Microbiome for protected status

Marc Kennedy, Edward Haynes
(Fera Science)

Project overview

- Defra/FSA funded research "Development of Metagenomic Methods for Determination of Origin" FA0141 (Phase1) – FA0160 (Phase 2)

- Overall Project Aim: develop a new non-targeted approach for determining origin and attribution of food products using microbial fingerprints

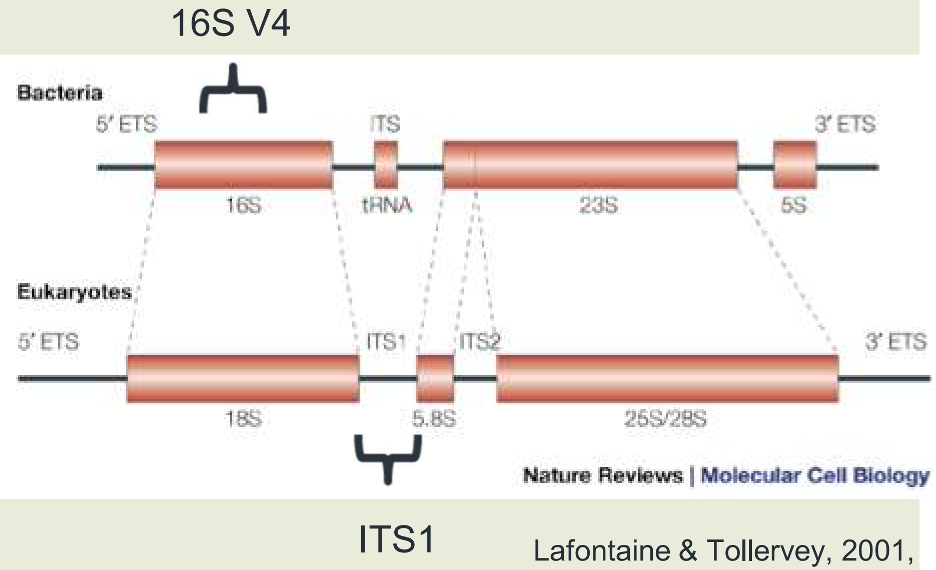
- Focussed on two foodstuffs, Pacific oysters (*Crassostrea gigas*) from the UK and France, and Stilton cheese (PDO)

Data – Oyster Samples

- UK oyster samples taken from four locations. Ten individual oysters sequenced from each location at a given time point (Cefas)
- Five time points sampled at each location in 2015 (except Cornwall), Six time points sampled at each location in 2016
- Two French locations sampled at three time points in 2016/17

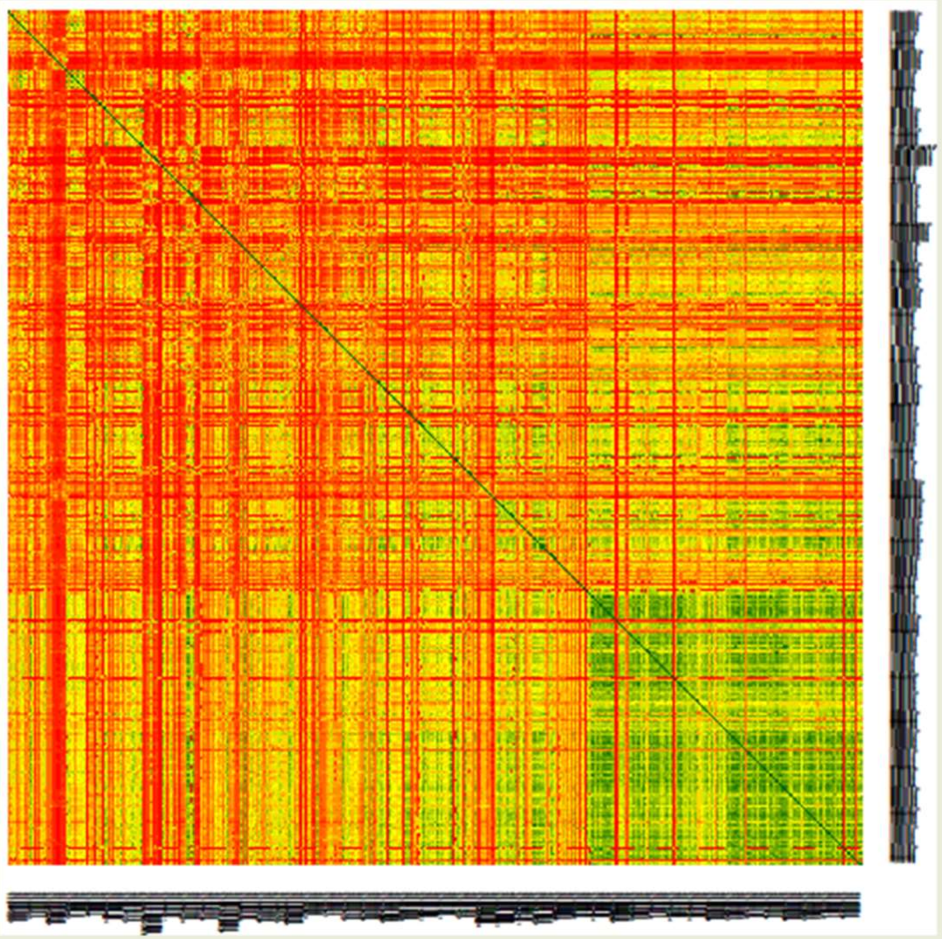
Methods - Next Generation Sequencing (NGS)

- DNA extracted from oyster gills/cheese samples
- Variable region V4 of 16S gene amplified using PCR for assessing bacterial populations in oysters and cheese
- Internal Transcribed Spacer (ITS) region 1 used for assessing fungal populations in cheese
- Products sequenced on illumina MiSeq platform



Lafontaine & Tollervey, 2001, doi:10.1038/35080045

Bray Curtis dissimilarity metric



Samples arranged by time of sampling

- Some areas of green (similarity between sites, within time points), but many areas of red
- ANOSIM supports the conclusion that samples are too diverse to reliably assign to location
- Large variations between time points were also observed

2 Classification methods tested

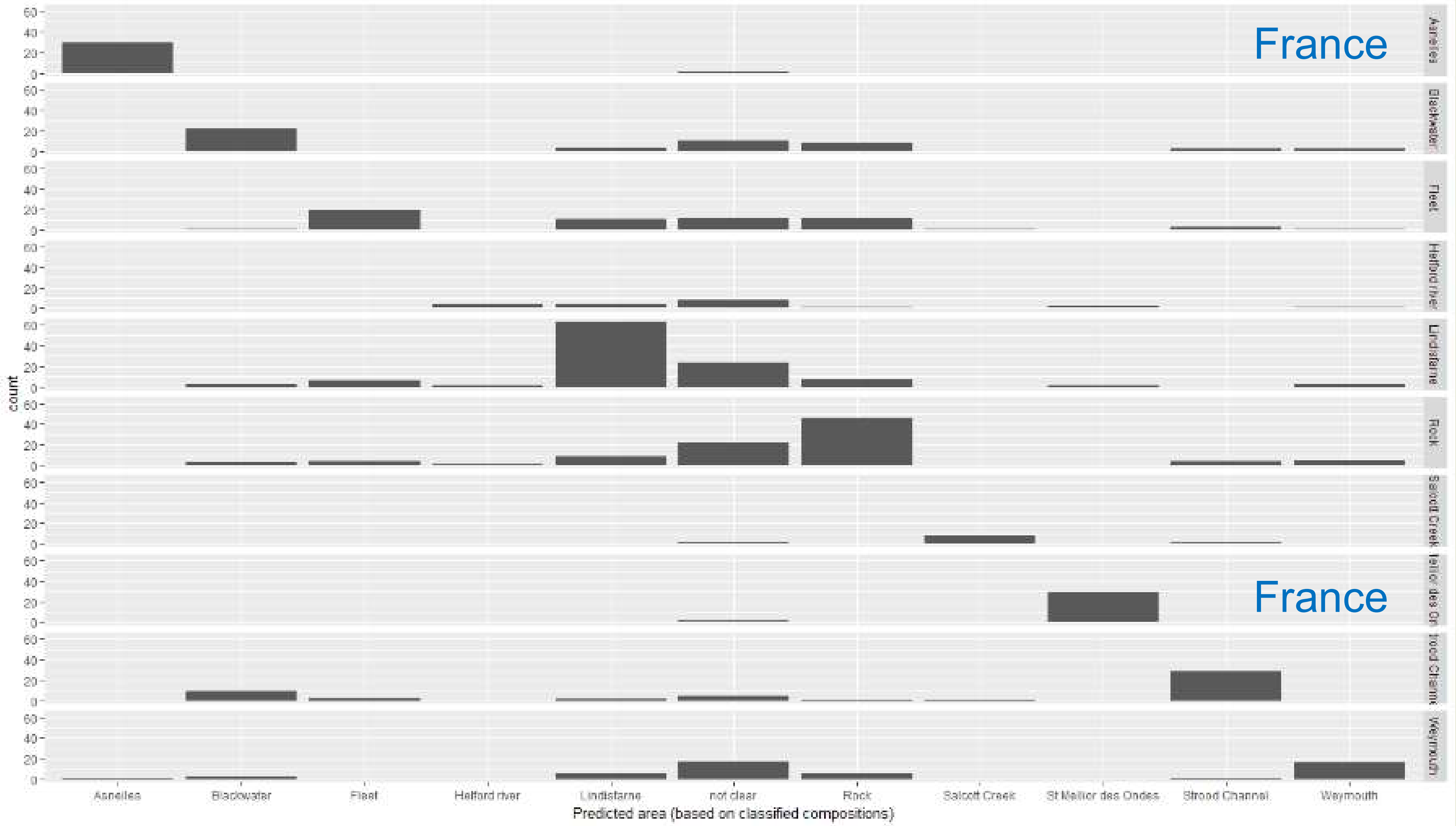
- Processed data represent counts of taxa (OTUs) – large matrix
  - Cheese data includes counts for 47,443 OTU in 101 samples
  - Sparse matrix, many zero counts

- High-Dimensional Regularized Discriminant Analysis (HDRDA)
  - Mathematical algorithm reduces a large numerical matrix to low-dimensional representation, does not assume count data
  - Fast implementation, suitable for sparse matrix with columns >> rows
- Latent Dirichlet Allocation
  - Infers properties of the underlying probabilistic mechanism generating the counts
  - CPU intensive, slow, but potentially more accurate

- Repeated cross-validation used to assess accuracy
  - Classify a subset (10%) of held-back data, after training the method on the remaining (90%) of data
  - Examine how many of the samples are correctly classified

Oyster Results – Location predictions

Uncertain cases



France

France

True source locations

Predicted locations

Summary - Oysters

- Ability to correctly assign origin varied by analysis method, and geographical area

- Approximately 35% correctly assigned to specific site, 50% to broad geographical region, 98% to sample country

- Country assignment may be an artefact of small sample size

Data – Cheese Samples

- Only six dairies are licensed to produce Blue Stilton cheese
- Cheeses sampled from producers (where possible) in 2016 and 2017
- Remaining samples obtained from retail locations

http://www.stiltoncheese.co.uk/the_stilton_producers

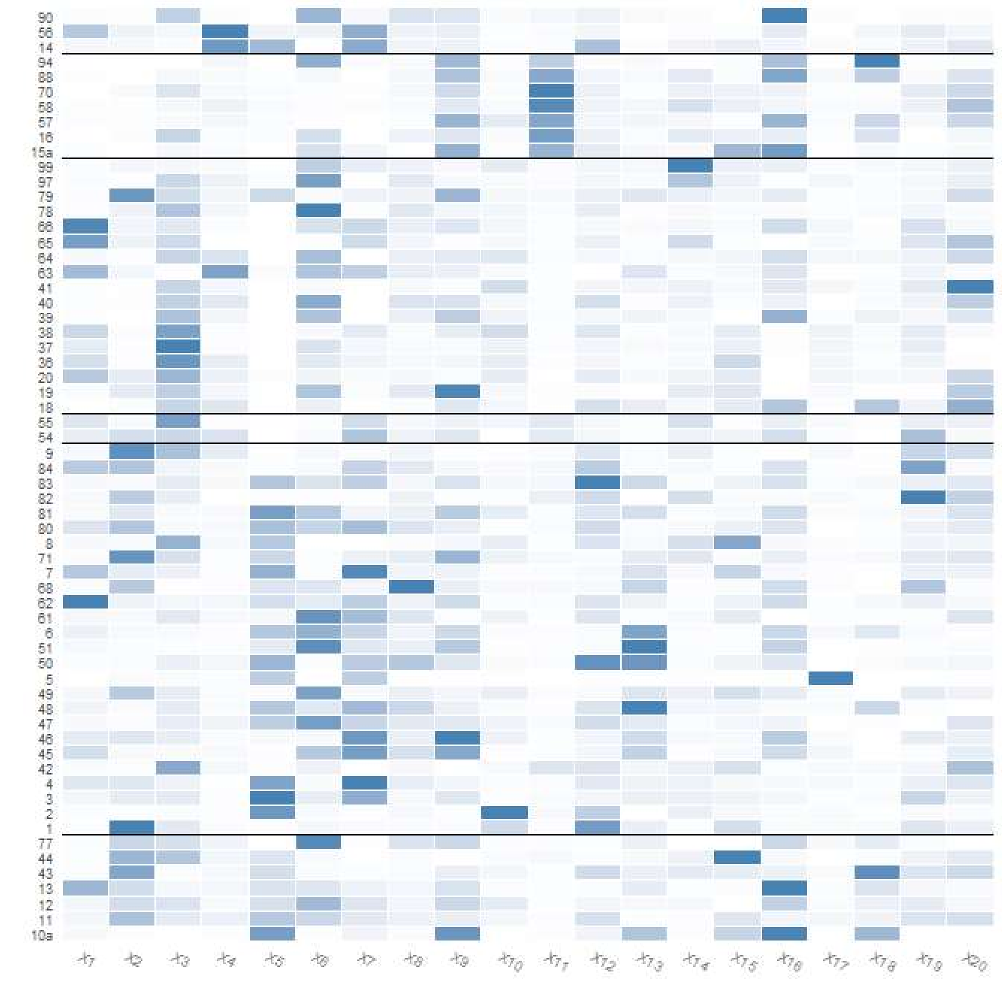
Location	Creamery
A	Colston Bassett Dairy
B	Cropwell Bishop
C	Hartington Creamery
D	Long Clawson Dairy
E	Tuxford & Tebbutt Creamery
F	Websters

Data – Cheese Samples

Cheese	Number of Samples	Number direct from Creamery
Colston Bassett	7	6
Cropwell Bishop	26	17
Hartington	2	0
Long Clawson	17	6
Tuxford and Tebbutt	7	0
Websters	3	0
Other UK	30	0
Other non-UK	9	0
Total	**101**	**29**

Results – Cheese Analysis

Bacterial Taxa



Websters

Tuxford and Tebbutt

Long Clawson

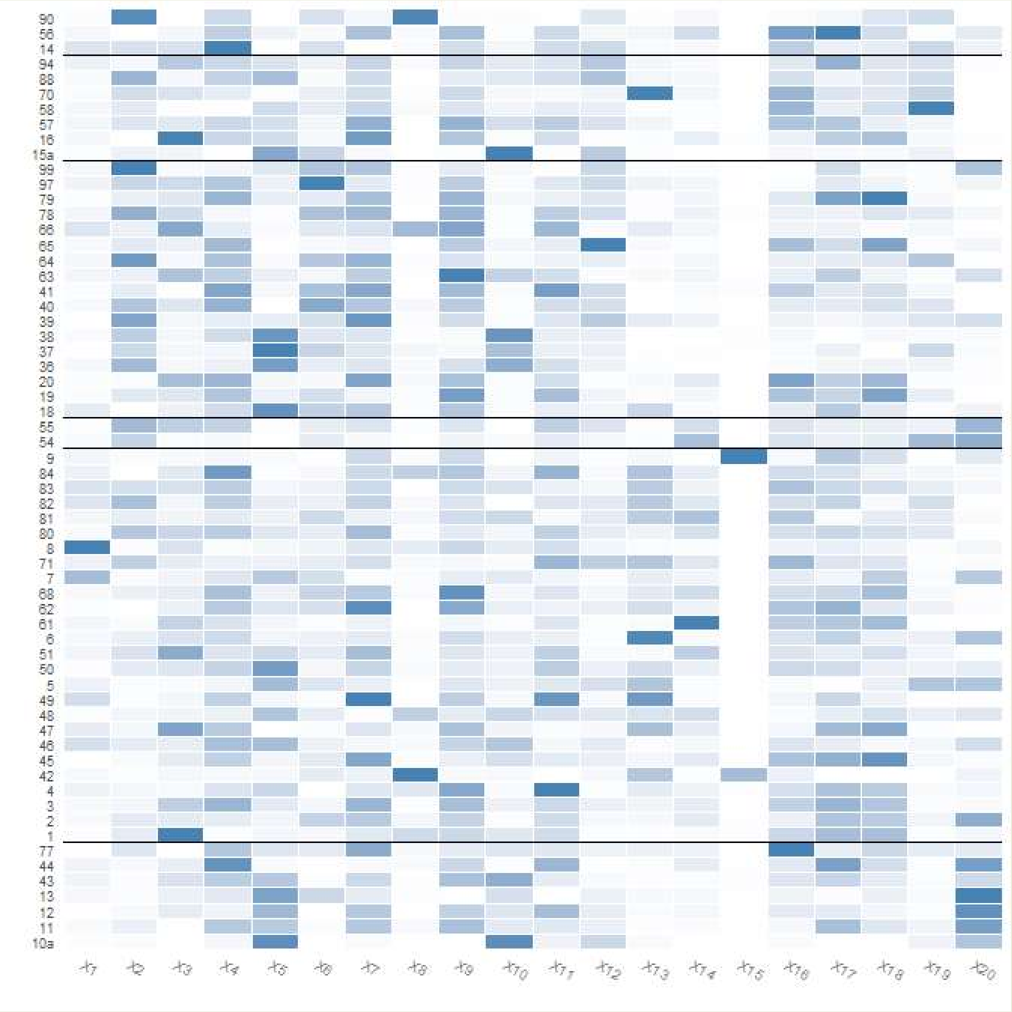
Hartington

Cropwell Bishop

Colston Bassett

Results – Cheese Analysis

Fungal Taxa



Websters

Tuxford and Tebbutt

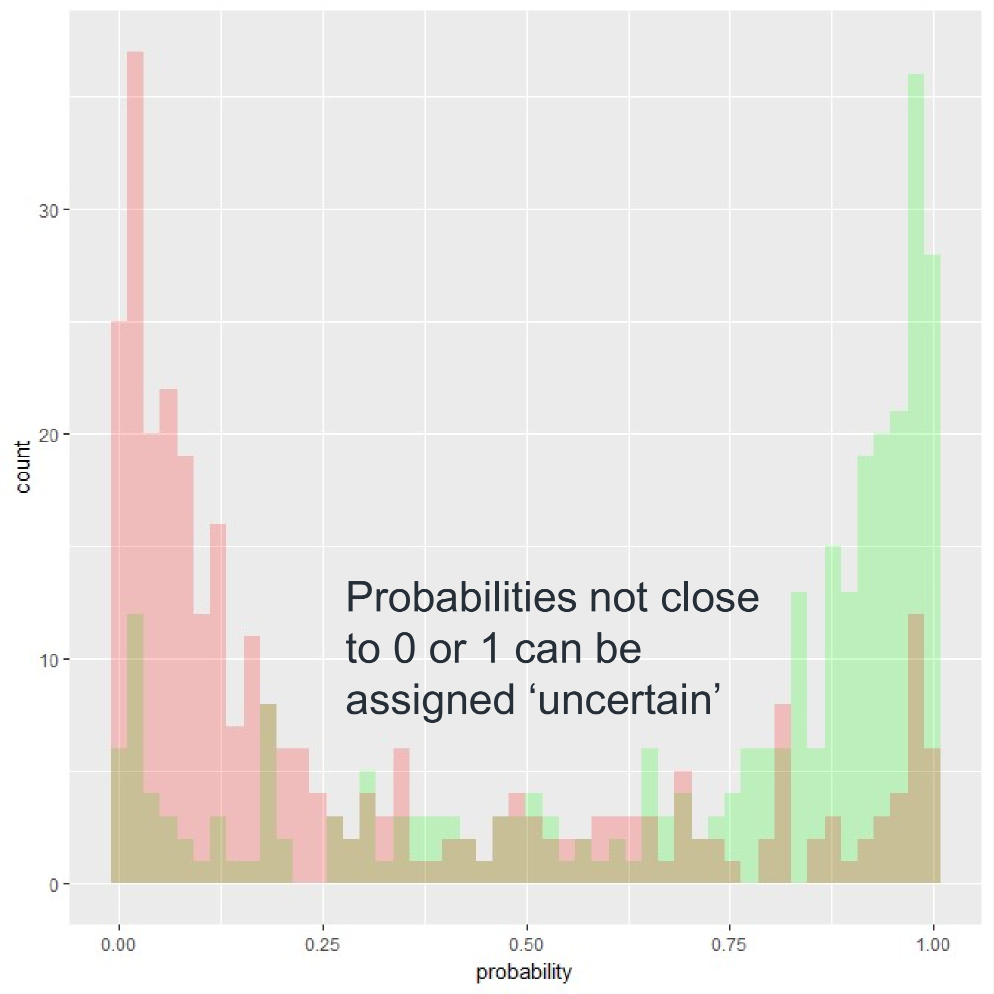
Long Clawson

Hartington

Cropwell Bishop

Colston Bassett

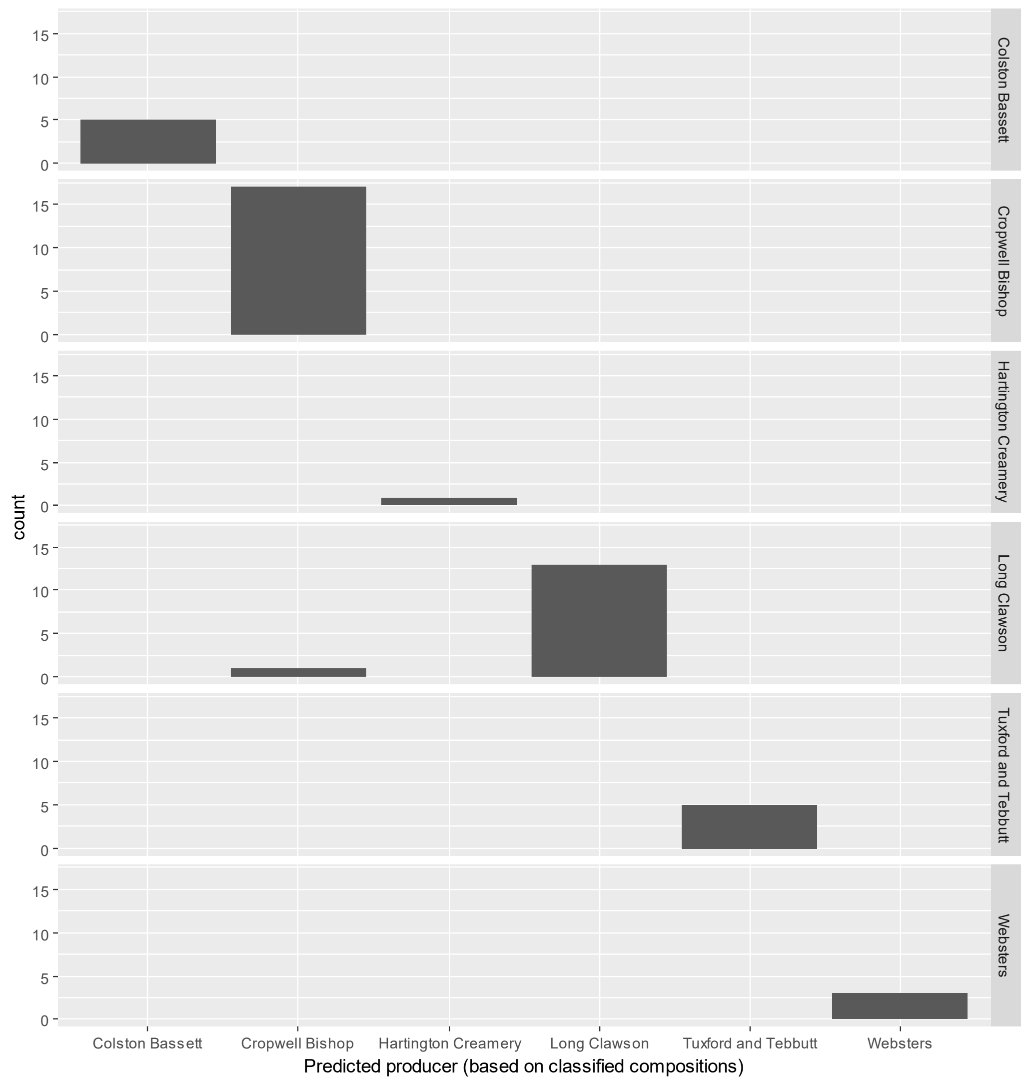
Can we distinguish Blue Stilton?



Probabilities not close to 0 or 1 can be assigned ‘uncertain’

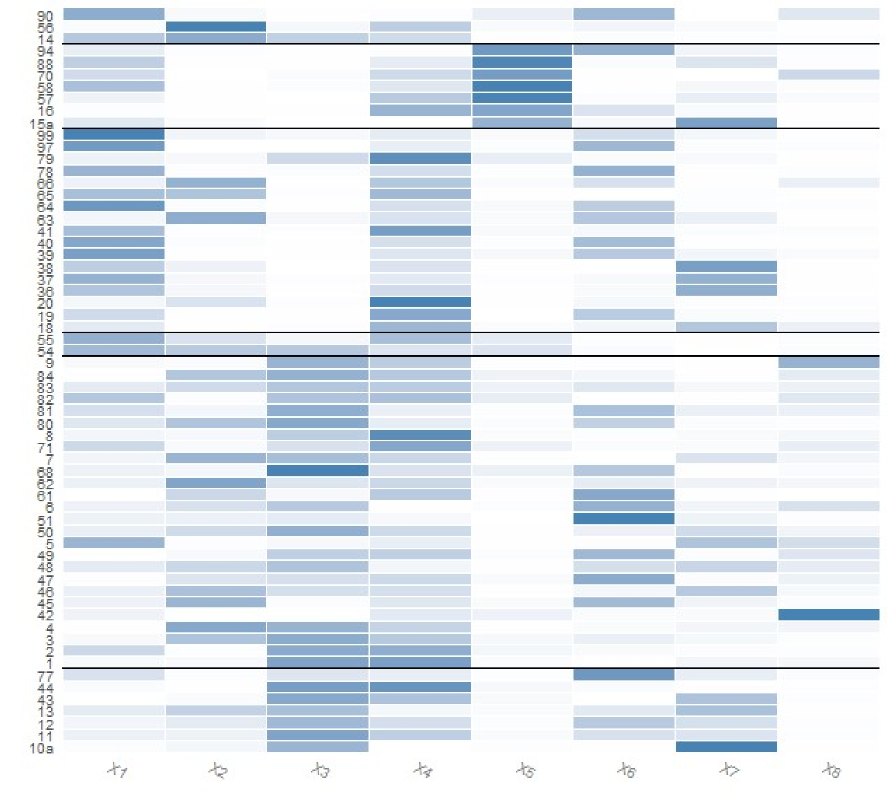
- Probability of predicting Blue Stilton, given a Blue Stilton sample (green) and other cheeses (red)

Can we distinguish Stilton Manufacturers?



Summary - Cheese Samples

- Blue Stilton can be distinguished from other blue cheeses with 80-90% accuracy
- Producer can be assigned with approx. 70% accuracy
- Non-UK Blue cheese can be distinguished from UK Blue cheese with approx. 90% accuracy
- Country assignment may be an artefact of small sample size



Different communities of fungi and bacteria (columns) are associated with different samples (rows, clustered by cheese producer). Some communities, such as “X5” in this example are strongly associated with particular producers.

Conclusions and Recommendations

- Microbial communities show promise as markers of origin, probably most relevant to fermented foods
- Oyster results: high variability within site, reduces ability to predict location
- Classification more accurate for cheese
  - Identifying Blue Stilton, or individual Stilton Manufacturers
- Unbiased survey design is essential to properly assess the methods
  - More samples required
- More experimentation possible
  - Fine-tuning e.g. to optimise the number of features/communities to search for
  - Consider new developments in machine learning

- Final report now published http://randd.defra.gov.uk

(Search FA0160 Development of Metagenomic Methods for Determination of Origin)