



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Theoretical Biology

journal homepage: [www.elsevier.com/locate/yjtbi](http://www.elsevier.com/locate/yjtbi)

# Fractional parentage analysis and a scale-free reproductive network of brown trout



Hitoshi Koyano<sup>a,\*</sup>, Dimitar Serbezov<sup>b</sup>, Hirohisa Kishino<sup>c</sup>, Tore Schweder<sup>b</sup>

<sup>a</sup> Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasyo, Uji, Kyoto 611-0011, Japan

<sup>b</sup> Center for Ecological and Evolutionary Synthesis, University of Oslo, P.O. Box 1066 Blindern, NO-0316, Oslo, Norway

<sup>c</sup> Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

## HIGHLIGHTS

- A new nonparametric approach to fractional parentage analysis.
- Simultaneous estimation of the true and null allele frequencies of alleles at a locus.
- A scale-free property of the reproductive system of wild brown trout.
- Sexual selection on body size as a preferential attachment in trout population.

## ARTICLE INFO

### Article history:

Received 5 November 2012

Received in revised form

9 May 2013

Accepted 26 June 2013

Available online 18 July 2013

### Keywords:

Fractional allocation

Null alleles

Reproductive systems

Scale-free networks

## ABSTRACT

In this study, we developed a method of fractional parentage analysis using microsatellite markers. We propose a method for calculating parentage probability, which considers missing data and genotyping errors due to null alleles and other causes, by regarding observed alleles as realizations of random variables which take values in the set of alleles at the locus and developing a method for simultaneously estimating the true and null allele frequencies of all alleles at each locus. We then applied our proposed method to a large sample collected from a wild population of brown trout (*Salmo trutta*). On analyzing the data using our method, we found that the reproductive success of brown trout obeyed a power law, indicating that when the parent-offspring relationship is regarded as a link, the reproductive system of brown trout is a scale-free network. Characteristics of the reproductive network of brown trout include individuals with large bodies as hubs in the network and different power exponents of degree distributions between males and females.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

Difficulties in obtaining behavioral observations hinder the investigation of reproductive systems in wild populations of species with no parental care. One approach for elucidating the reproductive system of these species is parentage analysis. Methods of parentage analysis have progressed from the early technique of exclusion through categorical allocation (Meagher and Thompson, 1986; Sancristobal and Chevalet, 1997; Marshall et al., 1998; Gerber et al., 2000) to fractional allocation (Devlin et al., 1988; Roeder et al., 1989; Smouse and Meagher, 1994; Nielsen et al., 2001; Neff et al., 2001; Signorovitch and Nielsen, 2002; Hadfield et al., 2006) and parental

reconstruction (Jones and Avise, 1997; Jones et al., 1998). Results of parentage analysis are significantly affected by genotyping errors resulting from null alleles and other causes. Therefore, much research has been conducted on the causes and consequences of genotyping errors and their solutions (Broquet and Petit, 2004; Pompanon et al., 2005; Hoffman and Amos, 2005; Dewoody et al., 2006). In particular, null alleles are a major problem in parentage analyses using microsatellite markers. Hence, various methods for estimating null allele frequency have been developed since Chakraborty et al. (1992) proposed the first method (Brookfield, 1996; Kalinowski and Taper, 2006; van Oosterhout et al., 2006; Dakin and Avise, 2004). All these methods were developed in a framework in which null allele frequencies are assumed to be equal for all alleles at one locus. This framework has been used in the problem of estimating null allele frequency for the last twenty years. However, this assumption is not always true because null allele frequency may vary depending on the allele. In Section 2, we report a new method of fractional parentage analysis, in which null alleles are handled at the allele level rather than at the locus level. Furthermore, we examined the utility of our

\* Corresponding author. Tel.: +011 81 7 7438 3020.

E-mail addresses: [koyano@kuicr.kyoto-u.ac.jp](mailto:koyano@kuicr.kyoto-u.ac.jp) (H. Koyano), [dimitar.serbezov@bio.uio.no](mailto:dimitar.serbezov@bio.uio.no) (D. Serbezov), [kishino@lbn.ab.a.u-tokyo.ac.jp](mailto:kishino@lbn.ab.a.u-tokyo.ac.jp) (H. Kishino), [tore.schweder@econ.uio.no](mailto:tore.schweder@econ.uio.no) (T. Schweder).

proposed method by applying it to a large sample, in which common null allele frequency cannot be assumed for all alleles at each genotyped locus. The data were collected from a wild population of brown trout living in Bellbekken in Norway. Results of data analysis are given in detail in Section 3.

## 2. Proposed method of fractional parentage analysis

In this section, we briefly survey previous studies on methods of parentage analysis and then systematically describe a method of fractional parentage analysis that we developed.

### 2.1. Survey of the previous studies

As stated in the previous section, null alleles are obstacles in performing parentage analysis using microsatellite markers, and consequently, various methods for estimating null allele frequency have been developed. Chakraborty et al. (1992) noted that some deficiencies in the number of observed heterozygotes can be explained by nondetectable null alleles in RFLP analyses, and they developed a method for estimating null allele frequency based on the number of observed and expected heterozygotes. The method of Chakraborty et al. (1992) was improved by Brookfield (1996). Genepop uses an EM algorithm by Dempster et al. (1977) to find a maximum likelihood estimate for null allele frequency. CERVUS uses an iterative algorithm by Summers and Amos (1997) and Marshall et al. (1998), in which the presence of homozygotes caused by null alleles is not allowed for initially, but is added during the process of optimization. Kalinowski and Taper (2006) proposed a new maximum likelihood estimator for null allele frequency and demonstrated the good performance of their estimator by computer simulation. van Oosterhout et al. (2006) presented a method for estimating null allele frequency for microsatellite data from nonequilibrium populations. Chybicki and Burczyk (2009) considered the simultaneous estimation problem of null allele frequencies and an inbreeding coefficient in a population that experiences inbreeding. All these methods were developed in a framework in which it was assumed that the null allele frequencies are uniform at a single locus. See for example, Dakin and Avise (2004) for a review on microsatellite null alleles in parentage analysis.

Results of parentage analysis are significantly affected by genotyping errors, including those due to causes other than null alleles. Therefore, much research has been conducted on the causes and consequences of genotyping errors and solutions to them. Broquet and Petit (2004) reviewed various methods for estimating the rate of genotyping errors and showed that error rates may be substantially underestimated when using an erroneous approach. Vandeputte et al. (2006) examined the effectiveness of allowing for mismatches at one or more alleles in parentage assignment as a method to recover assignment power when genotyping may be incorrectly performed. See Pompanon et al. (2005), Hoffman and Amos (2005), and Dewoody et al. (2006) for information on the causes of genotyping errors and methods for detecting them.

### 2.2. Basic model

In this section, we describe the fundamental framework of our nonparametric fractional parentage analysis. The part of the model that does not require the Hardy–Weinberg principle is described in the present section, and the part that requires it is described in the following section.

We first consider the case in which all individuals in a population are sampled and genotyped. Let  $F$ ,  $M$ , and  $O$  be the

set of potential fathers, mothers, and offspring, respectively. We denote the probability that a father and mother with genotype  $f$  and  $m$ , respectively, produce offspring with genotype  $o$  by  $p_{of,m}$ . If the pairs of fathers in  $F$  and mothers in  $M$  are equally likely, the conditional probability that the parents have genotype  $f$  and  $m$ , respectively, given the genotype  $o$  of offspring is

$$p_{f,m|o} = \frac{p_{of,m}}{\sum_{f' \in F, m' \in M} p_{of',m'}}.$$

Therefore, the expected number of offspring from parents with genotype  $f$  and  $m$  is

$$e_{f,m} = \sum_{o \in O} p_{f,m|o},$$

and the expected number of offspring from a father with genotype  $f$  and a mother with genotype  $m$  are

$$e_f = \sum_{m \in M} e_{f,m}, \quad e_m = \sum_{f \in F} e_{f,m},$$

respectively.

We next consider the case in which sampling is incomplete. Let  $F$ ,  $M$ , and  $O$  be the set of potential fathers, mothers, and offspring that are sampled and genotyped, respectively.  $F_u$ ,  $M_u$ , and  $O_u$  represent the set of unsampled potential fathers, mothers, and offspring, respectively. We assume that the sampling fraction depends on sex, but not on year, age, and individual, and write  $s_f = \#F / (\#F + \#F_u)$  and  $s_m = \#M / (\#M + \#M_u)$ , where  $\#F$ ,  $\#F_u$ ,  $\#M$ , and  $\#M_u$  represent the number of elements of  $F$ ,  $F_u$ ,  $M$ , and  $M_u$ , respectively. Our original method of simultaneously estimating sampling fraction and survival probability used in this study is described in Section 2.4. Let  $\hat{s}_f$  and  $\hat{s}_m$  be the estimators for  $s_f$  and  $s_m$ , respectively.  $p_{o|n,n}$  represents the probability that unsampled parents obtain offspring with genotype  $o$ .  $p_{o|n,m}$  and  $p_{of,n}$  are the probabilities that fathers that are unsampled and mothers with genotype  $m$  have offspring with genotype  $o$  and that fathers with genotype  $f$  and mothers that are unsampled have offspring with genotype  $o$ , respectively. These probabilities can be estimated by the corresponding means in the sample. Assuming the sampling to be random (with all individuals equally likely to be sampled), an unsampled couple is statistically identical to a sampled couple, and therefore, we estimate  $p_{o|n,n}$ ,  $p_{of,n}$ , and  $p_{o|n,m}$  by

$$\hat{p}_{o|n,n} = \frac{1}{\#F \times \#M} \sum_{f \in F, m \in M} p_{of,m},$$

$$\hat{p}_{o|n,m} = \frac{1}{\#F} \sum_{f \in F} p_{of,m}, \quad \hat{p}_{of,n} = \frac{1}{\#M} \sum_{m \in M} p_{of,m},$$

respectively. Let  $p_o$  be the probability of an offspring having genotype  $o$ , i.e., a randomly selected father–mother pair will have such an offspring. When the above estimators are used for the offspring probabilities for unsampled parents,  $p_o$  can be calculated from

$$\hat{p}_o = \frac{1}{\#F \times \#M} \sum_{f \in F, m \in M} p_{of,m}.$$

Then, the probability that an offspring of genotype  $o$  has a mother and father that are both unsampled,  $p_{n,n|o}$ , is estimated by

$$\hat{p}_{n,n|o} = \frac{1}{\hat{p}_o} (1 - \hat{s}_f)(1 - \hat{s}_m) \hat{p}_{o|n,n}.$$

For other parent combinations, we use the following estimators:

$$\hat{p}_{n,m|o} = \frac{1}{\hat{p}_o} \frac{(1 - \hat{s}_f) \hat{s}_m}{\#M} \hat{p}_{o|n,m}, \quad \hat{p}_{f,n|o} = \frac{1}{\hat{p}_o} \frac{\hat{s}_f (1 - \hat{s}_m)}{\#F} \hat{p}_{of,n},$$

$$\hat{p}_{f,m|o} = \frac{1}{\hat{p}_o} \frac{\hat{s}_f \hat{s}_m}{\#F \times \#M} p_{of,m}.$$

Download English Version:

<https://daneshyari.com/en/article/6370800>

Download Persian Version:

<https://daneshyari.com/article/6370800>

[Daneshyari.com](https://daneshyari.com)