Ruminal bacteria heritabilities: Which is the impact of different data processing?
Guillermo Martinez Boggio, Annabelle Meynadier, Charlotte Allain, Christel Marie-Etancelin

To cite this version:
Guillermo Martinez Boggio, Annabelle Meynadier, Charlotte Allain, Christel Marie-Etancelin. Ruminal bacteria heritabilities: Which is the impact of different data processing?.. EAAP 2021, Aug 2021, Davos, Switzerland. hal-03480858

HAL Id: hal-03480858
https://hal.inrae.fr/hal-03480858
Submitted on 14 Dec 2021

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.
Ruminal bacteria heritabilities: Which is the impact of different data processing?

Guillermo Martinez Boggio¹, Annabelle Meynadier¹, Charlotte Allain², Christel Marie-Etancelin¹
¹GenPhySE, Université de Toulouse, INRAE, ENVT, France. ²INRAE, Experimental Unit of La Fage, France

guillermo.martinez-boggio@inrae.fr

Introduction and objective

Microbiome data is compositional: the information is found in the operational taxonomic unit (OTU) ratios.

As high-dimensional data with zeros, it is necessary to replace the zeros, and apply value transformation.

To compare data processing methods in terms of heritability estimates ($h^2$).

Materials and methods

Rumen samples of 795 Lacaune ewes Metabarcoding 16s rRNA gene

Table 1. Data processing methods to be compare.

<table>
<thead>
<tr>
<th></th>
<th>GBM</th>
<th>C1</th>
<th>Z01</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>GBM-CLR</td>
<td>C1-CLR</td>
<td>Z01-CLR</td>
</tr>
<tr>
<td>Log</td>
<td>GBM-Log</td>
<td>C1-Log</td>
<td>Z01-Log</td>
</tr>
</tbody>
</table>

Zeros replacement:

<table>
<thead>
<tr>
<th></th>
<th>GBM= Geometric Bayesian Method</th>
<th>C1= Add constant equal 1</th>
<th>Z01= Replace only zeros with 0.001</th>
</tr>
</thead>
</table>

Data transformation:

<table>
<thead>
<tr>
<th></th>
<th>CLR= Centred Log-Ratio transformation</th>
<th>Log= Logarithm transformation</th>
</tr>
</thead>
</table>

Heritability estimates*:

$y = Xb + Za + e$

*The threshold of significance obtained by phenotype-animal permutation

Table 2. Number of heritable* OTUs by processing data method.

<table>
<thead>
<tr>
<th></th>
<th>GBM</th>
<th>C1</th>
<th>Z01</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>326</td>
<td>293</td>
<td>249</td>
</tr>
<tr>
<td>Log</td>
<td>294</td>
<td>274</td>
<td>254</td>
</tr>
</tbody>
</table>

But, the number of heritable OTUs defined by the methods were variable.

GBM > C1 > Z01

By replacing only the zeros we obtain lower number of heritable OTUs.

% OTUs in common between CLR and Log

78% 72%

Using GBM the differences between CLR and Log are lesser than with C1.

Conclusions

Ruminal bacteria $h^2$ are affected by data processing.

Zeros replacement impact the most the number of heritable OTUs.

GBM leads to less difference between the two data transformation.

Table 2. Number of heritable* OTUs by processing data method.

<table>
<thead>
<tr>
<th></th>
<th>GBM</th>
<th>C1</th>
<th>Z01</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>326</td>
<td>293</td>
<td>249</td>
</tr>
<tr>
<td>Log</td>
<td>294</td>
<td>274</td>
<td>254</td>
</tr>
</tbody>
</table>

But, the number of heritable OTUs defined by the methods were variable.

GBM > C1 > Z01

By replacing only the zeros we obtain lower number of heritable OTUs.

% OTUs in common between CLR and Log

78% 72%

Using GBM the differences between CLR and Log are lesser than with C1.

Results

Overall no differences in $h^2$ between the 6 methods.

Threshold of significance = 0.09

Overall no differences in $h^2$ between the 6 methods.