

Parentage Analysis

CERVUS 3.0

Available at: http://www.fieldgenetics.com/pages/aboutCervus_Overview.jsp

Cervus file format: Cervus files are simple to make! A Cervus file is a tab delimited text file. You can make a simple Excel file and “save as.” Each individual has one line of data and the alleles at each locus are separated. An example is below.

	A274B		B278B		C262B		SC23	
G1	272	275	186	202	222	234	272	318
G2	265	265	172	204	186	206	318	514
G3	272	293	204	242	218	238	360	380
G4	293	317	186	192	222	222	328	334

Cervus 3.0 is supposed to convert Genepop and Genetix files to Cervus format. However, I can not get this option to work. The program authors are aware of this and working on an update for this newest version to correct for this and other bugs.

Types of Input Files:

- **Offspring File:** This file designates which individuals are the offspring in question.
 - Must include the offspring ID (and can be a file that is ONLY the offspring ID)
 - Can also include: ID of a known parent (if available), the offspring genotype, a list of candidate parents
- **Candidate Parent File:** This file is a list of the potential parents of the offspring. It can also include demographic/biological information about each individual.
 - This file is not necessary if you include potential parents in your offspring file.
 - If you wish to include other information such as age or gender, you need this file (can't do that in the offspring file)
 - If all offspring have the same set of potential parents, this file can be a simple list of the parents' IDs in one column.
- **Genotype File:** This file includes genotypes of all individuals (parents and offspring).
 - Make sure the individual IDs are present and match the IDs given to the parents/progeny in the offspring and/or candidate parent file.

Analysis Steps:

- Calculate Allele Frequencies: The allele frequencies are critical to the simulations and later, parentage analyses you will perform.
 - When calculating allele frequencies, Cervus will also calculate: observed/expected heterozygosity, PIC value, HWE, estimate of the frequency of null alleles, and exclusion probabilities for each locus.
 - If you already have this information, you can input an “allele frequency” file. However, I find it just as easy or easier to let Cervus do it – and you also get the other summary statistics as a bonus.
- Simulation: The simulation is how Cervus assesses confidence of assignment when you do the actual parentage analysis.
 - Maternity analysis
 - Paternity analysis
 - Parent pair analysis
 - Sexes unknown
 - Sexes known
 - This is where you input information to account for inbreeding and/or selfing
- Parentage Analysis: Uses given data to assign parents or parent pairs.
 - Input your allele frequency files along with offspring and candidate parent files
 - Choose how much output you want: scores for all potential parents, just the most likely, non-exclusion probabilities, etc.
- Identity Check: This option will check the individual IDs and genotypes to identify if any individuals names and/or genotypes that are identical.

Output Files:

- Allele Frequency:
 - On screen output:
 - Summary statistics per locus
 - Estimate of frequency of nulls
 - If > 5% - recommended to exclude that locus
 - .alf file:
 - List of allele frequencies for input into simulation/parentage analysis
- Simulation:
 - On screen output:
 - Summary of assignment rate under strict/relaxed criteria for each parent and parent pairs
 - LOD/delta distributions for each parent and parent pairs (mean and st. dev.)
 - .sim file:
 - Input file for parentage analysis

- Parentage Analysis:
 - On screen output
 - Percent assignment, number of individuals tested, etc.
 - .csv file
 - Comma-delimited file you can open in Excel
 - All LOD scores for each parent and parent pair
 - Confidence/significance

Other Options for Parentage:

FAMAZ: Gerber et al. (2002)

<http://www.pierroton.inra.fr/genetics/labo/Software/Famaz/index.html>

- Calculates both exclusion and identity probabilities, log-likelihoods, for paternity or parent-pair
- Can test for sibships (only with codominant data)
- Can incorporate error rates and user can include departure from HWE by using Wright's F
- Like Cervus, uses simulations for statistical tests
- User defines the threshold LOD score for significance
- Can handle dominant, sex-linked, mitochondrial data
- Long time to download
- Can only handle a small set of offspring at a time (~55)
- Assumptions:
 - Independent loci
 - Cytoplasmic markers fully uniparental in inheritance
 - The pop is in panmixia
 - Allele frequencies samples represent those of the entire pop (the families sampled as well as potential parents not sampled)

PROBMAX3: Danzmann (1997): <http://www.uoguelph.ca/~rdanzman/>

- Complicated input files (.pro, .par, .cro files)
- Can handle dominant and codominant data (and both at the same time)
- Need to designate parents
- Can handle nulls – new version coming out is supposed to calculate their frequency
- User can change threshold for inclusion (i.e. account for genotyping error)
- Not made to consider self-fertilization
- Output:
 - .max file: a) lists each parent name and offspring assigned to him/her, and b) lists each offspring and the number of possible parent-pairs assigned to it
 - .prb file: lists parent pairs and then lists (in a hard to read format) the probability of each offspring being from that pair
 - .lft file: gives a summary of all potential genotypes of offspring that could have been produced by each parent-pair (assumes Mendelian segregation)
 - .mor file: lists each potential parent pair next to each offspring name (like .max file but opposite)

PARENTE: Cercueil et al. (2002): <http://www2.ujf-grenoble.fr/leca/membres/manel.html>
link no longer live!

- Can perform similar tests to Cervus
- Can incorporate age information to exclude individuals too young to be parents
- Uses allele frequencies to compute probability of the expected genotype of an offspring given the observed genotypes of the parents and offspring
- Assumptions:
 - Hardy-Weinberg equilibrium
 - No linkage
 - Wright-Fisher reproduction (finite and constant population size, random mating, non-overlapping generations)

GERUD2.0: Jones (2005): <http://www.bio.tamu.edu/USERS/ajones/parentage.html>

- Data has to include offspring from only one mother and it will calculate the minimum number of fathers from the data set
- Crashed several times!
- Can not handle nulls
- Can not deal with missing data

There are MANY others!! See the review by Jones and Arden (2003) for a start.

Helpful/Interesting Lit

Danzmann, 1997. PROBMAX: A computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *J. Hered.* 88:333.

Gerber et al. 2002. Program Note - FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Mol. Ecol. Notes.* 3:470-481.

Kalinowski et al. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16:1099-1106.

Marshall et al. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7:639-655.

Jones, 2005. GERUD2.0: A computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. *Mol. Ecol. Notes.* 5:708.

Jones and Arden. 2003. Review: Methods of parentage analysis in natural populations. *Mol. Ecol.* 12:2511-2523.

Pompanon et al. 2005. Review: Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics.* 6:847-858.