

Kinship Analysis

Inferring genealogical relationships among individuals within populations

Relatedness analysis

- estimation of the degree of relatedness (r) between pairs of individuals
- assignment of dyads to specific pedigree relationship categories (e.g. half-sib, avuncular, cousins)

Parentage analysis

- a unique application of relatedness analysis in which one searches among candidates for the most likely parents of a target offspring

What can we learn:

- Heritability estimates in wild populations
- Inferring gene flow into a population
- Estimating N_e & total # of breeders
 - population viability/maintenance of genetic variation
- Preventing inbreeding in captive populations
- Quantify variance in reproductive success
 - inferences about social structure & natural selection
- Adjust allele frequency estimates for the presence of related individuals in a sample

Three approaches:

1. Degree of relatedness
2. Category assignment
3. Partitioning

Relatedness estimators

- What is being ‘measured’?
 - **IBD** (identity by descent): *recently* descended from a single ancestral allele
- How to quantify it?
 - f_{ij} : is the probability that two alleles, one chosen randomly from each individual, are IBD.
 - r : ‘genome wide version of f_{ij} ’ – expected fraction of alleles shared by IBD – genetic similarity between two individuals *relative to two randomly chosen individuals from the same reference population*

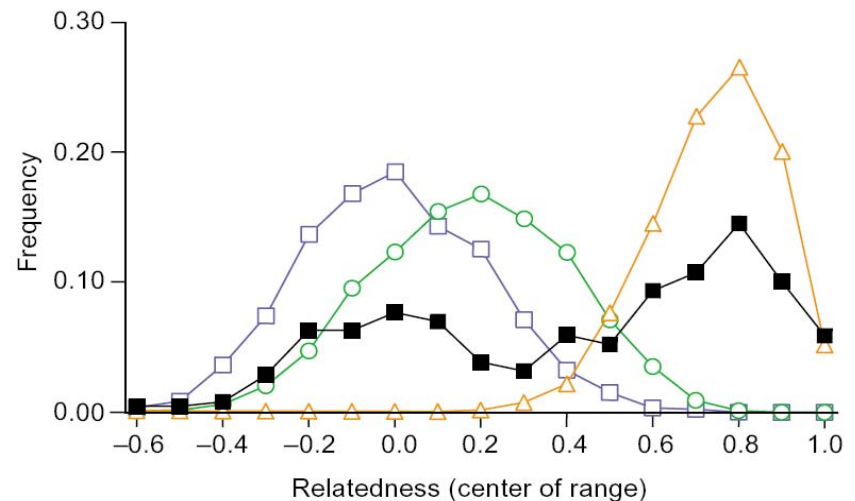
Important properties of r

- Continuously distributed
- Doesn't have to be symmetrical between individuals
- Many different algorithms for estimating:
 - Similarity index (Li et al. 1993 & Wang 2002)
 - Lynch-Ritland (1999)
 - Queller-Goodnight (1989)
- High variances = need >tens of microsats and many more SNP's to estimate r for dyads

Specific Questions:

- Three levels of query:
 - How related are individual X and Y? (repeat for all possible dyads)
 - What is the average relatedness between individual in this group?
 - What fraction of two more relationship categories is the group comprised of?

Queller, D.C. et al. (2000) Unrelated helpers in a social insect. *Nature* 405, 784–787



Assignment to Categories

- Defining Categories

- IBD coefficients:

- k_0 , $k_1(\Phi)$, $k_2(\Delta)$ are the probability that a dyad share 0, 1 or 2 alleles by IBD respectively.

Relationship category	k_0	k_1	k_2	r
Monozygotic twins or self	0	0	1	1
Parent-offspring	0	1	0	0.50
Full sibs	0.25	0.50	0.25	0.50
2° (e.g. half sibs, avuncular)	0.50	0.50	0	0.25
3° (e.g. first cousins)	0.75	0.25	0	0.125
Unrelated	1	0	0	0

For an outbred individual: expected $r = \Phi/2 + \Delta$

Assignment to Categories

- Restructuring the question:

- what is the probability that a dyad has the observed multilocus genotypes, given that they belong to a particular relationship category?
- Requires info of allele frequencies

Table I. Probabilities that a dyad has a pair of genotypes given that they share m alleles identical by descent

Genotype pair	Probabilities ^a		
	$m = 0$	$m = 1$	$m = 2$
$a_i a_i \ a_i a_i$	p_i^4	p_i^3	p_i^2
$a_i a_j \ a_i a_i$	$2 p_i^3 p_j$	$p_i^2 p_j$	0
$a_i a_i \ a_j a_j$	$p_i^2 p_j^2$	0	0
$a_i a_j \ a_i a_j$	$4 p_i^2 p_j^2$	$p_i p_j (p_i + p_j)$	$2 p_i p_j$
$a_i a_j \ a_j a_i$	$4 p_i p_j^2 p_i$	$p_i p_j p_i$	0
$a_i a_j \ a_i a_i$	$2 p_i p_j p_i^2$	0	0
$a_i a_j \ a_k a_i$	$4 p_i p_j p_k p_i$	0	0

^a p_i, p_j, p_k and p_l are the population allele frequencies of alleles a_i, a_j, a_k and a_l , respectively.

$$P(\text{data}|\text{model}) = P(\text{G}|\text{R}) = k_0 P_0 + k_1 P_1 + k_2 P_2$$

Methods: Maximum Likelihood

- Three types of question:
 - Given no hypotheses or other information about a dyad, what is their most likely relationship?
 - Test competing hypotheses using likelihood ratio.
 - Which of several individuals are most likely to have a given relationship (e.g. parentage or partitioning a cohort into sib-ships).

Methods: Allele Sharing

- Question: Does the # of shared alleles differ from that expected under the null hypothesis?
 - Generate expected distribution
 - number or proportion shared (r)
 - Define cut-off points
- Advantages:
 - Insensitive to genotyping errors
 - Fast analysis
 - Does not require specification of alternative hypothesis – two-tailed test.
- Disadvantage:
 - Low power relative to ML where alt. hypothesis is given.

Partitioning

Sib-ship partitioning

- Graphical - works well but subjective
- Maximum Likelihood of *each* alternative partition
 - Only possible w/ small samples
 - Allele frequency estimates unreliable
- MCMC to sample from the distribution of likelihoods
 - Allele frequencies can be iteratively re-estimated
 - Depends on assumptions of family size

Why?

- Estimate: QG parameters, family number`, variance in family size

Assumptions/Issues

- Hardy–Weinberg and gametic phase equilibrium
- Inbreeding
- Genotyping error and mutation
- Power to detect differences:
 - Depends on the # loci and r
 - Unlinked loci more powerful than same # of linked loci except...?
 - Not all relationships have different k 's so ML estimates of these are mainly useful for hypothesis generation.
 - If you don't include the right hypothesis ML ratios are less useful.

Genotyping Error & Mutation

- Typical rates of 0.25-2%
- Consequences:
 - Minimal effects except in the cases of:
 - Parent-offspring pairs, monozygotic twins, and triads where can cause likelihood to --> zero.
- Accounted for by
 - Assuming fixed error rate - all relationships have a non-zero likelihood
 - Specific models of mutation probabilities (small step vs. large step)

Blouin, S. 2003 TREE

DNA-based methods
for pedigree
reconstruction and
kinship analysis in
natural populations

Table 1. Software for implementing methods discussed in the text

Program	Description	Comments/Limitations	Web site	Refs
Relatedness RELATEDNESS 5.0	Pairwise or group average r via Queller–Goodnight method; Symmetrical or asymmetrical estimates for pairs; Standard errors via re-sampling over groups or over loci	Macintosh only; User friendly; Assumes unlinked loci	http://www.gsoftnet.us/GSoft.html	[32]
KINSHIP	Expected distribution of Queller–Goodnight r from simulated dyads	Macintosh only; User friendly; Assumes unlinked loci	http://www.gsoftnet.us/GSoft.html	[41]
DELRIIOUS	Pairwise r and Δ via Lynch–Ritland method; Standard errors via re-sampling over loci	Requires Mathematica software; Assumes unlinked loci	http://www.zoo.utoronto.ca/stone/delrious/delrious.htm	[33]
MER	r , Φ and Δ via Wang method	Only does one pair at a time; Assumes unlinked loci	http://www.zoo.cam.ac.uk/oz/software.htm	[30]
Likelihood of belonging to relationship category KINSHIP ^a	Estimates likelihood that a dyad belongs to a specified category of relationship; Likelihood ratio test for specified alternate hypotheses; Significance test via simulation	Unlinked loci only; No genotyping error; Flexible method for specifying any possible relationship	http://www.gsoftnet.us/GSoft.html	[41]
REPAIR 2.0	Estimates likelihood of specified relationship for each dyad, accounting for linkage among loci; Allele sharing statistics accounting for linkage	Incorporates X-linked loci; Accounts for genotyping error; Accepts putative pedigrees with input file; Eight possible relationships specifiable	http://www.sph.umich.edu/statgen/boehnke/repair.html	[36]
PREST	Estimates likelihood of specified relationship for each dyad, accounting for linkage among loci; Likelihood ratio test of specified null relationship versus a specified alternative or versus the most likely of ten possible alternative relationships; Significance estimated via simulation; Calculates maximum likelihood value of the IBD coefficients for each dyad; Tests of relationship via allele sharing statistics, accounting for linkage	Accounts for genotyping error only with parent-offspring and monozygotic twin pairs; Accepts putative pedigrees with input file; Eleven possible relationships specifiable	http://galton.uchicago.edu/~mcpeek/software/prest/	[37]
ECLIPSE (PANGAEA package)	Estimates likelihood of specified relationship for trios, accounting for linkage; Given a known pedigree, can also be used to identify mis-scored loci rather than to estimate likelihood of relationship	Accepts putative pedigrees with input file; Accounts for genotyping error	http://www.stat.washington.edu/thompson/Genepi/pangaea.shtml	[46]
Partitioning cohorts into sibships BOREL (PANGAEA package)	Exhaustive likelihood evaluation of sibship partitions		http://www.stat.washington.edu/thompson/Genepi/pangaea.shtml	[6]
Almudevar & Field methods	Sibship partitions via enumeration of genetically compatible groups		http://ace.acadiau.ca/~aalmudev/pedigree.htm	[58,68]
Thomas & Hill methods	Sibship partitions via likelihood evaluated with MCMC methods	Full sib or nested half sib (e.g. polygamous males, monogamous females) partitions	Software under development; Contact S. C. Thomas for further information (sthomas@srv0.bio.ed.ac.uk)	[2,13]
Smith <i>et al.</i> methods	Sibship partitions via likelihood evaluated with MCMC methods	Two methods available; Advantages of each method depend on details of the data set; Full sib partitions only	Software under development; Contact C. Herbinger for further information mail: (christophe.herbinger@dal.ca)	[12]
PARENTAGE 1.0	Full or half sib partitions via likelihood evaluated with MCMC methods; Infers parental genotypes or contributions per parent to a cohort; Estimates mutation rate	Conveniently incorporates any prior information on family sizes or numbers, and on parental identities or numbers; Accounts for genotyping error; Very flexible	http://maths.abdn.ac.uk/~ijw/	[57]

Other Resources

Programs:

GEURUD 2.0

Jones, A. G. 2005. Molecular Ecology Notes.

Literature:

Book: Molecular Markers, Natural History & Evolution (Avice, J. 2004)