

## Haplotype Inference from Genotype Data Using Evolutionary Method (ARO)

M. Moghaddam, A.R. khanteymoori

<sup>a</sup>Department of Computer, Zanjan Branch, Islamic Azad University, Zanjan, Iran

<sup>b</sup>Department of Electrical and Computer Engineering, University of Zanjan, Zanjan, Iran

\*luis.mahdi@gmail.com

**Abstract:** Single Nucleotide Polymorphism (SNP) is one of the most common structural changes in the genomic sequences. Haplotypes are sets of SNPs in each chromosome. Haplotypes in some organisms such as human are in diploid form i.e. there are two copies for each chromosome which one inherited from father and the other inherited from mother [1, 2]. Characterizing the haplotypes provide valuable information for researchers which can help in several domains such as drug designs and genetic diseases [3-5]. Since obtaining haplotypes from experimental methods are expensive, several computational methods have been proposed which reconstruct haplotypes based on genotype data [6-8]. It should be noted that the approaches which use genotype data are categorized as haplotype inference. The aim of these methods is to search the smallest number of distinct haplotypes that can generate the given genotypes [9] and it is proved to be NP-hard [10, 11]. In this paper, we propose a new algorithm based on the Asexual Reproduction Algorithm (ARO) [12]. High-convergence speed and less parameter setting against similar algorithms such as GA, are the main advantages of ARO. For this purpose, several new ideas are presented to improve the efficiency of the ARO in the haplotypes inference problems such as changing the structure of individual, improving the ProLarva operator and etc. In order to evaluate the method four real data sets has been used. The obtained results demonstrate that the proposed method often outperforms comparing methods both on switch error rate and genotype error rate.

**Keywords:** Haplotype; Genotype; Haplotype inference; NP-hard; ARO

### References

- [1] Y. Zhao, L. Wu, J. Zhang, R. Wang, X. Zhang, "Haplotype assembly from aligned weighted SNP fragments." *Comput Biol Chem*, 29(4) (2005) 281-7.
- [2] F. Geraci, "A comparison of several algorithms for the single individual SNP haplotyping reconstruction problem." *Bioinformatics*, 26(18) (2010) 2217.
- [3] S.R. Browning, B.L. Browning, "Whole-genome multilocus association mapping using localized haplotype clusters," *Genet. Epidemiol.*, 31 (2007) 455.
- [4] M. Stephens, N.J. Smith, P. Donnelly, "A new statistical method for haplotype reconstruction from population data," *Am. J. Hum. Genet.*, 68 (2001) 978-989.
- [5] J. Hugosson, E. Hemberg, A. Brabazon, M. O'Neill, "Genotype representations in grammatical evolution," *Appl. Soft Comput.*, 10 (2010) 36-43.
- [6] C. Burgdorf, P. Kepper, M. Hoehe, C. Schmitt, R. Reinhardt, H. Lehrach, S. Sauer, "Clone-based systematic haplotyping (CSH): a procedure for physical haplotyping of whole genomes," *Genome Res.* 13 (2003) 2717-2724.
- [7] C.M. Ding, C.R. Cantor, "Direct molecular haplotyping of long-range genomic DNA with M1-PCR," *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 7449-7453.
- [8] L. Tinini, P. Bertolazzi, A. Godi, G. Lancia, "CollHaps: a heuristic approach to haplotype inference by parsimony," *IEEE*, 7 (2010) 511-523.