Family Finder:
Looking under the Hood

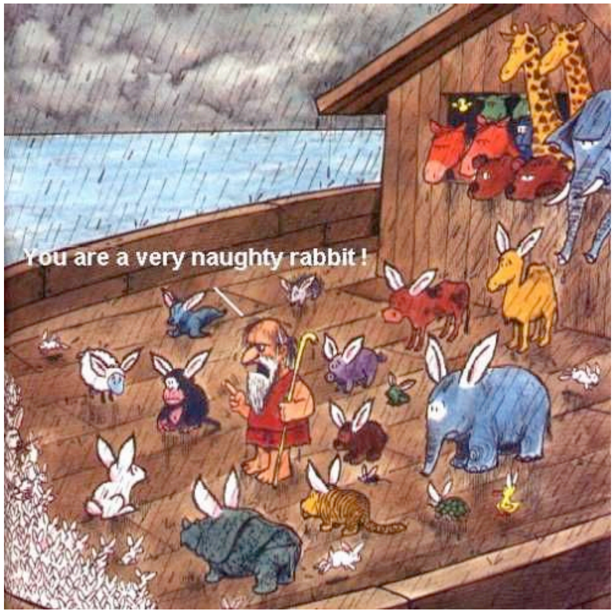
Bruce Walsh
University of Arizona

Lithophane leeae

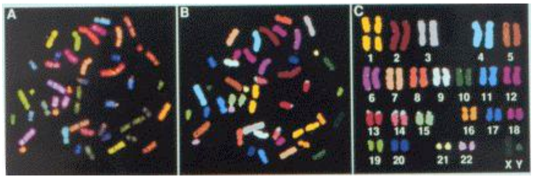
Outline

- Review of some genetics
  - Autosomes and sex chromosomes
  - recombination
- Using DNA to trace relatives
  - Y, mtDNA
  - autosomal
- The autosomal signature expected from shared relatives
- Finding this signature
- Complications and limitations

Review of Genetics



The Human genome

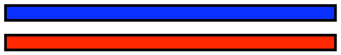


- Each human cell has 46 chromosomes plus multiple copies of mitochondrial DNA (mtDNA)
- 22 pairs of **autosomes** (chromosomes 1 to 22)
- One pair of sex chromosomes
  - XX female (X from both mom & dad)
  - XY male (X from mom, Y from dad)

- Each parent contributes one copy of each autosomal chromosome
- Y chromosomes only pass through males, hence a marker for **direct male lineages**
- mtDNA is **strictly maternally inherited**
- Hence, if male
  - Half your autosomes came from each parent
  - Your Y came from your dad, your X from mom
  - Your mtDNA came from mom
- If female
  - Again half your autosomes came from each parent
  - You got an X from both dad and mom
  - Your mtDNA came from mom

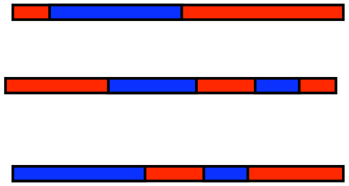
Recombination

- Autosomal chromosomes recombine.
- Hence, the copy of (say) chromosome one you got from your mom is a mixture of the copy she got from her mom (your maternal grandmother) and her dad (your maternal grandfather)
- What about the X and the Y?
  - No recombination in the Y (never two different Y copies in the cell)
  - Recombination in the X in females, but not males



Chromosome one in the mother

Suppose she has three offspring. Each time she contributes a copy of chromosome one to a child, it is usually a mixture of (roughly) half of each of her two copies:



Child one

Child two

Child three

Guinea pig harem says 'hello Sooty'

A GUINEA pig called Sooty had a night to remember after escaping from his pen and tunnelling into a cage of 24 females.

He romanced each of them in turn and was yesterday the proud father of 43 offspring.

Staff at Little Friend's Farm in Pontypridd, South Wales, have now secured Sooty's pen — and begun looking for homes for the guinea pigs.

His owner, Carol Feehan, 42, said: "I'm sure a lot of men will be looking at Sooty with envy.

"We knew that he had gone missing after wriggling through the bars of his cage.

"We looked for him everywhere but never thought of checking the pen where we keep 24 females. We did a head count and found 25 guinea pigs — Sooty was fast asleep in the corner.

"He was absolutely shattered. We put him back in his cage and he slept for two days."

Pen pal: Sooty after his one-night stand

Using DNA variation to find relatives

Polymorphisms

- DNA often varies between copies of the gene.
  - Two random people differ at roughly 20 million (out of 3 billion) bases of DNA
  - Different forms (DNA sequences) of a gene are called alleles
- STRs (Simple tandem repeats)
  - Each locus has many alleles (variation in repeat number)
  - Relatively unstable (high mutation rate)
- SNPs (Single nucleotide polymorphisms)
  - Each locus typically has only two alleles
  - Very stable (low mutation rate)

Using DNA to determine relationships. I. Identity

- In forensics, the question is did a suspect contribute a biological sample
  - Here, 13 autosomal STR markers (each of which has many alleles) are used
  - These are called the CODIS (Combined DNA Index system) markers
  - Odds of two random individuals matching in the trillions
  - Very informative, because all markers should match between contributor and sample

Using DNA to determine relationships. II. Paternity

- In paternity testing, the question is did a suspected father contribute one allele for each of the tested markers?
  - Again, the 13 CODIS loci are used
  - Exclusion occurs when the offspring has alleles at one (or more) loci that are not found in the father
  - Odds of a random individual (i.e., non relative) not being excluded are typically in the tens of millions.
  - Example: Mom is 1,2 at marker one, child is 1,3
    - Mother had to contribute the "1" allele, so the dad contributed the 3
    - If you lack allele 3 at this marker, you are excluded.
    - This is done over all 13 marker loci.

Using DNA to determine relationships. III. General relatives

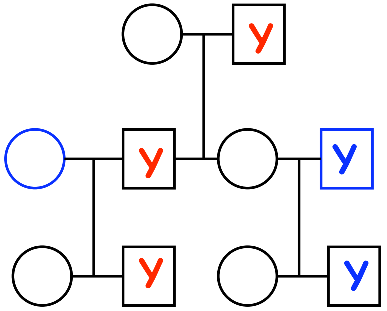
- Now the question is, given a DNA sample, what can we say about the degree of relationship between two individuals?
- If both are males, we can use the Y and ask how many generations back to a common (male) ancestor
  - Traces time back along male-male (paternal) lineages only.
- Independent of sex, we can use the mtDNA and also ask how many generations back to a common (female) ancestor
  - Traces time back along female-female (maternal) lineages only.

Tracing the Y chromosome

○ female

□ male

Blue = unrelated individual



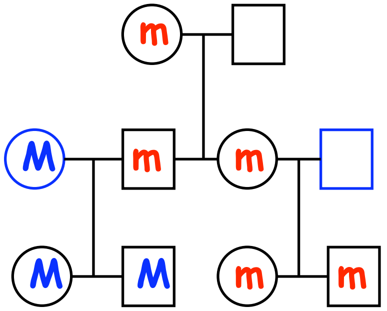
Individuals to be compared

Using Y, we missing the close relationship

Tracing mt DNA

○ female

□ male



Blue = unrelated individual

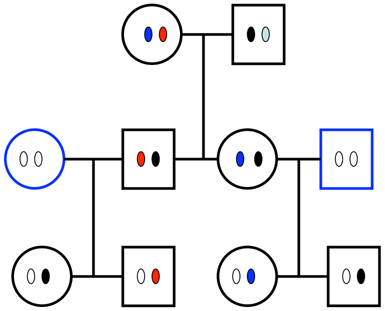
Individuals to be compared

Using mtDNA, we miss the close relationship

Tracing an autosome

○ female

□ male



Blue = unrelated individual

Individuals to be compared

PROVIDED that we can detect the shared allele,
a subset of the relatives show the relationship

Genetic markers

- Y chromosome, mtDNA
  - Only a single comparison -- the haplotype (or collection of markers being scored)
  - With no recombination, these are **inherited as a block**
- Autosomes:
  - A very large number of markers can be compared
  - Problem: each meiosis (generation), only half of the DNA is passed onto offspring, and two sibs may get different halfs!

What autosomal signal is expected?

CAT FOUND!!!

BLACK + TAN WITH GREY

-MALE -NO COLLAR

-Not VERY FRIENDLY, I think he might be scared.

-Not HOUSE BROKEN EITHER ☹

-Found on SUNSET BlVD.

-If HE IS yours Please call

[redacted]

Autosomal signal rapidly lost

- If two individuals share a common ancestor k generations back, then the chance they share the same allele from that ancestor is $(1/2)^{2k-1}$
  - For k =1, this is 50%
  - For k = 2, this is 12.5%
  - For k = 3 this is 3.1%
  - For k = 4, this is 0.78%
  - For k = 5, this is 0.19%
  - For k = 6, this is 0.05%
  - For k = 7, this is 0.01%

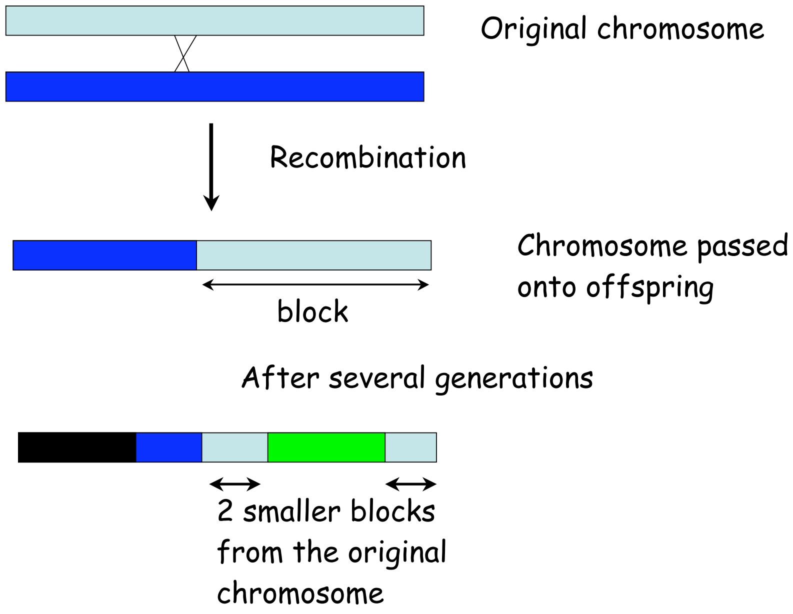
How much sharing?

- Suppose we follow a single marker on each of the 22 autosomes. What is the chance that any of these are shared among relatives?
- After 3 generations, there is a 50% chance that AT LEAST one allele is shared.
- More generally, ..

TMRCA (generations)	P(share at least one autosome)
1	0.99999976
2	0.94701204
3	0.50265502
4	0.15848371
5	0.04209892
6	0.01068729
7	0.00268211

More generally, look at blocks

- As we have seen, sections of chromosomes are linked together, and only broken up by recombination
- Hence, as generations proceed, we can think of a chromosome as a series of an increasing number of blocks, and some of these blocks can be shared



Original chromosome

Recombination

Chromosome passed onto offspring

block

After several generations

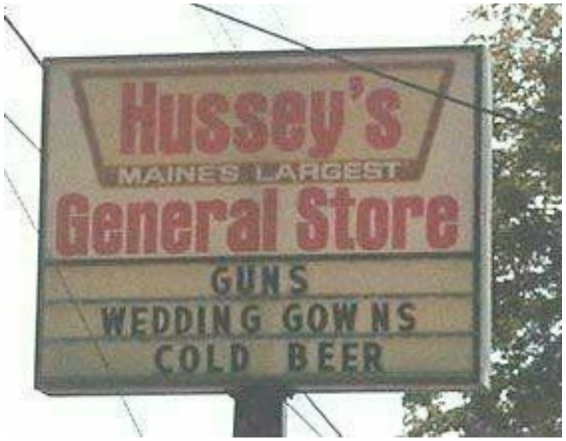
2 smaller blocks from the original chromosome

TMRCA	# independent blocks	Pr(share at least block)
1	44	100
2	88	100
3	212	99.9
4	272	88.2
5	331	47.6
6	391	17.4
7	451	5.4

tMRCA	Average size of block
1	44.06 cM
2	19.15
3	12.30
4	9.07
5	7.19
6	5.95
7	5.08

1 cM = centi Morgan, or a one % chance of recombination
1 cM corresponds to roughly 1 million DNA base pairs.

How can we detect this signal?

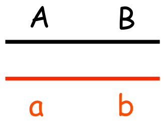


Finding blocks

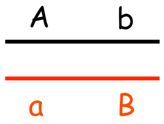
- The key is to look for signals of common blocks shared by two individuals.
- One signal for this would be a long run of identical alleles on a chromosome from each individual
- With very dense (close together) markers, a sufficiently long run of shared alleles indicates a block
- Complication: need to consider linkage phase

Linkage phase

Suppose an individual is AaBb, where A,a and B,b denote the two alleles at different loci



A & B on chromosome from (say) dad, and
a & b on mom's chromosome

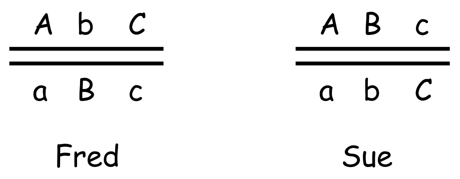


A & b on chromosome from (say) dad, and
a & B on mom's chromosome

Unphased data

Suppose we have unphased data (don't now which of the two copies of a chromosome the alleles are on)

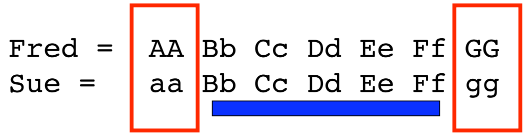
Suppose Fred is AaBbCc and Sue is also AaBbCc. Do they share a block of three markers? Can't tell:



With unphased data, heterozygotes are uninformative

Finding Blocks with Unphased data

When do we know for sure that a marker locus from two individuals DO NOT match? When they are different homozygotes



```
Fred =  AA Bb Cc Dd Ee Ff GG
Sue  =  aa Bb Cc Dd Ee Ff gg
```

Score as a run of 5 markers

Extend this simple idea

- The length of a run is scored by the number of markers that don't have different homozygotes between two individuals
- Obviously, what we have scored as a shared block of markers can occur by chance
- However, require (typically) hundreds of markers in a row to call this a block, making the odds very small
  - A random one cM block being called a match is < 5%
  - The odds of a run of 5 cM is 1/10,000,000
  - Odds of a run of 10 cM is 1×10^{-14}
- Data is scored for around 500,000 markers (SNPS)

Rough rules

- For fairly close relatives, (1-3 gens) simply use an estimate based on % shared
- For more distant relatives, the size of the largest block provides information on
  - these being relatives
  - The time to MRCA
  - Wide variation expected in the size of block
  - For distant (>5 gens) good chance all blocks have been lost

Complications and limitations

Loss of power for deep relatives

- Most blocks are lost
- After a few generations, only good signal is a long block
- Down in the weeds after 5 or so generations. Rarely, blocks will persist (by chance) in deeper relatives
- Lose of precision in dating TMRCA (relative to Y-based tests), as block length has a very high variance