Iranian J Publ Health, Vol. 38, No.4, 2009, pp.136-139

Case Report

The Investigation of Haplotype Phasing Efficiency at the *PAH* Gene Region in Iranian Family Trios

Z Fazeli, *S Vallian

Division of Genetics, Department of Biology, Faculty of Science, University of Isfahan, Iran

(Received 17 May 2009, accepted 13 Sep 2009)

Abstract

Background: The haplotype phasing is more useful than genotyping markers independently at carrier detection and prenatal diagnosis of diseases. The *PAH* gene region contains several markers used in detection of PKU disease. In the present study, the efficiency of *Bgl*II-*Eco*RI-VNTR haplotype phasing in Iranian family trios was investigated. Then, this information was compared with those obtained for unrelated individuals.

Methods: Blood samples were collected from 20 healthy family trios and 60 unrelated individuals. The genomic DNA was extracted by use of salting-out procedure. The two markers *Bgl*II and *Eco*RI were genotyped by use of PCR-RFLP. The genotype of VNTR marker was identified by use of PCR and electrophoresis. The genotyping data obtained from family trios was used to infer haplotype phase. We also compared this data with results obtained from a widely used method for haplotype frequency inference from unrelated individuals, the PHASE program.

Results: The haplotype phase of all members was only ascertained at eight family trios. The comparison of this data with the results obtained by use of PHASE program showed that eight haplotypes [211, 221, 215, 216, 214, 121, 225 and 111] were informative haplotypes in Iranian population.

Conclusion: Since diversity of *Bgl*II-*Eco*RI-VNTR haplotypes was high in Iranian population, haplotype phasing at family trios was difficult. The results of this study showed that the genotyping data obtained from family trios could not provide enough information for *Bgl*II-*Eco*RI-VNTR haplotype phasing at Iranian PKU families and the genotyping of other family members was necessary at most cases.

Keywords: *Phenylalanine hydroxylase (PAH), Phenylketonuria (PKU), Haplotype*

Introduction

The genotyping data in pedigrees or populations often consist of unordered genotypes and their haplotype phase is not directly observable. Haplotype data are valuable in the study of diseases. Indeed, haplotype phasing is a more powerful method than single-marker analysis at detection of status disease (1-3). The silico haplotyping, which infers haplotypes from observed genotype data by statistical and computational methods, is a valuable method at haplotype phase estimation (4). The haplotyping problem is to infer the two haplotypes of each individual from the unordered genotypes. Two types of genotype data, pedigree genotype data and population genotype data, can be used for haplotyping. Haplotype estimation from population data is often done using statistical algorithms such as maximum-likelihood (5-7) and Bayesian (8) algorithms. The haplotype phase from pedigree members is inferred using the genotyping data of their relatives. Nowadays, different programs were used to infer haplotype phase. One of them is PedPhase program which was ascertained haplotype phase at pedigrees by use of five different algorithms (9-10). In this study, the genotype data of 20 family trios was used to infer *Bgl*II-*Eco*RI-VNTR haplotype phase at phenylalanine hydroxylase (*PAH)* gene. Then, haplotype frequency obtained at these 20 families was compared with haplotype frequency estimated by use of PHASE software. The aim of this study was to evaluate the efficiency of *Bgl*II-*Eco*RI-VNTR haplotype phasing at Iranian PKU family.

Materials and Methods

The genotyping procedure

The genotyping data of three markers, *Bgl*II, *Eco*RI and VNTR, at the *PAH* gene was prepared

***Corresponding author:** Tel/Fax: +98 311 7932456, E-mail: svallian@biol.ui.ac.ir

by use of PCR and electrophoresis methods. Blood samples collected in EDTA tubes were obtained from 20 healthy family trios and 60 unrelated individuals. These individuals and families were from Iranian population. Then, the genomic DNA was extracted by use of salting-out procedure (11). The primers and PCR conditions were described earlier (12-14). After performing PCR, amplified product of *Bgl*II and *Eco*RI were digested with *Bgl*II and *Eco*RI restriction enzymes at 37 °C overnight, respectively. The PCR products of VNTR marker and enzyme digestion products for *Bgl*II and *Eco*RI markers were genotyped on 1.5% agarose gel and their size were determined by use of Gene Ruler TM 50 bp DNA ladder and Gene Ruler TM 100 bp DNA (MBI Fermentas).

The haplotype phasing procedure

The program PedPhase was used to infer haplotype phase from genotyping data obtained from a pedigree. The input to PedPhase is a text file. Each line represents a member, its parental and genotype information. Among five algorithms in PedPhase version 2.0 (9-10), we used member-based dynamic programming algorithm. This algorithm is useful to infer haplotype phase at small pedigrees. The genotyping data obtained from 20 family trios (60 individuals) were used to create input file. The input file contained of the pedigree information and the markers information for each member. The PedPhase was launch on a Command Prompt (DOS) window on windows. The output file contained the haplotype data for each member at pedigrees.

The haplotype frequency estimation

The haplotype frequency was estimated by using PHASE program (15). The input file used at this program contained the genotyping data obtained for three markers, *Bgl*II, *Eco*RI and VNTR, at the *PAH* gene in 60 unrelated individuals. This program implements the Bayesian algorithm.

Results

The nine alleles with core repeat unit of 3, 5, 6, 7, 8, 9, 10, 11 and 13 copies were observed at studied population. These alleles at VNTR marker were numbered from 1 to 9 at inputs of PedPhase and PHASE programs. For alleles of *Bgl*II and *Eco*RI markers, both of them are RFLPs, 1 and 2 numbers stand for presence or absence of restriction site, respectively. Each haplotype phase at output files was represented as a three-digit number, in which the order of the numbers indicated *Bgl*II, *Eco*RI and the VNTR marker, respectively. The determination of haplotype phase at 20 family trios was done by use of PedPhase program. The haplotype phase of all three members, two parents and their offspring, at eight families was only determined. The haplotype phasing of pedigree 9 was shown at Fig. 1a. At twelve remainder pedigrees, the haplotype of one, two or all three members at family trios remained phase-unknown. The haplotype phasing at these pedigrees required the genotyping data of maternal parents or parental parents or both of them. In Fig. 1b, the haplotype phasing of the pedigree 19, in which the haplotype of member 19-1 was phase-unknown, was shown. The haplotype phase of 32 individuals was only ascertained by use of this program. These phase-known haplotypes were used to estimate haplotype frequency at Iranian population. Then, the obtained haplotype frequency was compared with the results obtained by use of PHASE program. This program was able to estimate haplotype frequency by use of genotyping data obtained from unrelated individuals. The haplotype frequency estimated by use of these two programs, PedPhase and PHASE, was shown in Table 1.

Discussion

The genotyping data obtained by laboratory techniques provide unordered allele pairs for each marker (16). Haplotyping refers to the reconstruction of the haplotype phase from these observed genotype data. Currently, statistical methods were developed to infer the true haplotype phase (17). A haplotype phase is a transfer of ordered genotypes to all members in the pedigree at all loci, such that the transfer of haplotype phase is consistent with all observed genotype data and Mendelian segregation (18). Haplotype phasing in pedigrees provides useful information which could be used in carrier detection and prenatal diagnosis of related diseases.

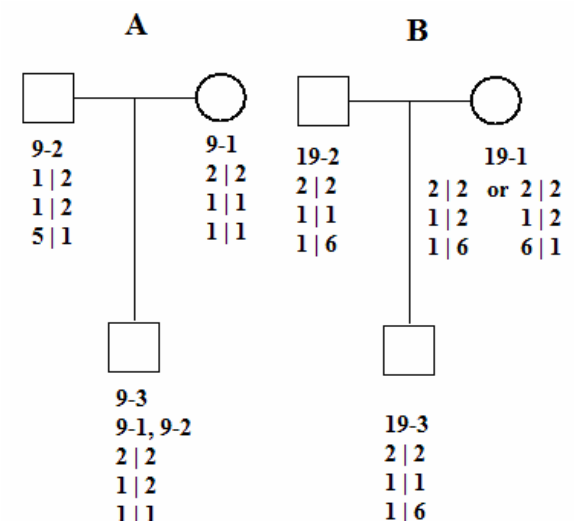

Fig. 1: In the pedigrees, the square and circle represent male and female, respectively. The member ID and haplotype data are placed under each member. In haplotype data, a | separates the parental allele from the maternal allele. A) The haplotype phase was ascertained at all three member of pedigree 9. B) In pedigree 19, the haplotype phase of member 19-1 was unknown. Therefore, we were not able to determine maternal haplotype and paternal haplotype at member 19-3. The haplotype phasing of member 19-1 requires the genotyping data of her parents

Table 1: Haplotype frequency obtained from PedPhase and PHASE programs

Haplotype1	PedPhase	PHASE
211	0.171875	0.175146
221	0.125000	0.104542
216	0.093750	0.059551
215	0.078125	0.086896
214	0.062500	0.063582
121	0.062500	0.055944
224	0.046875	0.043059
115	0.046875	0.039286
226	0.046875	0.048349
225	0.031250	0.060403
111	0.031250	0.072702
116	0.031250	0.028598
124	0.031250	0.023426
129	0.031250	0.002109
127	0.031250	0.010201
229	0.031250	0.006156
217	0.031250	0.010317
126	0.015625	0.021835
125	0.000000	0.013415
114	0.000000	0.036599

In this study, the haplotype phase of *Bgl*II-*Eco*RI-VNTR at 20 family trios was determined. This function was done by use of PedPhase program. The haplotype phasing at eight pedigrees was successfully done. In addition to haplotype phasing, the origin of haplotype phase (maternal or parental) was distinguished in these pedigrees. The genotyping data of reminder families was not enough for haplotype phasing. The estimation of haplotype frequency of *Bgl*II-*Eco*RI-VNTR by using PedPhase and PHASE programs showed that diversity of *Bgl*II-*Eco*RI-VNTR haplotype was high at studied population. Therefore, the inferring of *Bgl*II-*Eco*RI-VNTR haplotype phase at most pedigrees in Iranian population required genotyping data obtained from other family members.

The haplotype phase of only 32 individuals was determined by use of PedPhase program. This haplotype data was used to estimate haplotype frequency. This haplotype frequency was compared with the results obtained from PHASE program. The comparison of haplotype frequency obtained from these two programs indicated that eight haplotypes [211, 221, 215, 216, 214, 121, 225 and 111] had frequency higher than 5% (informative haplotypes). Although two haplotypes 225 and 111 had frequency <5% when the haplotype frequency was estimated by use of haplotype data obtained from PedPhase program, but it is likely that the low frequency of these two haplotypes was resulted from insufficient number of individuals which was studied at this estimation.

The PKU disease has high incidence at Iranian population. Since the genotyping data of single marker at *PAH* gene region was not informative at some families, the haplotype phasing could be used as a suitable method for detection of PKU disease. The high diversity of *Bgl*II-*Eco*RI-VNTR haplotype at Iranian population caused that the haplotype phasing was difficulty done at Iranian family trios.

This study showed that the genotyping of other family members is necessary at most cases.

Acknowledgments

The work was supported by grant number 851119 from the department of research of the University of Isfahan, Isfahan, Iran. The authors declare that they have no conflicts of interest.

References

1. Mailund T, Besenbacher S, Schierup M (2006). Whole genome association mapping by incompatibilities and local perfect phylogenies. *BMC Bioinformatics*, 7: 454.
2. Liu J, Papasian C, Deng HW (2007). Incorporating Single-Locus Tests into Haplotype Cladistic Analysis in Case-Control Studies. *PLoS Genet*, 3: e46.
3. Su SY, Balding DJ, Coin LJ (2008). Disease association tests by inferring ancestral haplotypes using a hidden markov model. *Bioinformatics*, 24: 972-78.
4. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, et al. (2006). Comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet*, 78: 437-50.
5. Excoffier L, Slatkin M (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*, 12: 921-927.
6. Fallin D, Schork NJ (2000). Accuracy of haplotype frequency estimation for biallelic Loci via the Expectation-Maximization algorithm for unphased diploid genotype data. *Am J Hum Genet*, 67: 947-59.
7. Qin ZS, Niu T, Liu JS (2002). Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms. *Am J Hum Genet*, 71: 1242-47.
8. Niu T, Qin ZS, Xu X, Liu JS (2002). Bayesian haplotype inference for multiple linked Single-Nucleotide Polymorphisms. *Am J Hum Genet*, 70: 157-69.
9. Li J, Jiang T (2003). Efficient rule-based haplotyping algorithms for pedigree data. *Proc. RECOMB*, 03: 197-206.
10. Li J, Jiang T (2003). Efficient inference of haplotypes from genotypes on a pedigree. *J Bioinfo and Comp Biol*, 1: 41-69.
11. Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16: 1215.
12. Goltsov AA, Eisensmith RC, Konecki DS, Lichter-Konecki U, Woo SLC (1992). Association between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am J Hum Genet*, 51: 627-36.
13. Dworniczak B, Wedemeyer N, Horst J (1991). PCR detection of the *Bgl*II RFLP at the human phenylalanine hydroxylase (*PAH*) locus. *Nucleic Acids Res*, 19: 1958.
14. Kidd KK. 2002. PAH *Eco*RI Polymorphism. Available: http://info.med.yale.edu/genetics/kkidd/PAH_EcoRI.html.
15. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, et al. (2006). A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet*, 78: 437- 50.
16. Gao G, Allison DB, Hoeschele I (2009). Haplotyping methods for pedigrees. *Hum Hered*, 67: 248-66.
17. Lin S, Speed TP (1997). An algorithm for haplotype analysis. *J Comput Biol*, 4: 535-46.
18. Baruch E, Weller JI, Cohen-Zinder M, Ron M, Seroussi E (2006). Efficient inference of haplotypes from Genotypes on a large animal pedigree. *Genetics*, 172: 1757-65.