

Dutt JE, 1973. A representation of multivariate probability integrals by integral transforms. *Biometrika* 60:637-645.

Hazel LN, 1943. The genetic basis for constructing selection indices. *Genetics* 28:476-490.

Meuwissen THE, 1989. A deterministic model for the optimization of dairy cattle breeding based on BLUP breeding value estimates. *Anim Prod* 49: 193-202.

Meuwissen THE, 1991. Reduction of selection differentials in finite populations with a nested full-half sib family structure. *Biometrics* 47:195-203.

Tallis GM, 1961. The moment generating function of the truncated multi-normal distribution. *J R Stat Soc B* 23:223-229.

Villanueva B, Wray NR, and Thompson R, 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim Prod* 57:1-13.

Woolliams JA and Bijma P, 2000. Predicting rates of inbreeding in populations undergoing selection. *Genetics* 154:1851-1864.

Woolliams JA, Bijma P, and Villanueva B, 1999. Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics* 153:1009-1020.

Wray NR and Hill WG, 1989. Asymptotic rates of response from index selection. *Anim Prod* 49:217-227.

Young SSY, 1964. Multi-stage selection for genetic gain. *Heredity* 19:131-143.

Received April 15, 2002

Accepted September 30, 2002

Corresponding Editor: Sudhir Kumar

## PARENTE: Computer Program for Parentage Analysis

A. Cercueil, E. Bellemain, and S. Manel

PARENTE is a user-friendly software package that conducts parentage inference using molecular data from diploid codominant markers. Based on the principle of genetic compatibility, PARENTE looks for maternity, paternity, or simultaneously for both potential parents, using multilocus genotypes and birth and death dates of individuals (if available). It also calculates the probability of successfully allocating an individual offspring to its parents. PARENTE is free and can be downloaded from: <http://www2.ujf-grenoble.fr/leca/membres/manel.html>.

### Parentage Analysis

Parentage analyses using multilocus genotypes are widely used to assess reproductive success, mating patterns, kinship, and fitness in natural popula-

tions and are gaining widespread use with the development of highly polymorphic molecular markers (Luikart and England 1999; Petrie and Kempnaers 1998; Queller et al. 1993; Sunnucks 2000). These analyses are especially useful when studying breeding systems with multiple mating, e.g., when individuals of one sex have more than one mating partner of the opposite sex. In this case, it is possible to investigate the "genetic mating system" of the species, to evaluate the actual reproductive success of males (Coltman 1999; Pemberton et al. 1992), and to construct pedigrees. Parentage analysis allows one to confirm monogamy in some species (Brotherton et al. 1997; Heller et al. 1993; Ribble 1991) and to demonstrate extrapair copulation in others (Girman et al. 1997; Goosens et al. 1998; Sillero-Zubiri et al. 1996). They are also useful when direct mating observations are difficult (Clapham and Palsboll 1997; McRae and Kovaks 1994).

The general principle of parentage analysis is to assign assumed parents to an individual based on the genetic compatibility between the candidate parents and individuals of the next generation. Many approaches, based on maximum likelihood estimations and/or Bayesian inference, have been developed recently, but they often remain theoretical and difficult for biologists to apply (Bernatchez and Duchesne 2000; Gerber et al. 2000; Neff et al. 2000a, 2000b, 2001). A few computer programs that perform parentage analysis are available (e.g., PROBMAX, Danzmann 1997; CERVUS, Marshall et al. 1998; NEWPAT, Amos, available online at <http://www.zoo.cam.ac.uk/zoostaff/amos/newpat.htm>; PATRI, Nielsen et al. 2001). Although they can take into account typing errors, mutation rates, null alleles (CERVUS), sex-linked loci (NEWPAT), or even an incomplete sampling of the candidate parents (PATRI and CERVUS), they do have some limitations. One major drawback is that most of these programs assign only one parent to an offspring, usually assuming the previous knowledge of the other parent (such situations apply only to species providing parental care). Additionally, compared with paternity analysis programs, parentage analysis software seldom provides the identification of offspring-parental pairs. PROBMAX is the only program able to find pairs of parents, but the previous knowledge of parental mating combination is required, and this is not often available

from field data. It is also possible to find both parents by running CERVUS twice—this can be a difficult and time-consuming process. Another weakness of the programs is that they are restricted to only one generation and do not take into account birth and death dates. As a consequence, they require that some individuals be defined as candidate parents and others as offspring.

The user-friendly application PARENTE (which means "parentage" in French) facilitates parentage analysis and allows consideration of frequent field situations, which is not possible with presently available software. PARENTE can determine paternity, maternity, and also both parents simultaneously (without a priori assumptions) for each offspring, since the mother is not always known from field data. In addition, our software uses the ages of individuals (if available), since it is necessary in parentage analysis to consider only individuals old enough to be potential parents.

### Program Description

#### Principle

PARENTE performs parentage analysis using data from all types of diploid codominant markers (i.e., microsatellites, SNP, allozymes, and so forth). It infers parentage for every individual in the genotype file, or for an individual of interest, considering the dates of birth and death of each one. As a consequence, only individuals old enough to reproduce and still be alive when the young are conceived are considered as potential parents. In this way, several generations from the sampled population can be considered in the same file.

First, from one single genotype file and for each individual, the program builds a list of potential parents (single mothers, single fathers, or pairs) according to birth and death dates. Then, in this list, the program looks for potential mothers and fathers for each individual according to the principle of genetic compatibility: for each locus, the young should inherit one allele from his mother and one from his father. Considering that errors are possible (e.g., mutations, typing errors, null alleles), some "allelic incompatibilities" can be allowed between the offspring and the parents. This error definition makes it possible to accept some loci for which one of the alleles cannot be inherited by any of the parents.

### Probability Computation

The probability of successfully allocating an individual offspring to its parents is based not only on its multilocus genotype but also on the error rate and the sampling rate of the population.

First, the program calculates the allelic frequencies for each locus in the sampled population. Next, it computes the probability of the expected genotype of an offspring, given the observed genotypes of the triplet (mother, father, offspring), assuming a uniform weight for all potential parents pairs (see the program documentation for further details). The program assumes that the loci are in Hardy-Weinberg equilibrium, do not show linkage disequilibrium, and exhibit a Wright-Fisher type of reproduction (finite and constant  $N$ , random mating with respect to the gene being studied, and nonoverlapping generations).

### Input Files

A single input file (text) is required for the data, containing the names and the multilocus genotypes of the individuals. If available, the program uses the following optional individual data: sex, date of birth, date of death, and multilocus genotype. A second file contains several parameters defined before running the software: the minimum age difference between a young and his potential father and mother, the error rate, the maximum number of missing alleles per individual that is acceptable to perform the analysis, an estimate of the proportion of the population sampled, and the maximum number of allelic incompatibilities between genotypes of offspring and parents.

### Output File

Four output files are produced: a first file with compatible parent pairs for each individual, two files with single parents (one for potential fathers and one for potential mothers), and a file with the

individuals for which no parent has been found.

### Obtaining the Program

PARENTE is free and can be downloaded from: <http://www2.ujf-grenoble.fr/leca/membres/manel.html>. It is available on a PC Windows platform. The distribution of the package includes executable programs, sample test data files, and documentation of the program. The binary file is also available upon request from the authors.

From the Laboratoire de Modélisation et Calcul, IMAG, B. P. 53, 38041 Grenoble Cedex 9, France (Cercueil), and the Laboratoire de Biologie des Populations d'Altitude, Université Joseph Fourier, Grenoble, France (Cercueil, Bellemain, and Manel). The authors give sincere thanks to Professor Swenson, Dr. Gordon Luikart, and Dr. Phillip England for advice and comments on the paper. Address correspondence to Alain Cercueil at the address above, or e-mail: [alain.cercueil@imag.fr](mailto:alain.cercueil@imag.fr).

© 2002 The American Genetic Association

### References

- Bernatchez L and Duchesne P, 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Can Fish Aquat Sci* 57:1-12.
- Brotherton PNM, Pemberton J, Komers P, and Malarkey G, 1997. Genetic and behavioural evidence of monogamy in a mammal, Kirk's dik-dik. *Proc R Soc Lond B Biol Sci* 264:675-681.
- Clapham PJ and Palsboll PJ, 1997. Molecular analysis of paternity shows promiscuous mating in female humpback whales (*Megaptera novaeangliae*, Borowski). *Proc R Soc Lond B Biol Sci* 264:95-98.
- Coltman DW, 1999. Male reproductive success in a promiscuous mammal: behavioural estimates compared with genetic paternity. *Mol Ecol* 8(7):1199-1209.
- Danzmann RG, 1997. PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from genotypic pools of parents and progeny. *J Hered* 88:333.
- Gerber S, Marlette S, Streiff R, Boénès C, and Kremer A, 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Mol Ecol* 9:1037-1048.
- Girman DJ, Mills MGL, Geffen E, and Wayne RK, 1997. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African

wild dog (*Lycan pictus*). *Behav Ecol Sociobiol* 40:187-198.

Goossens B, Graziani L, and Waits L, 1998. Extra-pair paternity in the monogamous Alpine marmot revealed by nuclear DNA microsatellite analysis. *Behav Ecol Sociobiol* 43:281-288.

Heller KG, Achmann R, and Witt K, 1993. Monogamy in the bat *Rhinolophus sedulus*? *Z Säugetierkunde* 58:376-377.

Luikart G and England P, 1999. Statistical analysis of microsatellite DNA data. *Trends Ecol Evol* 14:253-256.

Marshall TC, Slate J, Kruuk LEB, and Pemberton JM, 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639-655.

McRae SB and Kovacs KM, 1994. Paternity exclusion by DNA fingerprinting, and mate guarding in the hooded seal *Cystophora cristata*. *Mol Ecol* 3:101-107.

Neff BD, Repka J, and Gross MR, 2000a. Parentage analysis with incomplete sampling of candidate parents and offspring. *Mol Ecol* 9:515-528.

Neff BD, Repka J, and Gross MR, 2000b. Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? *Mol Ecol* 9:529-539.

Neff BD, Repka J, and Gross MR, 2001. A Bayesian framework for parentage analysis: the value of genetic and other biological data. *Theor Popul Biol* 59:315-331.

Nielsen R, Mattila DK, Clapham PJ, and Palsboll PJ, 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157:1673-1682.

Pemberton JM, Albon SD, Guinness FE, Clutton-Brock TH, and Dover GA, 1992. Behavioural estimates of male mating success tested by DNA fingerprinting in a polygynous mammal. *Behav Ecol* 3:66-75.

Petrie M and Kempnaers B, 1998. Extra-pair paternity in birds: explaining variations between species and populations. *Trends Ecol Evol* 13(2):52-58.

Queller DC, Strassmann JE, and Hugues CR, 1993. Microsatellites and kinship. *Trends Ecol Evol* 8(8):285-288.

Ribble DO, 1991. The monogamous mating system of *Peromyscus californicus* as revealed by DNA fingerprint. *Behav Ecol Sociobiol* 29:161-166.

Sillero-Zubiri C, Gottelli D, and Macdonald DW, 1996. Male philopatry, extra-pack copulations and inbreeding avoidance in Ethiopian wolves. *Behav Ecol Sociobiol* 38:331-340.

Sunnucks P, 2000. Efficient genetic markers for population biology. *Trends Ecol Evol* 15(5):199-203.

Received February 12, 2002

Accepted October 29, 2002

Corresponding Editor: Sudhir Kumar