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

Nathan Medina-Rodríguez1,2*, Angelo Santana1, Ana M Wägner3, José M Quinteiro2

Table 1 Possible allelic combinations in a parent-offspring pedigree

| Parents | 1 | 2 | Offspring1 | 1 | 2 | 3 | 4 | Case |
|---------|---|---|------------|---|---|---|---|------|
| Alleles | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | (variant) |
| a a a a | 1 | 1 | a a | 1 |
| a a b b | 2 | 2 | a b | 2 |
| a a a b | 3 | 3 | a a a b | 3 (I) |
| a b b b | 3 | 3 | a b b b | 3 (II) |
| a a b c | 4 | 4 | a b a c | 4 (I) |
| a c b b | 4 | 4 | a b b c | 4 (II) |
| a b c c | 4 | 4 | a c b c | 4 (III) |
| a b a b | 5 | 5 | a a a b b | 5 |
| a b a c | 6 | 6 | a a a b a c | 6 (I) |
| a b b c | 6 | 6 | a b a c b b c | 6 (II) |
| a c b c | 6 | 6 | a b a c b c c | 6 (III) |
| a b c d | 7 | 7 | a c a d b c b d | 7 (I) |
| a c b d | 7 | 7 | a b a d b c c d | 7 (II) |
| a d b c | 7 | 7 | a b a c b d c d | 7 (III) |

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

1Department 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

© 2014 Medina-Rodríguez et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
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 (≤2000) and SNPs (≤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

1Department of Mathematics, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain. 2IUMA - Information and Communication Systems, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain. 3Department 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.