group on the same copy, dark grey the same between two groups, light grey the same place with unknown side. Click a cell to see the place. Above 250 matches the matrix is not drawn in the window.

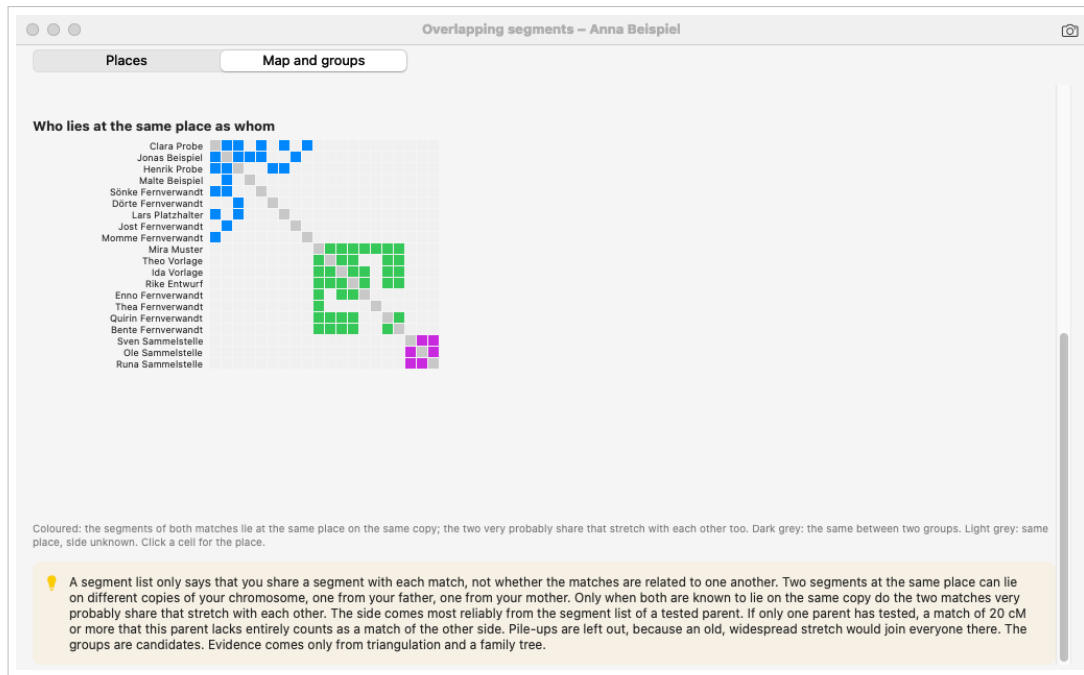


Figure 7.8: The “Who lies at the same place as whom” matrix.

Save as PDF ... saves the whole page as a PDF: settings, summary, map, each group with its members, their segments at the group’s places and the family tree details, plus the complete matrix on a page of its own (up to 500 matches).

The groups are candidates. Proof only comes from triangulation and a family tree.

Chapter 8

Family and family tree

8.1 People

The **People** window (toolbar or Family > People, ) lists all person folders on the left. On the right are the cards of the selected person.

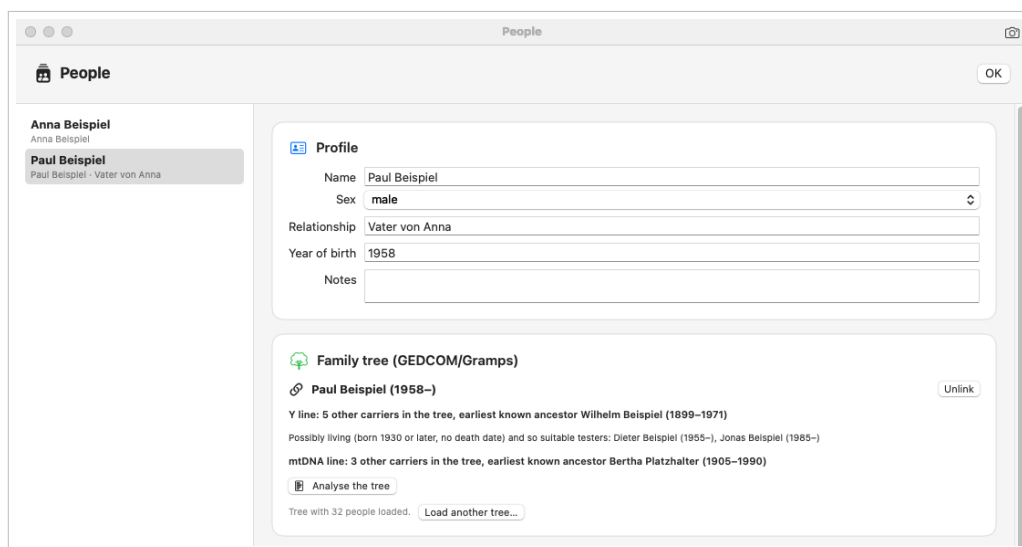


Figure 8.1: The “People” window with Paul Beispiel, linked to the family tree.

Profile

Name, sex, relationship (e.g. “father of Anna”), year of birth and notes. Everything is saved immediately. The sex controls test suggestions and X analyses. If it is not set and the person has a Y-DNA or Y-STR file, Clavis enters **male**.

Family tree (GEDCOM/Gramps)

Load a family tree and link the person to their entry (see below).

Which family lines are tested?

A small family tree over three generations: Y (blue) only follows the line father → father’s father, mt (pink) only the purely maternal line, autosomal DNA (green frame) covers all ancestors, with less from each generation. See below.

Who should test next?

Suggestions based on existing tests and sex, such as a Y test for a man of the paternal line or a brother of the mother.

Family tree ↔ DNA

Comparison between the family tree and the measured DNA (see below).

Family report (PDF)

See section 8.6.

8.1.1 Which family lines are tested?

The chart shows the tested person, father and mother and the four grandparents (**Grandfather (p)**, **Grandmother (p)**, **Grandfather (m)**, **Grandmother (m)**; p = paternal, m = maternal). Each box carries the badges of the tests that cover this line: “Y”, “mt” and “at” (autosomal). A blue seal marks a person who has tested directly.

If the data folder contains more than one person, there is a switch above the chart:

Own tests only

only the files of the selected person: their Y test covers the line of the father and of the grandfather (p), their mtDNA test that of the mother and of the grandmother (m).

Combined family

(default) also the tests of relatives in the same data folder. Clavis recognises father, mother and grandparents among the other people from the family tree (if both people are linked), from the **Relationship** field in the profile, from the FamilyTreeDNA match list of the selected person or from a folder name such as “father” or “mother”. In the relative’s **Relationship** field, “father” is enough; if it names someone (“father of Anna”), that must be the selected person. If the selected person’s field says “daughter of Paul”, Paul counts as the father, provided his profile states his sex (for a man a Y test is also enough). The words are also understood in German, French and Dutch (for example “Vater”, “père”, “vader”). From the match list, the “Linked Relationship” column counts with its direction: “Father” makes the match the father, “Son” or “Daughter” never makes it an ancestor. A parent–child range without a direction is only enough if the years of birth do not speak against it and the relative’s sex is known.

If the father is recognised, his box gets the seal, his name and his tests. His mtDNA test covers his mother: **Grandmother (p)** gets “mt” with the addition “mt via” and his name. If no mother is recognised, **Grandmother (m)** carries “mt via” with the name of the tested person, provided that person has an mtDNA test. Below the chart there is one line for each recognised relative with name, role, their tests and which line they cover. Match lists do not count as tests here: if a recognised relative has no tests of their own in the folder, their line reads “no own DNA tests in the folder (match lists do not count) – covers no line yet”. The choice of the switch is not saved.

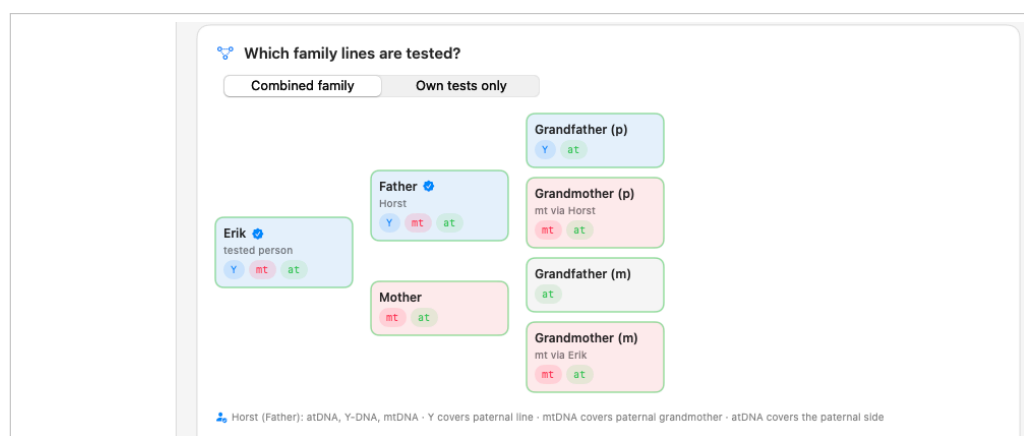


Figure 8.2: “Combined family”: the tested father covers the paternal line with his Y test and the line of his mother with his mtDNA test.

8.2 Loading and linking a family tree

Import tree (GEDCOM, Gramps)... loads a family tree in GEDCOM 5.5 or Gramps XML format. Clavis reads people, sex, year of birth and death and the parent-child relationships, and makes a private copy; there is one family tree per data folder. **Load another tree...** replaces it.

In the field **Search person in tree...** you search for the person and click the result. Empty profile fields (name, sex, year of birth) are filled in from the family tree. **Unlink** removes the link. The card then names the other carriers of the Y line (for men) and of the mtDNA line with the earliest known ancestor, and lists relatives who may still be alive (born 1930 or later, no date of death) and could therefore be asked to test.

8.3 Family tree ↔ DNA

If at least two people are linked, **Compare** checks, for each pair, the relationship according to the family tree against the measured cM from their autosomal raw data. The expected values come from the Shared cM Project 4.0. The verdict is “Consistent with the tree.”, “Less than expected – check the tree.” or “More than expected – closer or double relationship?”. Without raw data the measured value stays empty.

People

Anna Beispiel
Anna Beispiel

Paul Beispiel
Paul Beispiel - Vater von Anna

Relationship: e.g. father of Erik

Year of birth: 1988

Notes:

Family tree (GEDCOM/Gramps)

Anna Beispiel (1988-) Unlink

mtDNA line: 5 other carriers in the tree, earliest known ancestor Klara Notiz (1903-1988)

Possibly living (born 1930 or later, no death date) and so suitable testers: Marta Muster (1960-), Jürgen Muster (1957-)

Analyse the tree

Tree with 32 people loaded. Load another tree...

Which family lines are tested?

Combined family Own tests only

Anna Beispiel (tested person)

Father: Paul Beispiel

Mother:

Grandfather (p)

Grandmother (p)

Grandfather (m)

Grandmother (m)

Paul Beispiel (Father): no own DNA tests in the folder (match lists do not count) - covers no line yet

Y (blue) follows the father -> father's father line. mt (pink) follows the purely maternal line: the father's mtDNA test covers his mother (paternal grandmother). Autosomal DNA (green frame) covers all ancestors.

Who should test next?

- Take an autosomal test (e.g. MyHeritage, FTDNA Family Finder, AncestryDNA) - it finds relatives on all lines.
- Women have no Y chromosome: ask your father, a brother or a paternal uncle to take a Y test.
- Take an mtDNA test (mtFull Sequence) - it describes the direct maternal line.
- Have parents or grandparents tested while it is possible: their DNA reaches one generation further back and helps assign matches to the right side of the family.

Family tree $\leftrightarrow$ DNA

Compares, for all linked people, the relationship in the tree with the measured shared cM. Expected values: Shared cM Project 4.0 (Blaine Bettinger), rounded. A value outside the range points to an error in the tree or an unknown parentage - or to a second relationship.

Compare

Anna Beispiel $\leftrightarrow$ Paul Beispiel Parent/child 3,485 cM (2,376-3,720) —

Family report (PDF) Easy-to-read summary without raw data - to share with the family.

Figure 8.3: Anna Beispiel with “Family tree $\leftrightarrow$ DNA”: according to the tree, Anna and Paul are parent and child.

8.4 Analysing the family tree

Family > Analyse the tree () analyses the family tree from the point of view of a linked person, over 6 to 12 generations:

- **How complete is the tree?** – for each generation the possible and known ancestors, the share, distinct people, pedigree collapse and how many X paths are known. In practice autosomal DNA reaches back about five to seven generations; that is where filling gaps pays off.
- **Pedigree collapse** – ancestors who appear more than once, with their Ahnentafel numbers.
- **Who the X chromosome can come from** – the known ancestors up to the 6th generation from whom X-DNA can come. A man has his X only from his mother.

If no family tree is loaded or no person is linked, a note explains what is missing and takes you to the **People** window.

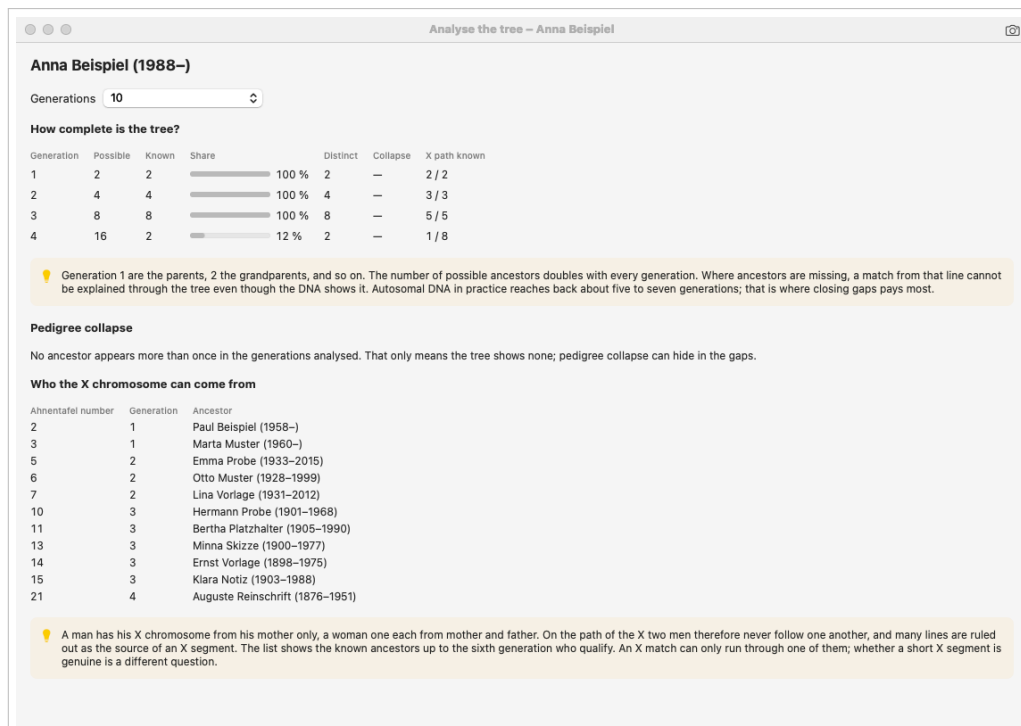


Figure 8.4: Analyse the tree for Anna Beispiel.

8.5 The chromosome map

Everyone has two copies of each chromosome: one from the father, one from the mother. The **Chromosome map** (Family > Chromosome map, ) records which segment comes from which ancestor – one map per person.

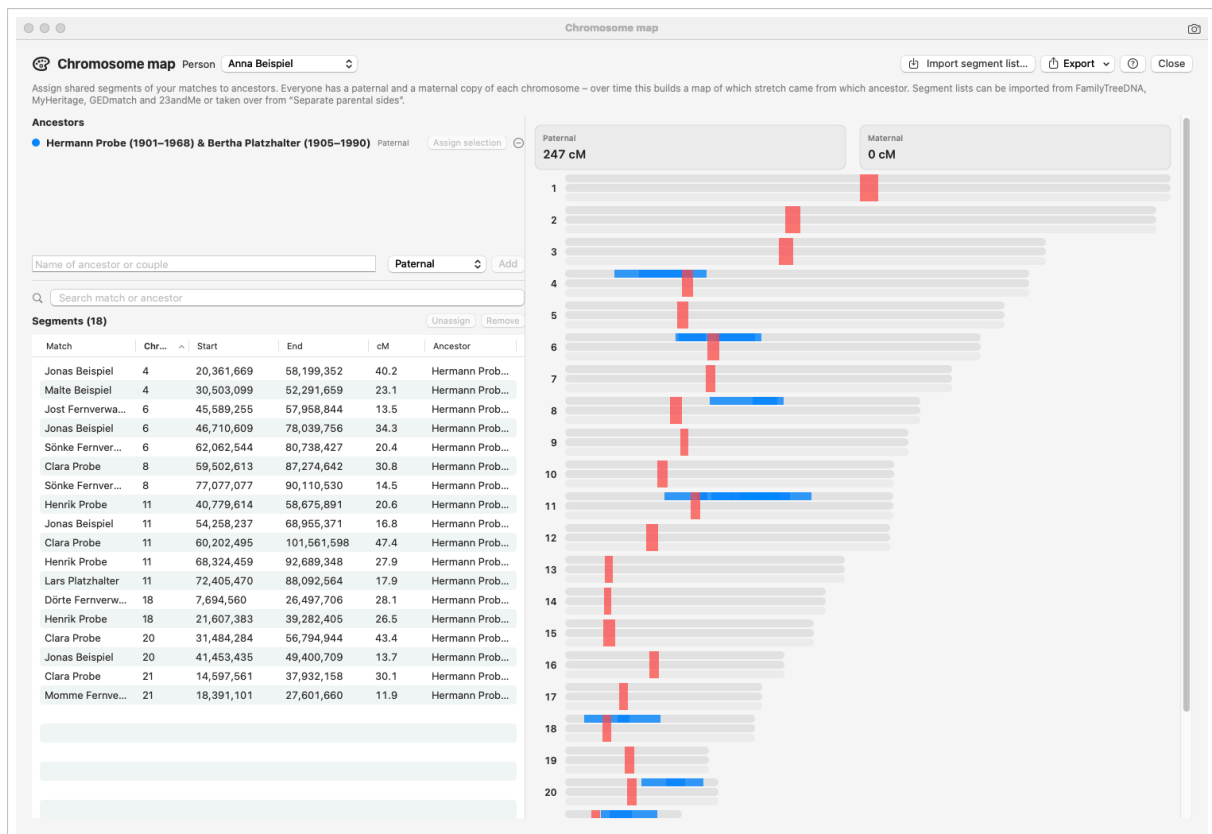


Figure 8.5: Anna's chromosome map with the segments of group 1 under the ancestral couple Hermann Probe & Bertha Platzhalter.

8.5.1 Layout

On the left are the **Ancestors** with colour and side and below them the table of **Segments** (match, chromosome, start, end, cM, ancestor). On the right the map shows chromosomes 1–22 and X with three tracks: at the top the paternal copy, in the middle the maternal one, at the bottom segments of unknown side. The colour depends on the ancestor; unassigned segments are grey. With chromosome bands loaded, the centromere is marked in red. At the top you see the painted cM for each side.

8.5.2 Step by step

1. **Get segments:** **Import segment list...** reads segment exports from FamilyTreeDNA (chromosome browser), MyHeritage, GEDmatch and 23andMe; duplicates are skipped. Comma, semicolon and tab are recognised as separators, as are the character sets UTF-8, UTF-16 and Latin-1 – so even a table from a German-language spreadsheet can be read in. Other sources: **Add to chromosome map** in phasing (section 6.3) and **Add to the chromosome map** in the segment groups (section 7.3.2).
2. **Create ancestors:** enter a name (person or couple), choose the side, click **Add**. Each ancestor gets a colour of their own.
3. **Assign:** select rows in the table (⌘ or ⌘ for several), then click **Assign selection** next to the ancestor. **Unassign** and **Remove** act on the selection.

4. Search: the search field filters by match or ancestor and highlights the results on the map.

Click a place on the map to see which segments lie there, with **Select in table** and **Zoom in**. Click the chromosome number to enlarge the chromosome: each segment becomes a labelled bar of its own, overlaps are stacked, with an Mb scale, bands and a zoom slider. **All chromosomes** takes you back.

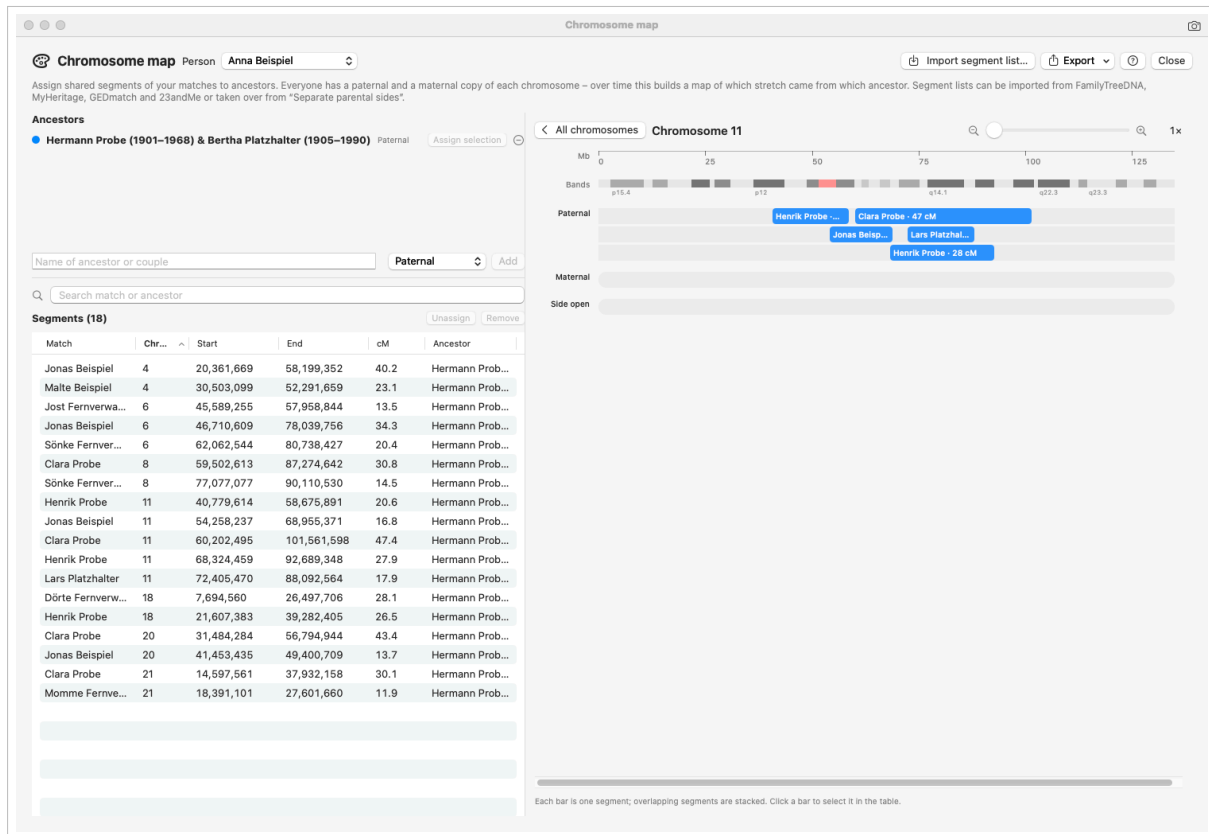


Figure 8.6: Chromosome 11 enlarged.

8.5.3 Exporting

The **Export** menu offers:

- **Segment table (CSV) ...** – all segments with match, chromosome, start, end, cM, SNPs, side, ancestor and source;
- **Export for DNA Painter ...** – DNA Painter's custom CSV format, with the ancestors as groups and their colours; there, read it in via "Import" as a custom CSV file;
- **Whole map as picture (PNG) ...** and **Whole map as PDF ...** – light on white, with title and a legend of the ancestors; in the enlarged view also the single chromosome as PNG or PDF.

The coordinates must be GRCh37, as at FamilyTreeDNA; the map does not convert them. The table shows the cM from the imported file; the totals are calculated by Clavis itself. The maps are stored only on this Mac.

8.6 The family report

Family report (PDF) (in the **People** window, in the File menu, in the **Data** menu or under “What should I do next?”) creates an easy-to-read A4 report about a person, without raw data – for passing on to the family. It contains the existing tests, the mtDNA and Y haplogroups with their meaning, the relationships to the other people in the data folder (from the autosomal raw data), test suggestions, background “explained simply” and a privacy note.

Research tools

9.1 The research log

Family › Research log (⌘⌘J) records which question you are pursuing and how far you have got. It has two parts.

Hypotheses: For each hypothesis the fields **The question, as precisely as possible**, **Data: provider, date of retrieval, measured values**, **Hypothesis**, **What speaks for it**, **What speaks against it**, **Next step** and **State** (open, hint, reasoned assumption, supported, proven, rejected) with a date. Rejected hypotheses belong in the log just as much as confirmed ones.

The screenshot shows a macOS-style window titled "Research log". At the top, there are tabs for "Hypotheses" and "Segment check sheets". To the right of the tabs are buttons: "Matches by hand", "+ New hypothesis", "PDF report...", and "Delete".

The main content area is divided into two columns. The left column contains a list of hypotheses, with the first one selected: "Wohin gehört Rike Entwurf (120 cM, ohne Seite des Anbieters)? open".

The right column shows the details of the selected hypothesis, with fields for:

- The question, as precisely as possible:** "Wohin gehört Rike Entwurf (120 cM, ohne Seite des Anbieters)?"
- Data: provider, date of retrieval, measured values:** "FamilyTreeDNA Family Finder, Abruf 10.10.2026: Rike Entwurf 120 cM, längster Block 30,5 cM, 5 Segmente. In Pauls Liste fehlt sie."
- Hypothesis:** "Rike ist mit Anna über die Mutterseite verwandt, etwa als Cousine 2. Grades."
- What speaks for it:** "Sie fehlt in der Match-Liste des Vaters; 120 cM liegen in der Spanne für Cousins 2. Grades."
- What speaks against it:** "Sie steht nicht im Stammbaum; ohne Vergleich mit einer Verwandten der Mutter bleibt die Seite offen."
- Next step:** "Rike nach ihrem Stammbaum fragen; Segmente mit Mira Muster vergleichen."
- State:** A dropdown menu showing "open" and a date field "State as of 10 October 2026".

Below these fields is a section titled "Compare candidates" with a paragraph of text explaining the process and a "+ Add candidate" button. At the very bottom, a small note states: "The log is stored on this Mac only. Initials rather than names are advisable for matches, especially if you pass a report on."

Figure 9.1: A hypothesis in the research log.

Under **Compare candidates** you enter, for each candidate, how the matches would be related to that candidate. Clavis checks each value against the observed range of the Shared cM Project and shows, for each candidate, how many values lie within the range; with two or more candidates a rough weighting follows as percentage bars. The weighting only appears if all candidates rest on the same matches: for every candidate, enter the same matches (the same initials

in the **Match (initials)** field) with their cM, each with the relationship that would hold for that candidate; otherwise a note explains what is missing. A value of 0 cM or below 7 cM counts as “not detectable”, not as an exact value: it speaks against a close relationship but does not rule out a distant one. A range is not a hard limit; the weighting is not proof, and if the right candidate is missing, it tells you nothing.

Segment check sheets: A check sheet records, for a single segment, what the finding rests on: match, data version, reference, position, size, method, side, triangulation, pile-up, genealogical path, alternative explanation and state. Check sheets can be created pre-filled from the overlapping segments and from the segment groups.

The screenshot shows the 'Research log' app interface. At the top, there are tabs: 'Hypotheses', 'Segment check sheets' (selected), 'Matches by hand', '+ New check sheet', 'PDF report...', and 'Delete'. Below the tabs, the 'Segment check sheets' form is displayed. On the left, a sidebar shows a list of hypotheses: 'Clara Probe, Jonas Beispiel, Henrik Probe, Malte Beispiel, Sönke Fernve...' with a 'hint' label. The main form contains the following fields:

- Match:** Clara Probe, Jonas Beispiel, Henrik Probe, Malte Beispiel, Sönke Fernverwandt, Dörte Fernverwandt, Lars Platz
- Date of data:** Date the list or file was downloaded
- Reference:** Reference genome version, e.g. GRCh37
- Location:** Chr 4, 30.5 – 52.3 Mb; Chr 6, 46.7 – 58.0 Mb; Chr 6, 62.1 – 78.0 Mb; Chr 8, 77.1 – 87.3 Mb; Chr 11, 54.3 – 58.7
- Size:** Length in centimorgans, number of shared positions
- Method:** FamilyTreeDNA, ≥ 10 cM
- Side:** paternal copy
- Triangulation:** Checked with whom; result; or not checked
- Pile-up:** Number of other matches at the same place
- Genealogical path:** Presumed ancestral couple and evidence in the tree
- Other explanation:** Strongest alternative not yet ruled out
- State:** hint (dropdown menu)

At the bottom, there is a note: 'The log is stored on this Mac only. Initials rather than names are advisable for matches, especially if you pass a report on.'

Figure 9.2: A new segment check sheet with its fields.

PDF report... saves the selected hypothesis or the selected check sheet as a PDF. For matches, initials are better than names, especially if you pass a report on.

9.2 Matches by hand

Some companies no longer hand out their match list as a file. In **Matches by hand** (Family or in the research log) you enter the important matches yourself: initials, person, company, shared cM, longest segment, date of retrieval, side, group and notes. The app then names the relationships whose range fits the value.

Matches by hand

+ Add match Delete 2 matches recorded

Person (whose matches) Paste from clipboard

Match (initials)	Person	Provider	Shared cM	Longest segment	Group	Retrieved on
K. M.	Anna Beispiel	MyHeritage	214	48,5	Linie Muster	10.10.2026
H. B.	Anna Beispiel	MyHeritage	96	31,2	Linie Probe	10.10.2026

K. M. Anna Beispiel Provider: MyHeritage

214 48,5 10.10.2026 Side: not assigned Linie Muster

Possible relationships according to the Shared cM Project
 1st cousins twice removed · 2nd cousins · 2nd cousins once removed · Half 2nd cousins · 3rd cousins · 1st cousins once removed

My notes
 Match aus der Liste bei MyHeritage; Stammbaum zeigt Muster aus Dorf A. Angeschrieben am 10.10.2026.

The log is stored on this Mac only. Initials rather than names are advisable for matches, especially if you pass a report on.

Figure 9.3: Matches by hand with two entries; the match “K. M.” is selected.

Paste from clipboard takes matches from a copied match page: in your browser, open the DNA matches page at MyHeritage or Ancestry, select all (⌘A) and copy (⌘C), enter whose matches they are in the **Person (whose matches)** field, and click the button. Name, cM and the company’s estimate are read, for MyHeritage also segments and longest segment. Existing entries are updated, not duplicated. Repeat this for further pages of the list. Clavis does not access your account.

9.3 Looking up shared cM

Compare › Look up shared cM (⌘L): enter the shared DNA in cM (1 to 8,000), and Clavis lists all relationships whose observed range in the Shared cM Project 4.0 includes the value, those with the closest average first.

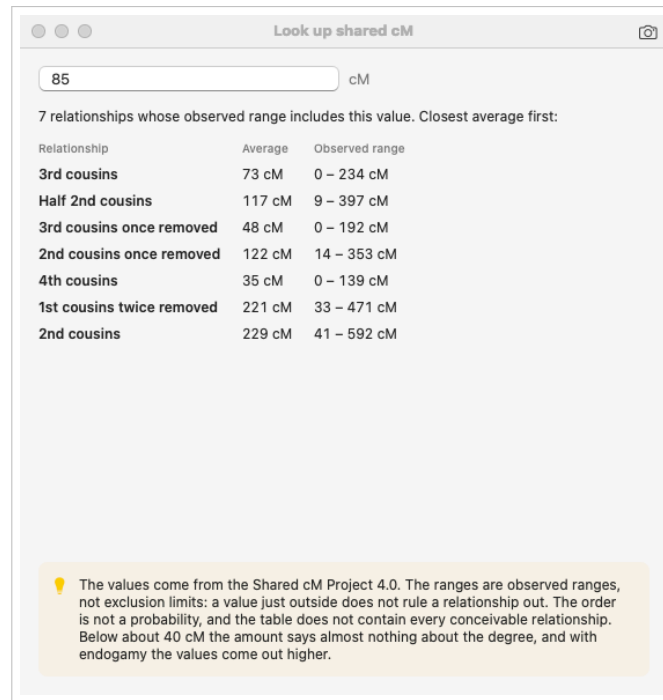


Figure 9.4: Look up shared cM for 85 cM.

The ranges are observed ranges, not hard limits. The order is not a probability; below about 40 cM the amount says almost nothing about the degree, and with endogamy the values come out higher.

9.4 Searching all files for a marker

Data › Find markers (F) searches all files for an rs number, an SNP name (e.g. M267), a DYS marker, an mtDNA mutation (e.g. 16519C) or a position. The results show person, file, marker, chromosome, position and value; **Open file** selects the file in the main window. Below the table you see the number of hits and the search term they belong to.

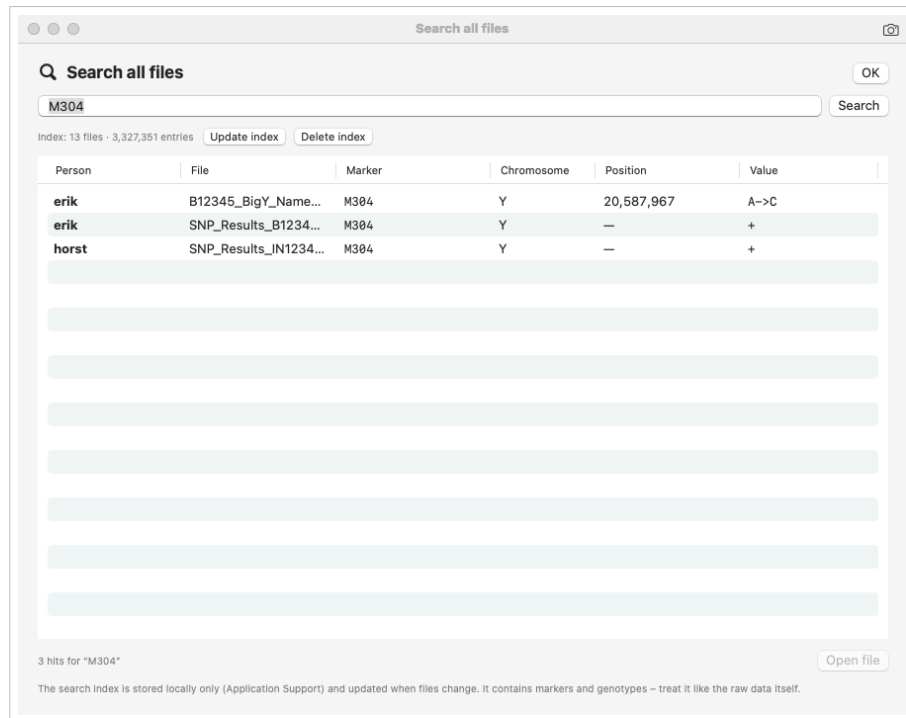


Figure 9.5: The search for the Y-SNP M304 finds it in three files.

The search uses a local index, which is extended with new or changed files when the window opens (**Update index**, **Delete index**). It covers chip raw data, Y variants (CSV and Y-VCF), Y-STR and mtDNA FASTA, but not genome VCFs, match and segment lists, BED or BAM. If you search while the index is still being built, the window shows “The search index is being built. The search for “...” starts as soon as it is complete.” and then searches by itself. The index contains markers and genotypes – treat it like the raw data themselves.

Data, privacy and settings

10.1 Reference data

Clavis comes with the most important reference data (PhyloTree Build 17 with rCRS for mtDNA, a basic Y tree with about 95 major branches, chromosome lengths). You can download further public data in the **Reference data** window (Data > Reference data). No DNA data are sent. Each row has a download button, **Import file...** for a file you obtained yourself, and after downloading **Update** and **Remove**. If a file is old, Clavis recommends downloading it again.

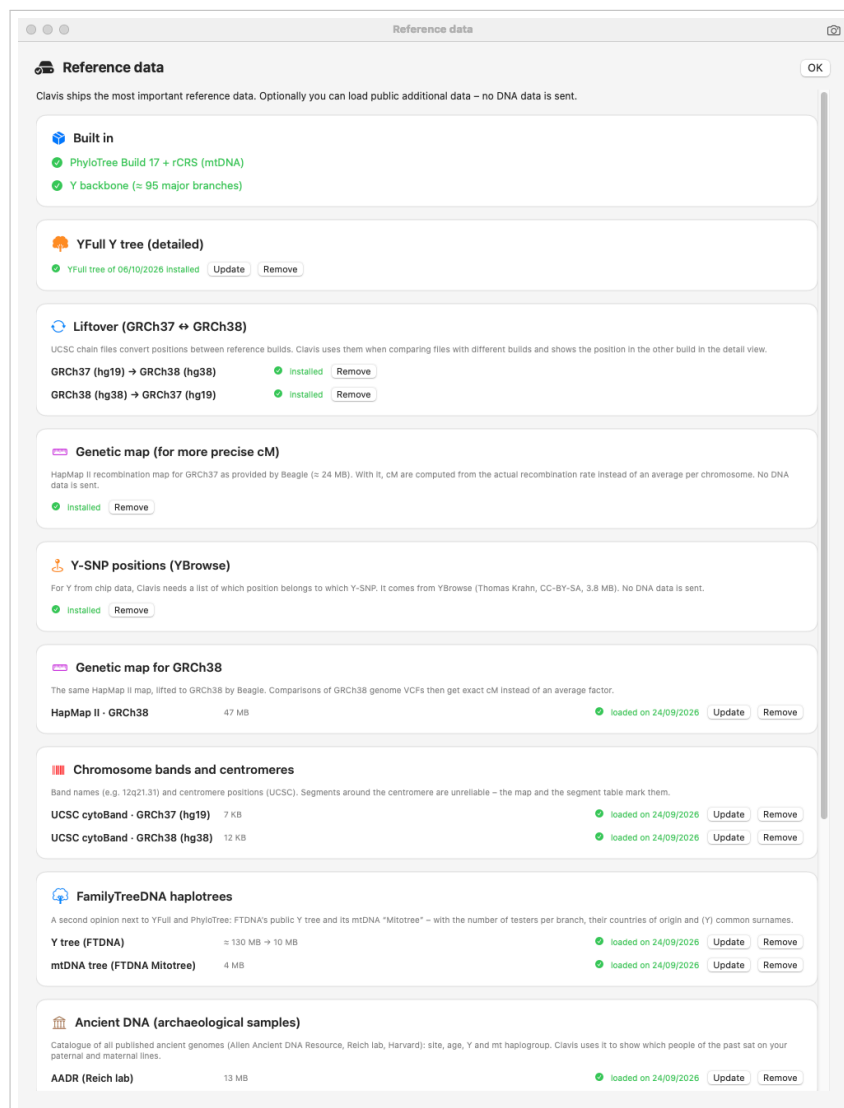


Figure 10.1: The “Reference data” window.

Data set	Used for	Size
YFull Y tree	precise Y terminal branch, ages	≈ 15 MB
Liftover GRCh37 ↔ GRCh38	comparisons across reference genomes, positions	per direction
Genetic map (GRCh37, GRCh38)	precise cM in comparison, phasing, runs of homozygosity	24 / 47 MB
Y-SNP positions (YBrowse)	Y haplogroup from chip data	3.8 MB
Chromosome bands and centromeres	bands in comparison and chromosome map	a few KB
FamilyTreeDNA haplotrees	second opinion for Y and mtDNA	10 / 4 MB
Ancient DNA (AADR)	“Ancient DNA on your line”	13 MB
Allele frequencies (1000 Genomes)	ancestry estimate, rare alleles	1.46 GB download
ClinVar, GWAS Catalog	only with medical data enabled	up to 200 MB each

10.2 Privacy & security

Data › Privacy & security opens four cards:

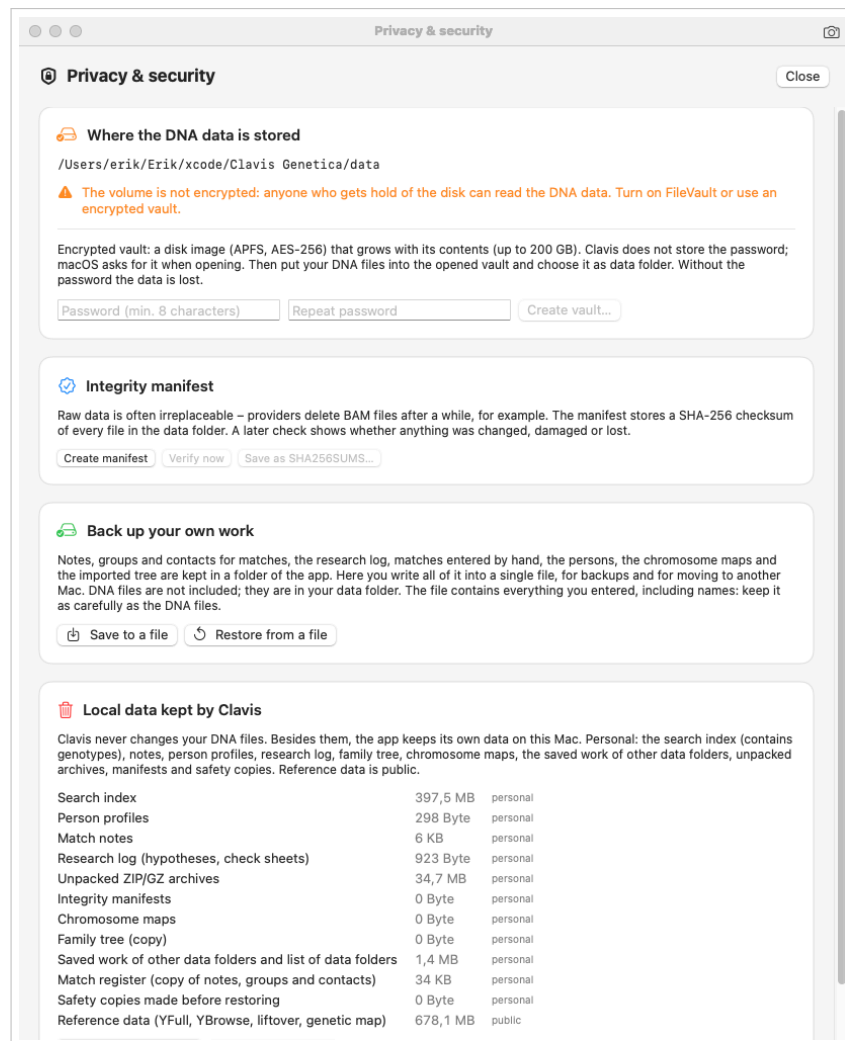


Figure 10.2: The “Privacy & security” window.

Where the DNA data is stored

checks where the data folder is and warns about iCloud Drive, other cloud folders, the Downloads folder or an unencrypted drive. With **Create vault...** you create an encrypted, growing disk image (AES-256) in which you can keep the DNA files. Clavis does not store the password – without the password the data are lost.

Integrity manifest

Create manifest saves an SHA-256 checksum of every file, **Verify now** reports new, changed and missing files, **Save as SHA256SUMS...** writes a file that can be checked with `shasum -c`. The Clavis file `clavis-project.json` is not included.

Back up your own work

Save to a file writes your notes, groups, contacts, name assignments, the research log, the matches entered by hand, the people, the chromosome maps and the family tree to a JSON file – without DNA files. **Restore from a file** brings them back; the current state is backed up automatically first. Before restoring, Clavis checks the file: if it contains parts that this version does not know or is damaged, it is rejected and nothing is changed. The message then reads “Restore failed: ... Nothing was changed.” with the reason. If writing fails partway through, Clavis brings back the previous state. The backup contains names and should be kept as carefully as the DNA files. Also available via File > Back up your own work....

Local data kept by Clavis

lists what Clavis itself has stored (search index, person profiles, notes, research log, unpacked archives, manifests, chromosome maps, the copy of the family tree, **Saved work of other data folders and list of data folders**, **Match register (copy of notes, groups and contacts)**, **Safety copies made before restoring**, reference data), with size and whether it is personal. **Delete personal data** deletes everything personal among them, including the saved work of all data folders; **Delete reference data** clears the reference data, which can be downloaded again at any time. Your DNA files stay untouched in both cases.

10.3 Converting for upload

File › Convert for upload... writes autosomal chip raw data (CSV, TSV or TXT) in a different format so that you can upload them to another service: **23andMe format**, **AncestryDNA format** or **FamilyTreeDNA/MyHeritage format (CSV)**. Then click **Convert and save....** Only the notation changes, nothing is added; files in GRCh38 are rejected. Other files (such as Y-STR, VCF or match lists) cannot be converted.

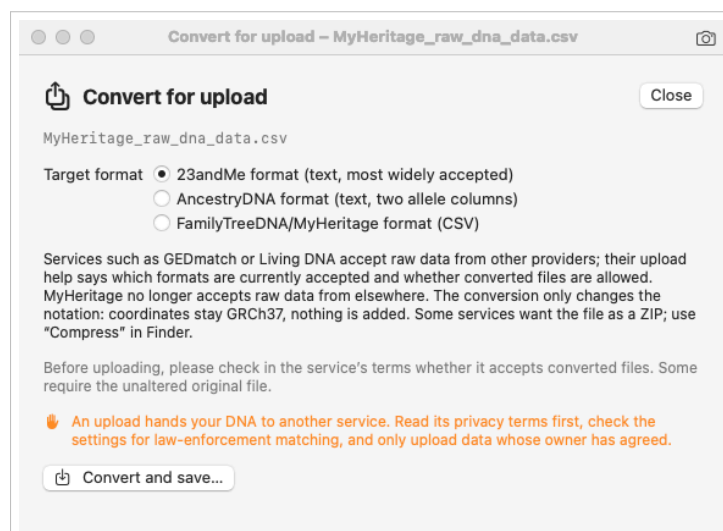


Figure 10.3: The “Convert for upload” window.

Please check for yourself which services accept converted files. FamilyTreeDNA does not accept converted files, and MyHeritage no longer accepts raw data from other companies. An upload passes your DNA on to another service: read its privacy terms and the settings for law-enforcement matching first.

10.4 Medical data

Clavis is not a medical device. Medical analyses are switched off. In the settings under **Medical data** you can enable them with **Show medical data?**; first a disclaimer appears, which you must accept. After that, Data › Medical variants becomes available, and under **Reference data** the ClinVar and GWAS Catalog data sets.

The **Medical variants** window analyses a chip file, a genome VCF or a complete mtDNA FASTA: ClinVar entries where the person carries the listed allele, filtered by significance and review

status, and hits from GWAS Catalog studies with effect and frequency.

Not a diagnosis, no guarantee. Findings from chip data are often wrong. Have every notable finding confirmed by a certified laboratory and discuss it with a doctor. The individual effects from studies are usually small and meaningless on their own. Switching the feature off withdraws your consent.

10.5 Settings

Clavis Genetica › Settings... (⚙️) or the gear:

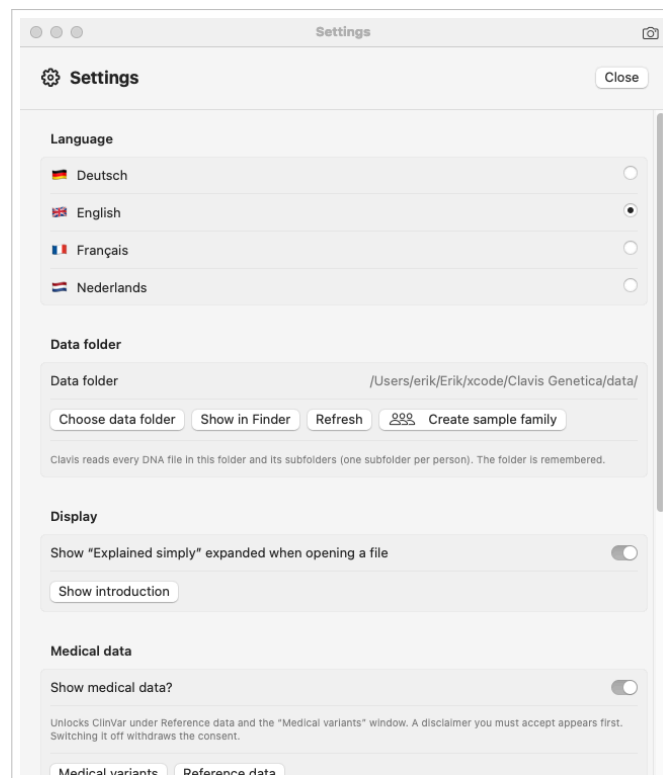


Figure 10.4: The settings.

Language

German, English, French, Dutch.

Data folder

Path and the buttons **Choose data folder**, **Show in Finder**, **Refresh** and **Create sample family**.

Display

Whether “Explained simply” is expanded when you open a file; **Show introduction**.

Medical data

See above.

Check for updates

Version, last check, button to check.

Stored data

The folder where Clavis keeps its own data, `~/Library/ApplicationSupport/be.f97.clavisgenetica`, with buttons for Finder, reference data and privacy.

Help in the app

DNA basics & help

(Help, ⓘ?) explains DNA, the three kinds of test, chip and sequencing, haplogroups, reference genomes, how the creator of a file is detected, the folder structure, privacy and the most important functions.

Glossary

(⌘ ⓘ G) with a search field; the entries come from the glossary of the book “Clavis Genetica”, extended by the app’s technical terms. In the app, small “?” buttons and the line “Terms:” open the explanation of a term directly.

Show introduction

The introduction in six steps.

Explained simply and What should I do next?

In every file view (chapter 3).

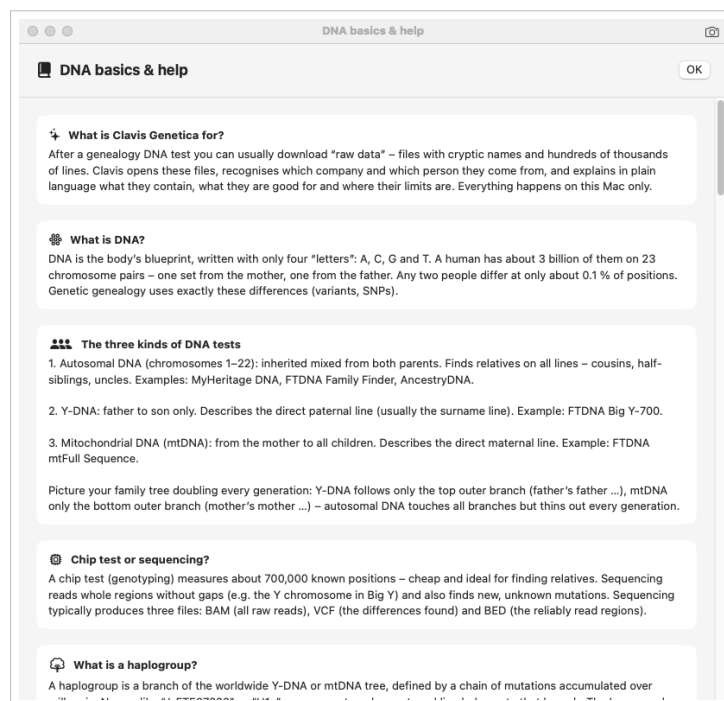


Figure 11.1: DNA basics & help.

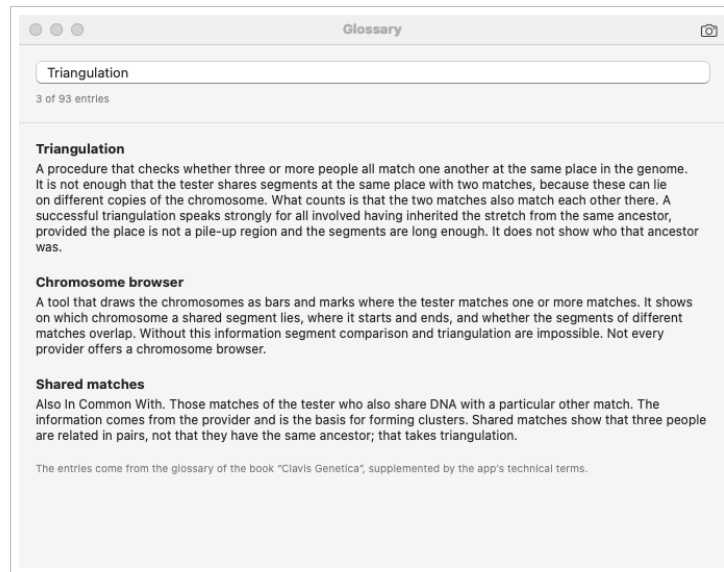


Figure 11.2: The glossary with the search “Triangulation”.

Quick Look in the Finder

Press the space bar in the Finder to see a preview of FASTA, BED and BAM files: for FASTA the length, GC content and the first bases, for BED the covered regions, for BAM the contigs, read group and programs. CSV and VCF files are deliberately excluded.

Menus and keyboard shortcuts

Command	Shortcut
<i>Clavis Genetica</i>	
About Clavis Genetica · Check for updates · Settings...	– · ⌘U · ⌘,
<i>File</i>	
Open file or folder...	⌘O
Refresh	⌘R
Convert for upload... · Family report (PDF)... · Back up your own work...	–
<i>Compare</i>	
Compare the selected file with a second one	⌘D
Multi-person comparison	⇧ ⌘M
Overlapping segments...	–
Look up shared cM	⇧ ⌘L
<i>Family</i>	
People	⇧ ⌘P
Analyse the tree	⇧ ⌘T
Chromosome map	⇧ ⌘K
Separate parental sides (phasing)	–
Research log	⇧ ⌘J
Matches by hand	–
<i>Data</i>	
Find markers	⌘F
Reference data · Privacy & security	–
Picture of the window	⌘S
Medical variants (if enabled)	–
<i>Help</i>	
DNA basics & help	⌘?
Glossary	⇧ ⌘G
Show introduction	–

Supported files

Type	Files
Chip raw data	MyHeritage, FamilyTreeDNA Family Finder, AncestryDNA, 23andMe, Living DNA (CSV, TXT, TSV)
Big Y	Named Variants, Private Variants, Matches (CSV), variants.vcf, regions.bed, BAM with BAI
Y-DNA	FamilyTreeDNA SNP results, Y-STR exports (Y-37 to Big Y-700)
mtDNA	FASTA (mtFull Sequence, mtPlus)
Match lists	FamilyTreeDNA Family Finder Matches, chromosome browser (segments), Y matches
General	VCF (also gVCF; for comparisons with the extension .vcf; with several samples only the first), BED, BAM/BAI, FASTA, ZIP and GZ archives
Family tree	GEDCOM 5.5, Gramps XML (in the People window)

Troubleshooting

A file does not appear in the list.

Check the extension (section 2.1); **Refresh**. Hidden files are skipped.

The folder shows only one person.

The chosen folder is probably one level too high; Clavis offers to open the correct one.

Match list and segment list are not linked.

Both must be in the same person folder and have their original FamilyTreeDNA names.

Names in the segment list are missing from the match list.

Assign them by hand in the **Name matching with the match list** box.

No Y haplogroup from chip data.

Load the Y-SNP positions under **Reference data**; the file must be GRCh37; many chips measure hardly any named Y-SNPs.

cM differ from the company's values.

Load the genetic map; differences of a few per cent are normal.

“Analyse the tree” reports that no person is linked.

In the **People** window, link the person to their entry in the family tree.

The quality pre-check shows no assessment.

The file has not been loaded yet: click **Load file** in the card.

The quality report shows “Limited” and a large gap.

More than 15 Mb without any measurement lie between two shared markers (centromeres do not count). This affects segments, triangulation and phasing, not the relationship estimate; check segment boundaries in this region by hand.

Two genome VCFs give “Too few shared markers”.

Variants-only VCFs share too few positions; use a gVCF or chip raw data.

Another data folder cannot be opened.

Clavis gives the reason and keeps the previous folder open with your work. Nothing changes in the folder or its DNA files.

A backup cannot be restored.

The file is damaged or contains parts that this version does not know. Nothing has been changed; choose a different backup or update the app.

The update cannot be downloaded.

Check your internet connection and, on a local network, the permission under Privacy & Security › Local Network.