Spotty Liver Disease 16S Metagenomics First Report.

Dana Stanley[1], Peter Scott[2], Robert Moore[1]

[1]CSIRO Livestock Industries, [2]Scolexia Pty Ltd

Summary

The microflora present in the gut of birds suffering from spotty liver disease has been investigated. Clear, statistically significant differences in the abundance of particular bacteria were noted. Some types of bacteria were seen only in the diseased birds and are candidates for the pathogen which is responsible for precipitating the disease.

Introduction

Spotty Liver Disease (SLD) has been recognised in Australian commercial poultry, particularly those maintained under barn and free range systems, for many. The disease has been both frustrating to producers ad veterinary professionals as its primary cause has never been elucidated. Diagnosis is based on typical clinical findings, including negative microbiological results, with treatment being the implementation of broad spectrum antibiotics. The finding that antibiotics can ameliorate the disease indicates that the disease is likely to be of bacterial origin or at least influenced by the bacterial milieu. The agent causing spotty liver disease is unknown. It has been suggested that *Campylobacter jejuni* may be involved but this now seems unlikely as the disease cannot be reproduced using *C. jejuni* and *C. jejuni* is only sporadically isolated from disease cases (and of course *C. jejuni* is very commonly found, even in healthy birds). Other attempts to culture a causative microorganism have been unsuccessful. The identification of the causative agent would allow a better understanding of SLD and a more strategic approach to the control of the disease, for example, by vaccination.

As culture methods have been unsuccessful in identifying the causative agent, alternative methods are required if the etiological agent is to be found. Recent advances in DNA sequencing technology have opened up new ways of analysing bacterial populations and identifying specific tags for particular bacteria. These high throughput sequencing technologies (Roche/454 pyrosequencing) have been used in this study to investigating the diversity of bacteria present in the gut of SLD affected birds. The clinical findings in SLD birds have lead us to hypothesise that the disease may be caused by a bacterial toxin that, when it enters the systemic system, can cause lesions within the liver. We further hypothesise that the most likely location for such a pathogenic bacterium is within the gut. In this study we have investigated the bacterial composition of the intestinal contents and the caeca.

Materials and Methods

Samples

Samples were collected from two SLD outbreaks. From each outbreak samples were obtained from 6 affected birds and 6 age-matched healthy control birds from an unaffected shed on the same property.

DNA preparation from gut samples

Chicken spotty liver disease samples were supplied by Dr Peter Scott. Material from the intestinal content samples and caecal samples were resuspended in 250 µl of phosphate buffered saline. Total DNA from these samples was then isolated using the method of Yu and Morrison (2004). Briefly, a sample was transferred to a tube with lysis buffer (500mm NaCl, 50mM Tris-HCl pH8.0, 50 mM EDTA and 4% sodium dodecyl sulfate) and sterile zirconium beads and then homogenized using a Qiagen TissueLyser at maximum speed 3 times for 10 seconds. Following centrifugation the supernatant was collected ammonium acetate was added and nucleic acid was precipitated with isopropanol, followed by ethanol wash. After centrifugation the pellet was resuspended in Tris-EDTA buffer and digested with

DNase-free RNase and proteinase K to remove RNA and protein. The DNA was finally purified on a QIAamp column (Qiagen) according to the manufacturer's instructions. DNA quantity and quality was measured on a NanoDrop ND-1000 spectrophotometer.

PCR amplification of 16S ribosomal RNA gene sequences

DNA derived from the bacteria of the birds was processed to amplify the 5' end (V2-V3 region) of the eubacterial 16S ribosomal RNA genes. DNA was amplified using Bio-Rad iProof DNA polymerase. Each PCR reaction contained 25 µl of iProof 2X master mix (containing buffer, nucleotides and iProof enzyme), 2 µl of each primer (final concentration 0.5 µM), 1.5 µl DMSO, 0.5 µl 50 mM $MgCl_2$ and template DNA made up to 19 µl in water. The primers used were designed to amplify the V2-V3 region of the 16S rRNA gene (forward primer, 5' AGAGTTTGATCCTGG 3'; reverse primer, 5' TTACCGCGGCTGCT 3'). Each primer also included sequences to facilitate the sequencing of products in the Roche/454 system and the reverse primers consisted of a related set of primers which differed in "barcode" sequences; specific sequences introduced into the primers to allow tagging of individual samples in a multiplex sequencing system. PCRs were performed in an Eppendorf Mastercycler using the following conditions: 98°C for 60 seconds then 25 cycles of 98°C for 5 seconds, 40°C for 30 sec, 72°C for 30 sec; elongation at 72°C for 10 min then hold at 5°C. The efficiency of PCR amplification of each sample was assessed by running 10 µl of the PCR mix on a 1.2% agarose gel.

High throughput sequencing and analysis of 16S amplicons

The amplified 16S samples from each animal were pooled using approximately equal amounts of each PCR product. There was 6 samples present for each of the conditions except for intestinal samples from outbreak one that was represented with 5 due to one sample sequencing failure. The pooled sample was sequenced using the Roche/454 FLX Genome Sequencer and the latest Titanium chemistry. The output sequence file was analysed using a number of publically available software packages and databases. Sff files were burst into fasta and qual files using PyroBayes (Quinlan et al., 2008) and data was filtered on qual file data to retain sequences with a minimum average sequence quality of 25 using MOTHUR (Schloss et al. 2009). The sequences were further filtered in MOTHUR to select for lengths between 300 and 600 bases, no ambiguous bases, and maximum homopolymer runs of 6. Further analysis of the dataset was done on this sequence file using QIIME (Caporaso et al. 2010), MOTHUR (Schloss et al. 2009), ARB (Ludwig et al. 2004) and the GreenGenes database (DeSantis et al. 2006).

Results

16S amplification

Samples amplified satisfactorily and gave clean products with very low amounts of non-target bands. An example of the gel analysis of the PCR amplification is shown in Figure 1. Template concentration and amplification conditions were varied to ensure that as many samples as possible produced amplification products for analysis.

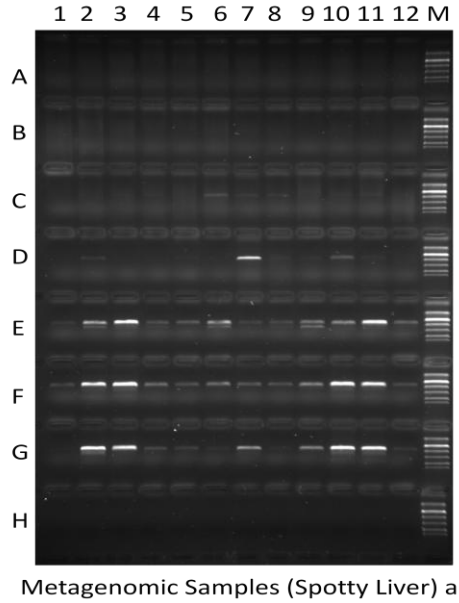


Figure 1: Example of gel analysis of amplification of 16S PCR products

Raw sequences were quality trimmed allowing a minimum average q-score of 25 for each individual sequence. Sequencing provided 382,743 sequences that were further quality trimmed to select sequence lengths between 300-600 bases, no ambiguities, and a maximum of 6 homopolymers. This trimming reduced the total number of sequences to 282,833. Mean sequence length was 466, lowest sequenced sample contained 1,773 sequences, mean number of sequences per sample was 6,148 and highest sequenced sample had 13,408 sequences. The rarefaction plot (Figure 2) shows adequate coverage of both healthy and diseased samples. The rarefaction plot was created by choosing a number smaller than the number of sequences in the lowest sequenced sample, in our case 1,620. The algorithm randomly samples 1,620 sequences from all samples and plots the estimated number of operational taxonomic units (OTUs – can be thought of as equivalent to species) on the y-axis as a function of the number of sequences sampled on the x-axis. Therefore, the plot shows the lowest coverage in the dataset, as the mean sample size in the present dataset was 6,148 sequences. The plot (Figure 2) demonstrates clear a difference in richness between caecal and intestinal content microbiota, with intestinal content (green in Figure 2) giving around 200 different OTUs, while caecal microbiota appears to be more diverse with more than 450 estimated bacterial species. Interestingly, diseased samples (labelled D in Figure 2) appear to have significantly more bacterial species, approximately 25 more species in intestinal content and 40 in caecum. This not only indicates that there is substantial microbiota perturbation in SLD, but also that the putative species or group of species that cause the diseases could be inducing wider changes in microbial environment that provides space and optimal growth conditions for some other, previously rare (bellow detection) or absent species to proliferate and colonise in more abundance.

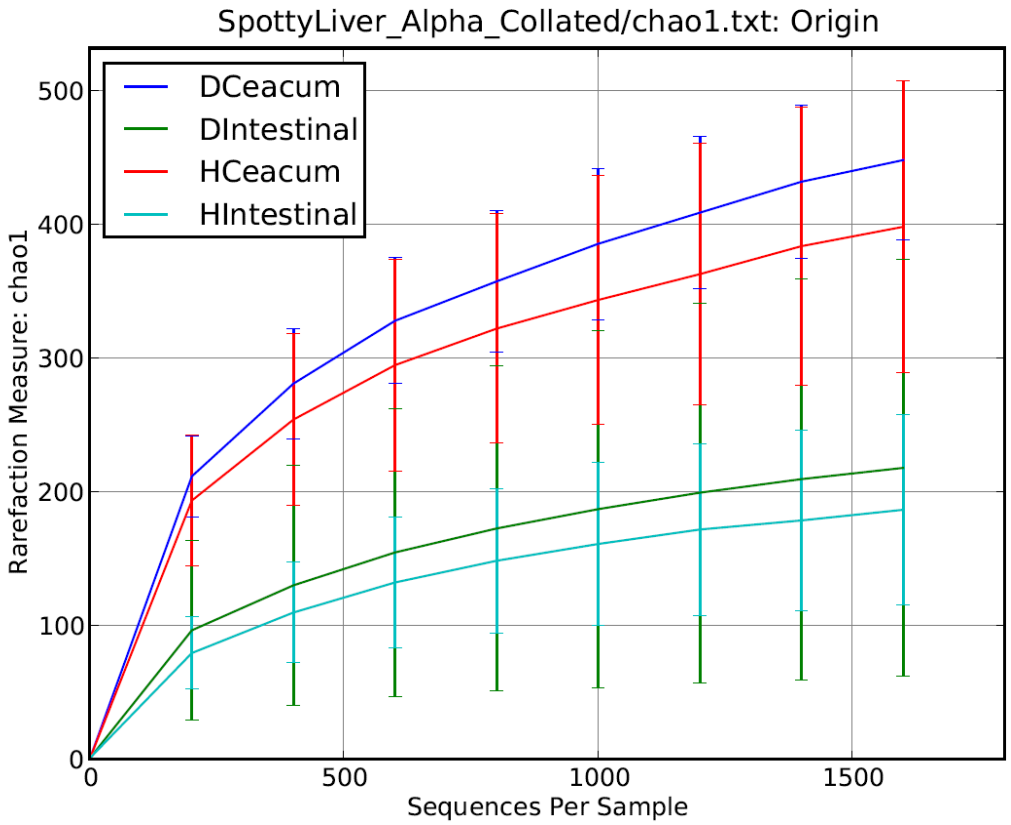


Figure 2: Rarefaction plot for samples from different origin (cecum and intestinal content) with healthy birds labelled as H and diseased with D.

UniFrac was used as a phylogenetic measure of between samples richness and diversity. Unweighted UniFrac, based on presence/absence of taxa, grouped diseased birds together (PCA plot, Figure 3) while healthy birds scatter through all PCA components. This indicates that it is presence of unshared species in diseased birds (not present in healthy) that makes them group together as the major difference component. The PCA plot based on weighted UniFrac, that takes into account presence/absence of an OTU and also weighs an OTU affect by taking into account the abundance (sequence number), does not group healthy nor diseased birds together. This may suggest that the cause of Spotty liver disease is not present in great numbers as it is influential with presence/absence but not when the numbers are taken into account. We may therefore look for an OTU, present in low numbers in diseased and absent (or bellow detection level) in healthy birds.

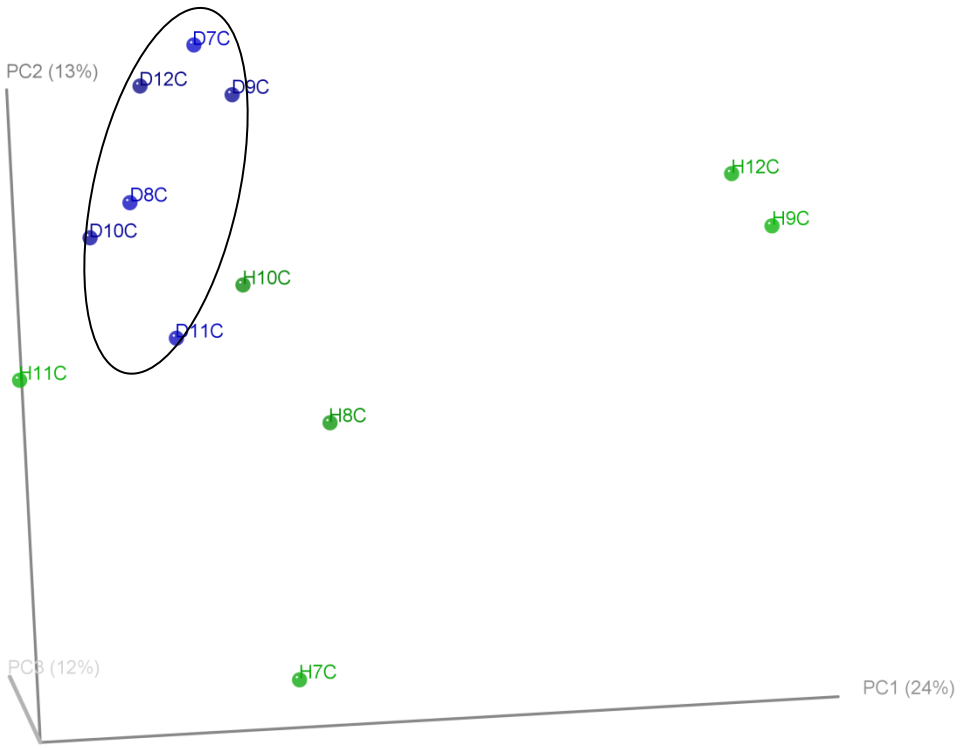


Figure 3: PCA view of unweighted UniFrac analysis of disease outbreak two samples. Diseased birds are grouping in PCA component 1 that accounts for most of the variation (24%) while healthy are randomly distributed.

Classification

Metagenomic softwares "bin" sequences, based on 97% sequence similarity (0.03 distance), into operational taxonomic units (OTUs), which for simplicity can be regarded as similar to "species". Each "bin" of similar sequences is then given a provisional number (OTU number) and a file is created listing all the sequences encompassed within that OTU. Assuming that 97% similarity is sufficient (and this is not always true) to ensure that the sequences assigned to the same OTU will belong to the same species, the softwares then complete most of the statistical and phylogenetic analysis using only OTU numbers, without actually calling it by the name of closest taxon. The current view among many metagenomic analysts is to first identify OTUs of interest, for example, the most abundant or differentially abundant between groups of replicates, and then, using manageable number of OTUs, attempt to classify them and actually call the OTU by microorganism species name. Qiime software provides an OTU table to estimate, at lower taxonomic level, the most reliable taxon assignment of the OTU and uses it to generate a number of interactive graphical tools. Its classification, however, does not go to the species level (Table 1). Somewhat closer classification can be achieved by placing aligned sequences into a phylogenetic tree, based on sequence similarity, using ARB (Figure 4). It is recommended to first identify potential candidate species based on such data inspection, and then continue further investigation by, for example sequencing the complete 16S sequence, fluorescently label the OTU of interest and enrich the culture for labelled species to continue with attempts to classify by using taxonomic and phylogenetic methods available.

Table 1: Bacterial OTUs (species) differentially represented in diseased and healthy birds.

OTUs present in diseased birds but not healthy birds are highlighted. O1 and O2 indicate data from outbreak 1 and 2 respectively while INT stands for intestinal content and CCM for caecum. For example, O1_CCM_151 represents OTU 151 from caecal origin collected in outbreak 1. Sample names of diseased birds start with D and healthy with H.

Sequence Name	**Tax asignment (ARB)**	**D1IC**	**D3IC**	**D4IC**	**D5IC**	**D6IC**	**H2IC**	**H3IC**	**H4IC**	**H5IC**	**H6IC**		
O1_INT_394	Lactobacillus helveticus	5	0	1	4	1	0	0	0	0	0		
O1_INT_168	Lactobacillus ingluvei	2	0	3	4	2	0	0	0	0	2		
Sequence Name	Tax asignment (ARB)	D1C	D2C	D3C	D4C	D5C	D6C	H1C	H2C	H3C	H4C	H5C	H6C
O1_CCM_1408	Ruminococcus productus	0	1	1	1	2	2	0	0	0	0	0	0
O1_CCM_811	Fecalibacterium prausnitzii	1	2	0	2	3	0	0	0	0	0	0	0
O1_CCM_24	Cytofaga sp	10	32	14	13	30	10	2	6	2	3	14	0
O1_CCM_151	Bacterioides tectus	4	0	0	2	2	1	0	0	0	0	0	0
O1_CCM_200	Parabacterioides merdae	8	10	4	21	26	3	0	0	0	10	1	0
O1_CCM_860	Clostridiaceae bacterium	4	27	12	5	18	13	2	0	0	1	0	9
O1_CCM_2183	Bacterioides sp	0	0	0	0	1	0	1	0	0	1	1	1
O1_CCM_80	Bacterioides sp	1	8	16	17	21	0	2	2	0	1	1	5
O1_CCM_838	Clostridiaceae bacterium	0	0	0	1	0	0	0	0	1	2	1	2
O1_CCM_232	Spyroplasma citri	6	0	6	0	15	2	0	1	0	0	0	0
Sequence Name	Tax asignment (ARB)	D10C	D11C	D12C	D7C	D8C	D9C	H10C	H11C	H12C	H7C	H8C	H9C
O2_CCM_593	clade1	38	22	20	50	26	38	12	11	5	9	4	0
O2_CCM_108	Bacterioides sp	1	1	0	2	2	2	0	0	0	0	0	0
O2_CCM_26	Bacterioides barnesiae	15	8	5	26	30	18	9	0	1	6	0	0
O2_CCM_877	Clostridiales bacterium	2	5	2	4	9	3	1	2	0	0	0	0
O2_CCM_498	Spyroplasma citri	2	1	1	1	0	1	0	0	0	0	0	0
O2_CCM_928	Clostridiales bacterium	1	0	1	3	4	2	0	1	0	0	0	0
O2_CCM_397	Anaerotruncus colihominis (Ruminococcus)	1	1	0	0	3	2	0	0	0	0	0	0
O2_CCM_764	Eubacterium contortum	0	0	2	0	0	0	4	0	4	2	3	3

O2_CCM_1042	Clostridiaceae bacterium	0	0	0	0	0	0	0	1	1	2	1	0
O2_CCM_1070	Clostridiaceae bacterium	2	1	0	2	4	2	0	1	1	0	0	0
O2_CCM_980	Clostridiales bacterium	2	1	1	0	6	6	0	0	0	0	0	0
O2_CCM_110	Bacterioides sp	2	1	1	4	6	8	9	2	6	18	14	9
O2_CCM_357	Pevotellaceae	0	0	0	0	0	0	0	5	0	4	2	1
O2_CCM_112	Bacterioides barnesiae	1	3	0	11	0	4	7	8	9	13	7	0
O2_CCM_226	Clostridiaceae bacterium	2	5	2	2	3	2	3	3	7	10	6	2
O2_CCM_369	Bacterioides sp (putedinis)	3	2	0	3	4	0	0	0	0	1	0	0
O2_CCM_622	Spyroplasma citri	1	11	9	14	0	6	1	2	1	2	0	0
O2_CCM_570	clade1	40	14	6	13	76	47	7	4	0	25	0	0
O2_CCM_75	Barnesiella viscericola	5	1	10	1	10	4	7	9	8	13	9	10
O2_CCM_1683	Bacterioides sp	1	0	1	4	6	1	0	1	0	0	0	0
O2_CCM_1621	Bacterioides sp	1	1	1	5	0	0	0	0	0	0	0	0
O2_CCM_532		29	1	68	74	0	50	7	6	0	0	0	3
O2_CCM_252	Bacterioides barnesiae	5	1	1	8	1	1	0	1	0	0	0	0
O2_CCM_188	Bacterioides coprocola	7	13	4	2	9	6	1	1	2	0	2	2
O2_CCM_950	Clostridiales bacterium	2	1	0	1	0	1	0	0	0	0	0	0
O2_CCM_1020	Ruminococcus productus	0	1	0	0	1	0	2	4	3	5	0	0
O2_CCM_409	Parabacterioides merdae	0	1	0	0	0	0	2	4	0	4	1	0
O2_CCM_219	Bacterioides sp	4	0	0	7	3	1	0	0	0	0	0	0
O2_CCM_77	Bacterioidetes vulgatus	36	30	49	195	284	44	9	37	11	14	30	4
O2_CCM_237	Fecalibacterium prausnitzii	1	3	0	1	1	1	0	0	0	0	0	0
O2_CCM_547	clade1	47	66	15	138	90	169	21	47	0	31	25	0
O2_CCM_415	Spyroplasma citri	1	0	1	2	2	2	0	0	0	0	1	0
O2_CCM_292	Spyroplasma citri	29	6	12	33	0	0	0	2	0	0	0	0
O2_CCM_623		5	0	10	6	0	5	1	0	0	0	0	0
O2_CCM_163	Alistipes sp	13	1	15	14	24	5	2	1	0	12	3	0
O2_CCM_538		5	0	17	22	0	19	0	2	0	0	0	2
O2_CCM_477	Bacterioidales genomosp	0	2	0	0	3	2	0	0	0	0	0	0

Sequence Name	Tax asignment (ARB)	D10I	D11I	D12I	D7I	D8I	D9I	H10I	H11I	H12I	H7I	H8I	H9I
O2_INT_457	Campylobacter sp	0	0	0	0	0	0	17	13	30	45	24	7
O2_INT_561	Campylobacter sp	0	0	0	0	0	0	3	2	7	11	6	3
O2_INT_960	Bacterioides sp	0	0	0	0	0	0	2	3	2	5	5	4
O2_INT_346	Campylobacter sp	3	0	0	1	0	0	136	169	374	650	295	82
O2_INT_517	Campylobacter sp	1	0	0	0	0	0	4	5	18	32	12	2
O2_INT_414	Campylobacter sp	0	0	0	0	0	0	4	6	17	44	14	2
O2_INT_386	Lactobacillus aviarius	1	0	0	0	0	0	10	3	11	20	0	1
O2_INT_1663	Bacterioides sp	1	0	2	3	0	1	0	0	0	0	0	0
O2_INT_79	Lactobacillus aviarius	2	0	0	37	0	3	271	41	402	493	6	33
O2_INT_430	Lactobacillus sp.	2	3	1	0	0	1	0	0	0	0	0	0
O2_INT_782	Clostridiales bacterium	0	0	0	0	0	0	3	0	6	1	1	1
O2_INT_408	Lactobacillus aviarius	4	0	0	7	2	3	0	0	1	0	0	0
O2_INT_767	clade1	10	9	2	0	4	2	0	2	0	0	0	0
O2_INT_117	Bacterioides plebius	23	18	4	0	23	0	0	2	1	0	0	0
O2_INT_344	Mogibacterium pumilum	0	0	0	0	0	0	5	0	2	2	2	0
O2_INT_66	Lactobacillus helveticus	47	0	36	97	5	20	0	2	11	0	0	0
O2_INT_257	Veillonella sp	12	0	15	0	0	15	0	0	0	0	0	0
O2_INT_107		0	0	0	0	0	0	6	0	1	6	2	0
O2_INT_35	Bacterioides plebius	45	26	9	0	50	0	0	3	0	1	0	2
O2_INT_806	Bacterioides sp	0	0	0	0	0	0	0	0	3	5	3	0
O2_INT_388	Streptococcus porcinus	0	0	0	0	0	0	3	0	4	1	1	0

Based on the data presented, two potential pathogen candidates emerge, both so far unknown and unclassified: a clade of OTUs closest to *Spyroplasma citri* (Figure 4) and the second clade of unclassified *Clostridiales* (Figure 5), both differential in caecum samples. There may be a number of other candidates that will also be worth considering. The reduction in numbers of *Lactobacillus* and unclassified *Bacterioides* and *Campylobacter* species is also evident in intestinal samples in outbreak 2 that show high level of changes in intestinal microbiota (Figure 6). Figure 7 demonstrates that the microflora seen in the gut contents is clearly different between diseased and healthy birds.

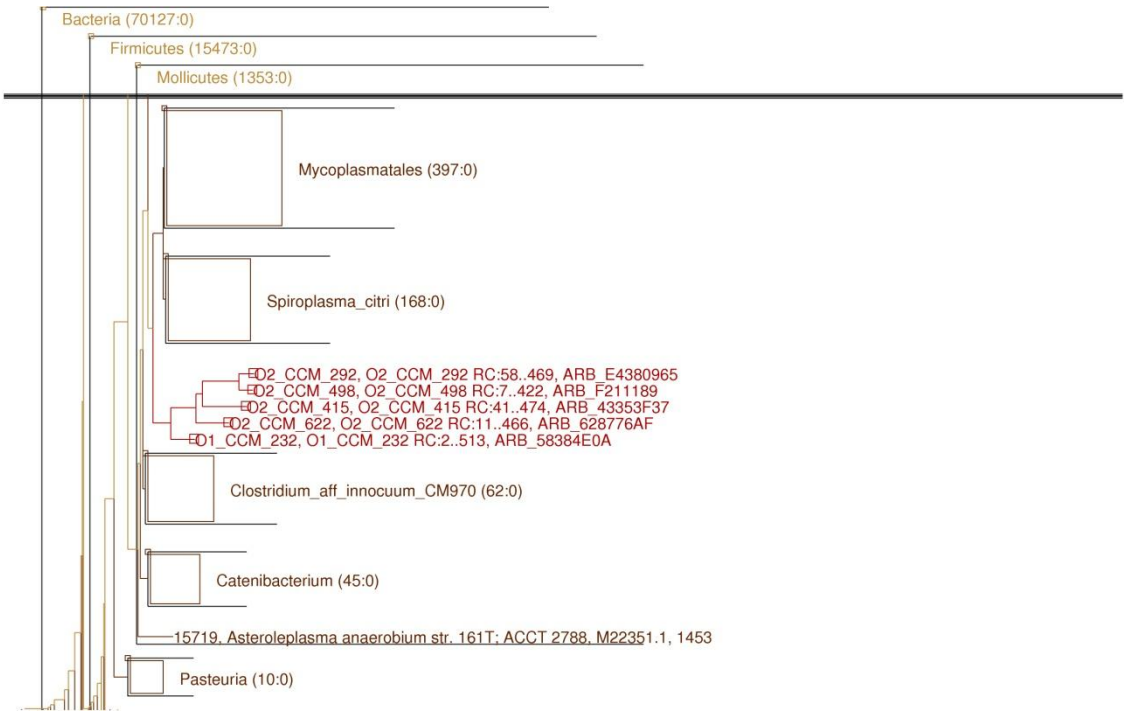


Figure 4: A clade of OTUs significantly more abundant in diseased samples, placed into Greene Gene – based tree using ARB software, mapping closest to *Spiroplasma citri*.

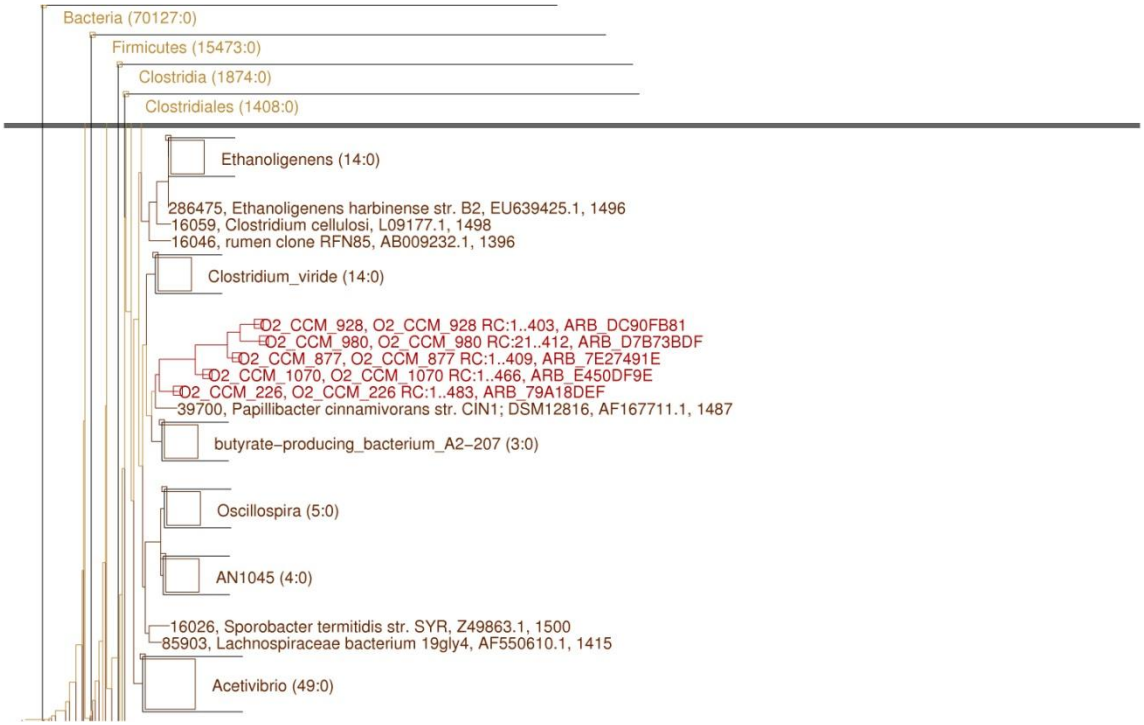


Figure 5: A clade of OTUs differentially abundant in cecal samples, with closest match to unclassified *Clostridiales*. There are number of other differential OTUs classified to the same level (unclassified *Clostridiales*), that do not belong to this particular clade.

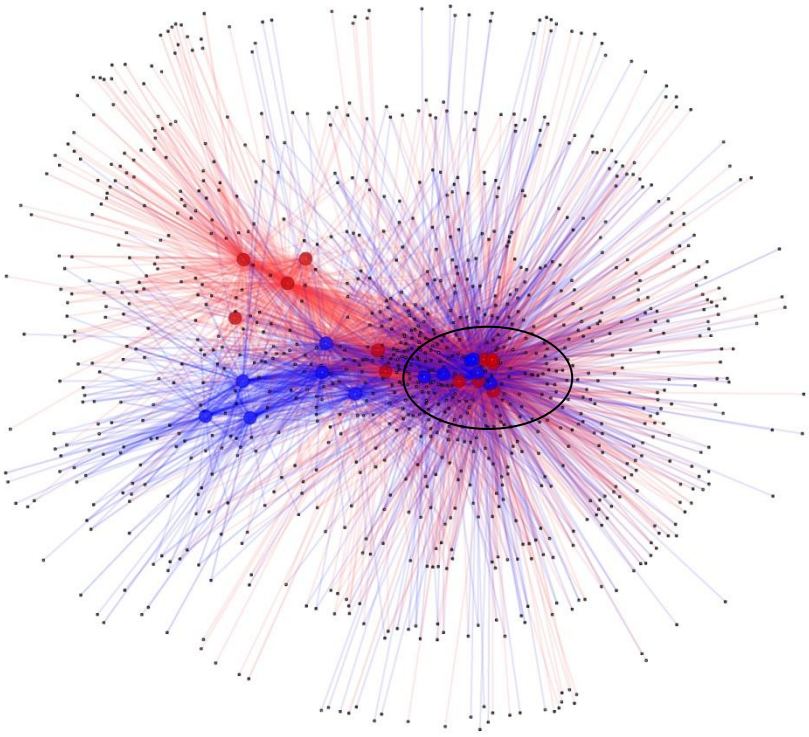


Figure 6: OTU network map (Qiime) presenting OTU interactions between samples originating from healthy (•) and diseased samples (•) from both locations (caecal and intestinal content) and their related OTUs from outbreak two.

Edges (lines) coming from healthy samples to OTUs present in that sample are colored blue and edges leading from diseased samples are red. The distance of an OTU (presented as black dots) to a sample is proportional to a number of OTU hits per sample, OTUs closer to the sample are more abundant and the ones furthest away are rare. OTUs can be linked to all samples in which it was detected. The black oval shape separates caecal samples to demonstrate the difference between the two sections of the gut. The network also demonstrates that differences are consistently present in both sections of the gut.

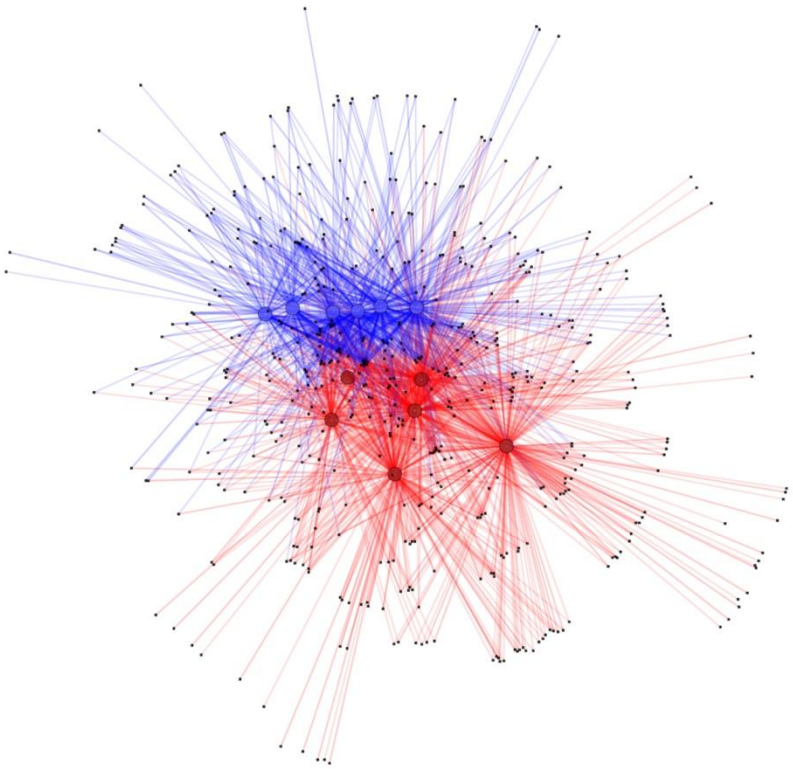


Figure 7: OTU network map (a subset of samples shown in Figure 6) presenting OTU interactions between samples originating from healthy (•) and diseased samples (•) from the intestinal content samples. The map has been rotated in three dimensions to emphasize the differences between the two groups of samples.

Conclusions

Based on the data obtained from the two outbreaks of the disease, it appears that the infectious agent is not among previously characterised bacteria. This is consistent with previously published data. A number of differentially abundant OTUs are identified in this study. Future research needs to validate these results in a wider selection of samples and it needs to be determined which of the OTUs (or group of OTUs) actually has a causative role in development of SLD and which of the OTUs are changing in abundance just because of physiological or micro-environmental changes in the gut resulting from the disease process. Gut microorganisms closely associated and dependant on the hosts history and environment, readily interact with one another and the host, contributing to a massive network of microbial interactions where knocking out or boosting any one of the nodes can significantly affect many other members of the network. The OTUs that are reduced in numbers due to the disease are also of interest: are they reduced because of toxic effect of the pathogen or is pathogen thriving because they are reduced? Can we manipulate microflora to compensate for the *Lactobacillus* and *Campylobacter* species that are reduced in disease to increase the survival rate?

Next steps

1. Analysis of samples from several more SLD outbreaks to determine if the changes seen in the current samples are seen in other outbreaks. This may help to narrow down the OTUs that we will target for further characterization.

2. Develop 16S gene probes for the *in vivo* identification of potential SLD candidate pathogens.
3. Investigate in vitro culture conditions to allow isolation of potential SLD candidate pathogens.
4. Test candidate SLD pathogens for ability to reproduce the disease.

References

Caporaso, J. G., J. Kuczynski, et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

DeSantis, T. Z., P. Hugenholtz, et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **72**: 5069-5072.

Ludwig, W., O. Strunk, et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research* **32**: 1363-1371.

Quinlan AR, Stewart DA, Stromberg MP, Marth GT (2008) Pyrobayes: an improved base caller for SNP discovery in pyrosequences. *Nature Methods* **5:** 179-181.

Schloss, P. D., S. L. Westcott, et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**: 7537-7541.

Yu, Z. and M. Morrison (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**: 808-812.

Spotty Liver Disease 16S Metagenomics – Second Report

Dana Stanley & Rob Moore

17/08/2011

Hypothesis

We hypothesise that spotty liver disease is caused by a bacterium, resident in the gut, which secretes a toxin which translocates from the gut to the liver where it causes lesions. By comparing the population structure of the microflora in the gut of diseased and healthy **birds it may be possible to identify the bacterial species that is responsible for the disease.**

Materials and Methods

DNA preparation from gut samples

Chicken spotty liver disease samples were supplied by Dr Peter Scott. Material from the intestinal content samples and caecal samples were resuspended in 250 µl of phosphate buffered saline. Total DNA from these samples was then isolated using the method of Yu and Morrison (2004). Briefly, a sample was transferred to a tube with lysis buffer (500mm NaCl, 50mM Tris-HCl pH8.0, 50 mM EDTA and 4% sodium dodecyl sulfate) and sterile zirconium beads and then homogenized using a Qiagen TissueLyser at maximum speed 3 times for 10 seconds. Following centrifugation the supernatant was collected, ammonium acetate was added and nucleic acid was precipitated with isopropanol, followed by ethanol wash. After centrifugation the pellet was resuspended in Tris-EDTA buffer and digested with DNase-free RNase and proteinase K to remove RNA and protein. The DNA was finally purified on a QIAamp column (Qiagen) according to the manufacturer's instructions. DNA quantity and quality was measured on a NanoDrop ND-1000 spectrophotometer.

PCR amplification of 16S ribosomal RNA gene sequences

DNA derived from the bacteria of the birds was processed to amplify the 5' end of the eubacterial 16S ribosomal RNA genes. DNA was amplified using Bio-Rad iProof DNA polymerase. Each PCR reaction contained 25 µl of iProof 2X master mix (containing buffer, nucleotides and iProof enzyme), 2 µl of each primer (final concentration 0.5 µM), 1.5 µl DMSO, 0.5 µl 50 mM $MgCl_2$ and template DNA made up to 19 µl in water. The primers used were designed to amplify the V2-V3 region of the 16S rRNA gene (forward primer, 5' AGAGTTTGATCCTGG 3'; reverse primer, 5' TTACCGCGGCTGCT 3'). Each primer also included sequences to facilitate the sequencing of products in the Roche/454 system and the reverse primers consisted of a related set of primers which differed in "barcode" sequences; specific sequences introduced into the primers to allow tagging of individual samples in a multiplex sequencing system. PCRs were performed in an Eppendorf Mastercycler using the following conditions: 98°C for 60 seconds then 25 cycles of 98°C for 5 seconds, 40°C for 30 sec, 72°C for 30 sec; elongation at 72°C for 10 min then hold at 5°C. The efficiency of PCR amplification of each sample was assessed by running 10 µl of the PCR mix on a 1.2% agarose gel.

High throughput sequencing and analysis of 16S amplicons

The amplified 16S samples from each animal were pooled using approximately equal

amounts of each PCR product. There were 6 samples present for each of the conditions except for intestinal samples from outbreak one that was represented with 5 due to one sample sequencing failure. The pooled sample was sequenced using the Roche/454 FLX Genome Sequencer and the latest Titanium chemistry. The output sequence file was analysed using a number of publically available software packages and databases. Sff files were burst into fasta and qual files using PyroBayes and data was filtered on qual file data to retain sequences with a minimum average sequence quality of 25 using MOTHUR (Schloss, Westcott et al. 2009). The sequences were further filtered in MOTHUR to select for lengths between 300 and 600 bases, no ambiguous bases, and maximum homopolymer runs of 6. Further analysis of the dataset was done on this sequence file using QIIME (Caporaso, Kuczynski et al. 2010), MOTHUR (Schloss, Westcott et al. 2009), ARB (Ludwig, Strunk et al. 2004) and the GreenGene database (DeSantis, Hugenholtz et al. 2006).

Results

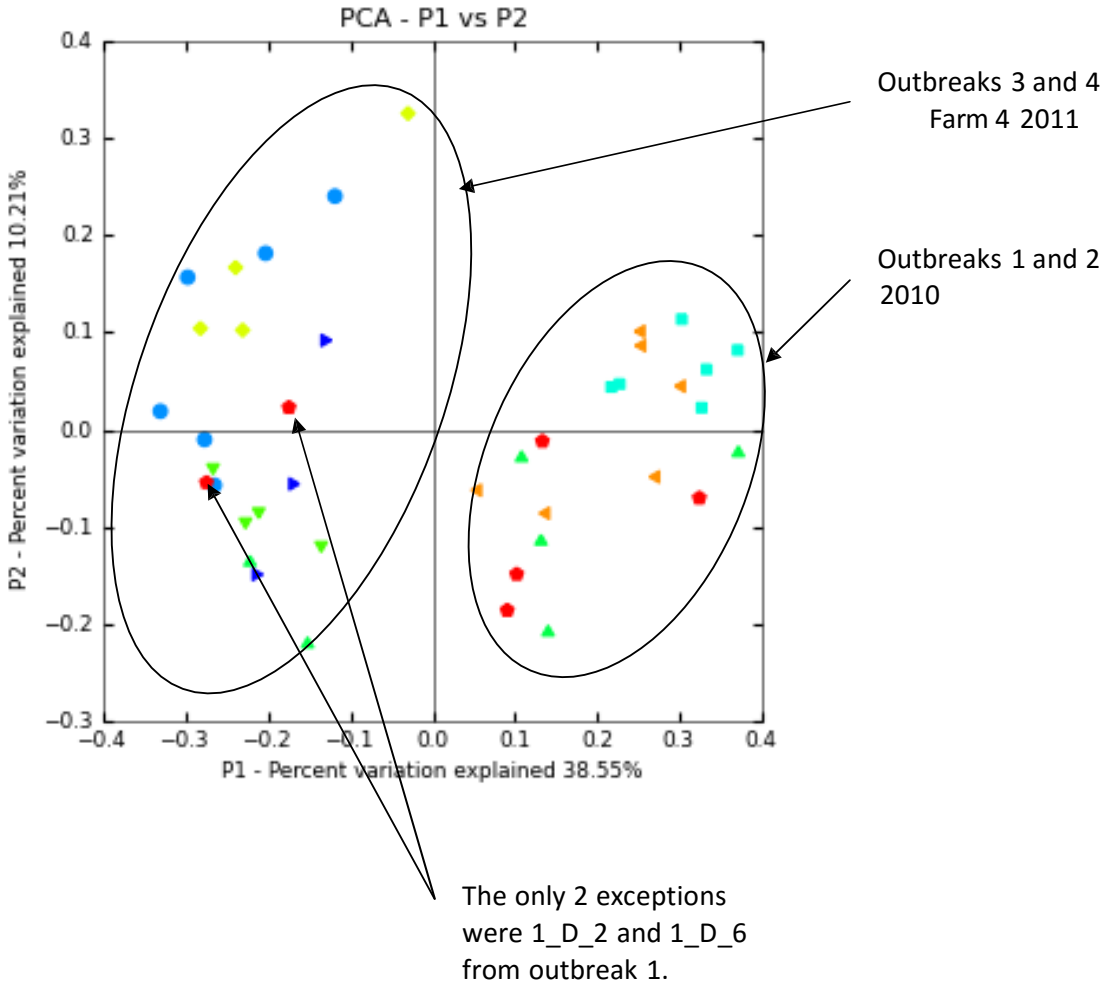
Data pre-processing

The latest dataset contained 2 outbreaks (Farms 4 and 10). Previously we analysed 2 other outbreaks provided by Dr P Scott. All 4 outbreaks were analysed individually and, separately, a merged experiment was created where all samples were analysed together. The previous 2 outbreaks were re-analysed to ensure comparability of data and use of the latest software updates and databases versions. Outbreaks 1 and 2 (from previous report) and the present outbreak 3 (Farm 4 shed 7a) had 12 birds, 6 healthy and 6 diseased, each. Outbreak 4 (Farm 4 shed 6) had 8 birds, 4 sick and 4 healthy. The quality of the sequencing data was ensured by removal of chimeric sequences and quality trimming as described above. The resulting fasta file was reduced from 565,160 to 159,071 of quality sequences. The lowest sequenced sample was 3_D_3 (outbreak 3, diseased bird 3) with 1079 sequences while highest sequenced sample (1_D_5) contained 8707 sequences. The rarefaction plot for the smallest sequence size sample showed that the sequencing effort was providing sufficient coverage. The coverage for the lowest sample at the distance of 0.03 (97% similarity) was 0.81. Sampling an additional 100 sequences, according to the efron calculator, would gain an additional 19 OTUs at 0.03 distance.

There are differences in the microbiota of healthy and diseased birds

When the whole set was analysed together weighted unifrac PCA plot (Figure 1) separates the birds into 2 groups with distinct caecal microbiota based on the origin and year of samples. The samples from farm 4 are grouping together regardless of health with only 2 samples from diseased birds from outbreak 1 showing some similarity to this group. This indicates that the microbiota in the gut of the birds in the first two outbreaks is significantly different to the microbiota in the outbreaks 3 and 4. There are many potential factors which may produce these distinctly different microfloras. Factors could include the influence of different physical environments, different feeds and additives, different sources of birds, different bedding materials and different management practices.

Figure 1: Unifrac PCA plot of samples from all 4 outbreaks.



The differences between healthy and diseased birds within each outbreak are significant. Table 1 shows weighted unifrac p-values for differences between healthy and diseased birds in all 4 outbreaks. The unifrac significance was calculated using a number of diversity calculators; all the different calculators agree that healthy and diseased samples are significantly different using weighted unifrac, while unweighted unifrac suggests no significant difference. Since unweighted unifrac takes into account only presence/absence of OTUs while weighted takes into account the abundance (number of sequences in each OTU) as well. This indicates that the differences between healthy and diseased birds in each outbreak are because of differential abundance in types of bacteria present rather than qualitative differences in the populations. These data also suggest that differential OTUs do not greatly influence the composition of accompanying microflora, which sometimes can be the case with aggressive species that produce products toxic for other microbiota or ones that change pH significantly.

Table 1: Unifrac significance values for groups healthy and diseased at distance of 0.03.

Mothur v.1.20.1	Calculator	Weighted Unifrac significance	UnWeighted Unifrac significance
OUTBREAK 1	sorabund	<0.001	0.402
	braycurtis	<0.001	0.741
	jabund	<0.001	0.396
	sorclass	<0.001	0.622
	sorabund	<0.001	0.417
	thetayc	<0.001	0.832
	morisitahorn	<0.001	0.842
OUTBREAK 2	sorabund	<0.001	1
	braycurtis	<0.001	0.103
	jabund	<0.001	1
	sorclass	<0.001	0.085
	sorabund	<0.001	1
	thetayc	<0.001	0.029
	morisitahorn	<0.001	0.042
OUTBREAK 3	sorabund	<0.001	0.036
	braycurtis	<0.001	0.004
	jabund	<0.001	0.069
	sorclass	<0.001	0.003
	sorabund	<0.001	0.05
	thetayc	<0.001	0.039
	morisitahorn	<0.001	0.047
OUTBREAK 4	sorabund	<0.001	0.337
	braycurtis	<0.001	0.102
	jabund	<0.001	0.32
	sorclass	<0.001	0.202
	sorabund	<0.001	0.318
	thetayc	<0.001	1
	morisitahorn	<0.001	1

- Note that the values are for each outbreak separately between high and low birds; both weighted and unweighted unifrac are significantly different between the 2 groups of outbreaks as in Figure 1

Figures 2 and 3 further support this observation. Results of classification show a number of OTUs abundant and significantly different between healthy and diseased birds in each of the outbreaks (Table 2). There tends to be a level of consistency between outbreaks 1 and 2 and between 3 and 4 but not across all 4 outbreaks.

Figure 2: 3D pca plots based on jclas diversity calculator. Healthy birds are colored blue and diseased red. From top to bottom, left to right: Outbreak 1, 2, 3 and 4.

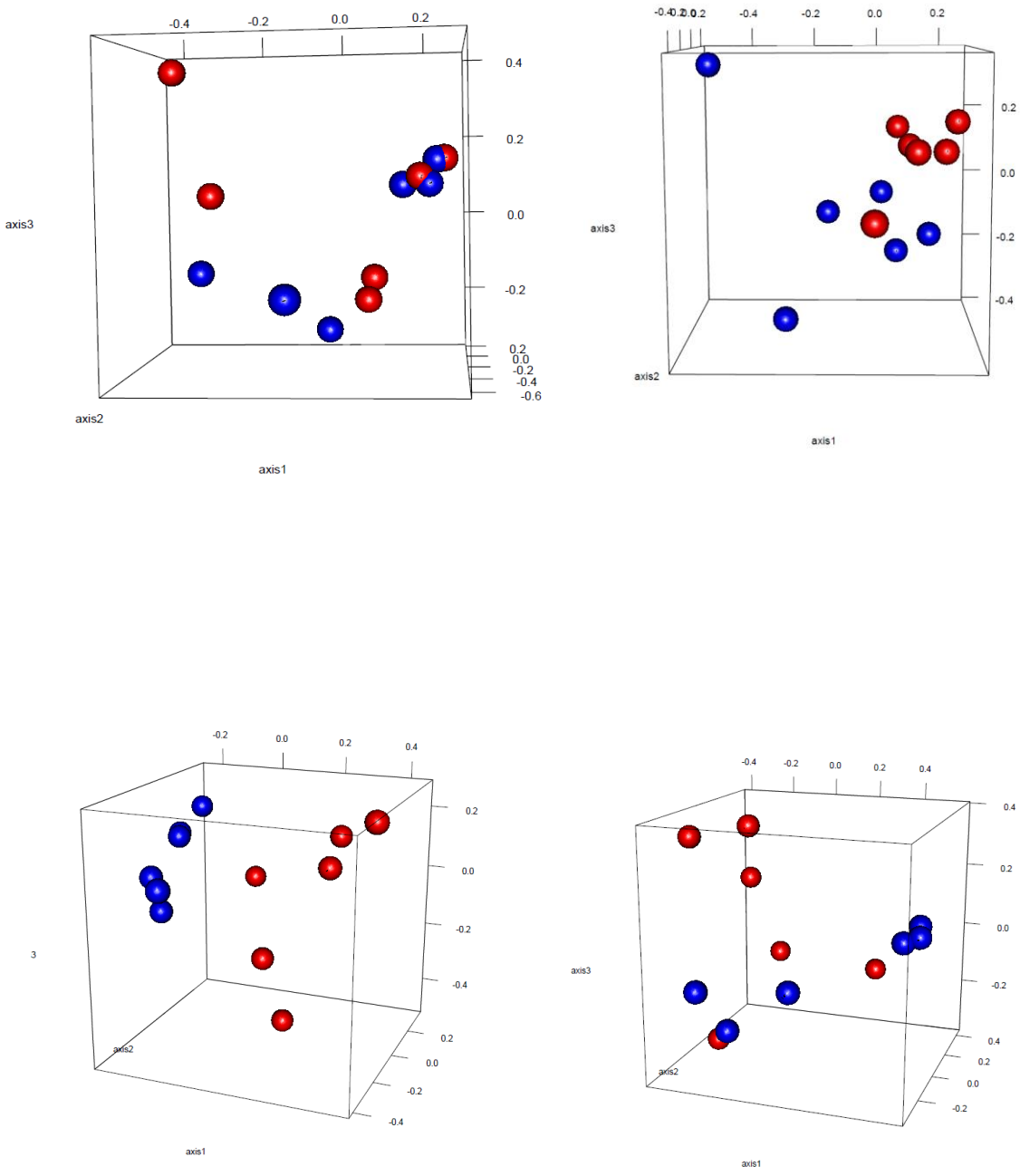
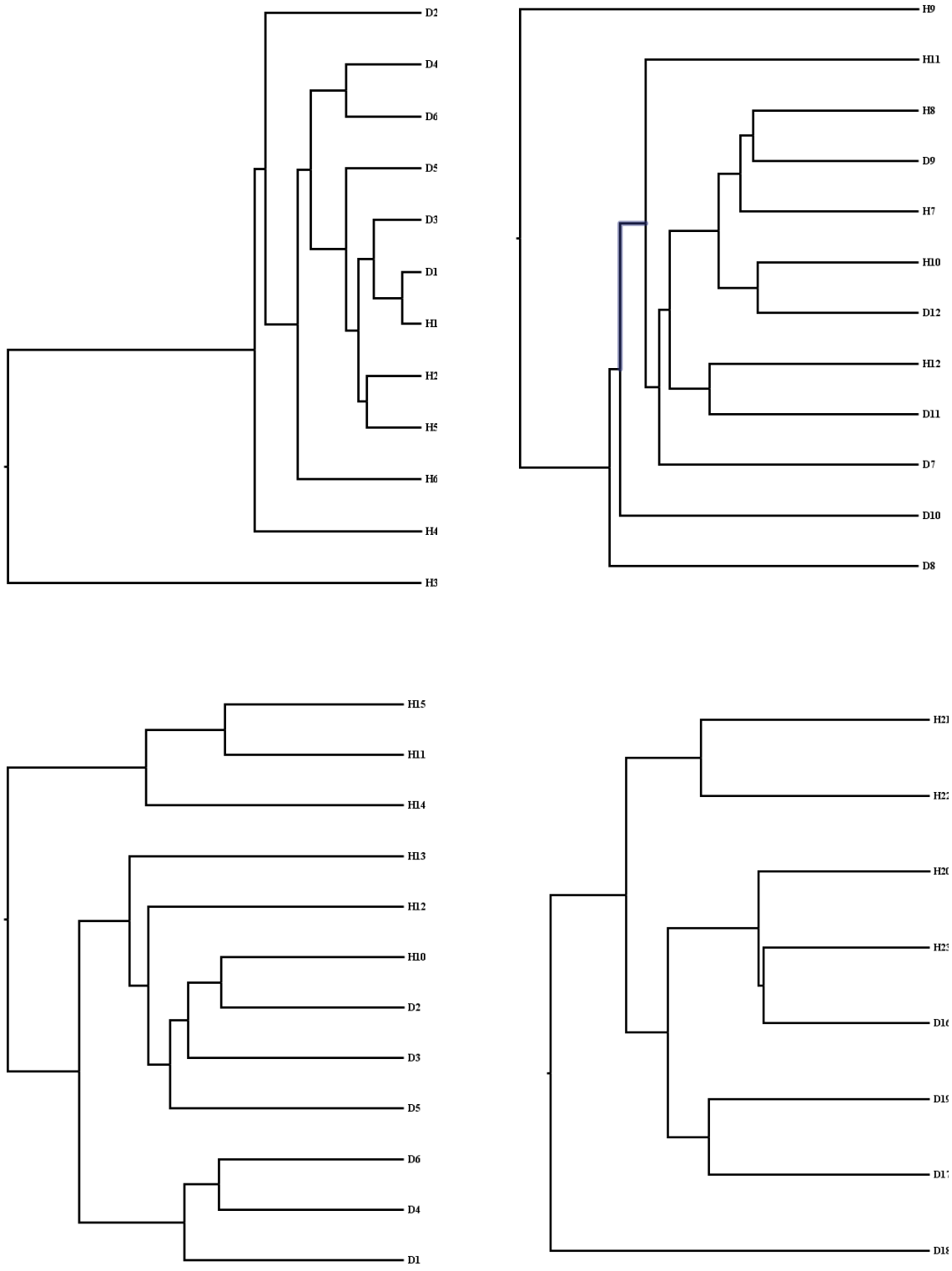


Figure 3: Comparison of the sorabund trees for all 4 outbreaks. From top to bottom, left to right: Outbreak 1, 2, 3 and 4.



Selecting candidates

The ideal candidate for the disease agent would be an OTU that is present in diseased birds in all 4 outbreaks but not present in any of the healthy birds. We clearly have not identified such a candidate in the current analysis. The best candidates that we can identify at the moment are OTUs that are more abundant in diseased birds than in healthy birds. There are no OTUs that are more abundant in diseased birds in all four outbreaks but there are several that are more abundant in the diseased birds in three of the four outbreaks. Based on the data presented in Table 2 the top candidates could be OTUs that are an unclassified Bacteroidetes, an unclassified Sphingobacteriales, an unclassified Alphaproteobacteria, an unclassified Proteobacteria and unclassified Bacteria.

The Sphingobacteriales are an order of bacteria that are not known as a pathogen in animals although certain representatives of the order are pathogenic in plants, where they can cause necrosis of leaves. It would therefore not seem to be a very likely candidate as the causative agent of Spotty Liver disease. All the other taxanomic groups identified as potential candidates do contain members that have some pathogenic potential in animals.

Candidates identified in first report

For the analysis reported here we have used the latest versions of the software and databases. The previous analysis of the first two outbreaks used much earlier versions of the software and databases. Despite the large changes that have occurred in the analysis software and databases we found that the results from the first round of analysis largely coincided with the more advanced analysis carried out in this report to compare all four outbreaks. The candidates tentatively identified in the first analysis were still seen as differentially abundant in outbreak 2, from which most of the candidates were chosen, but were not differentially abundant across all outbreaks. Mostly they were only differentially abundant in two of the four outbreaks. Therefore the new candidate OTUs outlined in the previous paragraph are likely to be better candidates.

Table 2: **OTUs differentially abundant across the 4 outbreaks.**

Red font indicates t-test significance, fold higher in diseased is highlighted red and higher in healthy in green. Folds under 1.3 are not highlighted. Total sequences column refers to total in all 4 outbreaks. Very high folds are mostly rare in that particular outbreak (say 0.5 average seqs in one and 5 in the other group gives 10 fold).

Tax level	taxon	Total Sequences	Fold 1	p-value_o1	Fold 2	p-values_o2	Fold 3	p-values_o3	Fold 4	p-values_o4
Strain	unclassified Bacteroides	62291	-1.1	1.29E-13	-1.1	7.44E-13	2.0	6.82E-73	-1.1	0.024528
FAMILY	Prevotellaceae	1772	1.5	0.000618	-3.8	1.02E-42	-3.1	3.33E-17	1.8	0.005541
Strain	unclassified Prevotellaceae	1350	1.4	0.073397	-4.6	4.16E-49	-2.8	1.62E-10	1.5	0.092725
Strain	unclassified Sphingobacteriales	2094	2.8	1.24E-38	2.0	4.16E-07	21.1	8.27E-49	-inf[@]	0.003904
Strain	unclassified Bacteroidetes	2503	2.4	7.52E-39	1.7	0.000122	4.2	3.48E-18	-1.3	0.451237
Strain	unclassified Candidate_division_TM7	3797	-3.2	1.81E-74	1.1	0.588555	-5.7	5.42E-85	-1.9	2.39E-12
CLASS	Bacilli	15862	2.4	7.6E-206	2.1	4.71E-23	-1.2	6.44E-06	-1.5	9.06E-14
ORDER	Lactobacillales	15714	2.4	3.3E-206	2.1	6.64E-23	-1.2	1.24E-05	-1.5	5.33E-14
Strain	unclassified Lactobacillus	15142	2.5	1.8E-208	2.1	1.62E-22	-1.2	1.55E-05	-1.5	4.73E-14
CLASS	Clostridia	26434	1.4	2.74E-45	1.2	0.000129	-1.2	2.51E-10	1.3	2.56E-08
ORDER	Clostridiales	26429	1.4	4.58E-45	1.2	0.000119	-1.2	2.75E-10	1.3	2.56E-08
Strain	unclassified Lachnospiraceae	5756	2.5	1.79E-46	1.4	0.005276	1.0	0.680896	-1.5	0.0002
FAMILY	Ruminococcaceae	14227	1.3	6.52E-17	1.2	0.000484	-1.3	1.7E-09	1.4	1.05E-09
Strain	unclassified Subdoligranulum	1032	2.2	5.79E-06	-1.4	0.211539	-1.9	3.65E-10	2.8	1.85E-05
Strain	unclassified Ruminococcaceae	8429	-1.0	0.291335	1.3	0.000294	-1.2	0.000829	1.1	0.097719
FAMILY	Veillonellaceae	1537	-1.8	6.11E-06	-1.1	0.325932	6.1	3.8E-27	3.0	0.003369
Strain	unclassified Phascolarctobacterium	876	-2.2	7.25E-09	-1.0	0.869867	23.0	1.82E-16	15.0	0.006343
Strain	unclassified RF9	2225	-1.5	6.38E-09	1.9	0.000104	1.3	0.109905	2.2	1.88E-05
Strain	unclassified Firmicutes	5091	-1.3	2E-09	1.8	1.04E-09	-1.4	0.004063	1.3	0.030014
CLASS	Alphaproteobacteria	889	-2.1	5.62E-05	1.8	2.86E-05	4.2	0.01181	3.0	0.289034
Strain	unclassified Alphaproteobacteria	886	-2.1	6.72E-05	1.8	2.86E-05	4.2	0.01181	3.0	0.289034
CLASS	Epsilonproteobacteria	2693	2.2	6.63E-10	-1.8	2.45E-16	1.2	0.499743	1.6	0.006695
ORDER	Campylobacterales	2693	2.2	6.63E-10	-1.8	2.45E-16	1.2	0.499743	1.6	0.006695
FAMILY	Campylobacteraceae	1011	2.3	6.58E-05	-3.8	1.04E-28	1.3	0.753895	-1.3	0.629016
Strain	unclassified Campylobacter	1004	2.2	0.000183	-3.9	4.29E-29	1.3	0.753895	-1.3	0.629016
Strain	unclassified Proteobacteria	3662	-24.6	0	1.9	8.31E-09	2.5	0.000175	2.3	0.093086
Strain	unclassified Bacteria	12821	-1.9	2.81E-81	2.0	1.94E-43	1.3	0.000534	1.3	0.000738

@ -inf had 0 seqs average in 4 diseased and 2 seqs average in healthy birds.

Discusion

The lack of a clear candidate that is only present in diseased birds suggest that if the disease is caused by a bacterium resident in the gut, as we have hypothesised, then it is likely to be caused by an opportunistic pathogen that can sit in the gut as part of the normal microbiota and only cause disease when either numbers increase or gene regulation is altered to produce a toxin. Such characteristics would be typical of a clostridial infection however we have not found a strong candidate with this taxonomy.

A number of candidates have been identified. These could be pursued by attempting to culture them and attempting to reproduce disease in a model system. To culture them we would need to develop specific 16S rRNA gene PCR assays to give some confidence that any cultured organism represents the actual targe species. Such PCR assays could also be used to specifically quantify the species in other independent samples from other disease outbreaks.

Other points for consideration

The causative bacterium could be hidden from our analysis if it is very closely related to other abundant bacterial species. Hence, even when the pathogen was present at elevated numbers, it may be hidden in the bigger population of closely related bacteria.

Is the pathogen present at the time of sampling? The liver damage may be caused by a transient infection which is no longer present by the time the birds are necropsied.

Were all the birds sampled as diseased specimens truly diseased? Any inclusion of doubtful birds might upset the group statistical analysis and thus disrupt our ability to identify the pathogen.

If the pathogen is a very minor component of the gut microflora then the depth of analysis that we have done may be insufficient to reliably detect it. It may be helpful to do deeper sequencing on selected or all samples.

References

Caporaso, J. G., J. Kuczynski, et al. (2010). "QIIM allows analysis of high-throughput community sequencing data." Nat Methods **7**(5): 335-336.

DeSantis, T. Z., P. Hugenholtz, et al. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." Appl Environ Microbiol **72**(7): 5069-5072.

Ludwig, W., O. Strunk, et al. (2004). "ARB: a software environment for sequence data." Nucleic Acids Res **32**(4): 1363-1371.

Schloss, P. D., S. L. Westcott, et al. (2009). "Introducing mothur: open-source, platform independent, community-supported software for describing and comparing microbial communities." Appl Environ Microbiol **75**(23): 7537-7541.

Yu, Z. and M. Morrison (2004). "Improved extraction of PCR-quality community DNA from digesta and fecal samples." Biotechniques **36**(5): 808-812.

Spotty Liver Disease 16S Metagenomics

Third Report

Dana Stanley[1], Peter Scott[2], Robert Moore[1]

[1]CSIRO Animal, Food and Health Sciences, [2]Scolexia Pty Ltd

SUMMARY

In an effort to find the causative agent of spotty liver disease the gut microbiota of diseased and healthy birds has been compared. Samples from eight independent disease outbreaks have been analysed. A candidate bacterium has been identified. An experimental infection trial was also analysed but did not provide any lead as to the pathogen involved in precipitating disease.

INTRODUCTION

Spotty Liver Disease (SLD), characterized by cream coloured lesions on the liver, has been recognised in Australian commercial poultry, particularly those maintained under barn and free range systems, for many years. The disease is of significant economic importance because of increased mortality and decrease in egg production. The disease has frustrated both producers and veterinary professionals as its primary cause has never been elucidated. Diagnosis is based on typical clinical findings, including negative microbiological results, with treatment being the implementation of broad spectrum antibiotics. The finding that antibiotics can ameliorate the disease indicates that the disease is likely to be of bacterial origin or at least influenced by the bacterial milieu within the bird.

The agent causing spotty liver disease is unknown but various potential culprits, including *Campylobacter jejuni* and a “vibrio”, have been suggested at different times.

Diseased livers have often been completely negative for bacterial isolation and microscopic analysis also sometimes fails to detect any sign of bacterial involvement within the liver. For this reason we hypothesise that the disease pathology within the liver may result from the action of a systemic toxin mobilised from a bacteria in the gut via the circulation to the liver. Therefore, we have investigated the bacteria within the gut as a potential source of toxin acting on the liver. The approach that has been taken is to characterize the structure of the gut microbiota in an attempt to identify bacterial species that are more abundant in diseased birds than in healthy birds.

MATERIALS AND METHODS

Samples

Samples from eight SLD outbreaks were collected. At each site samples were taken from both diseased birds and healthy birds. The number of samples from each outbreak included in the final analysis is shown in Table 1.

Table 1. Number of samples from each disease outbreak that were sequenced.

Outbreak No.	Healthy Bird Samples		SLD Bird Samples	
	Caecum	Intestine	Caecum	Intestine
1	6	5	6	5
2	6	6	6	6
3	6	0	6	0
4	4	0	4	0
5	5	3	8	7
6	4	3	3	2
7	3	3	3	0
8	2	3	4	0

Samples were also collected from an infection trial in which material from SLD affected birds was used to infect healthy birds. Control material from healthy birds was also used. One batch of the material used to infect birds was prepared anaerobically and a second batch was prepared under aerobic conditions. The number of birds sampled and sequenced from each group is shown in Table 2.

Table 2. Number of samples sequenced from infection trial material.

Sampling Point	SLD Material		Control Material	
	Anaerobic	Aerobic	Anaerobic	Aerobic
Day 3	9	6	0	3
Day 7	8	7	2	3
Day 11	11	11	2	3

DNA preparation from gut samples

Material from the intestinal content and caecal samples were resuspended in 250 µl of phosphate buffered saline. Total DNA was isolated using the method of Yu and Morrison (2004). Briefly, a sample was transferred to a 2 ml screw cap tube with lysis buffer (500mm NaCl, 50mM Tris-HCl pH8.0, 50 mM EDTA and 4% sodium dodecyl sulfate) and sterile zirconium beads and then homogenized using a Precellys 24 tissue homogenizer (Bertin Technologies) at maximum speed of 6500 rpm, twice, 3 x 10 seconds each time. Following centrifugation the supernatant was collected ammonium acetate was added and nucleic acid was precipitated with isopropanol, followed by ethanol wash. After centrifugation the pellet was resuspended in Tris-EDTA buffer and digested with DNase-free RNase and proteinase K to remove RNA and protein. The DNA was finally purified on a QIAamp column (Qiagen) according to the manufacturer's instructions. DNA quantity and quality was measured on a NanoDrop ND-1000 spectrophotometer.

PCR amplification of 16S ribosomal RNA gene sequences

DNA derived from the bacteria of the birds was processed to amplify the 5' end (V1-V3 region) of the eubacterial 16S ribosomal RNA genes. DNA was amplified using Bio-Rad iProof DNA polymerase. Each PCR reaction contained 25 µl of iProof 2X master mix (containing buffer, nucleotides and iProof enzyme), 2 µl of each primer (final concentration 0.5 µM), 1.5 µl DMSO, 0.5 µl 50 mM $MgCl_2$ and template DNA made up to 19 µl in water. The primers used were designed to amplify the V1-V3 region of the 16S rRNA gene (forward primer, 5' AGAGTTTGATCCTGG 3'; reverse primer, 5' TTACCGCGGCTGCT 3'). Each primer also included sequences to facilitate the sequencing of products in the Roche/454 system and the reverse primers consisted of a related set of primers which differed in "barcode" sequences; specific sequences introduced into the primers to allow tagging of individual samples in a multiplex sequencing system. PCRs were performed in an Eppendorf Mastercycler using the following conditions: 98°C for 60 seconds then 25 cycles of 98°C for 5 seconds, 40°C for 30 sec, 72°C for 30 sec; elongation at 72°C for 10 min then hold at 5°C. The efficiency of PCR amplification of each sample was assessed by running 10 µl of the PCR mix on a 1.5% agarose gel.

High throughput sequencing and analysis of 16S amplicons

The amplified 16S samples from each animal were pooled using approximately equal amounts of each PCR product. The pooled samples were sequenced using the Roche/454 FLX+ Genome Sequencer and Titanium chemistry. The output sequence file was analysed using a number of publically available software packages and databases. Sff files were burst into fasta and qual files using PyroBayes (Quinlan et al., 2008) and chimeric sequences removed using pintail (Ashelford et al., 2005). Sequence quality trimming settings were: sequence length 300-600 bases, no ambiguous sequences, minimum average quality score of 25 and maximum homopolymer run of 6 nucleotides, using QIIME (Caporaso et al., 2010). OTUpipe (Edgar, 2011), combining USEARCH and UCLUST scripts (Edgar, 2010; Edgar et al., 2011), was used to perform denoising error-correction, abundance and amplicon estimation and OTU picking. After OTUs were assigned, using 97% sequence similarity, all of the remaining analysis used QIIME software using QIIME defaults, unless stated otherwise. Taxonomy was assigned using a Blast method against the GreenGenes database (DeSantis et al., 2006) and further confirmed using the EzTaxon database (Chun et al., 2007). All samples represented by less than 1000 sequences were removed from the analysis.

RESULTS

16S ribosomal RNA gene amplification

Samples amplified satisfactorily and gave clean products with very low amounts of non-target bands. An example of the gel analysis of the PCR amplification is shown in Figure 1. Most samples amplified satisfactorily and progressed to sequencing. For a small proportion of the samples PCR amplicons could not be obtained and so they could not be included in the analysis.

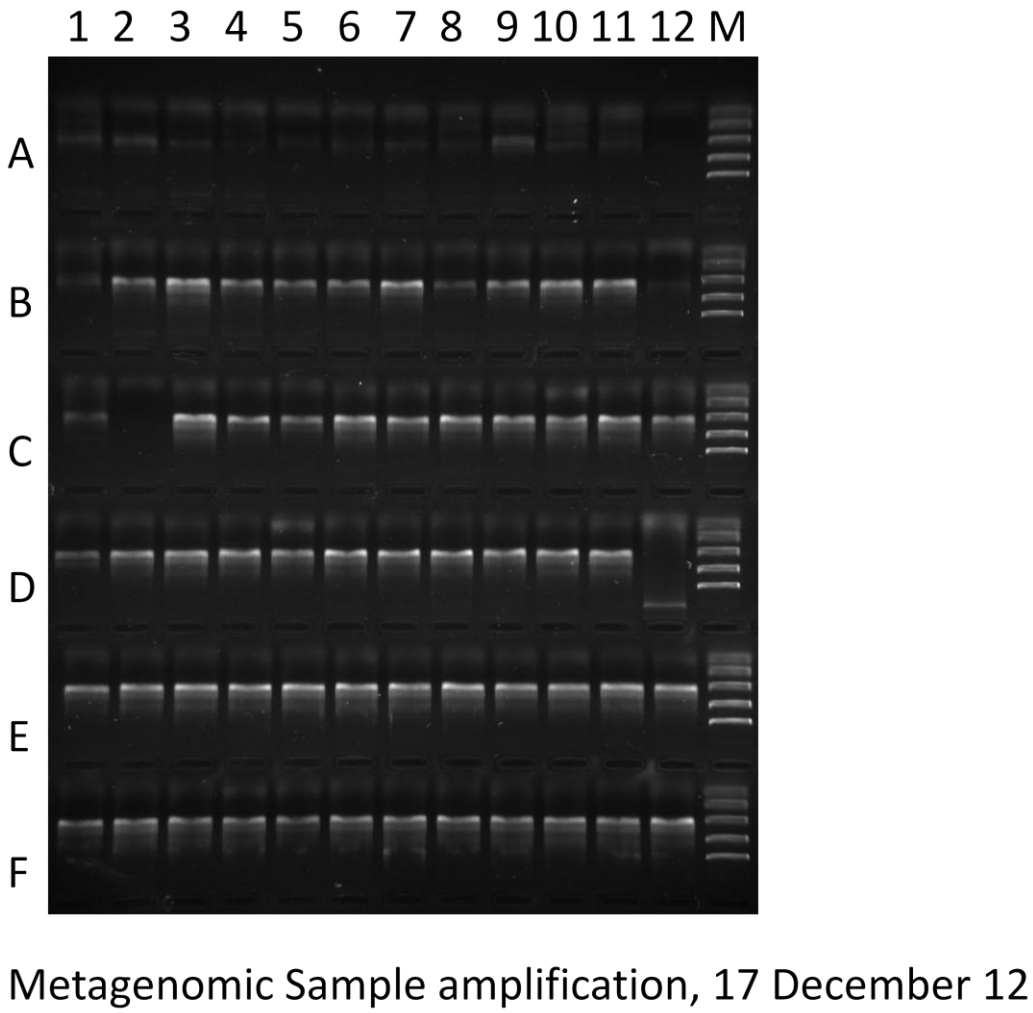


Figure 1. Example of gel analysis of amplification of 16S PCR products

Sequence output

Sequence data was recovered from 119 samples which gave a total of 773,921 raw sequence reads. After quality trimming 513,895 reads were retained; an average of 4,318 reads per sample.

For the experimental infection trial 157,032 raw sequence reads were obtained from the 65 samples. After quality trimming 133,716 sequence reads remained for analysis; an average of 2057 reads per sample.

Experimental infection trial

The experimental infection trial did not provide any indication of the bacterium that might be responsible for SLD. The only obvious sample clustering that could be seen was based only on the day samples were taken (Figure 2). No clustering of samples was seen based on the source of bacteria used to inoculate birds.

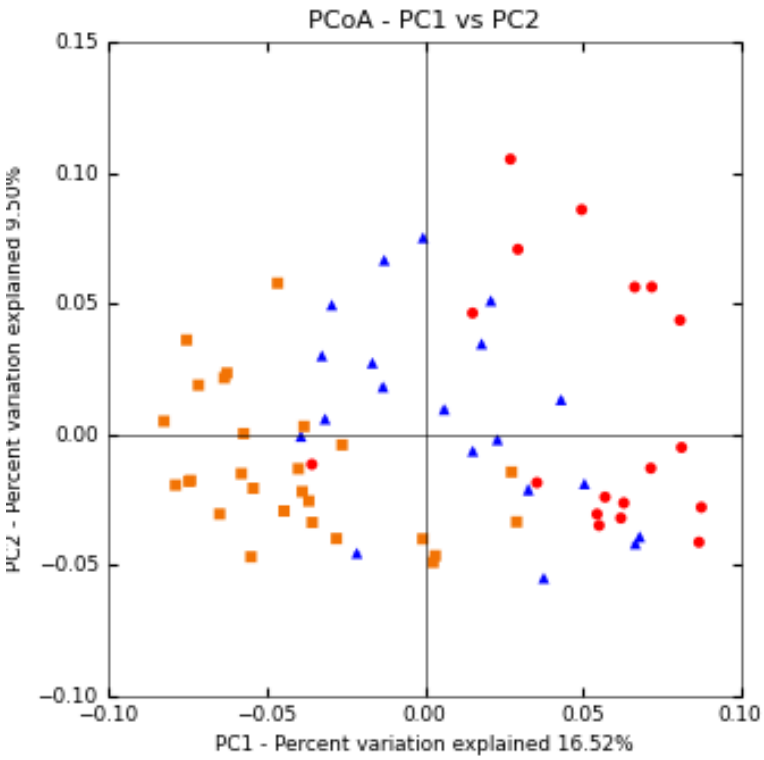


Figure 2. PCoA plot of Bray-Curtis results from analysis of experimental infection samples. The red circles indicate samples from day 3; the blue triangles are samples from day 7; the green squares are from samples on day 11.

Intestinal samples are dominated by *Lactobacillus* species

To explore the data principal component analysis (PCoA) plots were inspected (Figure 3). As expected the most obvious clustering of samples was based on the tissue origin of samples; that is whether derived from caecal content or intestinal content. Component 1 accounted for a large percentage (64.84%) of the variation seen between samples. In general the microbiota samples from intestinal content were more tightly clustered than the caecal derived samples.

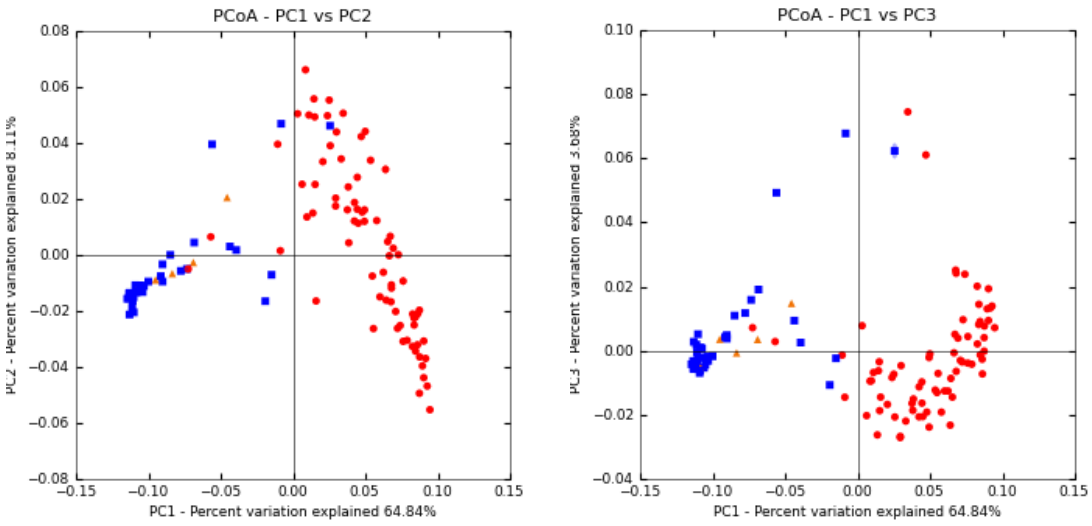


Figure 3. PCoA plots using weighted Unifrac results. Red circles represent 16S samples derived from caecal content. Blue squares represent samples derived from intestinal content and the green triangles are from ileal content.

Background: *PCoA plots are designed to show the overall relatedness of samples, the closer samples are plotted to each other the more similar they are. There are a range of algorithms that can be used to define the similarity (or dissimilarity) of samples that are used in PCoA plots. Within the software packages the plots are interactive such that mousing over a symbol reveals which sample it is derived from. 3D plots can also be produced.*

The ileal samples were quite distinct from the caecal samples because they were strongly dominated by *Lactobacillus* species. A phylogenetic analysis (Figure 4) shows that most ileal samples consisted of greater than 90% *Lactobacillus* species whereas the caecal samples rarely had more than 20% *Lactobacillus*. The domination of the intestinal samples by *Lactobacillus* species meant that very little depth of data was seen for other bacterial species. Therefore, the rest of the analysis concentrated on the caecal derived samples.

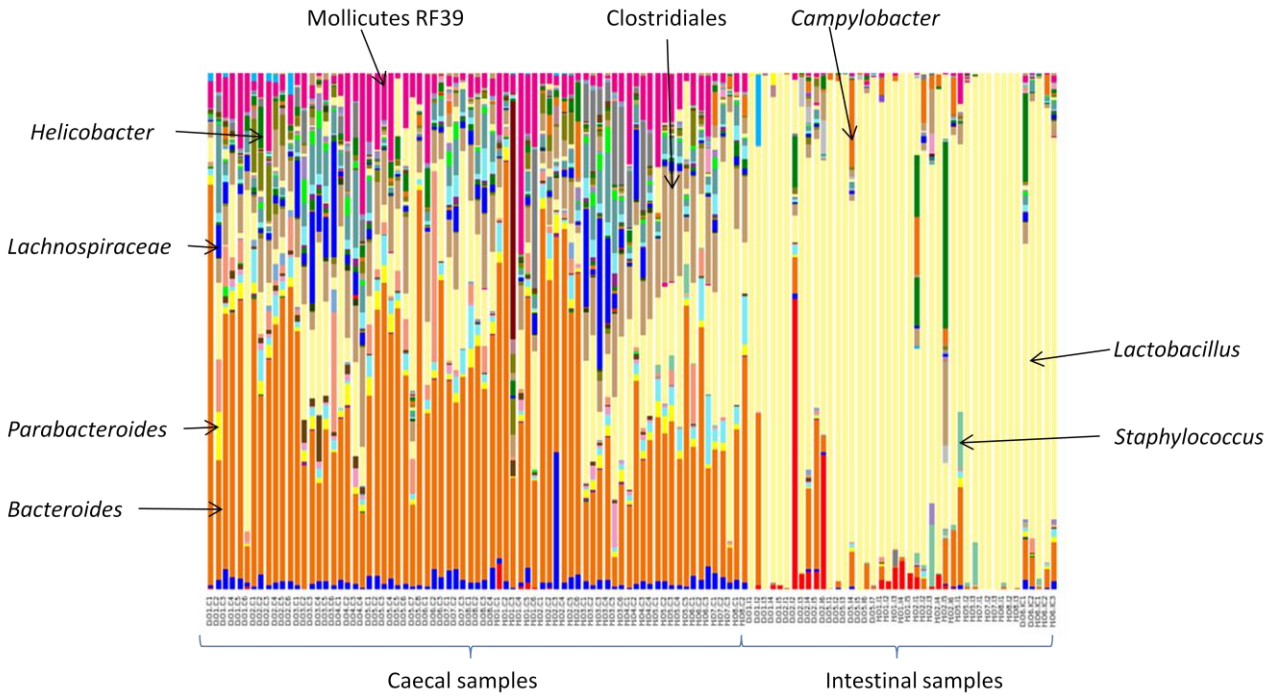


Figure 4. Phylogenetic analysis of all samples at the genus level. The caecal samples are on the left and the intestinal samples are on the right. Each column represents the total microbiota for a particular sample; coloured and divided proportionally to the type of bacteria present. The colour representing a number of the different bacterial groups have been identified by the labels and arrows.

Samples cluster according to the disease outbreak from which they are derived

The PCoA analysis also clearly showed that samples clustered according to the disease outbreak from which they were derived (Figure 5). Also, outbreaks 1 to 4 were somewhat clustered together as were outbreaks 5 to 8. Outbreaks from the same properties, but sampled at different times were clustered together. For example, outbreaks 1 and 2 were both from Ooroolong, sampled 12 days apart; outbreaks 3 and 4 were both from Orchards/Lethbridge, sampled 1 week apart; and outbreaks 5 and 6 were both from Sunny Acres, sampled 7 weeks apart. It is to be expected that such clustering would be seen as the caecal microbiota of birds from one property is likely to have been influenced by such factors as common origin and in some cases a common batch of birds, similar environments and husbandry practices, as well as similar feed and water.

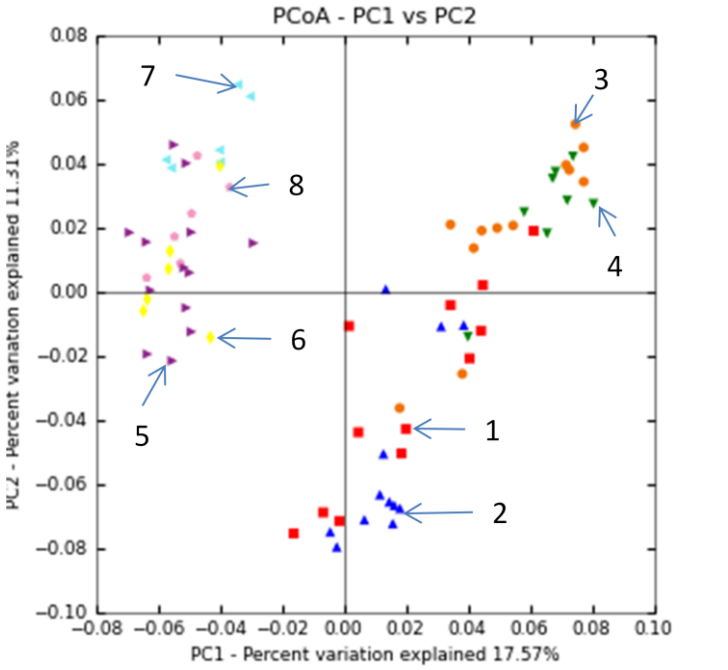


Figure 5. PCoA plot using Bray-Curtis results from the caecal samples. The coloured symbol corresponding to samples from each outbreak are indicated by the numbers and arrows.

Samples do not cluster according to health status

We further examined the clustering analysis for any evidence of clustering based on health status. No clustering could be discerned (Figure 6). Because of the differences in microbiota composition between outbreaks the global analysis to compare healthy and diseased birds may be confounded by the wide spread of results. Therefore clustering by health status was further examined by considering the clustering analysis of the caecal samples from individual outbreaks (Figure 7). Clearly there are no gross systematic differences in the overall structures of microbiota from healthy and diseased birds. This is consistent with the hypothesis that a single pathogenic bacterial species, perhaps of fairly low abundance, is responsible for disease pathogenesis. To identify such a hypothesised bacterium it was necessary to look in more detail at the microbiota analysis, down to the level of each bacterial species identified.

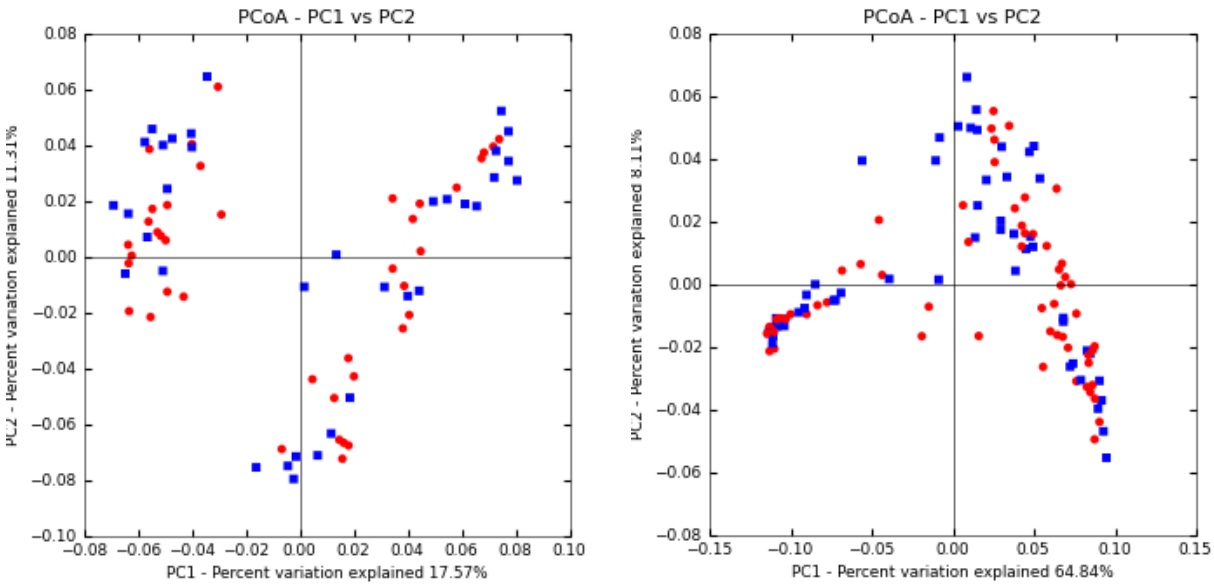


Figure 6. PCoA plots colour coded for health status. The left panel plots the Bray-Curtis results for the caecal samples (related to Figure 5). The right panel plots the weighted Unifrac results for all samples (related to Figure 3). The Blue squares indicate samples from healthy birds and the red circles indicate samples from SLD birds.

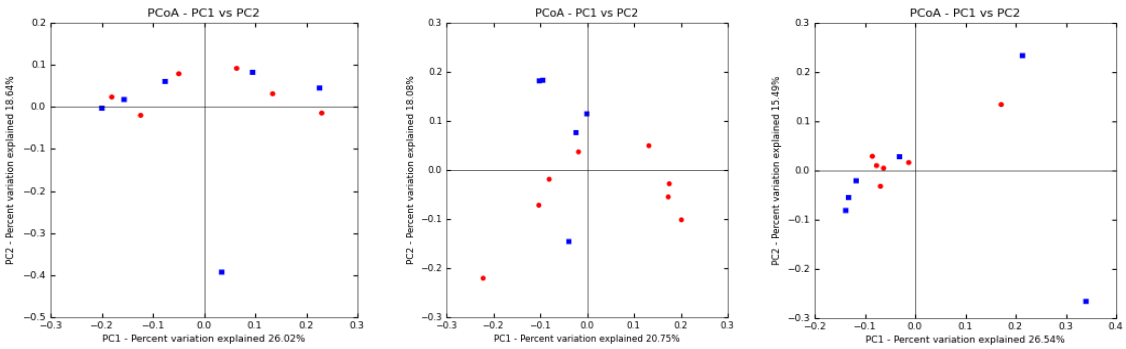


Figure 7. PCoA plots of Bray-Curtis results from caecal samples of outbreaks 1, 2 and 5. The Blue squares indicate samples from healthy birds and the red circles indicate samples from SLD birds.

Background: *In this type of high throughput 16S rRNA based analysis of microbiota the analysis is based on operational taxonomic units (OTU). For the purposes of discussion the OTU can be regarded as roughly equivalent to a bacterial species when the sequence clustering to produce the OTUs is set at a similarity level of 97%. This means that all the sequences that are classified within a particular OTU have a similarity of at least 97%. If the OTUs are based on a lower level of similarity then, depending on the percentage similarity, the OTU would be more equivalent to a genus, family or order. Once sequences are clustered into OTUs a representative sequence can then be compared with phylogenetic databases to determine if the OTU is related to a known bacterial species.*

Statistical analysis of OTU abundance identifies a potential pathogen

The analysis of the sequence data results in a table in which the abundance of each OTU is mapped against each sample. Statistical tools can then be used to identify OTUs in which variations in abundance across samples correlates with variations in some other sample characteristic, for example which outbreak the sample is from or the health status of the bird. When ANOVA was used to interrogate the data from each outbreak in isolation it was found that for each outbreak there were a number of OTUs that correlated with health status (healthy or SLD) at a statistically significant level ($p<0.05$) however no single OTU correlated at statistically significant levels across

more than 3 of the 8 outbreaks. When all the data was pooled to create a single data set, and therefore give greater statistical power, ANOVA identified 12 OTUs that were differentially abundant between healthy and SLD birds at $p<0.05$; these OTUs are shown in Table 3. If the working hypothesis is correct then the pathogen causing SLD would be expected to be in higher abundance in the diseased birds. Therefore, a relevant OTU should have a D/H ratio of greater than 1. Only 3 of the 12 differential OTUs had a ratio of greater than one. OTUs 1229 and 267 were classified by reference to the GreenGenes database (DeSantis et al., 2006) down to genus level and identified as *Bacteriodes*. OTU 47 was classified to the genus *Helicobacter*. When the data set was investigated in more detail it was seen that OTU 1229 was only identified in 3 of the outbreaks (2, 5, and 6) and OTU 267 was seen in 5 of the outbreaks (1, 2, 3, 4, and 7) (data not shown). OTU 47 was identified in all 8 outbreaks (Table 4) and on this basis is a candidate as the potential pathogen causing SLD.

Table 3. OTUs identified as differing in abundance between healthy and diseased birds

OTU	Probability	H_mean[a]	D_mean[a]	D/H[b]	Consensus Lineage[c]
90	0.0078	0.0034	0.0014	0.4136	p__Firmicutes; c__Clostridia; o__Clostridiales
222	0.0094	0.0051	0.0026	0.5113	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides
19	0.0150	0.0235	0.0064	0.2730	p__TM7; c__TM7-3; o__I025; f__Rs-045
657	0.0203	0.0014	0.0005	0.3500	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__Bacteroides plebeius
1229	0.0215	0.0002	0.0006	4.0250	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides
94	0.0216	0.0018	0.0005	0.2625	p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus coleohominis
316	0.0224	0.0010	0.0003	0.3387	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales
1059	0.0237	0.0090	0.0022	0.2487	p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus helveticus
32	0.0240	0.0085	0.0013	0.1490	p__Firmicutes; c__Clostridia; o__Clostridiales
97	0.0271	0.0084	0.0038	0.4600	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides
267	0.0352	0.0003	0.0012	4.2778	p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides

47	0.0408	0.0026	0.0073	2.7897	p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter

[a] percentage of OTU in data set, [b] ratio of proportion in diseased birds compared to healthy birds, [c] taxonomic assignment of OTU by reference to GreenGenes database

Table 4. Prevalence of OTU 47 across the outbreaks

Outbreak	Probability[a]	H_mean	D_mean	D/H ratio
CO1	0.3693	0.0004	0.0030	8.5000
CO2	0.7547	0.0097	0.0072	0.7407
CO3	0.1646	0.0005	0.0025	4.6667
CO4	0.0841	0.0013	0.0167	12.4000
CO5	0.0493	0.0000	0.0164	Inf
CO6	0.6612	0.0022	0.0039	1.8333
CO7	0.3027	0.0054	0.0007	0.1333
CO8	0.3552	0.0000	0.0008	Inf
All	0.0408	0.0026	0.0073	2.7897

[a] Probability that the difference between the abundance in healthy and diseased birds is by chance.

Although OTU 47 was detected in all outbreaks it was only more abundant in the diseased birds in 6 of the 8 outbreaks and the difference between abundance in diseased and healthy birds was only statistically significant in a single outbreak, although the overall pooled results for OTU 47 also reached statistical significance. In some outbreaks (e.g. 1 and 4) substantial differences were seen in the abundance of OTU 47 but did not reach statistical significance because of the high bird-to-bird variation within a group and the low number of samples that were available from some outbreaks –

note that the outbreak that did show statistical significance was also the outbreak with the greatest number of samples.

OTU 47 is closely related to *Helicobacter pullorum*

A sequence representative of OTU 47 was used to interrogate the EzTaxon database (Chun et al., 2007) to identify the most closely related cultured bacteria. OTU 47 had 98.8% identity to *Helicobacter pullorum* NCTC 12824. This high degree of homology suggests that OTU 47 is indeed a strain of *H. pullorum*. *H. pullorum* NCTC 12824 was isolated from a chicken and is the type strain for this species.

DISCUSSION

OTU 47, tentatively identified as *H. pullorum*, has been identified as the potential pathogen responsible for the development of Spotty Liver Disease. *H. pullorum* was first recognised and named in 1994 (Stanley et al., 1994). It has been commonly found in healthy chickens (Atabay et al., 1998; Ceelen et al., 2006a; Zanoni et al., 2007; Basaran Kahraman and Ak, 2103; Wai et al., 2012; Manfred et al., 2011) and turkeys (Zanoni et al., 2011). There have been tentative suggestions of a link between *H. pullorum* and SLD but no convincing evidence has been reported (Stanley et al., 1994; Burnens et al., 1996) and no signs of liver pathology have been reproduced in an infection model (Ceelen et al., 2007; Neubauer and Hess, 2006). *H. pullorum* is difficult to differentiate from other Helicobacters and Campylobacters and so this is consistent with early suggestions that *Campylobacter* and a "vibrio" like organism may be involved with SLD. *H. pullorum* may be a zoonotic agent as infections in humans have been implicated in gastrointestinal diseases and in liver disease. *H. pullorum* also causes liver disease in mice. A toxin, cytolethal distending toxin, has been identified in avian and human isolates of *H. pullorum* (Young et al., 2000; Ceelen et al., 2006b)

The evidence produced in this study, that has lead to the conclusion that *H. pullorum* may be involved in SLD, is certainly not overwhelming but it seems a remarkable coincidence that the single candidate identified has previously been suggested by other workers to be involved in disease. OTU 47 was seen in healthy birds as well as diseased birds. This suggests that OTU 47 can be non-pathogenic; this could be because a critical population level is required to induce disease, other predisposing factors are required, or strain differences determine pathogenicity. There is certainly good evidence that there are significant levels of strain diversity, both in human and chicken isolates (Gibson et al., 1999). In six out of eight sets of outbreak samples OTU 47 was detected at higher levels in the SLD birds than in healthy control birds, as our hypothesis would predict. The finding that in two sets of outbreak samples OTU 47 was more abundant in healthy birds than in SLD birds does not necessarily argue strongly against the hypothesis as it is not clear which part of the gastrointestinal tract (GIT) should ideally by assayed for the pathogen. This study has concentrated on samples from the caecum but other areas of the gut may be of equal or greater importance as the place of residence of the pathogen.

FUTURE WORK

The priority of future work should be to obtain definitive proof of the importance of *H. pullorum* in the pathogenesis of SLD. While that work proceeds all opportunities should be taken to sample further outbreaks to determine if the findings reported here can be reproduced. It would always be helpful to have more samples from each outbreak so that the statistical power of the analysis can be increased.

Two lines of research that could be pursued in order to provide more evidence of the proposed role of *H. pullorum* is to undertake a large survey of outbreak samples using PCR and culturing and infection studies to reproduce the disease.

A PCR study would be relatively straightforward. A specific PCR test has already been published (González et al., 2008) and if it proved reliable could be used to rapidly assay samples from disease

outbreaks. Ideally samples from throughout the GIT would be assayed to determine the population of *H. pullorum* in various niches throughout the gut to determine if there is a preferred location or if population levels in particular regions of the gut correlate more strongly with disease status. Such studies, if successful, would add much stronger circumstantial evidence for the involvement of *H. pullorum*.

Definitive proof that *H. pullorum* causes SLD would be achieved if the disease could be reliably reproduced using cultured bacteria. Previous attempts to do this have been unsuccessful (Ceelen et al., 2007; Neubauer and Hess, 2006) indicating that either non-pathogenic isolates were used or other unidentified predisposing factors are required to effectively reproduce disease. The finding that *H. pullorum* is commonly found in samples from healthy birds may indicate that some strains are pathogenic and others are not. It may be necessary to screen many isolates in order to find one that is pathogenic – obviously isolating bacteria from diseased birds rather than healthy birds may increase the probability of finding a pathogenic strain. It would be worthwhile to first attempt to develop an in vitro assay that may indicate the pathogenic potential of *H. pullorum* isolates. An obvious thing to try would be to see if *H. pullorum* culture supernatants are cytotoxic to cultured cells, ideally chicken liver cells. Candidate bacterial isolates should then be used to infect birds. Some thought needs to be devoted to determining if there are any predisposing factors (e.g. stressed or immune-compromised birds) that may increase the likelihood of disease.

If definitive proof of the involvement of *H. pullorum* can be found then it opens up the possibility of devising specific intervention strategies, such as vaccination, to alleviate the burden of disease.

REFERENCES

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* 71: 7724-7736.

Atabay HI, Corry JEL, On SLW (1998) Identification of unusual Campylobacter-like isolates from poultry products as *Helicobacter pullorum*. *J Appl Microbiol* 84:1017–1024.

Basaran Kahraman B, Ak S (2013) Investigation of *Helicobacter pullorum* occurrence in chicken in the Marmara region of Turkey. *Istanbul Univ Vet Fak Derg* 39:63-66.

Burnens AP, Stanley J, Nicolet J (1996) Possible association of *Helicobacter pullorum* with lesions of vibrionic hepatitis in poultry. In: Newell DG, Keatly JM, Feldman RA (Eds) *Campylobacters, Helicobacters* and Related Organisms. Plenum Press, New York, pp 291-293.

Caporaso, J. G., J. Kuczynski, et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

Ceelen LM, Decostere A, Van den Bulck K, On SLW, Baele M, Ducatelle R, Haesebrouck F (2006a) *Helicobacter pullorum* in Chickens, Belgium. *Emerg Infect Dis* 12:263–267.

Ceelen LM, Haesebrouck F, Favoreel H, Ducatelle R, Decostere A (2006b) The cytolethal distending toxin among *Helicobacter pullorum* strains from human and poultry origin. *Vet Microbiol* 113:45–53.

Ceelen LM, Decostere A, Chiers K, Ducatelle R, Maes D, Haesebrouck F (2007) Pathogenesis of *Helicobacter pullorum* infections in broilers. *Internat J Food Microbiol* 116:207–213.

Chun J, Lee JH, Jung Y, Kim M, Kim S, et al. (2007) EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 57: 2259-2261.

DeSantis, T. Z., P. Hugenholtz, et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069-5072.

Edgar RC (2011) Otupipe - OTU clustering pipeline for next-gen reads. Available: http://drive5.com/otupipe/otupipe_manual1.1.pdf. Accessed 18 June 2013.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194-2200.

Gibson JR, Ferrus MA, Woodward D, Xerry J, Owen RJ (1999) Genetic diversity in *Helicobacter pullorum* from human and poultry sources identified by an amplified fragment length polymorphism technique and pulsed-field gel electrophoresis. *J Appl Microbiol* 87:602–610.

González A, Piqueres P, Moreno Y, Cañigral I, Owen RJ, Hernández J, Ferrús MA (2008) A novel real-time PCR assay for the detection of *Helicobacter pullorum*-like organisms in chicken products. *Int Microbiol* 11:203–208.

Ludwig, W., O. Strunk, et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* 32: 1363-1371.

Manfreda G, Parisi A, Lucchi A, Zanoni RG, De Cesare A (2011) Prevalence of *Helicobacter pullorum* in Conventional, Organic, and Free-Range Broilers and Typing of Isolates. *Appl Environ Microbiol* **77**:479–484.

Neubauer C, Hess M (2006) Tissue tropism of *Helicobacter pullorum* in specific pathogen-free chickens determined by culture and nucleic acid detection. *Avian Dis*. 50:620–623.

Quinlan AR, Stewart DA, Stromberg MP, Marth GT (2008) Pyrobayes: an improved base caller for SNP discovery in pyrosequences. *Nature Methods* 5: 179-181.

Schloss, P. D., S. L. Westcott, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.

Stanley J, Linton D, Burnens AP, Dewhirst FE, On SLW, Porter A, Owen RJ, Costas M (1994) *Helicobacter pullorum* sp. nov. - genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. *Microbiol* 140:3441–3449.

Wai SS, Saleha AA, Zunita Z, Hassan L, Jalila A (2012) Occurrence of co-infection of *Helicobacter pullorum* and *Campylobacter* spp. in broiler and village (indigenous) chickens. *Pak Vet J* 32:503-506.

Young VB, Chien C-C, Knox KA, Taylor NS, Schauer DB, Fox JG (2000) Cytolethal Distending Toxin in Avian and Human Isolates of *Helicobacter pullorum*. *J Infect Dis* 182:620–623.

Yu Z and M. Morrison (2004) Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36: 808-812.

Zanoni RG, Rossi M, Giacomucci D, Sanguinetti V, Manfreda G (2007) Occurrence and antibiotic susceptibility of *Helicobacter pullorum* from broiler chickens and commercial laying hens in Italy. *Internat J Food Microbiol* 116:168–173.

Zanoni RG, Piva S, Rossi M, Pasquali F, Lucchi A, De Cesare A, Manfreda G (2011) Occurrence of *Helicobacter pullorum* in turkeys. *Vet Microbiol* 149:492–496.

Spotty Liver Disease 16S Metagenomics

Fourth Report

Robert Moore[1], Theo Allnutt[1], Dana Stanley[2], and Peter Scott[3]

[1]CSIRO Biosecurity Flagship, [2]Central Queensland University, [3]Scolexia Pty Ltd

SUMMARY

In previously work we reported the use of high throughput 16S rRNA gene sequencing to characterise the microbiota populations present in the gut of birds affected by spotty liver. By comparing the microbiota composition to that seen in healthy birds we were able to identify a bacterial species that was more prevalent in diseased birds and was observed in all eight disease outbreaks investigated. In the latest work we have extended the analysis of the candidate bacterium by carrying out whole metagenome analysis of the sample in which the candidate organism was most prevalent. We have shown that the candidate is related to but distinct from *Helicobacter pullorum*. There is no close match to the genome in the publicly available sequence databases and so we conclude that the candidate bacterium represents a new species. The derivation of genome sequence data has allowed the design of a series of PCR assays for the organism. These PCR assays should allow a more detailed survey of samples from disease outbreaks and will allow us to test the hypothesis that this new organism is the causative agent of Spotty Liver Disease.

INTRODUCTION

OTU 47 is closely related to *Helicobacter pullorum*

The candidate bacterium that has been identified, initially as OTU 47, had 98.8% sequence identity to the 16S rRNA gene of *Helicobacter pullorum* NCTC 12824. On this basis it was tentatively identified as a strain of *H. pullorum*.

H. pullorum was first recognised and named in 1994 (Stanley et al., 1994). It has been commonly found in healthy chickens (Atabay et al., 1998; Ceelen et al., 2006a; Zanoni et al., 2007; Basaran Kahraman and Ak, 2013; Wai et al., 2012; Manfreda et al., 2011) and turkeys (Zanoni et al., 2011). There have been tentative suggestions of a link between *H. pullorum* and SLD but no convincing evidence has been reported (Stanley et al., 1994; Burnens et al., 1996) and no signs of liver pathology have been reproduced in an infection model (Ceelen et al., 2007; Neubauer and Hess, 2006). *H. pullorum* is difficult to differentiate from other helicobacters and campylobacters and so this is consistent with early suggestions that *Campylobacter* and a "vibrio" like organism may be involved with SLD. *H. pullorum* may be a zoonotic agent as infections in humans have been implicated in gastrointestinal diseases and in liver disease. *H. pullorum* also causes liver disease in mice. A toxin, cytolethal distending toxin, has been identified in avian and human isolates of *H. pullorum* (Young et al., 2000; Ceelen et al., 2006b).

Our initial attempts to culture a *H. pullorum* like organism from the gut of birds affected by spotty liver disease have been unsuccessful. Therefore we sought other ways to advance our knowledge of the organism without having a pure cultured isolate. We used a metagenomic approach to garner information about the genome of the candidate organism. We anticipated that comprehensive genomic information would certainly allow us to design new PCR assays to detect the organism and may give us sufficient information about the organism's metabolism to allow the design of more robust culturing methods to allow its isolation from clinical cases.

MATERIALS AND METHODS

Metagenomics

Whole metagenome analysis involves the sequencing of all the genomes in a complex microbial population, such as that recovered from the gut of a spotty liver affected bird. Our initial 16S rRNA gene sequencing targeted the characterisation of a single gene across the many different bacteria present but whole metagenome analysis attempts to characterise all genes across all the genomes. Because metagenome analysis generates datum that is thousands of times more complex than 16S rRNA gene analysis it is only possible to do one or a few samples in sufficient detail to give useful results.

The 16S rRNA gene analysis demonstrated that the candidate organism, OTU 47, makes up only a small proportion of the total microbial population in any sample. To maximise the chances of recovering useful information from a metagenomic analysis we used the specific DNA sample isolated from SL outbreak material which the 16S rRNA gene analysis had demonstrated to carry the highest level of OTU 47. We anticipated that even when using this sample the amount of OTU 47 sequence identified would be low with very sparse coverage across its genome. To increase the value of recovered DNA sequence information we chose to us a specialised sequencing method on our Roche/454 sequencing machine. Long paired-end sequencing allows sequences that are several kilobases apart to be linked to each other and hence assists in assembling sequence data. This information was supported by a 2 x 250

bp sequencing run on an Illumina MiSeq instrument.

RESULTS

Metagenome sequencing

The Roche/454 sequencing run produced 1.17 million sequence reads and the Illumina MiSeq run produced about 15 million reads. Database searches indicated that no more than 0.7% of this sequence information had homology to helicobacter sequences. Figure 1 indicates the phylogenetic composition of the sample as indicated by MG-RAST analysis of a random subset of 2.4 million sequences of MiSeq data.

The low level of the candidate organism has meant that only a small amount of fragmentary genomic DNA sequence information has so far been mined from the metagenomic data. This has not allowed the construction of any form of whole genome assembly but has given sufficient information to allow the design of new PCR assays to assist in ongoing characterisation of clinical material derived from spotty liver cases.

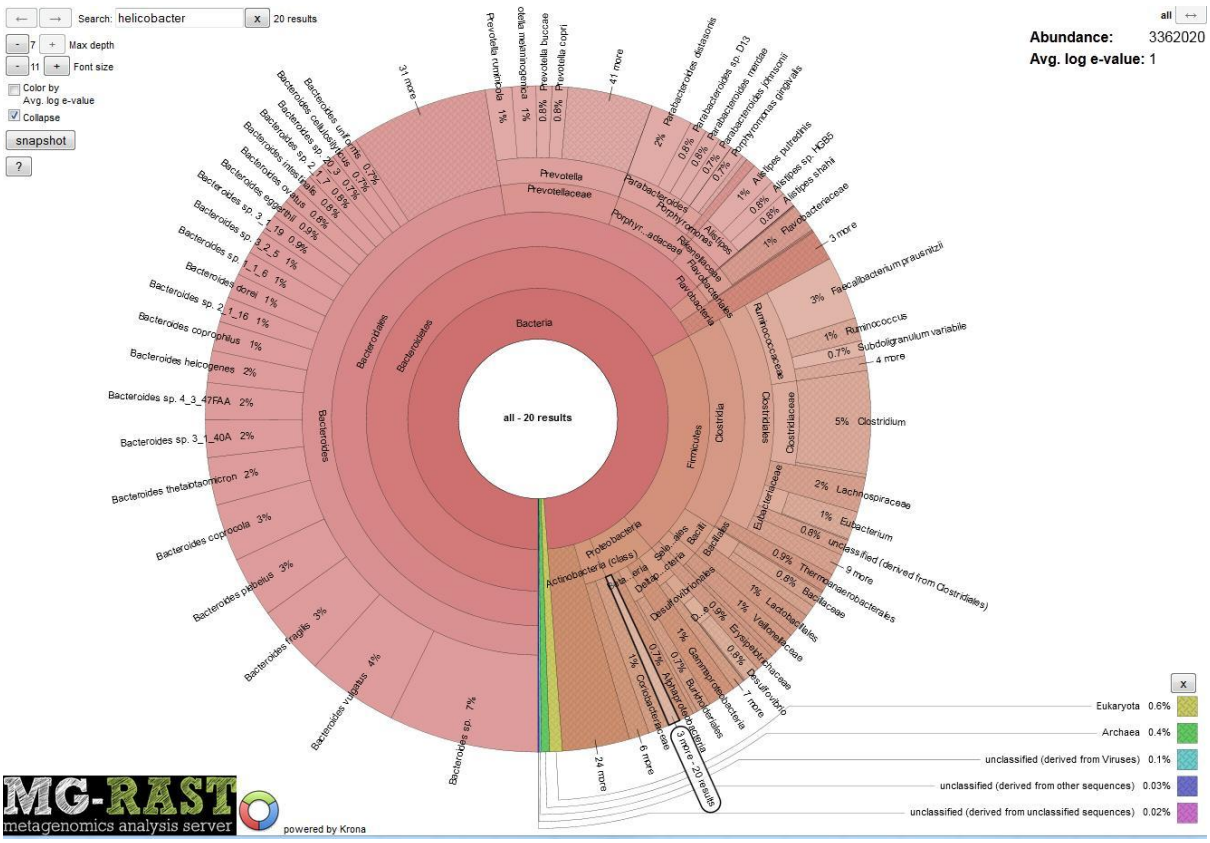


Figure 1. Phyogenetic composition of the sequenced DNA sample (derived from the caecal content of a spotty liver affected hen) as determined by MG-RAST analaysis (Glass et al., 2010). The small helicobacter related proportion is contained within the highlighted segment at about 5 o'clock on the diagram.

Using paired-end sequence reads from the Roche/454 run it was possible to select out all those sequence reads that corresponded to the OTU 47 16S rRNA gene sequence. By using the other, non-16S, segment of each paired-end read it was possible to "walk out" into the genome to find other genomic regions that were linked to 16S in the target genome. When the new sequences were investigated by searching against the public sequence databases it was commonly found that the sequence had no good match to anything previously characterised. Other regions showed significant homology to helicobacter genomes but generally not to *H. pullorum* but rather *H. hepaticus*, *H. cinaedi*, and other helicobacters. This indicates that the candidate organism is unlikely to be a strain of *H. pullorum* but rather some other previously uncharacterised helicobacter.

PCR development

Some of the newly identified genomic regions of the candidate organism were used to design PCR primer pairs. The aim was to develop a specific and reliable PCR assay to monitor for the presence of the organism in DNA extracts prepared from clinical material. Five primer pairs were designed. All pairs produced PCR products when used on the DNA sample that was used for metagenomic analysis (Figure 2).

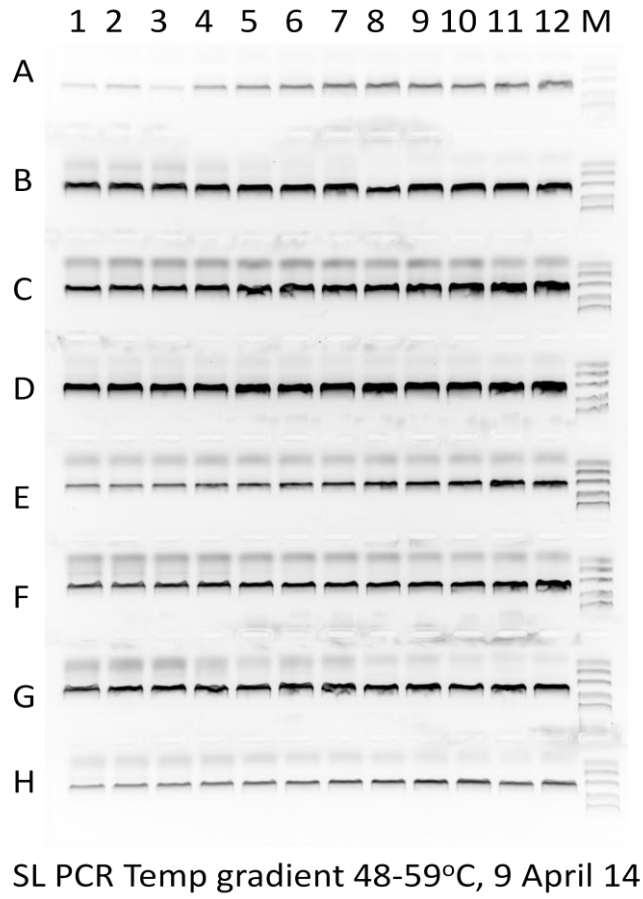


Figure 2. Gradient PCR to determine temperature optimum for PCR reactions. PCRs were carried out in a 96-well plate and samples run on a 1.5% agarose gel for visualisation. Temperature increased in 0.5°C increments across the plate from wells 1 to 12. Five primer pairs were used; T1-T2 in row A, T3-T4 in row B, T5-T6 in rows C and F, T7-T8 in rows D and G, and T15-T16 in rows E and H. Bioline Taq polymerase was used in rows A to E and New England Biolabs Q5 polymerase was used in rows F to H. The template DNA in all tracks was from the DNA sample that had been used for metagenome analysis. Molecular weight markers are loaded in track M.

Each of the PCR primer pairs was used on a series of samples, including known positive and negative controls. An example of the results is shown in Figure 3. The primers appear to provide specificity for the candidate organism and under the PCR cycling conditions used the assay is semi-quantitative in that samples that, from deep 16S sequencing, were expected to have the highest levels of the candidate gave the strongest PCR bands while samples with lower levels gave less intense PCR bands. Negative controls did not produce a band. Interestingly, the candidate organism was readily

detected in caecal samples from a series of birds (Figure 3, positions C6 to C10) but was not detected in the DNAs prepared from small intestine content samples from the same birds (Figure 3, positions C1 to C5). This indicates that culturing efforts should be concentrated on caecal samples rather than small intestine samples.

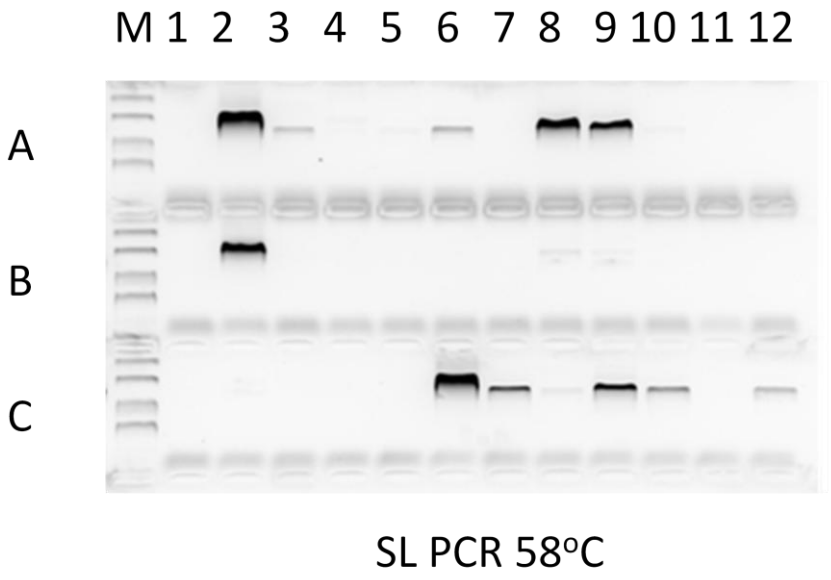


Figure 3. PCR using primer pairs T7-T8 (rows A and C) and T15-T16 (row B). Position 2A and 2B use the same template as used in Figure 2. Positions 7A and 7B are no template controls and positions 11A and 11B are a mouse caecal DNA negative control. Note that the T7-T8 primer pair amplifies a band in a series of clinical sample DNAs derived from caecal samples (C6 to C10) but not in the intestinal sample DNAs from the same birds (C1 to C5).

DISCUSSION

OTU 47, which had been tentatively identified as *H. pullorum*, has been identified as the potential pathogen responsible for the development of Spotty Liver Disease. Culturing of the candidate organism has proving difficult so alternative approaches to advance our knowledge of the organism have been sought. Metagenomic analysis has been carried out to generate genomic sequence information beyond the 16S rRNA gene sequence originally used to identify the candidate organism. Because there is only low level coverage of the candidate organism, OTU 47, in the metagenomic data only fragmentary genomic information has so far been defined. It is clear from the fragmentary genomic data that the candidate organism, although probably a helicobacter, is not currently represented in the publically available sequence databases and is likely to be distinct from *H. pullorum*. Because of this clear separation from *H. pullorum* it was inadvisable to use the currently published PCR assay (González et al., 2008) for further analysis of clinical samples. The fragmentary genomic sequence data was sufficient to allow the design of new PCR assays which, in initial work, appear to be specific for the target bacterium.

FUTURE WORK

There is a clear path forward for future work with two key areas of activity – the first directed towards more confidently establishing the identity of the pathogen via survey work of clinical material from disease outbreaks and the second aimed at culturing the candidate organism and then going on to experimentally reproduce the disease.

The new PCR assays can now be used to interrogate clinical material from cases of spotty liver disease, with several goals in mind.

- Firstly, to test how strong the correlation is between disease and presence of the candidate organism.

- Secondly, to identify new clinical samples, which carry a higher load of OTU 47, which could be subjected to metagenomic sequencing in order to achieve a more complete characterisation of the genome.
- Thirdly, the PCR could be used to monitor OTU 47 growth in enrichment cultures – it has proven difficult to isolate potential colonies of OTU 47 in a single step primary plating procedure but it might be possible to enrich in liquid culture with the assistance of the PCR assays.

From the current analysis only fragmentary pieces of the OTU 47 genome have been recovered from the metagenomic sequencing. It would be valuable to analyse the existing data in more detail and also generate more complete information from new sequencing efforts. A good draft genome sequence of the organism would allow an analysis of its biochemical potential and may reveal ways in which it could be cultured. Successful culturing of the organism is likely to give a much higher probability of experimentally reproducing the disease. A draft genome would also facilitate a search to discover key virulence factors, including extracellular toxins, which might be responsible for disease pathogenesis. Even without successfully culturing the organism it may be possible to design vaccines based on such virulence factors but it would be difficult to test experimental vaccines without having also developed an in vivo disease induction model to reproduce the disease.

REFERENCES

Atabay HI, Corry JEL, On SLW (1998) Identification of unusual Campylobacter-like isolates from poultry products as *Helicobacter pullorum*. *J Appl Microbiol* 84:1017–1024.

Basaran Kahraman B, Ak S (2013) Investigation of *Helicobacter pullorum* occurrence in chicken in the Marmara region of Turkey. *Istanbul Univ Vet Fak Derg* 39:63-66.

Burnens AP, Stanley J, Nicolet J (1996) Possible association of *Helicobacter pullorum* with lesions of vibrionic hepatitis in poultry. In: Newell DG, Keatly JM, Feldman RA (Eds) *Campylobacters, Helicobacters* and Related Organisms. Plenum Press, New York, pp 291-293.

Ceelen LM, Decostere A, Van den Bulck K, On SLW, Baele M, Ducatelle R, Haesebrouck F (2006a) *Helicobacter pullorum* in Chickens, Belgium. *Emerg Infect Dis* 12:263–267.

Ceelen LM, Haesebrouck F, Favoreel H, Ducatelle R, Decostere A (2006b) The cytolethal distending toxin among *Helicobacter pullorum* strains from human and poultry origin. *Vet Microbiol* 113:45–53.

Ceelen LM, Decostere A, Chiers K, Ducatelle R, Maes D, Haesebrouck F (2007) Pathogenesis of *Helicobacter pullorum* infections in broilers. *Internat J Food Microbiol* 116:207–213.

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harbor Protocols* 2010: pdb.prot5368.

González A, Piqueres P, Moreno Y, Cañigral I, Owen RJ, Hernández J, Ferrús MA (2008) A novel real-time PCR assay for the detection of *Helicobacter pullorum*-like organisms in chicken products. *Int Microbiol* 11:203–208.

Manfreda G, Parisi A, Lucchi A, Zanoni RG, De Cesare A (2011) Prevalence of *Helicobacter pullorum* in Conventional, Organic, and Free-Range Broilers and Typing of Isolates. *Appl Environ Microbiol* **77**:479–484.

Neubauer C, Hess M (2006) Tissue tropism of *Helicobacter pullorum* in specific pathogen-free chickens determined by culture and nucleic acid detection. *Avian Dis*. 50:620–623.

Stanley J, Linton D, Burnens AP, Dewhirst FE, On SLW, Porter A, Owen RJ, Costas M (1994) *Helicobacter pullorum* sp. nov. - genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. *Microbiol* 140:3441–3449.

Wai SS, Saleha AA, Zunita Z, Hassan L, Jalila A (2012) Occurrence of co-infection of *Helicobacter pullorum* and *Campylobacter* spp. in broiler and village (indigenous) chickens. *Pak Vet J* 32:503-506.

Young VB, Chien C-C, Knox KA, Taylor NS, Schauer DB, Fox JG (2000) Cytolethal Distending Toxin in Avian and Human Isolates of *Helicobacter pullorum*. *J Infect Dis* 182:620–623.

Zanoni RG, Rossi M, Giacomucci D, Sanguinetti V, Manfreda G (2007) Occurrence and antibiotic susceptibility of *Helicobacter pullorum* from broiler chickens and commercial laying hens in Italy. *Internat J Food Microbiol* 116:168–173.

Zanoni RG, Piva S, Rossi M, Pasquali F, Lucchi A, De Cesare A, Manfreda G (2011) Occurrence of *Helicobacter pullorum* in turkeys. *Vet Microbiol* 149:492–496.