The Lander-Green Algorithm

Biostatistics 666
Lecture 22

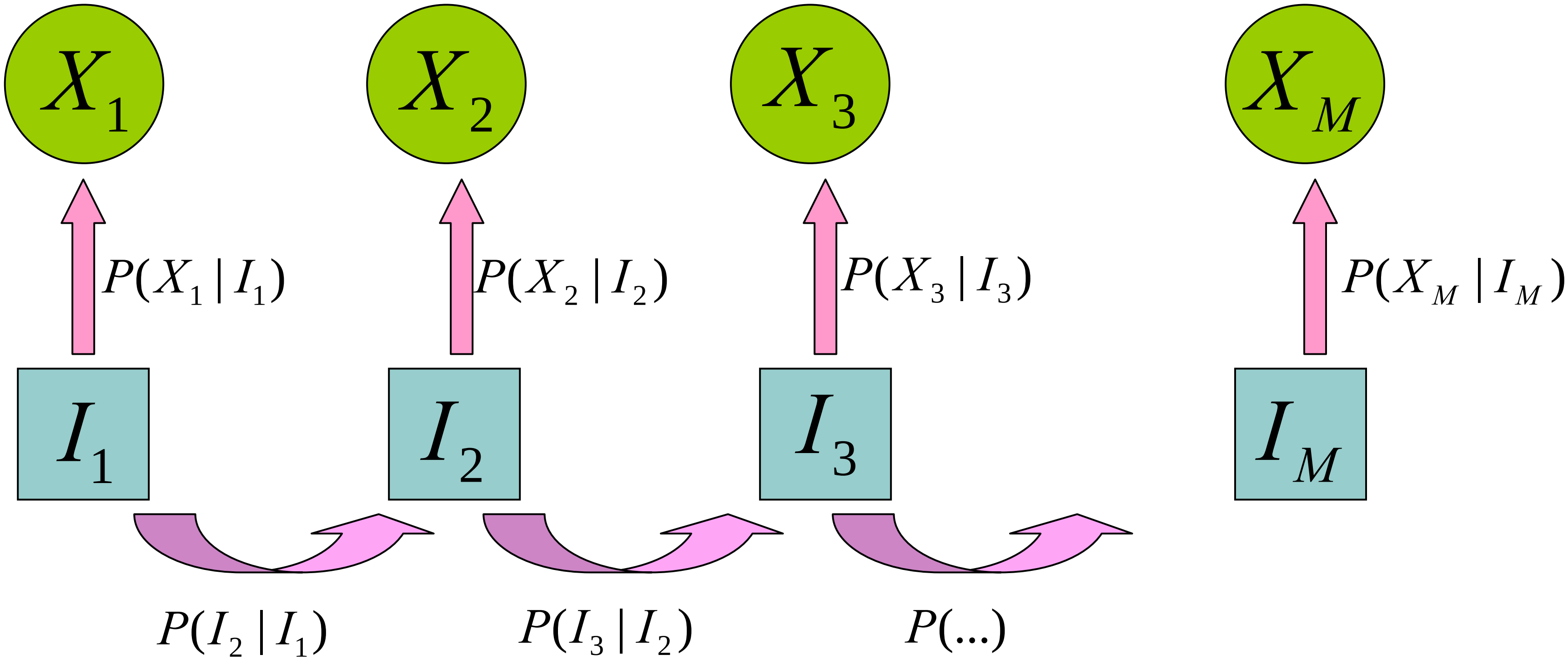
Last Lecture…
Relationship Inferrence

- Likelihood of genotype data
- Adapt calculation to different relationships
  - Siblings
  - Half-Siblings
  - Unrelated individuals
- Importance of modeling error

Today ...

- The Lander-Green Algorithm
- Multipoint analysis in general pedigrees
- The basis of modern pedigree analysis packages

Hidden Markov Model



IBD states form a Hidden Markov Chain along the chromosome …

Fundamental Calculations

- Enumerate possible IBD states
- Transition probability for neighboring IBD states
- Probability of genotype data given IBD state

Lander-Green Algorithm

$$L = \sum_{I_1} \cdots \sum_{I_m} P(I_1) \prod_{i=2}^{m} P(I_i \mid I_{i-1}) \prod_{i=1}^{m} P(X_i \mid I_i)$$

- More general definition for I, the "IBD vector"
- Probability of genotypes given “IBD vector”
- Transition probabilities for the “IBD vectors”

Part I

“IBD Vectors”

Inheritance Vectors

Descent Graphs

Gene Flow Pattern

“IBD Vector” Specifications

- Specify IBD between all individuals
- Must be compact
- Must allow calculation of:
  - Conditional probabilities for neighboring markers
  - Probability of observed genotypes

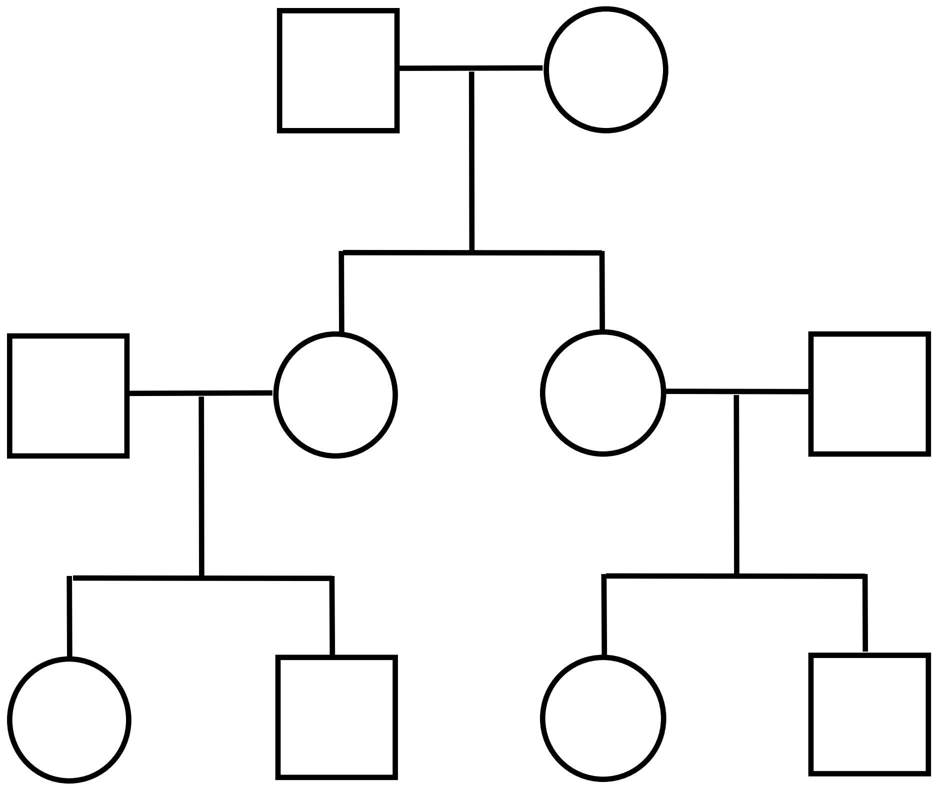
“IBD Vector”

- Specify the outcome of each meiosis
  - Which of the two parental alleles transmitted?
- Implies founder allele carried by each individual
- Implies whether a pair of chromosomes is identical-by-descent

For any pedigree, consider …

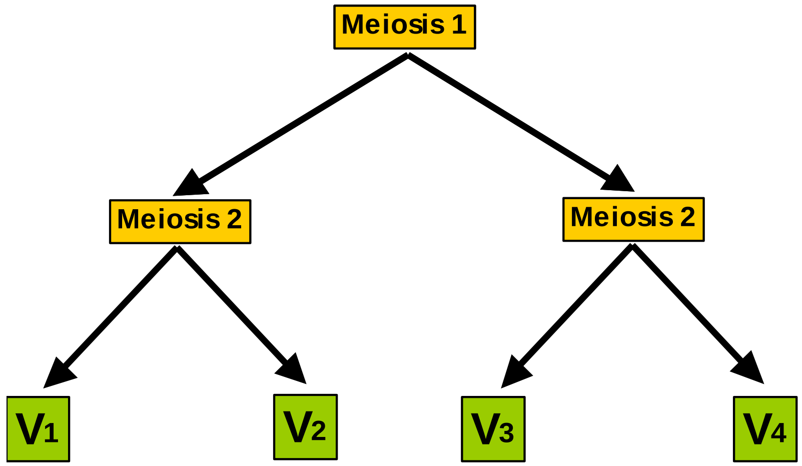
- What are the meioses?
- What are the possible outcomes for the entire set of meioses?

Example …

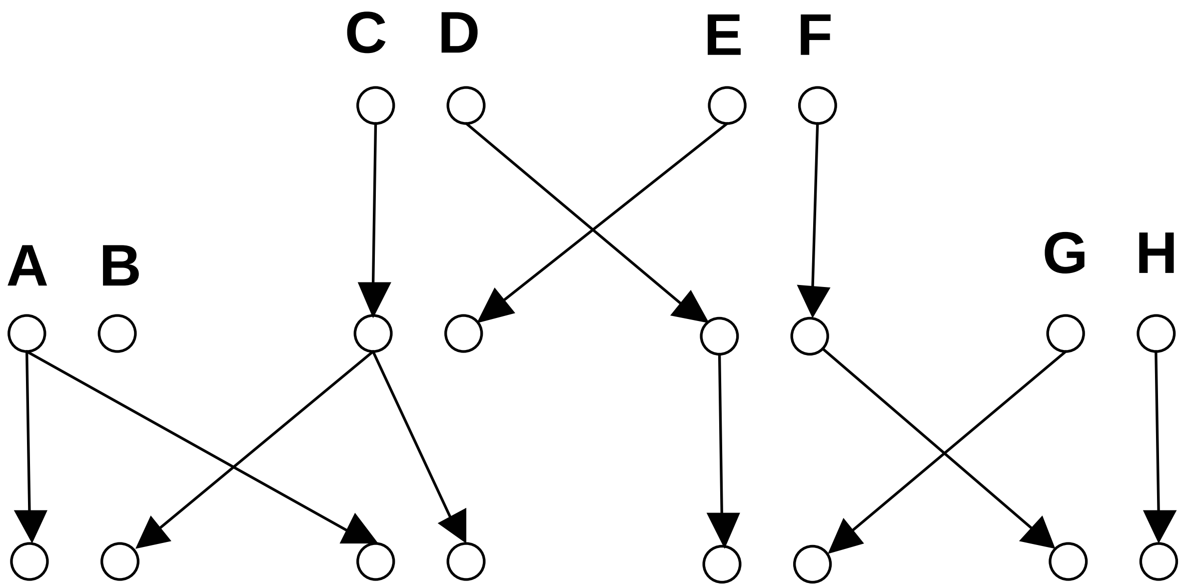


What we are doing …

- Listing meioses
- Alternating outcomes
- The outcomes of all meioses define our “IBD vector”



Example ... Descent Graph



This is one representation of inheritance in the pedigree. The 8 founder alleles are labeled A-H and we their descent through the pedigree is specified when we fix the outcome of each meiosis.

So far …

- A set of $2n$ binary digits specifies IBD in a pedigree with n non-founders
- There are 2^{2n} such sets …
- Next, must calculate the probability of the observed genotypes for each one…

Part II

Probability of Observed Genotypes

Founder Allele Graph

Founder Allele Frequencies

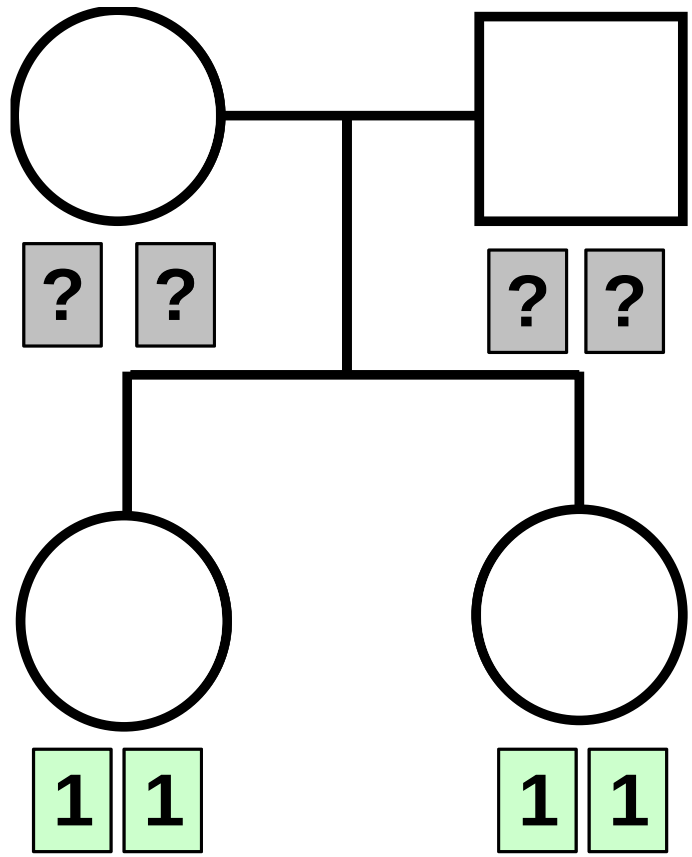
Founder Allele Graphs / Sets

- Calculated for each marker individually
- List of founder alleles compatible with:
  - Observed genotypes for all individuals
  - A particular gene flow pattern
- Likelihood of each set is a product of allele frequencies

Observed Genotypes

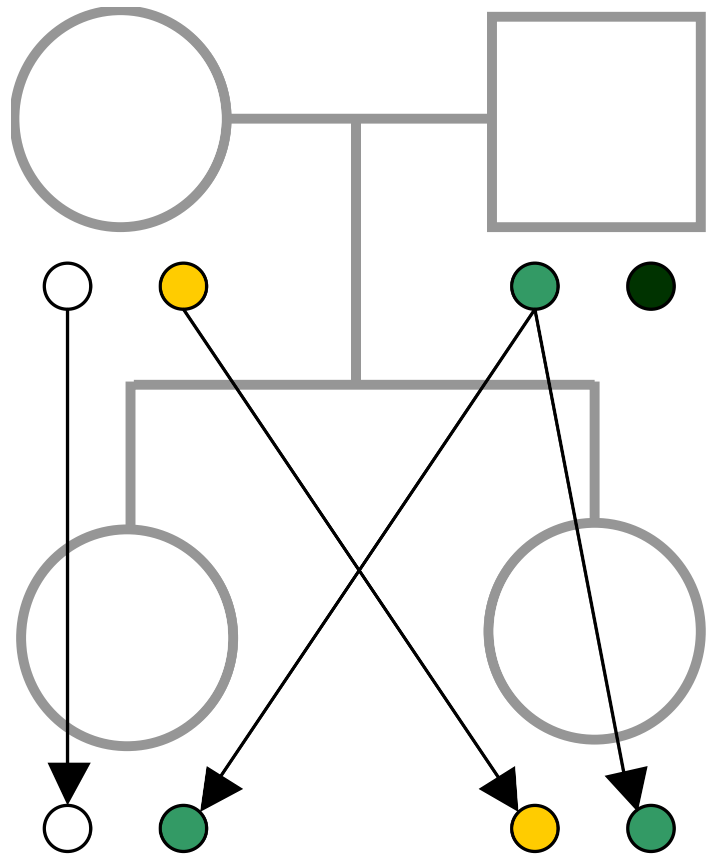
- For each family
- For each marker

- Some pattern of observed genotypes



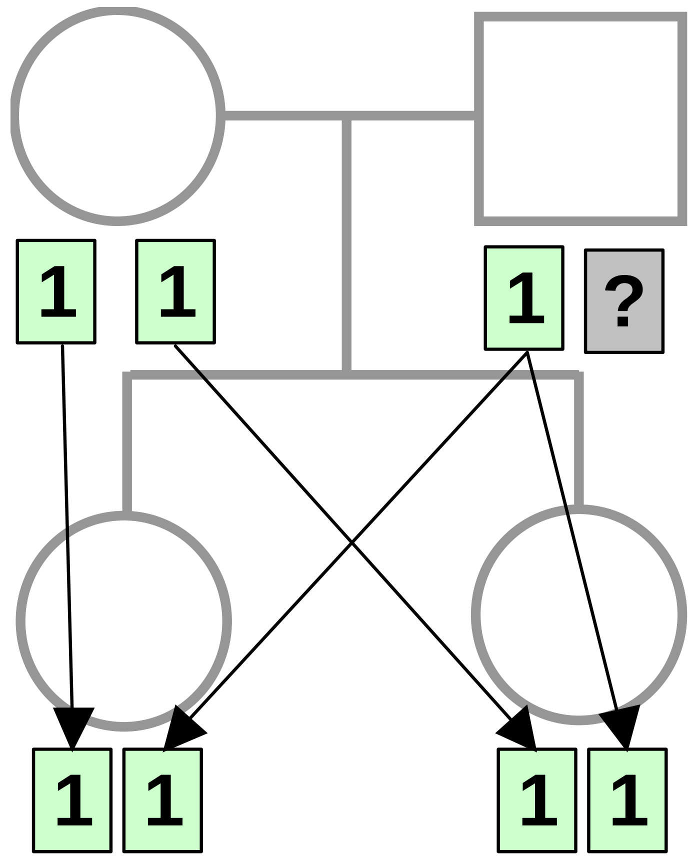
Gene flow pattern

- In turn, specify gene flow throughout the pedigree
- For each individual, we know precisely what founder allele they carry



Combine the two…

- Conditional on gene flow…
- Founder allele states are restricted
  - In this case, there is only one founder allele set: {1, 1, 1, ?}
- Likelihood is a product of allele frequencies
  - P(allele 1)^3 P(any allele)



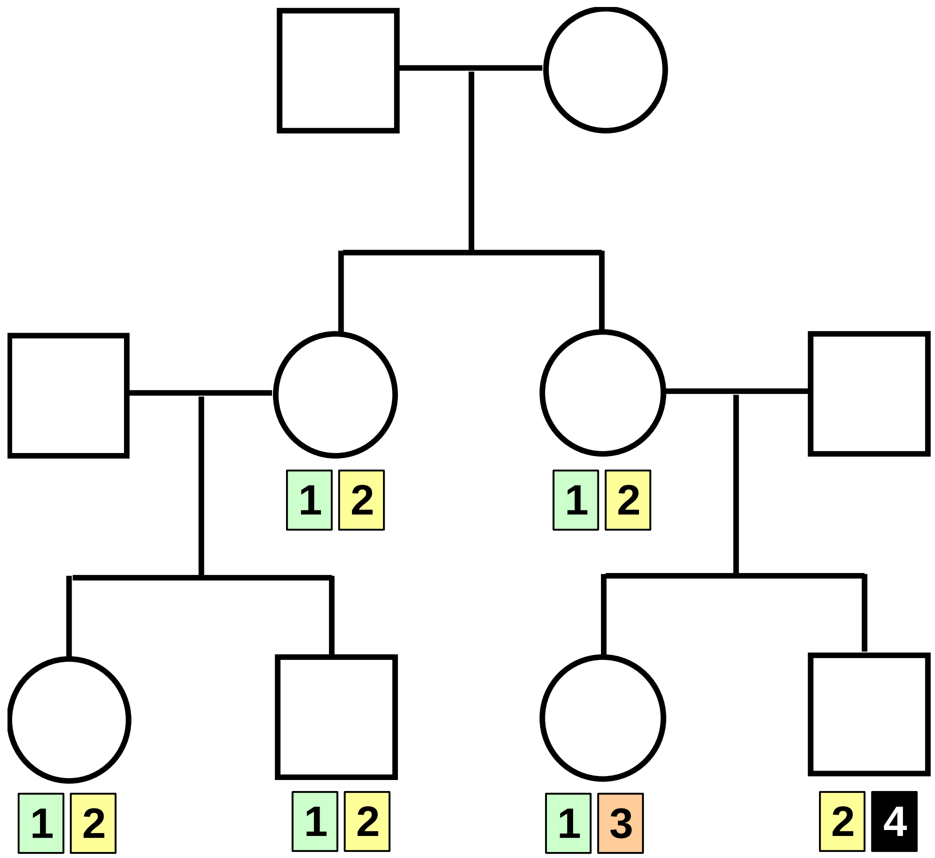
Finding founder allele sets

- Group founder alleles transmitted to the same genotyped individuals
- If a founder allele passes through an homozygote or different heterozygotes
  - Its state is either fixed or impossible
  - Fixes state of other alleles in the group

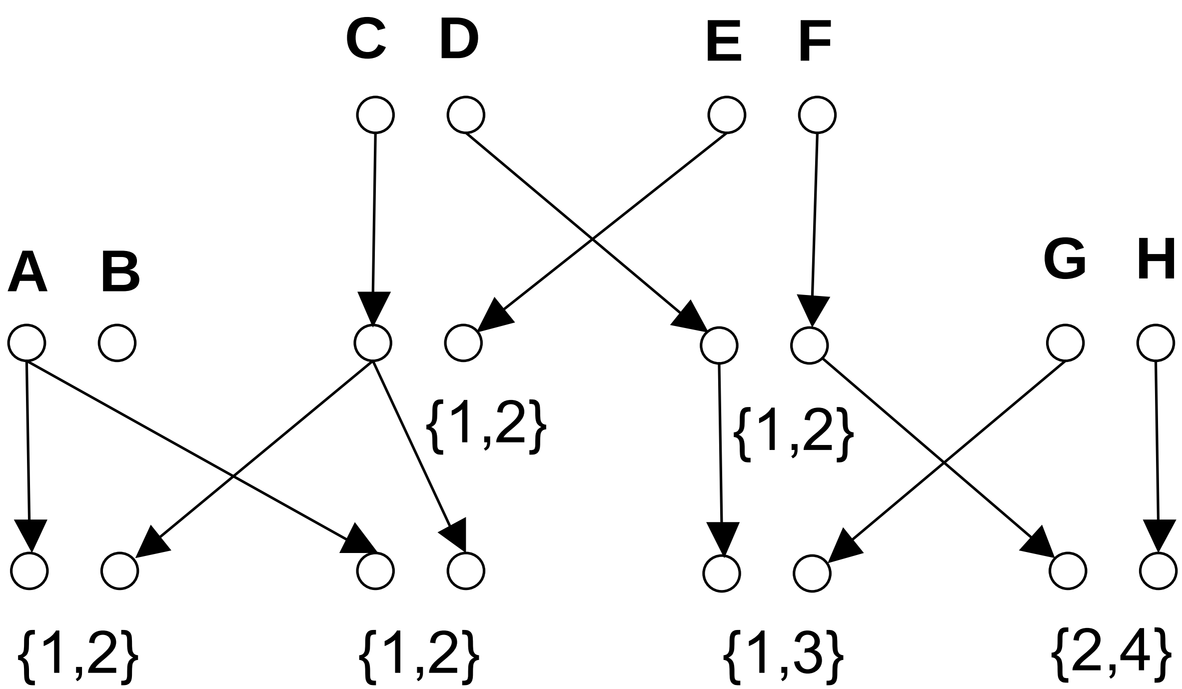
No. of Possible States for Grouped Founder Alleles

- No compatible states
- One Possible State
  - If >1 founder allele passes through a person with a homozygous genotype or two heterozygous persons
- Two Possible States For Each Allele
  - Observed genotypes are all identical and heterozygous
- Every marker allele is possible
  - Founder alleles that are not passed to genotyped persons

Example: Observed Genotypes



Example …
Descent Graph

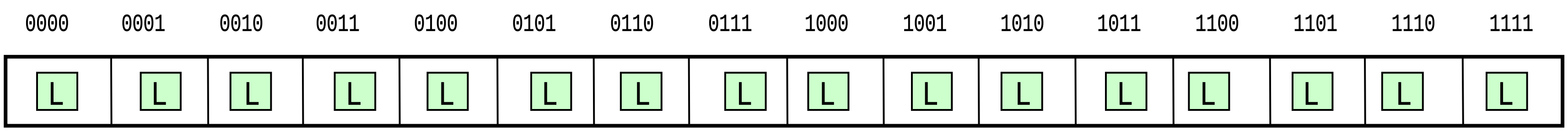


Possible founder allele states…

Founder Alleles in Group	Corresponding Allele States	Probability
(B)	(any allele)	1
(A,C,E)	(1,2,1) or (2,1,2)	$P(1)^2P(2)+P(2)P(1)^2$
(D,F,G,H)	(1,2,3,4)	$P(1)P(2)P(3)P(4)$

Lander-Green inheritance vector

- 2^{2n} elements
- Meiotic outcomes specified in index bit
- Stores probability of genotypes for each set of meiotic outcomes



So far ...

- Generalized the “IBD vector”
- Probability of observed genotypes
- Next step: Transition probabilities
  - HMM to combine information along the genome

Part III

Transition Probabilities

Recombination Fraction

Changes in IBD Along Chromosome

With one meiosis

$$T = \begin{bmatrix} (1-\theta) & \theta \\ \theta & (1-\theta) \end{bmatrix}$$

With two meioses

$$T^{\otimes 2} = \begin{bmatrix} (1-\theta)T & \theta T \\ \theta T & (1-\theta)T \end{bmatrix}$$

With two meioses

$$T^{\otimes 2} = \begin{bmatrix} (1-\theta)^2 & (1-\theta)\theta & \theta(1-\theta) & \theta^2 \\ (1-\theta)\theta & (1-\theta)^2 & \theta^2 & \theta(1-\theta) \\ \theta(1-\theta) & \theta^2 & (1-\theta)^2 & (1-\theta)\theta \\ \theta^2 & \theta(1-\theta) & (1-\theta)\theta & (1-\theta)^2 \end{bmatrix}$$

With three meioses

$$T^{\otimes 3} = \begin{bmatrix} (1-\theta)T^{\otimes 2} & \theta T^{\otimes 2} \\ \theta T^{\otimes 2} & (1-\theta)T^{\otimes 2} \end{bmatrix}$$

With three meioses

$$T^{\otimes 3} = \begin{bmatrix}
(1-\theta)^3 & (1-\theta)^2\theta & (1-\theta)^2\theta & \theta^2(1-\theta) & (1-\theta)^2\theta & \theta^2(1-\theta) & \theta^2(1-\theta) & \theta^3 \\
(1-\theta)^2\theta & (1-\theta)^3 & \theta^2(1-\theta) & (1-\theta)^2\theta & \theta^2(1-\theta) & (1-\theta)^2\theta & \theta^3 & \theta^2(1-\theta) \\
(1-\theta)^2\theta & \theta^2(1-\theta) & (1-\theta)^3 & (1-\theta)^2\theta & \theta^2(1-\theta) & \theta^3 & (1-\theta)^2\theta & \theta^2(1-\theta) \\
\theta^2(1-\theta) & (1-\theta)^2\theta & (1-\theta)^2\theta & (1-\theta)^3 & \theta^3 & \theta^2(1-\theta) & \theta^2(1-\theta) & (1-\theta)^2\theta \\
(1-\theta)^2\theta & \theta^2(1-\theta) & \theta^2(1-\theta) & \theta^3 & (1-\theta)^3 & (1-\theta)^2\theta & \theta^2(1-\theta) & (1-\theta)^2\theta \\
\theta^2(1-\theta) & (1-\theta)^2\theta & \theta^3 & \theta^2(1-\theta) & (1-\theta)^2\theta & (1-\theta)^3 & (1-\theta)^2\theta & \theta^2(1-\theta) \\
\theta^2(1-\theta) & \theta^3 & (1-\theta)^2\theta & \theta^2(1-\theta) & \theta^2(1-\theta) & (1-\theta)^2\theta & (1-\theta)^3 & (1-\theta)^2\theta \\
\theta^3 & \theta^2(1-\theta) & \theta^2(1-\theta) & (1-\theta)^2\theta & (1-\theta)^2\theta & \theta^2(1-\theta) & (1-\theta)^2\theta & (1-\theta)^3
\end{bmatrix}$$

In general ...

- Transition matrix is patterned
- Transition probability depends on:
  - No. of meiosis were outcome changed
  - No. of meiosis were outcome did not change
- Product of powers of θ and $(1 - \theta)$

Recursive Formulation

$$T^{\otimes n+1} = \begin{bmatrix} (1-\theta)T^{\otimes n} & \theta T^{\otimes n} \\ \theta T^{\otimes n} & (1-\theta)T^{\otimes n} \end{bmatrix}$$

Lander-Green Markov Model

- Transition matrix $\mathbf{T}^{\otimes 2n}$

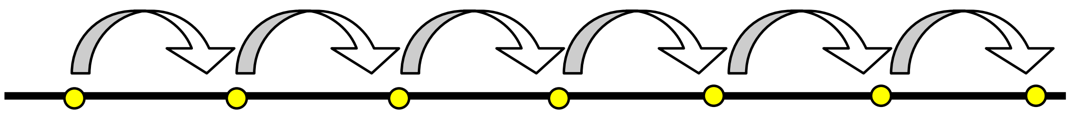
$$\mathbf{T} = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix}$$

- $\mathbf{v}_{\ell|1..\ell} = \mathbf{v}_{\ell\text{-}1|1..\ell\text{-}1}\, \mathbf{T}^{\otimes 2n} \circ \mathbf{v}_{\ell}$
- $\mathbf{v}_{\ell|\ell..m} = \mathbf{v}_{\ell+1|\ell+1..m}\, \mathbf{T}^{\otimes 2n} \circ \mathbf{v}_{\ell}$
- $\mathbf{v}_{\ell|1..m} = (\mathbf{v}_{1..\ell\text{-}1}\, \mathbf{T}^{\otimes 2n}) \circ \mathbf{v}_{\ell} \circ (\mathbf{v}_{\ell+1..m}\, \mathbf{T}^{\otimes 2n})$

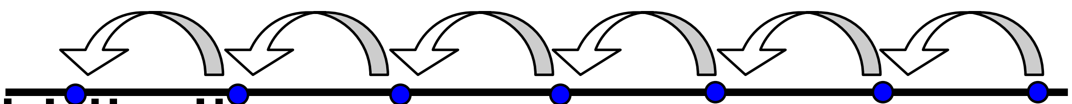
All The Ingredients To …

- Single Marker

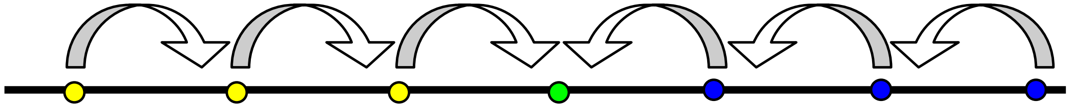
- Left Conditional



- Right Conditional



- Full Likelihood



Appropriate Problems

- Large number of markers
  - Analysis of >5,000 markers possible

- Relatively small pedigrees
  - 20-30 individuals
  - 2x larger pedigrees for the X chromosome. Why?

So far …

- Key components for Lander-Green

- Extending definition of IBD vector
- Probability of genotypes given IBD
- Transition probabilities

- Next: Practical applications!

Lander-Green Algorithm

$$L = \sum_{I_1} \cdots \sum_{I_m} P(I_1) \prod_{i=2}^{m} P(I_i \mid I_{i-1}) \prod_{i=1}^{m} P(X_i \mid I_i)$$

- More general definition for I, the "IBD vector"
- Probability of genotypes given “IBD vector”
- Transition probabilities for the “IBD vectors”

Reading

- Historically, two key papers:
  - Lander and Green (1987)
    PNAS **84**:2363-7
  - Kruglyak, Daly, Reeve-Daly, Lander (1996)
    Am J Hum Genet **58:**1347-63