



The Future of Genetic Genealogy 2014 Forecast

Brad Larkin

Sponsored by GeneticHomeland.com



Prepared for the
Genetic Genealogy Ireland Conference 2014



Overview

The growth of genetic genealogy has been rapid but exciting new possibilities are forecast from several major developments:

1. Sampling coverage to approach 100% of the genetic lineages of Ireland and the British Isles
2. Continued expansion of the number of markers commercially feasible to test
3. Growth of reference databases with phased results integrated with pedigrees and geography

These elements in combination will one day allow an individual's raw test results to be linked to an established genetic lineage and its ancestral geography in a matter of moments.

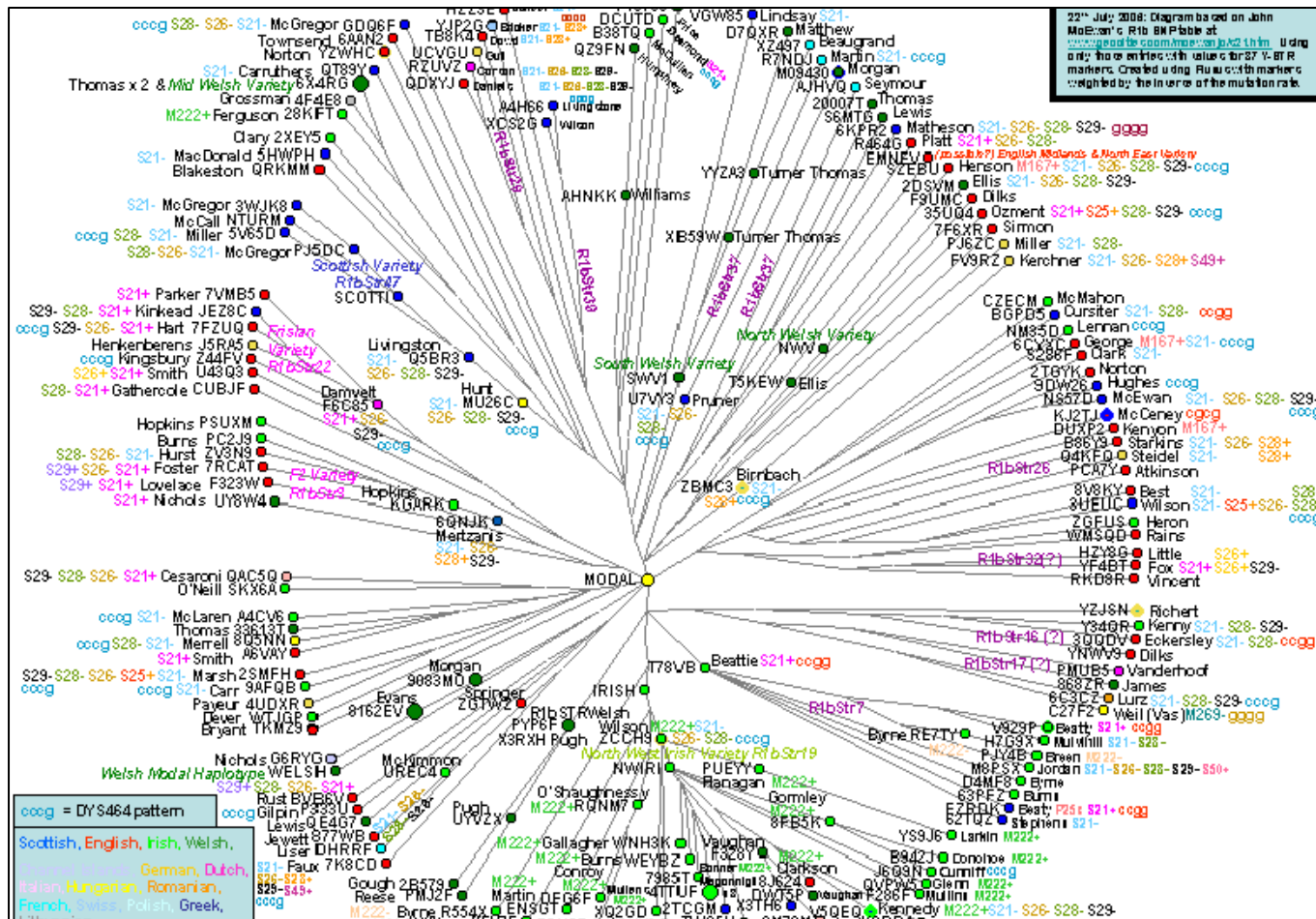
This discussion is intended for those interested in science, technology, and how genetic genealogy will develop as a field.

DNA Sampling Coverage

- The number of lineages and the number of variations in those lineages is finite.

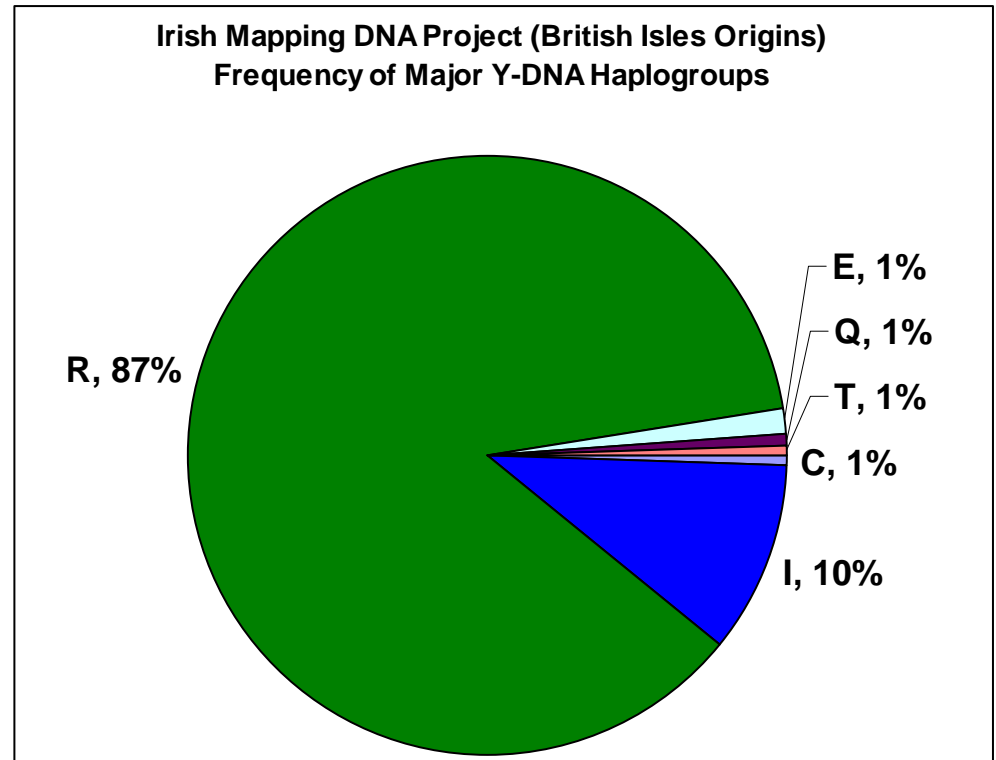
=> Sampling coverage will approach 100% of the genetic lineages in heavily sampled areas.

Ireland and the British Isles.



DNA Lineage Variation

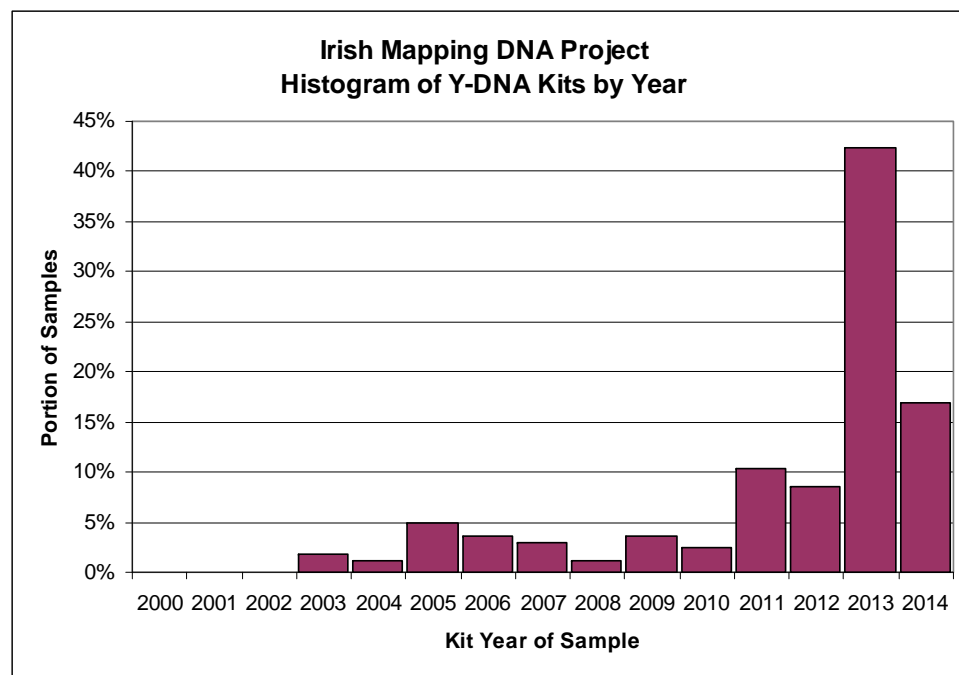
- Ireland Population History
 - ~ 8000 bc – First humans in Ireland
 - ~ 2000 bc – Arrival of R1b males
 - ~ 1000 ad – Viking era
 - 2000 ad – Genetic genealogy
- How close are we today to 100% Y-DNA lineage coverage?



Brad Larkin, [Irish Mapping DNA Project](#), 2014

Coverage Measurement

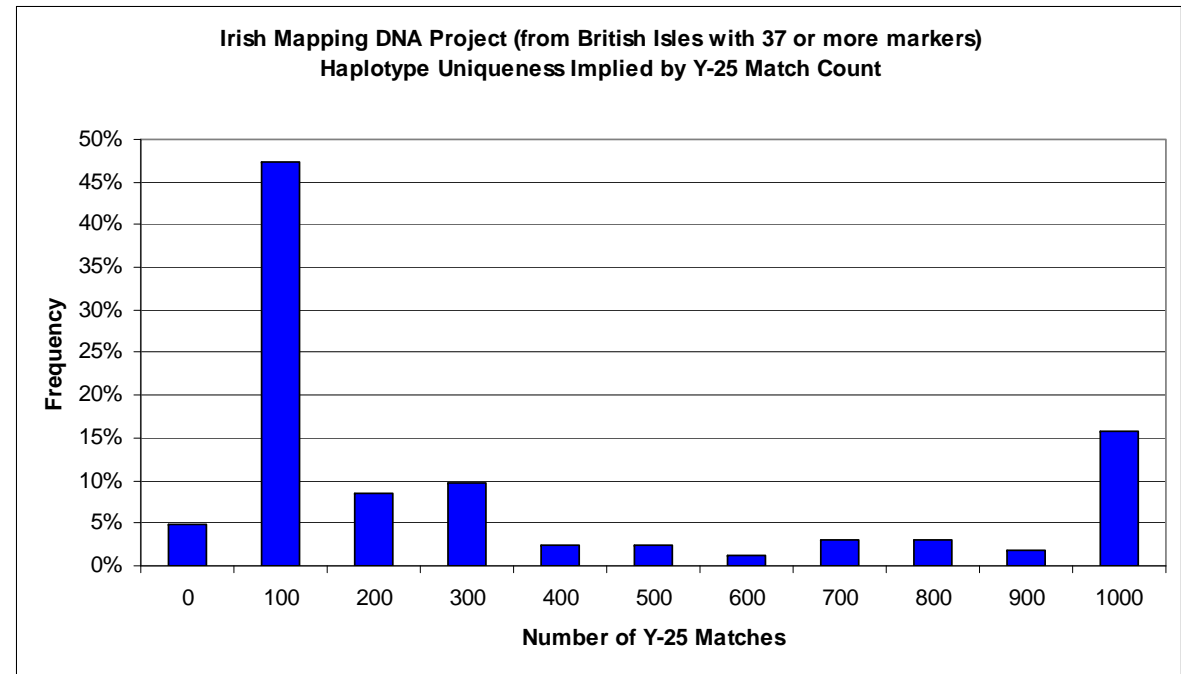
- Hypothesis: *for everyone still in the FTDNA database with no matches at 25 Y-STR markers, there is probably another lineage which has not been tested at all*
- Illustrated with sample set chart from the Irish Mapping DNA Project
 - Fairly random self-selection
 - Screened for ancestral attribution to British Isles and at least 37 Y-STR markers
 - $n = 165$
- Critical Mass: most of the samples came in the past 3.5 years



Brad Larkin, [Irish Mapping DNA Project](#), 2014

Match Count Distribution

- Histogram of Y-25 Match Count¹
 - Bimodal Shaped Distribution
 - 5% of samples with NO 25 marker matches
 - 13% < 3 matches
 - 27% > 500 matches
- FTDNA overall²:
 - 161,000 samples with Y-25 in Haplogroups I & R
 - 4% have no Y-25 matches



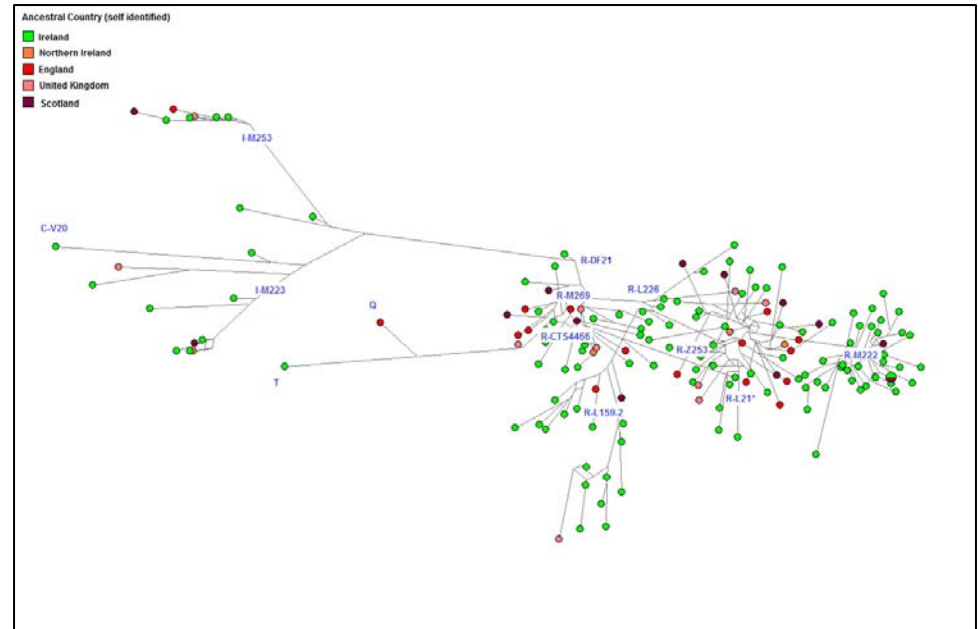
¹ Brad Larkin, [Irish Mapping DNA Project](#), 2014

² Bennett Greenspan, [Family Tree DNA](#), Sept 18, 2014

⇒ Suggests at least 95% lineage coverage in only 14 years of commercial testing

Autosomal Convergence

- Y-DNA
 - More commercially established
 - Easier analysis without recombination
 - ‘Map’ getting well-established
- Autosomal Testing
 - Many times more DNA base pairs to measure
 - Very early stages but growing rapidly
 - Twice as many potential samples
 - Needs chromosome mapping & phasing to be most genealogically meaningful



Brad Larkin, [Irish Mapping DNA Project](#), 2014

⇒ Sampling Coverage of Autosomal DNA testing for genetic genealogy will approach 100% one day as Y-DNA is doing

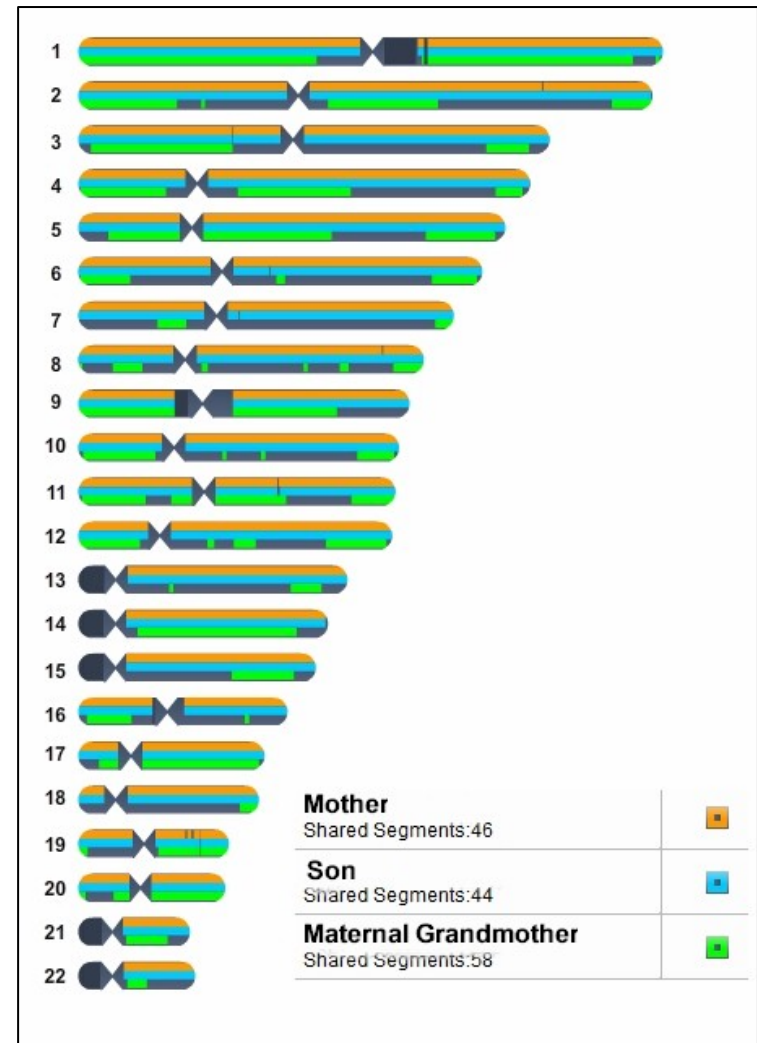
“*Moving from survey to census*” Bennett Greenspan, FTDNA

Phasing Simplified

- Identify the ancestral source of pieces of Autosomal DNA
 - Testing multiple persons with shared segment and ancestor-in-common
- Good references on phasing
 - Whit Athey –
 - [Phasing the Chromosomes of a Family Group When One Parent is Missing](#)
 - Dr. Tim Janzen
 - [Excel-based Phasing Program](#)
 - ISOGG Wiki on [Phasing](#)

Phasing Illustration

- FTDNA Chromosome Browser
 - Sections of chromosomes that seem to be shared with up to 5 others.
 - High level, ‘impressionist phasing’



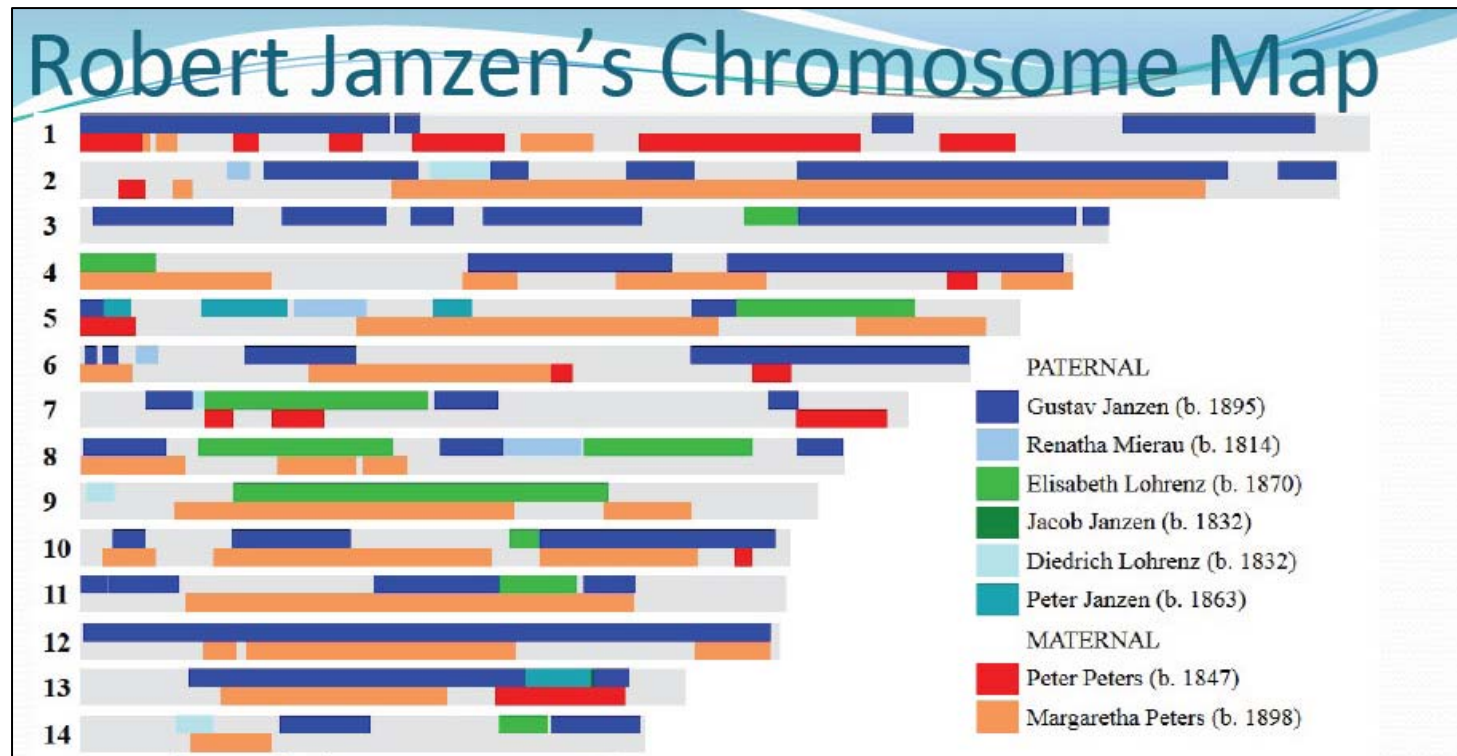
Real Phasing

- Uses individual base pairs of autosomal DNA
 - For best results: Test all available relatives
 - parents, siblings, relatives, and cousins
 - => phasing by inference as well as homozygosity and heterozygosity

SNP Numbers	Sib1 From Dad	Sib2 From Dad	Sib3 From Dad	Sib4 From Dad	Dad Informative Pattern	Sib1 From Mom	Sib2 From Mom	Sib3 From Mom	Sib4 From Mom	Mom's Informative Pattern
SNP1	C	C	C	C		C	C	C	C	
SNP2	A	A	G	G	AAGG	A	A	A	A	
SNP3	T	T	T	T		C	T	T	C	CTTC
SNP4	C	C	C	C		C	C	C	C	
SNP5	G	G	G	G		G	T	T	G	GTTG
SNP6	C	C	C	C		T	T	T	T	
SNP7	C	C	T	T	CCTT	T	T	T	T	
SNP8	G	G	A	A	GGAA	G	G	G	G	
SNP9	A	A	G	G	AAGG	G	G	G	G	

Phasing Integrated with Genealogy

- Real world example of powerful genealogical insights from Tim Janzen's mapping work.



Tim Janzen MD, Advanced Techniques for Use of Autosomal DNA Tests to Break through Genealogical Brick Walls, 2014 <http://tinyurl.com/ny4k9de>

Universal Human Phasing (UHP)

- As phasing coverage increases and we link it to genealogical pedigrees
 - It gets very powerful.
- If every child born were mandated to get a genetic sequencing test
 - Phased results along with those of parents and siblings

=> We will be able to phase the DNA of every ancestral person who ever lived and who has living descendants

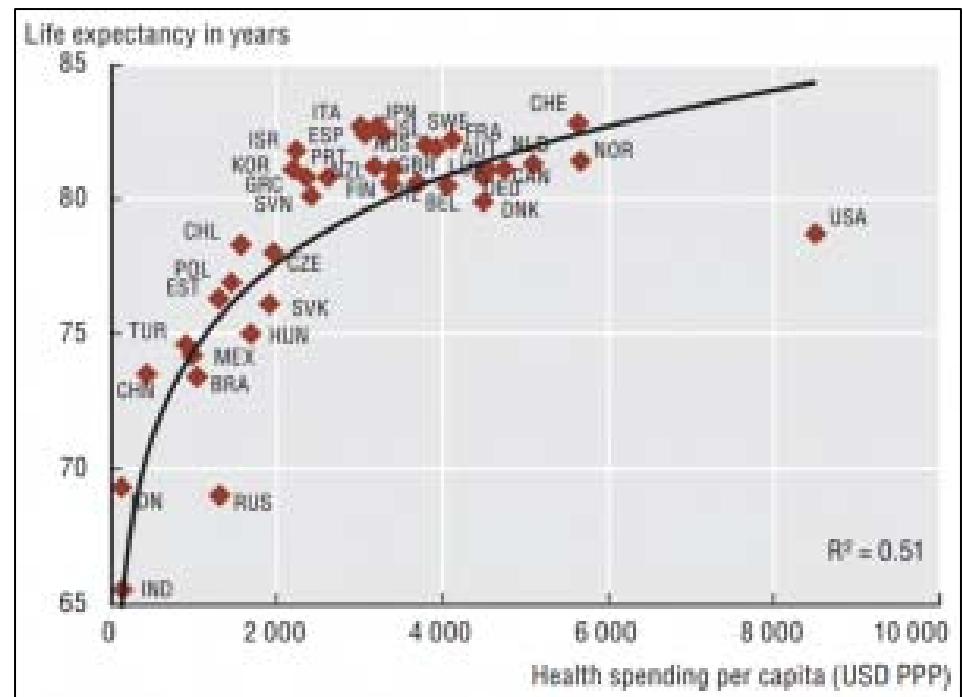


The Path to UHP

- Drivers
 - Health Care Cost Savings
 - Genetically-targeted medicines & vaccines
 - Integration with wearable, real-time health informatics
- Stages
 - Persons Tested
 - Persons Phased
 - Phased Databases



ENCyclopedia Of Dna Elements



Implications of UHP

- Traditional genealogy pedigrees become better informed
 - key part in linking the phased results to our ancestors names & identities
- Even if we do not know their names, we would know the DNA segments of our ancestors
- The power to connect new participants to existing lineages becomes tremendous
- Ancient DNA tests of exhumed remains would provide even more insight into human migration

Sequencers: Faster than Moore's Law

- **Applied Biosystems AB370**
 - Year: 1987
 - Technique: Sanger
 - Size: 39x29x35, 396 lbs
 - Cost: \$2,400 / Mb
 - Yield: 0.5 Mb per day
 - 12 bases per sq inch per day
- **Illumina HiSeq 2000**
 - Year: 2010
 - Technique: Synthesis Method
 - Size: 47x37x30 inches, 488 lbs
 - Cost: \$0.07 / Mb
 - Yield: 750 Mb per day
 - 14,376 bases per sq inch per day
- **Future**
 - ?



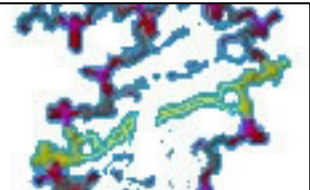
“NGS data output has increased at a rate that outpaces Moore's law -- more than doubling each year. “

technology.illumina.com

Reference Databases

- Commercial Genealogy DNA Testing Labs
- dbSNP – U.S. Government Model
- Genealogical Databases
- University & Medical DNA Labs
- Atlas and Geographical Data
- Independent Databases

dbSNP
Short Genetic Variations



Database Dimensions

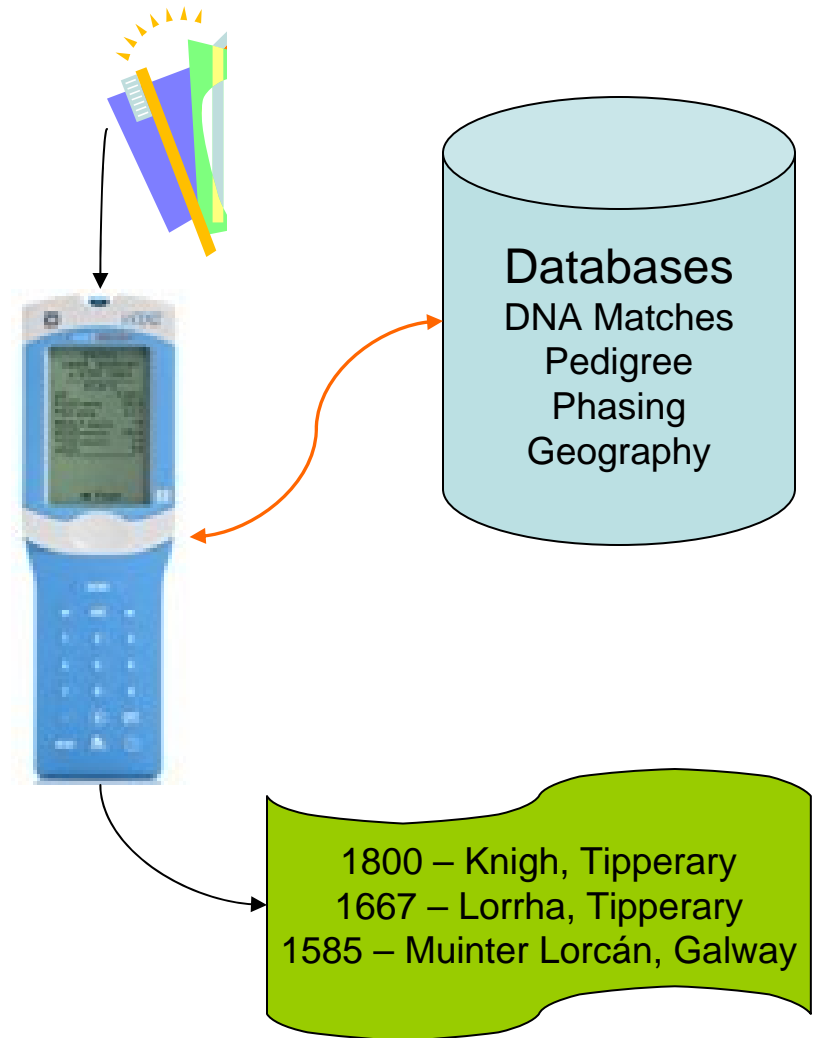
- Raw DNA Values
 - SNPs, STRs
 - Shared Segments
- Matches
- Phasing Analysis
- Ancestral Geography
 - “Where Do I Come From”
- Mobile Platform



Instant Ancestry Field Meter

Envision a portable DNA field sampler

- carry around your neck
- input sample material
- emails a result report
 - list with hyperlinks to your ancestors, available photos, & geographical locations



Available Now

- *Freedom4*
 - University of Otago, New Zealand
 - PCR-based
- FBI moving to Next Generation Identification (NGI) systems and new database
 - Trying to get laws changed to allow rapid testing (outside of accredited lab)



Photo by Sharron Bennett, [University of Otago](#)

What Now?

- It will take time to get to the Instant Ancestry Field Meter.
 - “*Instant coffee wasn't instantly discovered*”
- Practical solution in this generation is to sample as many people as possible
 - “It's a lot cheaper to test folks while they are still alive.”
 - All Available Family
 - Ancestral Geography
- Your sampling documents the human tree
 - ISOGG [Y-DNA Haplogroup Tree](#)¹
 - 2013: 4,115 SNPs
 - 2014: 14,328 SNPs
 - As resolution of phylogenetic SNP tree increases
 - ⇒ Number of markers that are phylogenetically meaningful reaches a plateau



¹ Alice Fairhurst, ISOGG Y-DNA Haplotree Coordinator, 2014

Summary

- Future of Genetic Genealogy
 - More DNA sequencing, performed faster, on smaller machines.
 - Sampling coverage approaching 100%
 - Bigger databases but integrated with more types of data.

Questions?



The Future of Genetic Genealogy 2014 Forecast

Brad Larkin

Sponsored by GeneticHomeland.com

