Detecting past contraction in population size using haplotype homozygosity
C Merle, Jean-Michel Marin, F. Rousset, Raphaël Leblois

To cite this version:
C Merle, Jean-Michel Marin, F. Rousset, Raphaël Leblois. Detecting past contraction in population size using haplotype homozygosity. Mathematical and Computational Evolutionnary Biology 2016, Jun 2016, Montpellier, France. hal-02932275

HAL Id: hal-02932275
https://hal.inrae.fr/hal-02932275
Submitted on 7 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Detecting past contraction in population size using haplotype homozygosity

C. Merlo, J-M. Marin, P. Roussel and R. Leblois
1 - Institut de Montpellier Alexander Grothendieck (IMAG, UM); 2 - Centre de Biologie pour la Gestion des Populations (CBGP, INRA); 3 - Institut de biologie computationnelle (IBC); 4 - Institut des Sciences de l’Evolution (ISEM, CNRS)

Next generation genetic data

Classical inference methods of the demographic history (likelihood based methods, IS, MCMC), approximate bayesian methods and Site Frequency Spectrum, suitable for polymorphism data sets consisting in some loci and assuming the genealogies of different loci are independent, do not exploit the recombinetation.

The patterns of linkage disequilibrium (LD) between polymorphic markers are shaped by the ancestral population history as we can see on haplotype homozygosity curves which measure the LD.

Demographic history inference from genome wide sequence data

Haplotype Homozygosity: $HHi(i)$ : probability for $i$ adjacent markers drawn at random in the whole genome sequence to be homozygote.

Theoretical $HHi$ of [3], denoted $HH_{0i}$ : coalescent based computation, assuming the mutation and recombination rate known and constant along the genome.

Empirical $HHi$, denoted $HH$ computed from the observed data, is the segment proportion of at least $i$ homozygotes adjacent markers.

We estimate $\theta_i$ and $\theta_i$ by:

$$\theta_i \in \arg \min_{\theta_i} \sum_{i=0}^{L} \left( HHi(i) - \tilde{HH}(i) \right)^2$$

$\tilde{HH}(i)$ : estimate of Sobs's sensitivity index of order one of a given parameter $\phi$ computed for $i$ adjacent markers under the more complex model (sensitivity R package).

On this simulated data set with constant population size $\theta_i = 500$, the non penalized mean square criterion would choose $M_{0i}$ whereas the penalized mean square criterion choose $M_{i}$.

Penalized Adj. $	heta_i \in \arg \min_{\theta_i} \sum_{i=0}^{L} \left( HHi(i) - \tilde{HH}(i) \right)^2$

Numerical results

<table>
<thead>
<tr>
<th>Data sets simulated under $M_{0i}$, $\theta_i = N_{E0} = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_{pen}</td>
</tr>
<tr>
<td>Data sets</td>
</tr>
<tr>
<td>Adjustment</td>
</tr>
<tr>
<td>Penalized Adj.</td>
</tr>
</tbody>
</table>

Holstein data sets

<table>
<thead>
<tr>
<th>Data sets simulated under $M_{0i}$, $\theta_i = N_{E0} = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_{pen}</td>
</tr>
<tr>
<td>Data sets</td>
</tr>
<tr>
<td>Penalized Adj.</td>
</tr>
</tbody>
</table>

References


Acknowledgements