



Day 3. Estimating male fitness

Mario Vallejo-Marín
University of Stirling

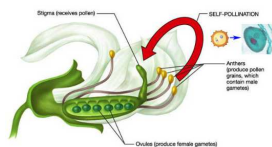
- Fitness in hermaphrodites: male and female functions
- Male fitness: number of sired seeds
- Seeds can be sired on the same individual (selfing) or on different individuals (outcrossing)
- Estimating seed paternity
 - (1) Estimating the proportion of seeds sired through self-fertilisation
 - (2) Outcrossing. How to identify fathers in outcrossed seeds?
- Indirect methods: Pollen dyes and nanodots
- Visible genetic markers, e.g., flower colour, seed flavour...
- Direct genetic methods
 - 2.1 Paternity exclusion methods
 - 2.2 Maximum likelihood methods
- Examples and limitations

Sexual fitness

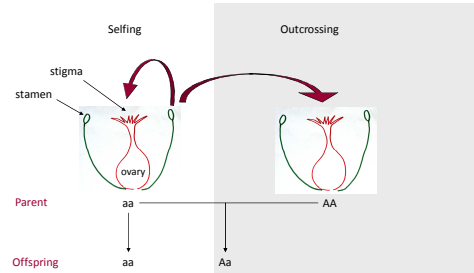
- Life-time offspring production (number of gene copies transmitted to next generation)
- Seed quantity and seed quality
- Seed quality: Affected by genetic constitution and inbreeding history
- Inbreeding depression

Sexual fitness in hermaphrodites

- Multiple pathways for reproduction in hermaphrodites:
 - Male and female pathways to fitness
 - In self-compatible species: Selfing and outcrossing



Different genetic outcomes of selfing and outcrossing



Lower genetic quality of inbred offspring

- Inbred offspring is more likely to be homozygous
- Deleterious alleles tend to be recessive

- Inbreeding depression:
 - Reduction in fitness of inbred offspring relative to outbred

$$\delta = 1 - \frac{\text{selfed}}{\text{outcrossed}}$$

- Widespread in nature



Estimating seed paternity

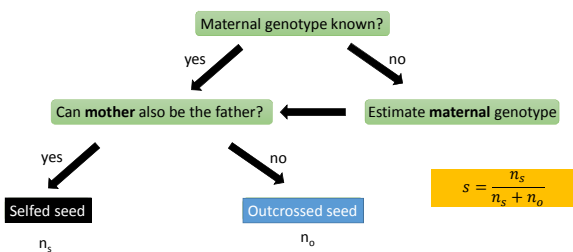
- Female fitness: Count seed numbers and weight them by seed quality
- Male fitness: Estimate individual contribution to seed paternity
- (1) Estimating the proportion of seeds sired through self-fertilisation
- (2) Identify fathers in outcrossed seeds

Seeds sired via selfing

- Selfing and outcrossing rates
 - s = Proportion of seeds produced by self fertilisation
 - t = Pr. of seeds produced through outcrossing
- $$s + t = 1 \qquad t = 1 - s$$
- Calculating s

Calculating the selfing rate (s)

- Genotype of the offspring



Calculating the selfing rate (s)

- Some mother-offspring combinations are easy to classify:

- Mom = AA; offspring = Aa paternal allele

Paternity exclusion

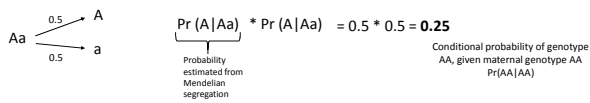
- Other combinations are ambiguous:

- Mom = AA; offspring = AA
- Mom = Aa; offspring = AA or Aa

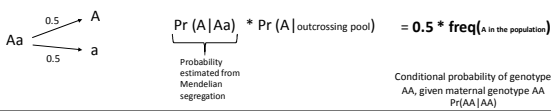
- Mom = Aa; offspring = AA or Aa

Ritland (2002)

Case 1. if offspring is produced by **selfing** (Mom = Aa; offspring = AA):



Case 2. if offspring is produced by **outcrossing** (Mom = Aa; offspring = Aa):



Maximum likelihood estimation of s

Ritland (2002)

$$s \prod_{loci} p_{kl}^{ij,s} + t \prod_{loci} p_{kl}^{ij,t}$$

$p_{kl}^{ij,s}$ = Pr. of progeny genotype $A_i A_j$, given that parent is $A_i A_j$ and self-fertilises

$p_{kl}^{ij,t}$ = Pr. of progeny genotype $A_i A_j$, given that parent is $A_i A_j$ and outcrosses

- Maximum likelihood estimates the values of s and t given the genotypes of mother and offspring

MLTR: A ML program to estimate s

Ritland (2002)

Multilocus Estimates

Parental F estimate (SD) = 0.435 (0.096)
Multilocus t estimate (SD) = 0.069 (0.015)
Single locus t estimate (SD) = 0.043 (0.013)
Difference tm-ts (SD) = 0.025 (0.008)
Correlation of t (or s) estimate (SD) = 0.027 (0.290)
Multilocus correlation of p estimate (SD) = 0.999 (0.088)
Single locus correlation of p estimate (SD) = 0.999 (0.633)
Difference [p(s)-p(m)] of p estimate (SD) = 0.000 (0.630)
Correlation of s (or t) among loci (SD) = 0.655 (0.135)

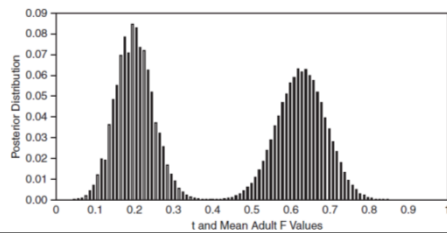
Most likely maternal genotypes:

Family	55b4GA-015b4GA-05b4GH-045b4GH-045b5-236	5b5-236	5b4GE-015b4GE-015b4-22	5b4-22						
Fam11	2	2	2	8	7	7	1	1		
Fam12	6	7	4	4	1	2	6	8	2	2
Fam13	2	10	3	3	8	1	2	5	2	2
Fam14	6	6	4	4	3	10	6	6	2	2

BORICE: A Bayesian approach to estimate s

Koelling et al. (2012)

- Robust to small sample sizes (families <8 offspring)
- Does not calculate correlated paternity



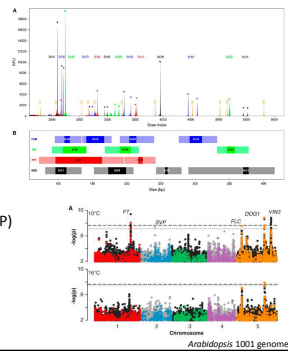
Parentage analysis

Jones et al. (2009)

- Parentage: Maternal and paternal parent estimation
- Paternity analysis: Male parent estimation
- Identifying parents of individual offspring
- Infer parentage based on the genotype of single individuals or groups of individuals

Genetic markers

- Microsatellites
- Single Nucleotide Polymorphism (SNP)



Six approaches to parentage analysis

1. Exclusion
2. Categorical allocation
3. Fractional allocation
4. Full probability parentage analysis
5. Parental reconstruction
6. Sibship reconstruction

1. Exclusion

- Eliminate candidate parent who fails to share at least one allele with offspring

Offspring = AA
 Mom = AA
 Father 1: AA
~~Father 2: aa~~
 Father 3: Aa

- Genotyping error can be accounted for by increasing the required number of mismatches

Offspring = AAbb
 Mom = AAbb
 Father 1: AABb
~~Father 2: aaBB~~
 Father 3: AaBB

2. Categorical allocation

Meagher (1986)

- Excluding all but one parent may not be possible
- How to choose among the remaining candidates?
- Statistical selection of most likely parent among pool of candidates
- Most commonly used
- Maximum likelihood
- Bayesian approach (posterior probability)

Maximum Likelihood

- Probability of the data given the model
 $P(\text{data}|\text{model})$
- Likelihood of the data given the model (parameters)
 $Likelihood \propto P(\text{data}|\text{parameters})$
- Probability:
Knowing the parameters -> Prediction of the outcome
- Likelihood:
Observation of the data -> Estimation of the parameters

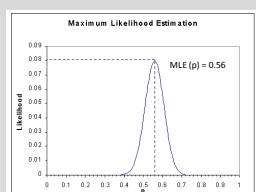
Maximum Likelihood

- $n = 100$ coin tosses
- $k = 56$ heads

$$L(\text{data}|p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$L(\text{data}|0.5) = \frac{100!}{56!44!} 0.5^{56} 0.5^{44} = 0.0389$$

$$L(\text{data}|0.52) = \frac{100!}{56!44!} 0.52^{56} 0.48^{44} = 0.0581$$



2. Categorical allocation: Maximum Likelihood

- Is parent j the most likely parent of a given offspring?

- H1: The probability that individual i is the father
- H2: The probability that parent is randomly chosen from the population
- Calculate the likelihood ratio (H1/H2)

$$L(H_1, H_2 | D) = \frac{P(D | H_1)}{P(D | H_2)}$$

Mendelian probability of offspring genotype, given maternal and paternal genotypes

$$L(H_1 | g_m, g_a, g_o) = T(g_o | g_m g_a) \cdot P(g_m) \cdot P(g_a)$$

Frequency of alleged father's genotype

Mendelian probability of offspring genotype, given maternal genotype ONLY

$$L(H_2 | g_m, g_a, g_o) = T(g_o | g_m) \cdot P(g_m) \cdot P(g_a)$$

Frequency of alleged father's genotype in the population

g_m = maternal genotype
 g_a = alleged father genotype
 g_o = offspring genotype

Mendelian probability of offspring genotype, given maternal and paternal genotypes

$$L(H_1, H_2 | g_m, g_a, g_o) = \frac{T(g_o | g_m g_a) \cdot P(g_m) \cdot P(g_a)}{T(g_o | g_m) \cdot P(g_m) \cdot P(g_a)} = \frac{T(g_o | g_m g_a)}{T(g_o | g_m)}$$

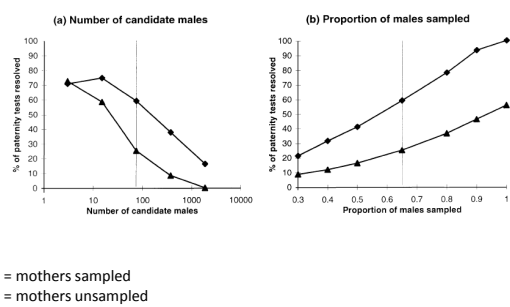
g_m = maternal genotype
 g_a = alleged father genotype
 g_o = offspring genotype

Table 1 Likelihood ratios for all compatible mother-alleged father-offspring trios. X represents any allele other than B; Y represents any allele that is neither B nor C. The frequencies of alleles B and C are denoted p and q . The likelihood ratio, $L(H_1/H_2)$, is the probability of the offspring's genotype given the mother's and alleged father's genotypes, $T(g_o | g_m g_a)$, divided by the probability of the offspring's genotype given the mother's genotype, $T(g_o | g_m)$. A similar table is shown in more condensed form in Brenner (1997)

Offspring's genotype (g_o)	Alleged father's genotype (g_a)	Mother's genotype (g_m)	$T(g_o g_m g_a)$	$T(g_o g_m)$	$L(H_1/H_2)$
BB	BB	BB	1	p	$1/p$
BB	BX	BB	$1/2$	p	$1/2p$
BB	BX	BX	$1/2$	$1/2$	$1/p$
BB	BX	CC	$1/4$	$1/2$	$1/2p$
BC	BB	CC	1	p	$1/p$
BC	BX	CC	$1/2$	p	$1/2p$
BC	BB	CY	$1/2$	$1/2$	$1/p$
BC	BX	CY	$1/4$	$1/2$	$1/2p$
BC	BB	BC	$1/2$	$(p+q)/2$	$1/(p+q)$
BC	BX	BC	$1/4$	$(p+q)/2$	$1/2(p+q)$
BC	BC	BC	$1/2$	$(p+q)/2$	$1/(p+q)$

Marshall et al. (1998)

- Likelihood ratios (LOD scores) for each alleged father are ranked, and the difference between the most likely and second most likely are compared
- If the difference is large enough, the offspring is assigned to the most likely father
- Categorical assignment:
Each offspring is assigned to ONE father



2. Categorical allocation: Bayesian approach

- Calculate the probability that potential father k is the actual father of offspring i

Bayesian statistics



- Bayesian probabilities: degree of belief
- "Probability is orderly opinion, and the inference of data is the revision of such opinion in the light of new information."

$$P(\text{parameters}|\text{data}) = \frac{P(\text{data}|\text{parameters}) * P(\text{parameters})}{P(\text{data})}$$

posterior
likelihood
prior

2. Categorical allocation: Bayesian approach

$$P(I_k(i)|\mathbf{M}_i, \mathbf{F}_k, \mathbf{A}, N) = \frac{P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{F}_k)}{\sum_{j=1}^n P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{F}_j) + (N-n)P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{A})}$$

Posterior probability that offspring i is sired by father k
Pc of offspring genotype given mother i and father k

Pc offspring summed over all fathers
Unsampled males
Pc offspring sired by randomly sampled male

\mathbf{O}_i , \mathbf{M}_i , and \mathbf{F}_i = multilocus genotypes of offspring i , known mother and alleged father
 \mathbf{A} = Matrix of frequencies at all loci
 n = number of males sampled out of N total males in breeding population

- Compare posterior probabilities for each alleged father, choose the most probably father
- Categorical assignment:
Each offspring is assigned to ONE father

3. Fractional allocation of parentage

- Each offspring is partially allocated to potential fathers in proportion to their siring probability
- Can be calculated with ML or Bayesian approaches
- Seldom used

4. Full probability analysis

- Elaboration of the Bayesian estimation of parentage
- Estimates both parentage and other parameters of interest simultaneously
- Allows incorporating uncertainty when estimating paternity into the estimation of other parameters of interest
- Example:
Estimate if flower size influences siring success

- Categorical or fractional analysis:

1. Assign parentage to each offspring
2. Analyse the relationship between paternal phenotype (e.g., flower size), and number of offspring sired

Two separate statistical models

- Full probability analysis:

1. Build a single model to estimate paternity and relationship with phenotype

Single model

$$P(I_k(i)|\mathbf{M}_i, \mathbf{F}_k, \mathbf{A}, N, z_k) \propto \underbrace{f(\theta, z_k)}_{\text{Function relating phenotype } z \text{ to paternity (flower size)}} * \frac{P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{F}_k)}{\sum_{j=1}^n P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{F}_j) + \bar{z}(N-n)P(\mathbf{O}_i|\mathbf{M}_i, \mathbf{A})}$$

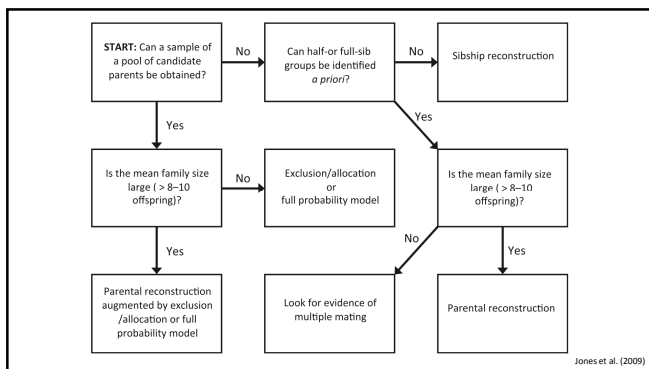
5,6. Parental and sibship reconstruction

Parental reconstruction

- Reconstruct maternal or paternal genotypes
- Requires family groups
- Most effective when >8-10 siblings are available

Sibship reconstruction

- Family groups are unknown
- Use genetic data to partition individuals into family groups



Software

- Mating system:
 - MLTR. <http://genetics.forestry.ubc.ca/ritland/programs.html>
 - Borice. <https://eeb.ku.edu/john-kelly>
- Paternity:
 - Cervus. http://www.fieldgenetics.com/pages/aboutCervus_Overview.jsp
 - Colony. <https://www.zsl.org/science/software/colony>
 - ColonyR. <https://github.com/mdjbru-R-packages/colonyR>
 - MasterBayes. <http://jarrod.bio.ed.ac.uk/software.html>

Future challenges

Patterns of pollen flow using fluorescent markers

- Can fluorescent powder be used to track pollen in flowers with poricidal anthers?
- Assessing safe-sites in pollinators
- Points of contact of anthers, stigma, and pollinators

Simulating outcrossing rates
