*Bioinformatics*, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Application Note

|  |
| --- |
| Genome analysis  Supplementary Material  HapSampler: a Metropolis-Hastings sampler to calculate the probability that a haplotype contains an allele of interest  Andrew George1, John M. Henshall2,3, Laercio R. Porto-Neto2, Yutao Li2, Sigrid A. Lehnert2, Antonio Reverter2,\* and William Barendse2  1CSIRO Data61, 41 Boggo Road, Dutton Part, Queensland 4102, Australia, 2CSIRO Agriculture & Food, 306 Carmody Road, St. Lucia, Queensland 4067, Australia, 3Present Address, Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, USA.  \*To whom correspondence should be addressed.  Associate Editor: XXXXXXX  Received on XXXXX; revised on XXXXX; accepted on XXXXX  Abstract  **Motivation:** Haplotypes offer the potential of identifying whether an individual carries an allele of a quantitative trait locus (QTL) when causative mutations are unknown. Most methods of analysis of haplotypes are deterministic and analyses are neither uniform nor standardized.  **Results:** Here we present a method that uses a Metropolis-Hastings sampler to assign a probability that an individual carries the QTL allele of interest and the probability that a haplotype contains the QTL. This method easily updates as more haplotypes are discovered. It is suitable for the analysis of both discrete and quantitative traits.  **Availability and Implementation:** The HapSampler software embodies the method and is available on GitHub as an R package (https://github.com/geo047/HapSampler).  **Contact:** [tony.reverter-gomez@csiro.au](mailto:tony.reverter-gomez@csiro.au) |

# Supplementary Tables

**Supplementary Table 1**. Example of a penetrance function for Polled phenotypes and genotypes

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Diploid Genotype** | | |
| **Phenotype** | PP | PH | HH |
| **Horned** | 0.05 | 0.10 | 0.85 |
| **Polled** | 0.49 | 0.49 | 0.02 |
| **Scurred** | 0.20 | 0.79 | 0.01 |

**Supplementary Table 2.** Example of an input data file of haplotypes of animals of Polled and horned phenotypes

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Phen** | **hap1** | **hap2** |
| **1** | horned | 0000 | 1111 |
| **2** | polled | 1111 | 1010 |
| **3** | scurred | 1010 | 0101 |
| **4** | polled | 0000 | 1010 |

**Supplementary Table 3.** Example of a set of candidate samples (S1 being the first sample) given the input data file in Supplementary Table 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Candidate Samples** | | | | | |
| **Hap** | S1 | S2 | S3 | S4 | S5 | S6 |
| **0000** | H | H | P | H | H | H |
| **1111** | H | H | H | H | H | H |
| **1010** | P | P | P | P | H | P |
| **0101** | P | H | H | P | P | P |
| **Lkld** | 0.04 | 0.16 | 0.02 | 0.04 | 0.00 | 0.04 |

**Supplementary Table 4.** Candidate sample likelihoods given the penetrance function and the phenotypes and haplotypes given the set of candidate samples in Supplementary Table 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Candidate Sample Likelihoods** | | | | | |
| **Hap** | L1 | L2 | L3 | L4 | L5 | L6 |
| **0000** | 0.85 | 0.85 | 0.10 | 0.85 | 0.85 | 0.85 |
| **1111** | 0.49 | 0.49 | 0.49 | 0.49 | 0.02 | 0.49 |
| **1010** | 0.20 | 0.79 | 0.79 | 0.20 | 0.49 | 0.20 |
| **0101** | 0.49 | 0.49 | 0.49 | 0.49 | 0.02 | 0.49 |

**Supplementary Table 5.** Accumulated samples and their probability given the candidate sample likelihoods in Supplementary Table 4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Accumulated Samples** | | | | | | |
| **Hap** |  |  |  |  |  |  | Prob(P) |
| **0000** | H | H | H | H | H | H | 0/6 |
| **1111** | H | H | H | H | H | H | 0/6 |
| **1010** | P | P | P | P | P | P | 6/6 |
| **0101** | P | H | H | H | H | P | 2/6 |

**Supplementary Table 6.** Genetic and phenotypic variance of the simulated phenotypes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trait** | **Size** | **NoiseSD** | **Varg** | **Varp** | **Diff** | **SD** |
| **S1.1** | 1.0 | 0.10 | 0.26 | 0.27 | 0.998 | 0.511 |
| **S1.2** | 1.0 | 0.51 | 0.26 | 0.52 | 0.946 | 0.729 |
| **S1.3** | 1.0 | 0.71 | 0.26 | 0.78 | 1.107 | 0.855 |
| **S2.1** | 1.0 | 0.10 | 0.08 | 0.09 | 0.159 | 0.381 |
| **S2.2** | 1.0 | 0.29 | 0.08 | 0.17 | 0.174 | 0.498 |

The contents of Supplementary Table 6 are the size of the phenotypic effect, the amount of Gaussian noise added to the phenotype, the actual genetic variance due to the SNP genotype modelled as a phenotype, the total expected variance given the size of the noise, the observed difference between homozygous genotypes identified by HapSampler for the QTL, and the observed standard deviation of the sample of individuals separated into haplotype groups. The difference in genetic variance between the two phenotypes is due to different allele frequencies at the two SNP.