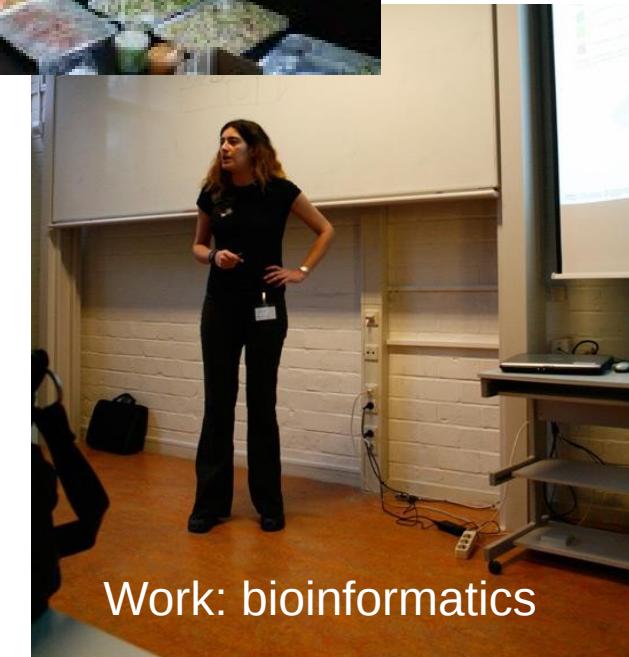
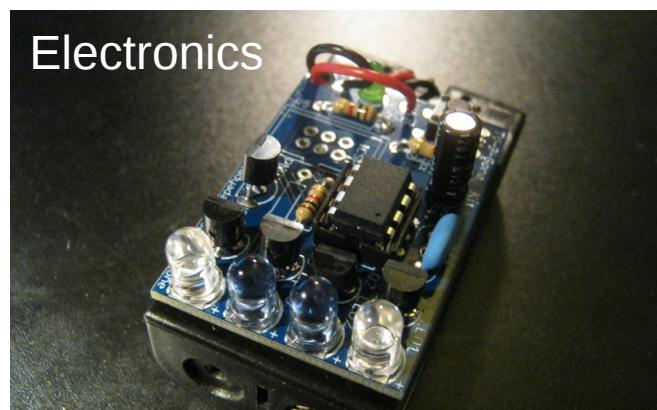
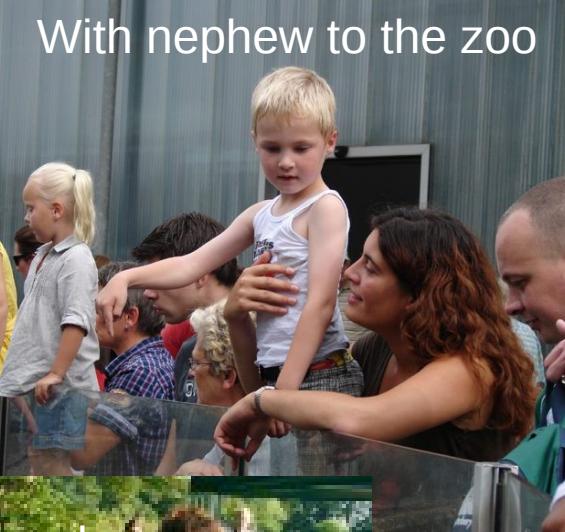




Hacking the DNA

30 March 2013 - Technologia Incognita
Barbera van Schaik (Narya)



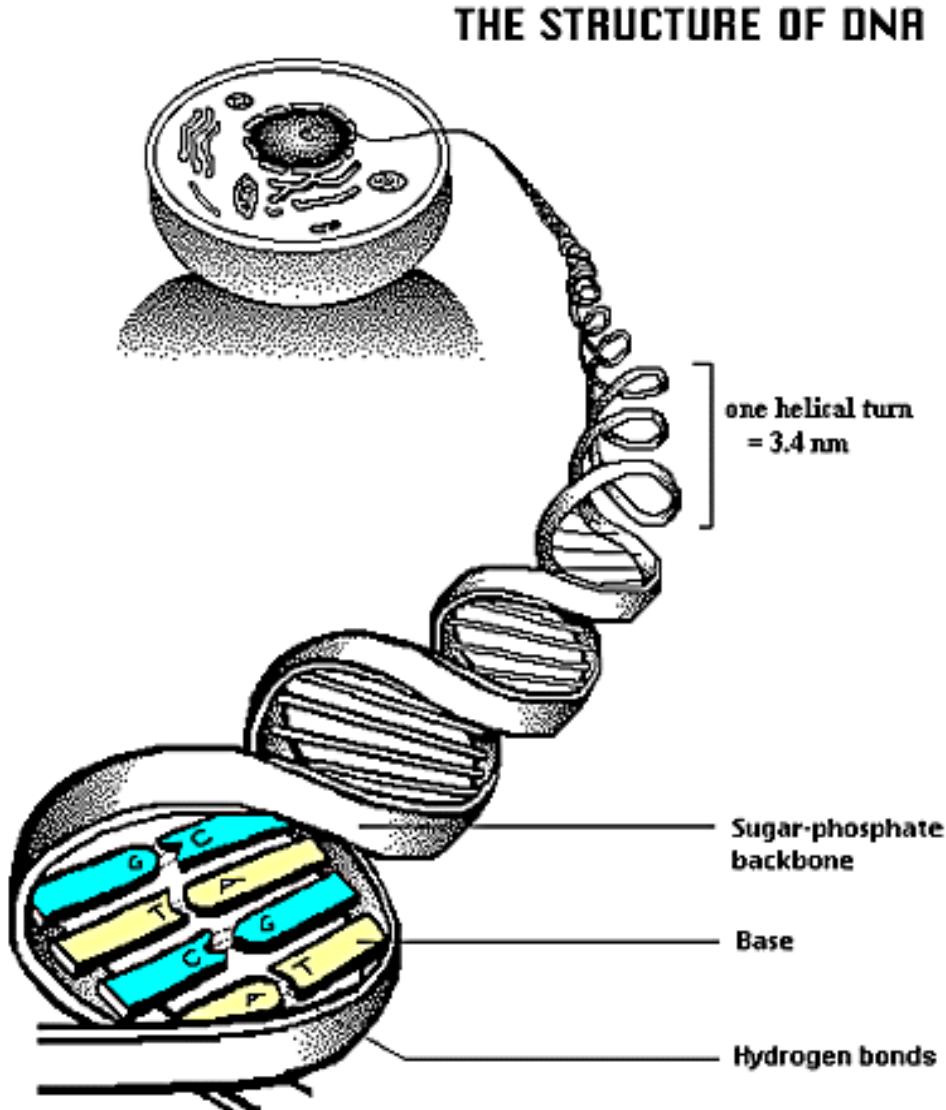
Hacking the DNA

DNA where
DNA what
DNA how



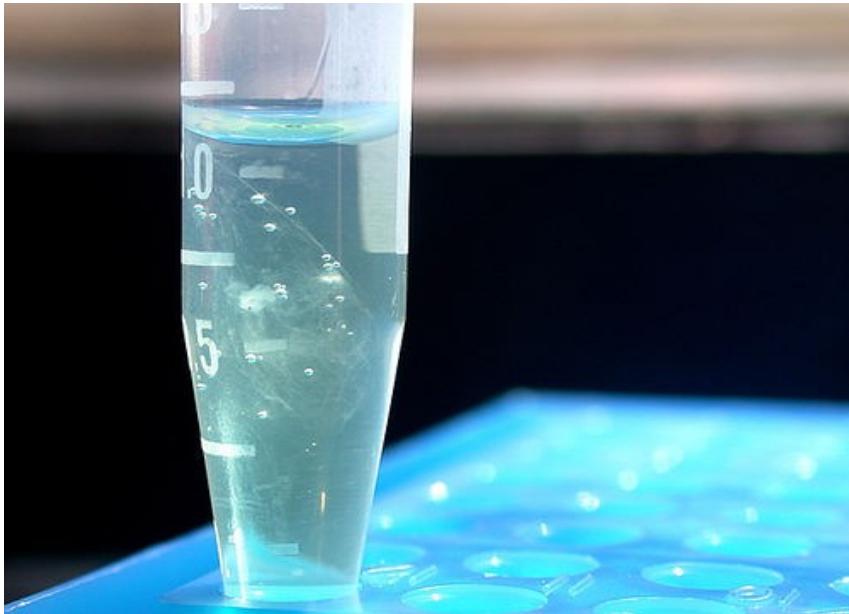
Large hadron collider versus DNA experiments
What researchers do with DNA
What YOU can do with DNA

DNA where



4 building blocks
(nucleotides)

And a bunch of
proteins to wrap it
in the nucleus



DNA what

```
>hg19_dna range=chr16:226703-227520 5'pad=0 3'pad=0 strand=+
GAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGG
CCGCCTGGGTAAGGTGGCGCGCACGCTGGCGAGTATGGTGCAGGGCC
CTGGAGAGGTGAGGCTCCCTCCCTGCTCCGACCCGGGCTCTGCCCGC
CCGGACCCACAGGCCACCCCTAACCGTCCTGGCCCCGGACCCAAACCCA
CCCCTCACTCTGCTTCTCCCCGAGGATGTTCTGTCTCCCCACCACC
AAGACCTACTTCCGCACCTCGACCTGAGGCCACGGCTTGCCCAGGTTA
GGGCCACGGCAAGAAGGTGGCCGACCGCGCTGACCAACGCCGTGGCGC
TGGACGACATGCCAACCGCGCTGTCCGCCCTGAGCGACCTGCACGCC
AAGCTCGGGTGGACCCGGTCAACTCAAGGTGAGCGGGGGGGGGGGGG
GATCTGGTCGAGGGCGAGATGGCGCTTCCTCGCAGGGCAGAGGAT
CGCGGGTTGCAGGGAGGTGTAGCGCAGGGCGGCTGCAGGGCTGGG
CGGCCCACTGACCCCTCTCTGCACAGCTCTAACGCCACTGCCGTGCT
GTGACCCCTGGCCGCCACCTCCCCGCCAGTTCACCCCTGCCGTGAC
CTCCCTGGACAAGTTCTGGCTTCTGTGAGCACCGTGCTGACCTCCAA
ACCGTTAAGCTGGAGCCTCGGTGGCCATGCTTCTGCCCTGGGCCTC
CCCCAGCCCCCTCTCCCCCTGCACCCGTACCCCGTGGTCTTGAAT
```

split20120109_human_t - schaik@amc-app1.amc.sara.nl - WinSCP

File Commands Mark Session View Help

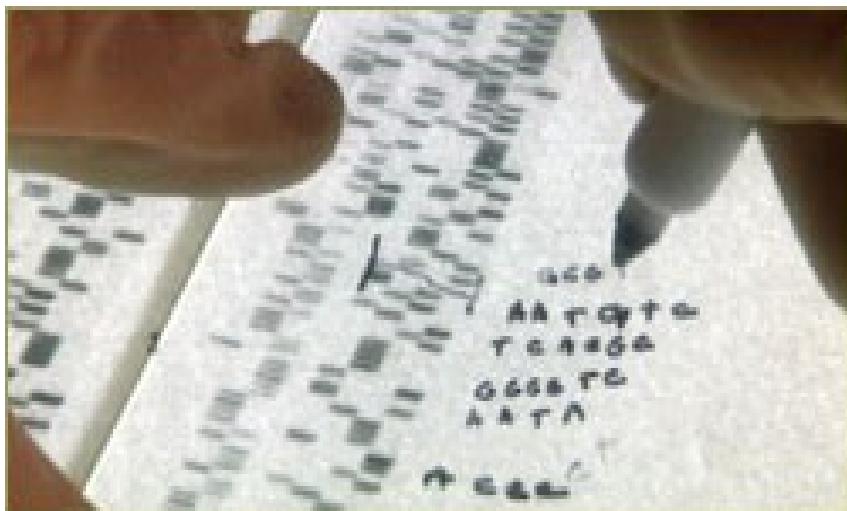
Address /data/home/schaik/Projects/nextgen_sequencing/DATA/20120109_run112_paul_marieke/split20120109_human_t

Default

Name Ext

Name	Ext
fasta_1_ACAACTGCTC.seq	
fasta_1_ACAGGGCTCA.seq	
fasta_1_ACATGAGCTC.seq	
fasta_1_ACTAGGGCTC.seq	
fasta_1_ACTGAGCTCA.seq	
fasta_1_AGAACGGCTC.seq	
fasta_1_AGACGGCTCA.seq	
fasta_1_AGCTCAGCTC.seq	
fasta_1_AGTCACTCA.seq	
fasta_1_ATAACGCTCA.seq	
fasta_1_ATGAGGCTC.seq	
fasta_1_CCGAGGCTCA.seq	
fasta_1_CCTGAGGCTC.seq	

DNA how



1 sequence

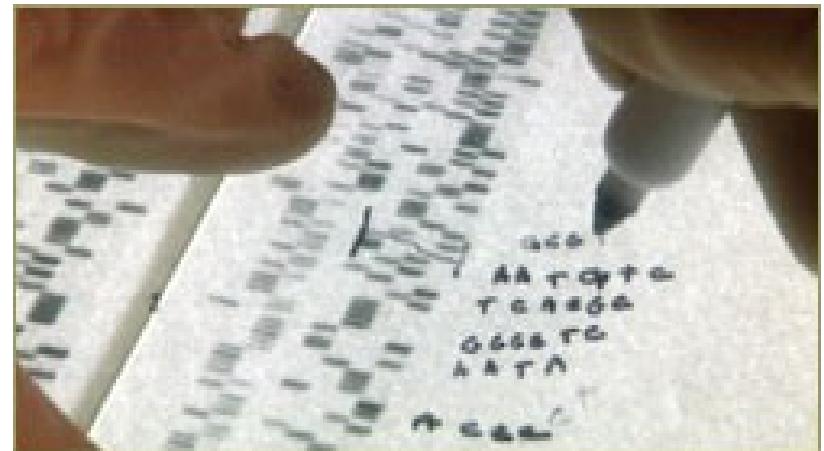
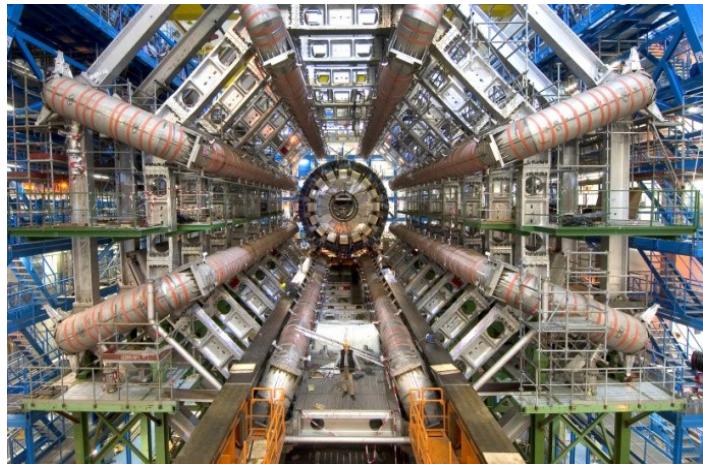


~400 sequences



©2010, Illumina Inc. All rights reserved.

~3,000,000,000 sequences



LHC

VS

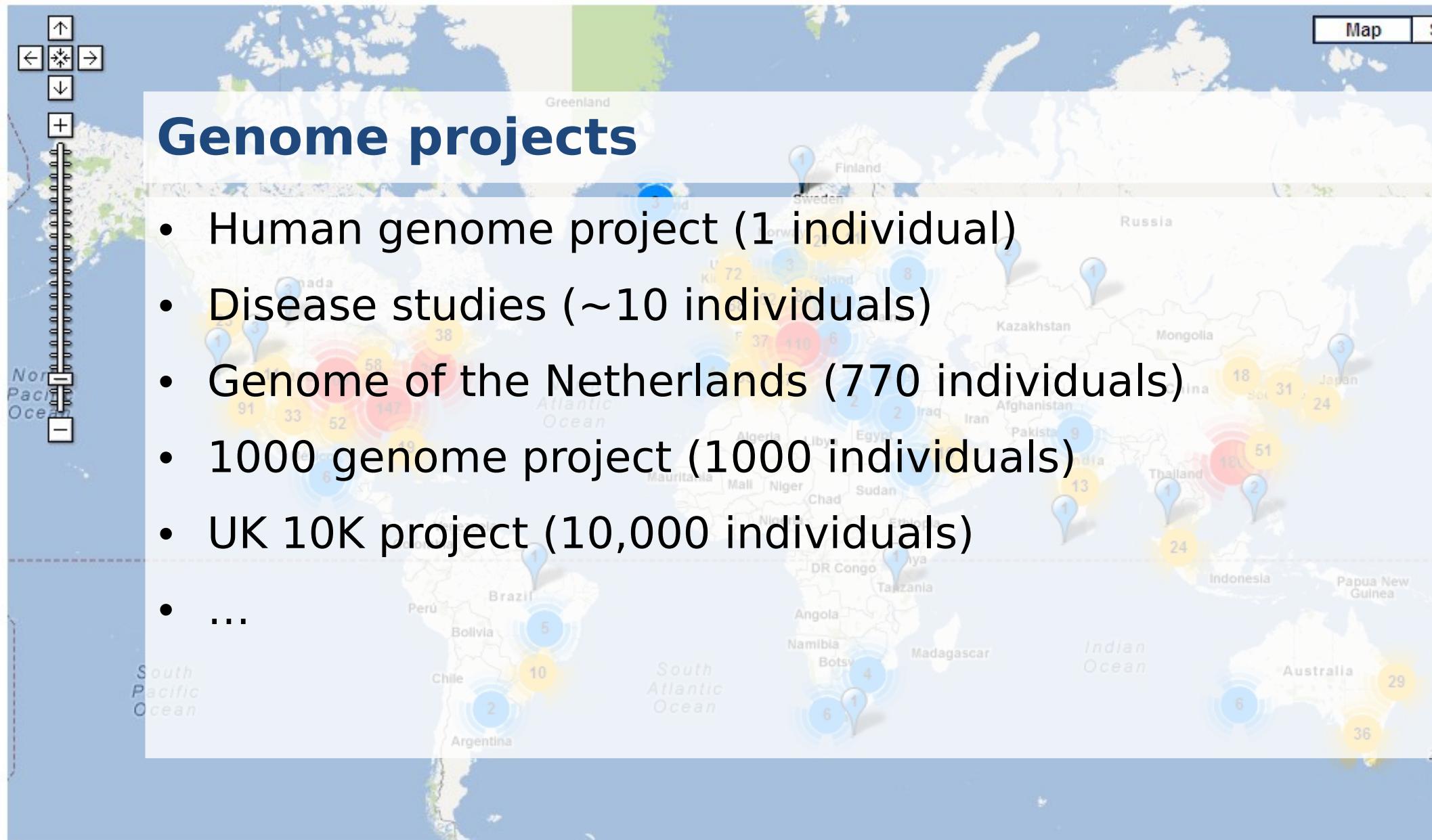
DNA

15 petabyte (10^{15}) per year
Generated on one location

Tens of terabytes (10^{12}) per year
Generated per location

Next Generation Genomics: World Map of High-throughput Sequencers

Show all platforms Illumina GA2 Illumina HiSeq Illumina MiSeq Ion Torrent PacBio Polonator Roche/454 SOLiD Service Provider



How bioinformaticians sometimes feel



“oh, shit, that's 320TB!”

Tony Cox, The Guardian, Feb 28 2008

How to process all that data?



Cloud

Hadoop

Cluster

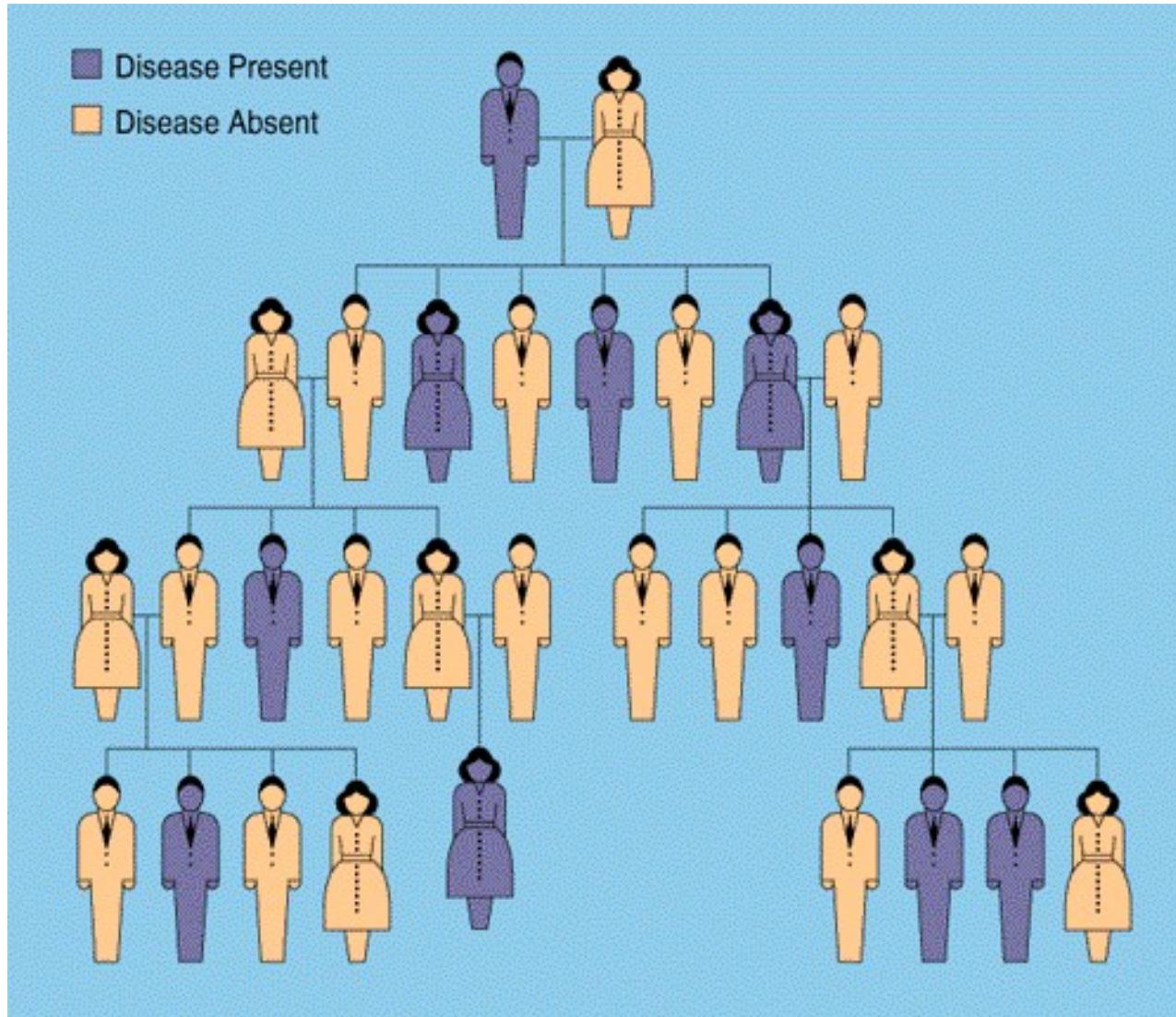
Grid



What researchers do with DNA



Which gene(s) causes disease X?

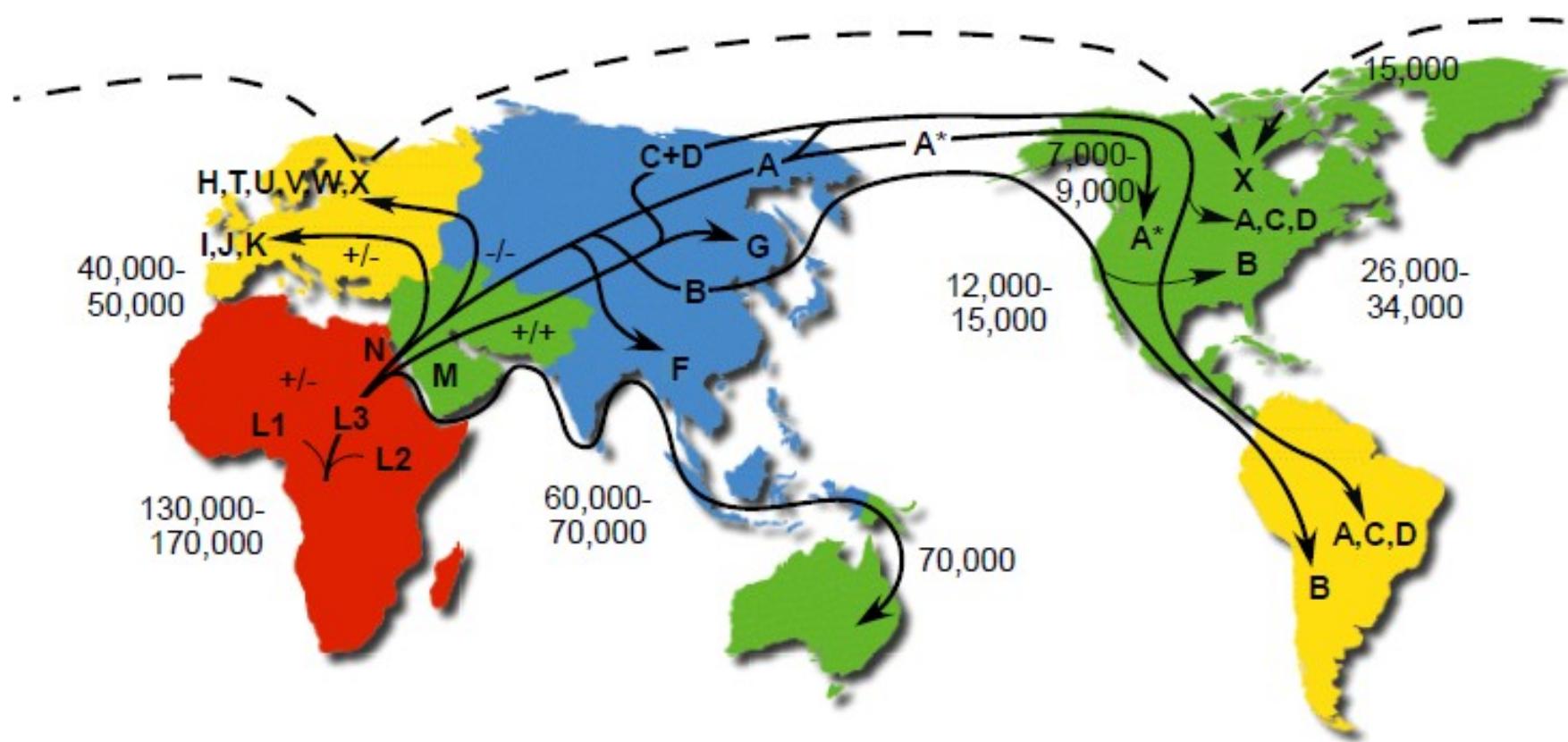


Study migration

Human mtDNA Migrations

<http://www.mitomap.org/pub/MITOMAP/MitomapFigures/WorldMigrations.pdf>

Copyright 2002 © Mitomap.org



+/-, +/+, or -/- = Dde I 10394 / Alu I 10397
* = Rsa I 16329

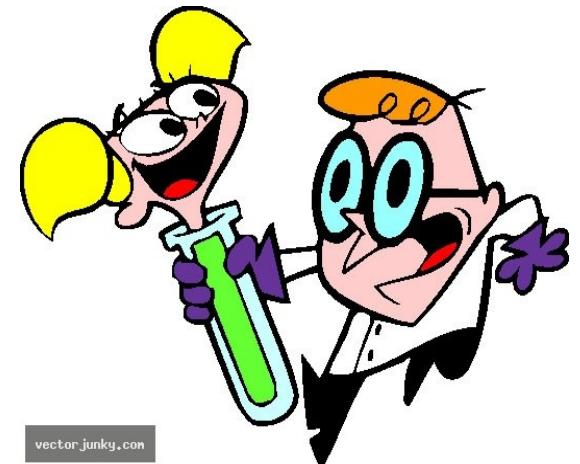
Mutation rate = 2.2 - 2.9 % / MYR
Time estimates are YBP

Study evolution



What YOU can do with DNA

- Download the genomes of 1000 people
 - <http://www.1000genomes.org/>
- Discuss privacy issues
 - Contribute to security
- Contribute to bioinformatics software
 - Most of it is open source



Biohacking



DIY laboratory

Ellen Jorgensen: Biohacking -- you can do it, too
<http://on.ted.com/p4tf>