

# Genome hacker uncovers largest-ever family tree

Gigantic genealogy linking 13 million people sheds light on how genes influence complex traits.

Heidi Ledford

28 October 2013 | Corrected: 30 October 2013



Andrew Bret Wallis/Getty Images

Using data pulled from online genealogy sites, a renowned 'genome hacker' has constructed what is likely the biggest family trees ever assembled. The researcher and his team now plan to use the data — including a single uber-pedigree comprising 13 million individuals, which stretches back to the fifteenth century — to analyse the inheritance of complex genetic traits, such as longevity and fertility.

In addition to providing the invitation list to what would be the world's largest family reunion, the work presented by [computational biologist Yaniv Erlich](#) at the American Society of Human Genetics annual meeting in Boston could provide a new tool for understanding the extent to which genes contribute to certain traits. The pedigrees have been [made available to other researchers](#), but Erlich and his team at the Whitehead Institute in Cambridge, Massachusetts, have stripped the names from the data to protect privacy.

The structures of the trees themselves could provide interesting information about human demographics and population expansions, says Nancy Cox, a human geneticist at the University of Chicago, Illinois, who was not involved in the study. But more interesting, she says, is the possibility that such data may one day be linked to medical information or to DNA sequence data as more people have their genomes sequenced and deposit that information in public databases.

"We've really only begun to scratch the surface of what these kinds of pedigrees can tell us," she says.

## Putting down roots

Pedigrees provide clues about genetic inheritance. For instance, by comparing an individual to their more distant relatives on the family tree, the change in frequency of a given trait, such as fertility, can indicate to what extent the trait has its roots in genetics. It can also provide clues as to whether the trait is controlled by a few genes that have large effects, or by many genes that each make smaller contributions.

But it takes years to assemble genealogical data for even just a few thousand individuals, said Erlich during a presentation at the meeting on 24 October. In the past, researchers have painstakingly gathered such data from church records and individual volunteers. Erlich and his team decided to streamline the process by collecting data from more than 43 million public profiles on the genealogy

website [geni.com](http://geni.com). The profiles typically included birth and death dates, as well as locations.

The team assembled the data into family trees that ranged from a few thousand individuals up to 13 million people in size. Erlich says that pedigrees previously available for genetic studies contained hundreds of thousands of family members at best.

Lisa Cannon-Albright, a geneticist at the University of Utah in Salt Lake City, urges caution when using self-reported genealogical data. She has worked extensively with a large Utah genealogy database that is linked to some medical information. "Everyone wants to trace their family back to royalty," she says. "For these giant pedigrees, we just don't believe them beyond a certain date." Cannon-Albright says that she cuts off her data at the year 1500.

Ultimately, the value of a pedigree is in the information you can link it to, she adds. At the same meeting in Boston, Cannon-Albright presented data from the Utah database suggesting that the Y chromosome, which only passes from father to son, can carry risk factors for prostate cancer. She has also recently launched a new programme to link genealogical data to medical records from the federal Veterans Health Administration.

The Reykjavik-based genetics company deCODE leveraged Iceland's extensive genealogical data to streamline genome-wide sweeps for genetic signatures that influence a variety of traits, including diseases. Understanding the population structure allowed the company to target which DNA samples should be sequenced for its studies. "It's an incredibly powerful approach," says the company's founder, Kári Stefánsson, who has traced his own family tree back to the birth in 910 of a famously unattractive Icelandic warrior and poet.

For now, it is unclear how the huge pedigrees generated by Erlich and his team will be useful. Some scientists at the meeting expressed enthusiasm for the project, but were hard-pressed to come up with a specific experiment using the data.

But Stefánsson is confident that genealogical analysis will play a big part in genetic studies in the future. "People are becoming more willing to contribute data and medical records," he says. "It's an exciting possibility."

*Nature* | doi:10.1038/nature.2013.14037

## Corrections

---

**Corrected:** This article originally suggested that researchers were also studying the inheritance of facial morphology. They are not. The text has been corrected to reflect this.