

Hum Hered 2009;67:174–175
DOI: 10.1159/000181155

Haplotype Reconstruction and Estimation of Haplotype Frequencies from Nuclear Families with One Parent Available and Varying Numbers of Children Using the Exact Likelihood

Xiang Dong Ding^{a,b}, Qin Zhang^a, Henner Simianer^b

^aState Key Laboratories of Agrobiotechnology, Key Laboratory of Animal Genetics and Breeding, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing, China; ^bUniversity of Goettingen, Institute of Animal Breeding and Genetics, Goettingen, Germany

Recently we proposed a method for haplotype inference in nuclear families with only one parent available [1]. Our approach, which we call PCHAP (parent-child haplotyping) here for convenience, is exact for families consisting of one parent and one child only, in which parent and child share one haplotype. While the original PCHAP approach was designed to handle a single child only, we suggested as an approximation to handle two children by splitting the one parent and two children in two parent child haplotype pairs treated as independent.

Obviously the decomposition of families with multiple children makes no correct and effective use of the family information. This is illustrated by the following example: consider a parent with diplotype (H_i, H_j) ($i \neq j$), the first child with diplotype (H_i, H_k) ($k \neq i, j$), and the second child with diplotype (H_i, H_j). Correctly, H_i is the joint haplotype of all three individuals and the correct number of occurrences of H_i, H_j , and H_k is 2, 1, 1. If

this family is divided into two independent parent-child pairs $\{(H_i, H_j), (H_i, H_k)\}$ and $\{(H_i, H_j), (H_i, H_j)\}$ as described above, the number of occurrences of H_i, H_j , and H_k is 3, 2, 1. Therefore, the approximation used in Ding et al. [1] for multiple children families will bias the frequency estimates of some haplotypes. Despite this approximation was empirically shown to be still more efficient than GENEHUNTER [table 5 in 1] in the multiple children case.

We suggest here a theoretically more justified method to handle the multiple children case. Recently we proposed another EM-based approach to handle full-sib families with two or more children and missing parents, termed FSHAP [2], which can handle families with one parent and multiple offspring as well. Our simulation study (table 1) demonstrates that the performance of the PCHAP is not significantly improved by increasing the number of children, while the performance of FSHAP leads to results that (a) improve significantly with an increasing number of children and (b) significantly outperform the results obtained with PCHAP in almost all situations.

FSHAP can handle one parent families with multiple children only and PCHAP can handle one parent families with a single child, but both of them are similar in the likelihood function and implemented via the EM algorithm. Thus we combine PCHAP and FSHAP in one framework to handle families with one available parent and 1 to n children.

For the new method, named PMCHAP, the likelihood function of the population haplotype frequencies is the one used in PCHAP and described in [1] for single-child families or the one used in FSHAP and described in [2] for multiple-children families. In the implementation of the EM algorithm, the probabilities of parent-child haplotype pairs and full-sib haplotype sets are calculated according to the expectation step described in PCHAP and FSHAP, respectively [1, 2]. In the M-step a weighted average

Table 1. Comparison of efficiency of our original approach (PCHAP) and FSHAP for one-parent families with multiple children (60 one parent families and 8 SNPs)

	Number of children					
	2		3		6	
	PCHAP	FSHAP	PCHAP	FSHAP	PCHAP	FSHAP
I_F	0.9783 ^{b,c}	0.9746 ^d	0.9778 ^{c,d}	0.9814 ^b	0.9753 ^{c,d}	0.9898 ^a
I_H	0.9752 ^b	0.9743 ^b	0.9805 ^b	0.9842 ^{a,b}	0.9866 ^{a,b}	0.9938 ^a
Error rate	0.0199 ^a	0.0128 ^{b,c}	0.0163 ^{a,b}	0.0077 ^{c,d}	0.0210 ^a	0.0044 ^d
I_R	0.9260 ^c	0.9190 ^{c,d}	0.9183 ^{c,d}	0.9498 ^b	0.9146 ^d	0.9823 ^a

^{a–d} Means with different superscript within one row differ significantly at the $\alpha < 0.01$ significance level.
 I_F = Haplotype frequency discrepancy, I_H = Identification rate, I_R = haplotype reconstruction reliability.

Table 2. Comparison of efficiency of the original approach PCHAP and the new approach PMCHAP for one-parent families with varying numbers of children (60 families and 8 SNPs)

	Number of children					
	1–2		1–4		1–6	
	PCHAP	PMCHAP	PCHAP	PMCHAP	PCHAP	PMCHAP
I _F	0.9738 ^c	0.9729 ^c	0.9720 ^c	0.9780 ^b	0.9735 ^c	0.9835 ^a
I _H	0.9733 ^{a, b}	0.9713 ^{a, b}	0.9652 ^b	0.9749 ^{a, b}	0.9644 ^b	0.9832 ^a
Error rate	0.0206 ^{a, b}	0.0164 ^{a–c}	0.0205 ^{a, b}	0.0122 ^{b, c}	0.0239 ^a	0.0104 ^c
I _R	0.9074 ^c	0.8958 ^d	0.9123 ^c	0.9293 ^b	0.9043 ^{c, d}	0.9481 ^a

is used to update the population haplotype frequencies. For more details see Ding et al. [3] where a more general framework for haplotype inference is described.

A simulation study assuming variable family sizes shows that PMCHAP performs equivalently to or significantly better than PCHAP (table 2). The approach will be especially helpful for research in human populations and in multiparous domestic animal species like pigs, dogs, fish, and many lab animals where it is easy to collect families with multiple siblings. We therefore suggest to use the PCMHAP approach in such situations. The respective software is available from the authors.

Acknowledgements

We are grateful to Ji-yuan Zhou and Yue-qing Hu for constructive suggestions. This research was supported by the FUGATO (Functional Genome Analysis in Animal Organisms) program of the German Federal Ministry of Education, the Förderverein Biotechnologieforschung e.V. Bonn, Lohmann Tierzucht GmbH Cuxhaven, and The National Key Basic Research Program of China (Grant No. 2006CB102104).

References

- 1 Ding XD, Zhang Q, Flury C, Simianer H: Haplotype reconstruction and estimation of haplotype frequencies from nuclear families with only one parent available. *Hum Hered* 2006;62:12–19.
- 2 Ding XD, Simianer H, Zhang Q: A new method for haplotype inference including full sib information. *Genetics* 2007;177:1929–1940.
- 3 Ding XD, Zhang Q, Simianer H: PMCHAP: a new approach for haplotype inference using nuclear families with only one parent and multiple offspring. *J Anim Breed Genet* (in press).

Henner Simianer
Albrecht-Thaer-Weg 3
DE–37075 Goettingen (Germany)
Tel. +49 551 39 5640, Fax +49 551 39 5587
E-Mail hsimian@gwdg.de