

Joint RuminOmics/Rumen Microbial Genomics Network Workshop

Microbiome analysis - Amplicon sequencing

Dr. Sinéad Waters

**Animal and Bioscience Research Department, Teagasc
Grange, Ireland**

Prof. Leluo Guan

**Agricultural, Food and Nutritional Science Department,
University of Alberta, Canada**



DEPARTMENT OF
**AGRICULTURAL, FOOD
& NUTRITIONAL SCIENCE**
UNIVERSITY OF ALBERTA

Teagasc Grange Animal Bioscience

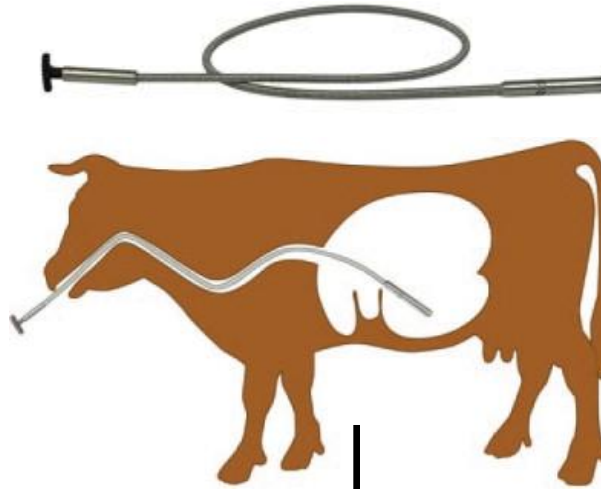


Sample collection

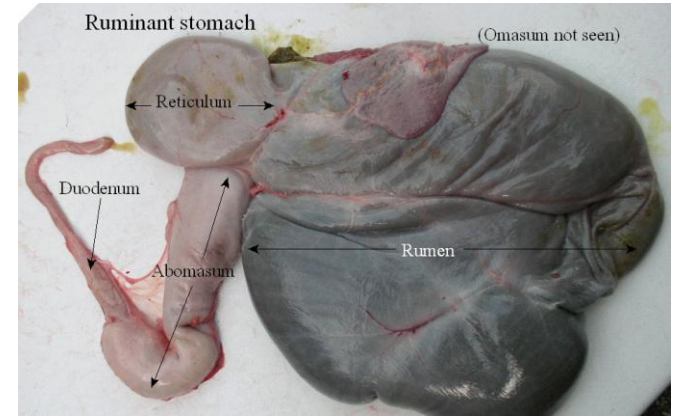
Rumen Fistulation



Stomach tubing



Slaughter

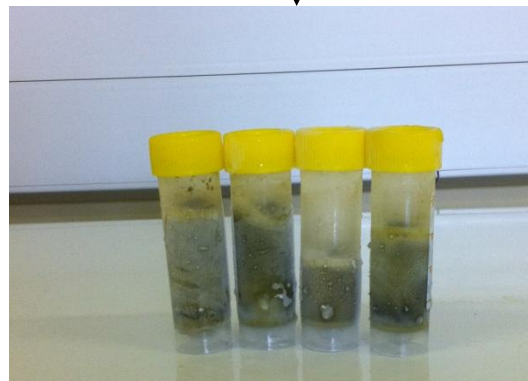


Fistulation: Collect samples from 5 different locations and pool?

Which Fraction?

Solid
Liquid
Epimural

Is the rumen sample representative?

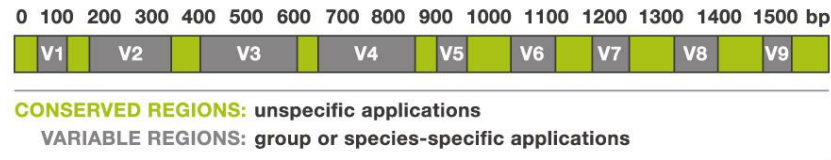


Snap frozen at -80°C

Rumen metagenomics

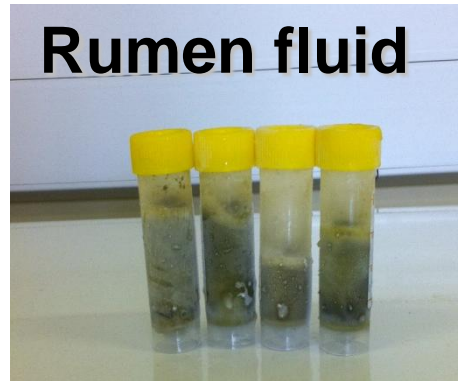
Study of the metagenome or genetic material isolated from the rumen

- Amplicon sequencing
 - 16S ribosomal RNA (rRNA) gene



- 1500 bp – interspersed conserved and variable regions
- Facilitates sequencing and phylogenetic classification
- Targeted amplicon sequencing

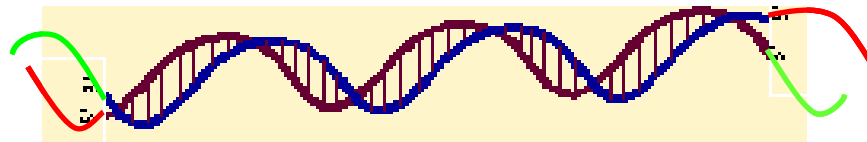
DNA extraction



Beat beating protocol



DNA



Extraction and Quantification

- Modification of repeated bead beating and column purification (RBBC) method described by Yu and Morrison (2004)
- Efficient recovery of PCR-quality microbial DNA
 - Liquid and solid - crushed to a powder under liquid nitrogen
 - Zirconia beads
 - Cells lysed using Fast prep tissue and cell homogeniser
 - Supernatant – protein precipitated (10M ammonium acetate)
 - DNA precipitated – isopropanol
 - DNA pellet washed using ethanol and re-suspended in TE
 - RNA removal – DNase free RNase
 - DNA purified using the QIAmp stool Mini Kit Clean up Qiagen kit
 - Integrity of high molecular weight DNA and removal of RNA verified using agarose gel electrophoresis
 - Concentration and purity of DNA determined using the nanodrop

DNA spike-in standards

- Spike with synthetic external standard
- For 16S amplicon sequencing, spike DNA with 16S/ERCC fusion amplicons made from ERCC synthetic cDNA
- Recalibrate base quality scores

Amplicon sequencing workflow

- Design region of interest specific primers with over hang adapters
- Generate template library by PCR
- Attach indices and sequence adapters to libraries by PCR
- Normalise and pool libraries
- Perform automated cluster generation and sequencing (on MiSeq system)
- Data analysis

Amplicon sequencing

- **Targets:** usually a region of the 16S or 18S rRNA gene
- **Groups:** Bacteria, archaea, fungi and protozoa
- Barcoded/indexed fusion primers
- PCR products purified using a magnetic purification kit (Ampure®)
- Concentration determined: qPCR, agilent bioanalyser, Qubit fluorimeter and picogreen assay
- 15-30% Phix run with amplicon pool – control library

Taq

- Finding a GC bias free 16S rDNA free Taq
- 16S rDNA-free Taq (Bioron DFS) for bacterial amplicon sequencing

Adapters

- Ligate Illumina Truseq adapters to 16S amplicon
 - expensive, time consuming, problems with adapter dimers etc..
- Adapter/16S fusion primers
 - Illumina method
 - *Caporaso method

*Caporaso et al. (2012) The ISME Journal. 6, 1621-1624.

Caporaso method:

One PCR step

Full adapter and 16S target on one primer pair

Nine extra bases in the adapter region of forward primer - support paired end sequencing

2,168 x 12 bp indexes on reverse primers

2,168 samples in one 250 bp PE Miseq run

Cannot mix different targets in one run because sequencing primers are complementary to targets

Illumina method:

2 PCR steps

Partial Truseq adapters and 16S target on first primer pair

96 x Dual 6 bp indexes (on forward and reverse primers added using Nextera primers in second PCR)

96 samples in one run

Can mix different targets in one run because sequencing primers are complementary to Truseq adapter

Illumina fusion primer method

PCR amplify template out of genomic DNA using region of interest-specific primers with overhang adapters

Forward primer overhang adapter:
5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

Region of interest-specific primer

Reverse primer overhang adapter:
5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'



Attach indices and Illumina sequencing adapters using the Nextera® XT Index Kit



Normalize and pool libraries



Sequence

Targets

- Bacteria: Ba27F + Ba338R 16S (365 bp) Kittelmann et al (2013)
- Methanogens: amplify an approximately 550 bp region of the 16S rRNA gene using forward (Hegarty et al., 2007) and reverse (Zarda et al., 1997); Ar915aF Ar1386R (16S gene) (492 bp) Kittelmann et al (2013)
- Protozoa: RP841F + Reg1302R (18S gene) (511 bp) Kittelmann et al (2013)
- Fungi: MN100F + MNGM2 (ITS1 gene) (250 bp) Kittelmann et al (2013)

TAXANOMIC SEQUENCING

Sequencing Platforms

Individual PCR amplicons

Sanger sequencing



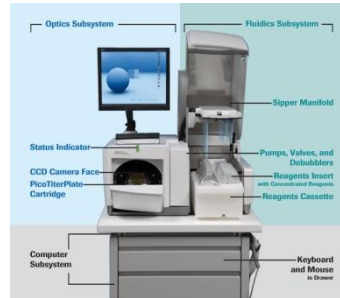
Chain termination, ~800 bp, 99.999%,
1.9-84 kb



Species/strain level

Pooled PCR amplicons

454 Sequencing (pyrosequencing)



GS FLX Titanium: up to 600 bp, 1M reads, 10hr
GS FLX plus: up to 700 bp, 0.56 Gb, 1M reads , 23 hr

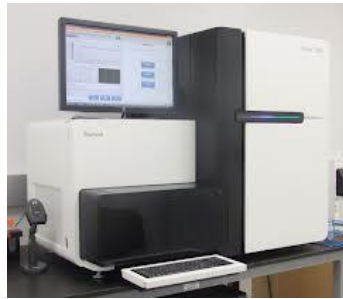


99.9% accuracy

Species level

Pooled PCR amplicons

ILLUMINA SEQUENCING (Sequencing by synthesis)



MiSeq: paired end 250 bp, 39 hr, ~8Gb,

HiSeq 2000: paired end 100 bp, 11 d, 600Gb, 3 billion reads

HiSeq 2500: rapid run, paired 150 bp, 25 hr, 120Gb, 600M reads



98% accuracy

Family/Genus level

Sequencing Platforms-others

Pooled PCR amplicons (continued)

SOLiD Sequencing: Sequencing by ligation, 50+50 bp, 99.94%, 120 Gb

Ion Torrent Sequencing: Detecting dNTP protons (addition of a dNTP releases a H⁺ ion) , 400 bp, 98.2%, 1Gb

PacBio Sequencing: Single molecules sequencing, 600 bp, 88.14%, 100 Mb

Data Analysis

Taxonomy Identification

16S rRNA databases and tools

- Ribosomal Database Project (RDP)
 - <http://rdp.cme.msu.edu/>, 2.76M sequences
- Greengenes
 - <http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>, 408k sequences
- SILVA
 - <http://www.arb-silva.de/>, 3.19M sequences
- Global Alignment for Sequence Taxonomy (GAST)
 - <http://vamps.mbl.edu/resources/software.php>, improved SILVA database

Data Analysis

Sequence-based Community Analysis

Microbial communities analysis

- Quantitative Insights Into Microbial Ecology (QIIME)
 - <http://qiime.org>
- Mothur
 - <http://www.mothur.org/>
- UniFrac
 - <http://bmf.colorado.edu/unifrac/>,
- VEGAN within R
 - <http://cran.r-project.org/web/packages/vegan/index.html>

Targeted data output

How many sequence reads?

Community comparison for screening purpose: ~1000 sequences (reads) per sample (community)

Community identification: ~10,000 sequences (reads)?

Rumen: 6,000 for bacteria, 2000 for archaea,
1000 for protozoa, 1000 for fungi

The more the better?