

MEETING ABSTRACT

Open Access

# alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees

Nathan Medina-Rodríguez<sup>1,2\*</sup>, Angelo Santana<sup>1</sup>, Ana M Wägner<sup>3</sup>, José M Quinteiro<sup>2</sup>

From Ninth International Society for Computational Biology (ISCB) Student Council Symposium 2013 Berlin, Germany. 19 July 2013

## Background

Haplotype inference is an essential stage in genetic linkage analysis and estimation methods are also very frequently used to reconstruct haplotypes in current genetic association studies. Most of the latter are focused on haplotype phasing from recombinant DNA areas of unrelated individuals and use likelihood-based methods to infer the presence of alleles in several loci with very time-consuming probabilistic algorithms.

So far, literature does not analyze haplotypes using deterministic techniques, and there are hardly any alternative methods for constructing haplotypes from non-recombinant

DNA areas, despite the fact that computational inference by probabilistic models may cause a large number of incorrect inferences.

## Description and results

We have developed an algorithm called alleHap, which is able to impute alleles from parent-offspring pedigree databases with missing family members, to later construct their corresponding, unambiguous haplotypes.

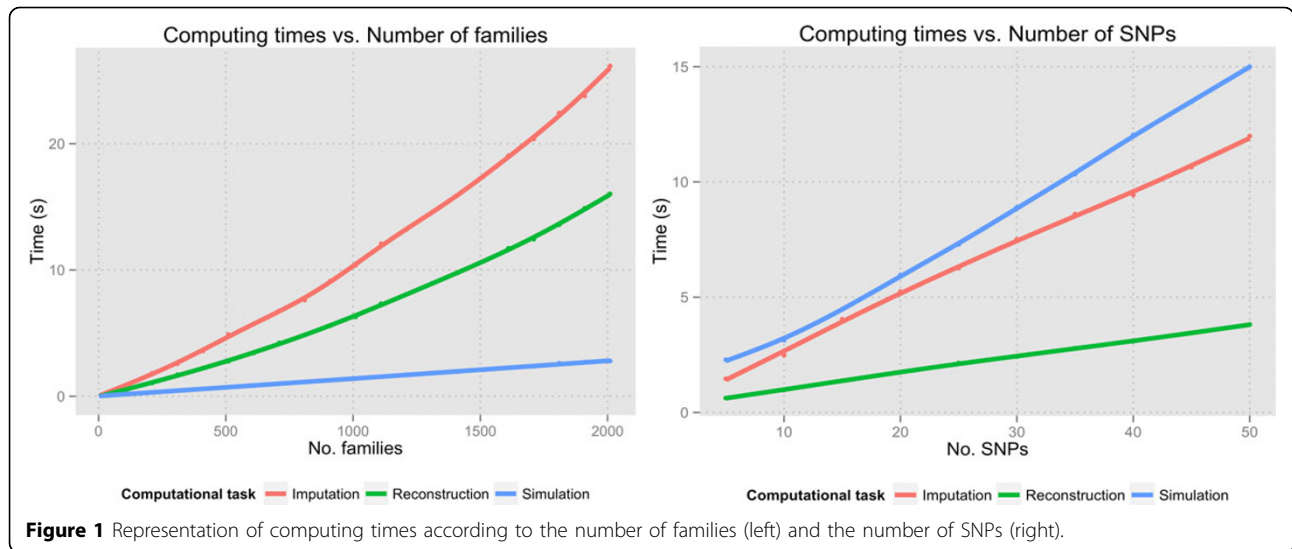
The alleHap algorithm is based on a preliminary analysis of all possible combinations that may exist in the genotyping of a family, considering that each member,

**Table 1 Possible allelic combinations in a parent-offspring pedigree**

Parents Alleles	1		2		Offspring* Alleles	1		2		3		4		Case (Variant)
	1	2	1	2		1	2	1	2	1	2	1	2	
a	a	a	a	a	a	a								1
a	a	b	b	b	a	b								2
a	a	a	b	b	a	a	a	b						3 (I)
a	b	b	b	b	a	b	b	b						3 (II)
a	a	b	c	c	a	b	a	c						4 (I)
a	c	b	b	c	a	b	b	c						4 (II)
a	b	c	c	c	a	c	b	c						4 (III)
a	b	a	b		a	a	a	b	b	b				5
a	b	a	c		a	a	a	b	a	c	b	c		6 (I)
a	b	b	c		a	b	a	c	b	b	b	c		6 (II)
a	c	b	c		a	b	a	c	b	c	c	c		6 (III)
a	b	c	d		a	c	a	d	b	c	b	d		7 (I)
a	c	b	d		a	b	a	d	b	c	c	d		7 (II)
a	d	b	c		a	b	a	c	b	d	c	d		7 (III)

\*Considering all allele combinations, the maximum number of "unique" children and alleles is four.

<sup>1</sup>Department of Mathematics, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain  
 Full list of author information is available at the end of the article



due to meiosis, should unequivocally have two alleles, one from each parent. The analysis was founded on the differentiation of seven cases, as described in [1], but some of them divided into a maximum of three variants, representing a different combination of alleles of the family members (Table 1).

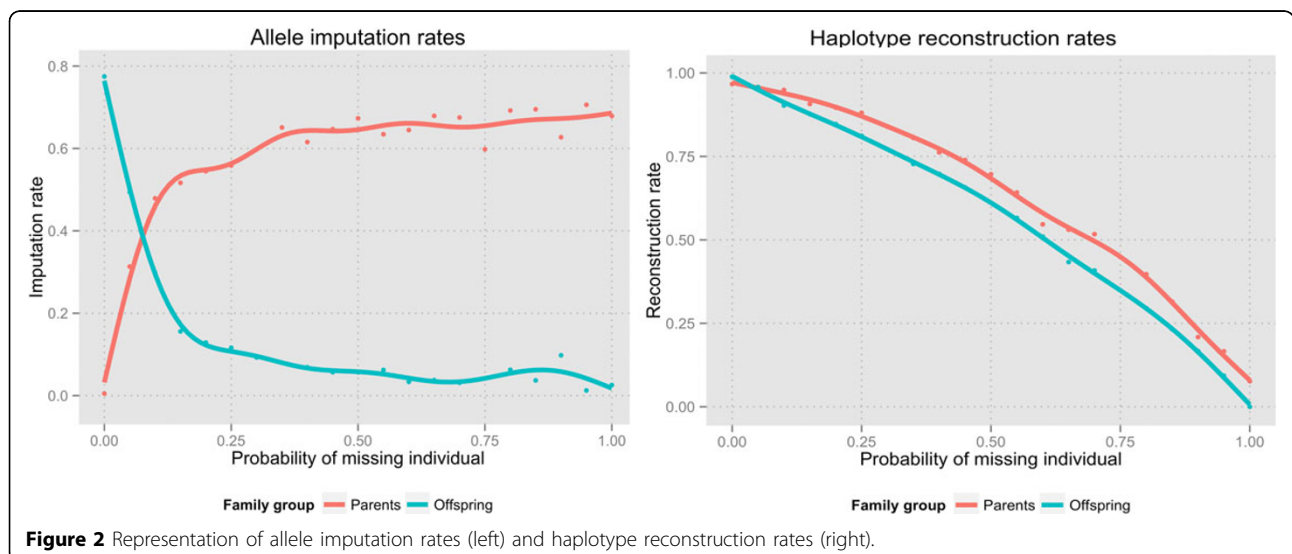
The classification by cases and variants allows the algorithm to impute missing values efficiently in the loaded database to proceed afterwards to the conformation of corresponding unambiguous haplotypes. Furthermore, the algorithm allows the construction of haplotypes, without any limitation in terms of the number of SNPs, i.e. enables the construction of haplotypes of more than two SNPs.

By analyzing all possible combinations of a parent-offspring pedigree in which parents may be missing, as

long as one child has been genotyped, theoretically an unequivocal imputation of three possible parent haplotypes is possible in 92.3% of cases even when one parent is missing. When neither parent has been genotyped, in 36.4% of cases at least two haplotypes can be constructed. Regarding offspring allele imputation with both parents fully genotyped, a minimum of one haplotype for each child may be successfully reconstructed in 6.1% of possible cases.

Evaluation of the results (Figure 1) reveals an optimum performance of alleHap computational tasks, namely Simulation, Imputation and Reconstruction. Their corresponding execution times are quite low even when considering a large number of families ( $\leq 2000$ ) and SNPs ( $\leq 50$ ).

Figure 2 shows how our algorithm has high allele imputation rates (about 65%) even when the probability



of missing parents in each family is high (>50%). Regarding haplotype reconstruction rates, there is an almost linear relationship between reconstruction rates and the number of missing individuals per family. This is because alleHap is mainly based on the information included in the offspring, so the more children that are missing the more difficult it is to reconstruct the family haplotypes.

## Conclusions

alleHap has been tested by simulations and also with the Type 1 Diabetes Genetics Consortium [2] database. Our algorithm is very robust against inconsistencies within the genotypic data and consumes very little time, even when handling large amounts of data. The missing data imputation may improve results in numerous epidemiological and/or genetic linkage studies.

Our algorithm could be a useful instrument for information retrieval and knowledge discovery in genetics, since it would allow epidemiological specialists to discover new intergenic patterns by studying zero-recombinant haplotypes with a larger number of SNPs from family-based databases.

## Authors' details

<sup>1</sup>Department of Mathematics, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain. <sup>2</sup>IUMA - Information and Communication Systems, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain. <sup>3</sup>Department of Medical and Surgical Sciences, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain.

Published: 11 February 2014

## References

1. Berger-Wolf TY, *et al*: Reconstruction sibling relationships in wild populations. *Bioinformatics* 2007, **23**:i49-i56.
2. Rich SS, *et al*: The Type 1 Diabetes Genetics Consortium. *Ann N Y Acad Sci* 2006, **1079**:1-8.

doi:10.1186/1471-2105-15-S3-A6

Cite this article as: Medina-Rodríguez *et al*: alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. *BMC Bioinformatics* 2014 **15**(Suppl 3):A6.

Submit your next manuscript to BioMed Central  
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

