

Quantitative analysis of population-scale family trees with millions of relatives

Family trees have vast applications in multiple fields from genetics to anthropology and economics. However, the collection of extended family trees is tedious and usually relies on resources with limited geographical scope and complex data usage restrictions. Here, we collected 86 million profiles from publicly-available online data shared by genealogy enthusiasts. After extensive cleaning and validation, we obtained population-scale family trees, including a single pedigree of 13 million individuals. We leveraged the data to partition the genetic architecture of longevity by inspecting millions of relative pairs and to provide insights into the geographical dispersion of families. We also report a simple digital procedure to overlay other datasets with our resource in order to empower studies with population-scale genealogical data.

<http://science.sciencemag.org/content/early/2018/02/28/science.aam9309>

Short bio:

Dr. Yaniv Erlich is the Chief Science Officer of MyHeritage.com and an Associate Professor of Computer Science and Computational Biology at Columbia University (leave of absence). Prior to these positions, he was a Fellow at the Whitehead Institute, MIT. Dr. Erlich received his bachelor's degree from Tel-Aviv University, Israel (2006) and a PhD from the Watson School of Biological Sciences at Cold Spring Harbor Laboratory (2010). Dr. Erlich's research interests are computational human genetics. Dr. Erlich is the recipient of DARPA's Young Faculty Award (2017), the Burroughs Wellcome Career Award (2013), Harold M. Weintraub award (2010), the IEEE/ACM-CS HPC award (2008), and he was selected as one of 2010 Tomorrow's PIs team of Genome Technology.