Identifying genomic signatures of niche specialisation in the rumen microbiome

Prof. Chris Creevey

Chair of Computational Biology,
Institute for Global Food Security,
School of Biological Sciences,
Queens University Belfast.

Email: c.creevey@qub.ac.uk
www.creeveylab.org
Twitter: @hairy_llama @creeveylab

The Institute For Global Food Security
Does niche specialisation for plant cell wall degradation exist?

What drives Niche specialisation in the Rumen?

The Conundrum

- Somehow niche specialisation of specific groups of organisms is maintained even under the continued influence of horizontal gene transfer.

- One explanation may be that the horizontal acquisition of a single isoform of a novel gene is not enough to maintain a competitive advantage in fluctuating environmental conditions where a range of isoforms may be required to, for instance, maintain enzymatic activity.

- This could be described as negative frequency dependent selection.
Frequency-dependent selection.

- In **positive** frequency-dependent selection, the fitness of a phenotype increases as it becomes more common.

- In **negative** frequency-dependent selection, the fitness of a phenotype decreases as it becomes more common.
Drivers of Niche Specialisation

• Can we use genetic information from the microbiome to identify the drivers of niche specialization?

• We should be able to use adaptive evolution and population genetics approaches developed in other systems to address these questions.

• This will allow us to test whether negative frequency dependent selection drives niche specialization.
Metagenomic data can result in “average” assemblies
14 X Rumen samples

Prepare DNA, sequence randomly

Assemble Contigs

Identify SNPs

Align reads to assembly

Predict Genes

Calculate sequence Diversity/Evolutionary rate
  i.e. pN/pS
  Schloissnig et al 2012

Map to Metabolic Processes

Rubino & Creevey Mgkit: Metagenomic Framework For The Study Of Microbial Communities
https://bitbucket.org/setsuna80/mgkit
Estimating adaptive diversity estimates from metagenomic data using SNPs

\[ \frac{pN}{pS} = \frac{oN}{eN} \frac{oS}{eS} \]

- \( oN \) = number of observed non-synonymous SNPs
- \( eN \) = number of “expected” non-synonymous SNPs
- \( oS \) = number of observed synonymous SNPs
- \( eS \) = number of “expected” synonymous SNPs

Enzymes categories with a higher adaptive diversity in Prevotella

Isomerases in Prevotella may be specialised for the action on xylose (a major component of hemicellulose) to release sugars.

Enzymes categories with a higher adaptive diversity in Clostridium

Glycosylases (in particular cellulases – EC 3.2.x) in Clostridium are specialised for the breakdown of Cellulose.

Identifying population-level paths in sequencing reads from metagenomic data

Hansel: A graph-inspired data structure for determining likely chains of sequences from breadcrumbs of evidence
https://github.com/SamStudio8/hansel

An algorithm for recovering haplotypes from metagenomes
https://github.com/SamStudio8/gretel
In vitro validation of predicted haplotypes

(a) G31  (b) G90  (c) G123  (d) G152  (e) G251

(a) Careful loading of prepared library  (b) MinION in use, reporting real-time status to connected laptop (excited person for scale)  (c) A MinION flow cell
Butyrivibrio proteoclasticus
Glycosyl hydrolase family 10 haplotypes
Microbial community composition data from the Global Rumen Census overlaid with the 16S rRNA gene sequences (red dots) from the Hungate genomes.
Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome

Weibing Shi, Christina D. Moon, Sinead C. Leahy, Dongwan Kang, Jeff Froula, Sandra Kittelmann, Christina Fan, Dragana Gagic, Henning Seedorf, William J. Ke Carrie Sang, Priya Soni, Dong Li, Cesar S. Pina Peter H. Janssen, Feng Chen, Axel Visel, Graeme T. Attwood, and Edward M. Rubin

1 Department of Energy, Joint Genome Institute, Walnut Creek, California 94597, USA; 2Agrisearch Limited, Grassland Research Institute, New Zealand; 3School of Natural Sciences, University of California, Merced, California 95343

Summary

Ecological and evolutionary concepts can be applied to microbiomes, but require adjusting for the unique properties of these systems.

Suggests that niche specialisation and homogeneity in the rumen microbiome is driven by an interplay between HGT, pangenomes and frequency-dependent selection.

Provides a set of hypotheses that can be tested and a mechanism for identifying functions under selection in novel datasets or microbiomes.
# Acknowledgements

@Creeveylab present and past members:

![Image of lab members](image)

<table>
<thead>
<tr>
<th>Karen Siu Ting</th>
<th>Hilario Mantovani</th>
<th>Sinead Waters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sam Nicholls</td>
<td>James McInerney</td>
<td>Graeme Attwood</td>
</tr>
<tr>
<td>Ben Thomas</td>
<td>Mary O’Connell</td>
<td>Matthias Hess</td>
</tr>
<tr>
<td>Nick Dimonaco</td>
<td>Kai Hilpert</td>
<td>Diego Morgavi</td>
</tr>
<tr>
<td>Jess Friedersdorff</td>
<td>Toby Wilkinson</td>
<td></td>
</tr>
<tr>
<td>Toby Wilkinson</td>
<td>Martin Hughes</td>
<td></td>
</tr>
<tr>
<td>Martin Hughes</td>
<td>Max Friedersdorff</td>
<td></td>
</tr>
<tr>
<td>Tom Hitch</td>
<td>Francesco Rubino</td>
<td></td>
</tr>
<tr>
<td>Francesco Rubino</td>
<td>Bruce Moran</td>
<td></td>
</tr>
<tr>
<td>Tom Hitch</td>
<td>Anthony Doran</td>
<td></td>
</tr>
</tbody>
</table>

Our collaborators:

<table>
<thead>
<tr>
<th>Sharon Huws</th>
<th>Hilario Mantovani</th>
<th>Sinead Waters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alison Kingston-Smith</td>
<td>James McInerney</td>
<td>Graeme Attwood</td>
</tr>
<tr>
<td>Jamie Newbold</td>
<td>Mary O’Connell</td>
<td>Matthias Hess</td>
</tr>
<tr>
<td>Joan Edwards</td>
<td>Kai Hilpert</td>
<td>Diego Morgavi</td>
</tr>
</tbody>
</table>

**Contact:**
- c.creevey@qub.ac.uk
- www.creeveylab.org
- @hairy_llama
- @creevey_lab