Introduction

Understanding the enzymology of plant biomass conversion is a key issue in the world’s desire to establish a sustainable bio-based economy. Whilst current available enzyme technology is insufficiently effective, many free-living organisms readily deconstruct plant biomass by enzyme-driven hydrolysis to take advantage of this material as a nutrient source. In particular, obligate herbivores have evolved to maintain a symbiotic relationship with a specialized consortium of gut microbes (microbiomes) that underpins lignocellulose deconstruction. Current paradigms for microbial lignocellulose degradation in gut microbiomes are centered on well-known key cellulolytic enzymes (i.e., GH5, GH6, GH7, GH9, and GH48), see [1] for enzyme classification) and multi-enzyme cellulosome complexes [2].

Accumulating knowledge indicates that these paradigms are not unique and that nature has additional, yet poorly understood tools to accomplish lignocellulose degradation. For instance, the gut bacterium *Fibrobacter succinogenes* has long been known to degrade crystalline cellulose and other plant structural polysaccharides at a rate exceeding that of most other microorganisms. However, genome studies indicate this bacterium lacks both GH6, processive GH9, and GH48 representatives and cellulosome structures [3]. The absence or poor representation of key enzymes and cellulosomes in cellulosytic gut microbiome communities has been further exemplified by recent metagenome sequencing projects in the hindgut of termites [4], the foregut of marsupials [5] and the rumen of cows [6,7]. Here we have explored this further by studying the rumen microbiome of the Svalbard reindeer, a herbivore whose extreme habitat and diet are very different compared to the herbivores studied so far.

*Svalbard reindeer* (*Rangifer tarandus platyrhynchus*) live under austere nutritional conditions on the high-arctic archipelago of Svalbard (74–80°N lat.), where snow and ice cover most

*E-mail: phil.pope@umb.no*
vegetation for more than eight months of the year. In winter time, the reindeer feed on poor quality forages that are high in lignocellulose. Body reserves are not sufficient for winter survival, and it has been estimated that only 10–30% of the daily energy expenditure during the dark part of winter can be covered by mobilization of fat [8]. Therefore, optimal utilization of the feed during winter time is crucial for survival and metabolic capabilities of the rumen microbiome are likely to play a central role. Indeed, studies on cultivable members of the microbiome showed high levels of bacteria capable of degrading various forms of cellulose and heteroxylans [9], whilst a limited cultivation-independent study suggested that the rumen microbiomes of Svalbard reindeer are dominated by novel species [9].

Svalbard reindeer research to date indicates the presence of a microbiome with high microbial diversity, powerful functional capabilities in terms of cellulose degradation, and a considerable degree of novelty. We present here a compositional and comparative analysis of metagenomic data for the rumen microbiome of Svalbard reindeer during winter. Novel bacterial lineages were identified and nucleotide composition-based sequence binning using PhyloPythiaS [10] facilitated the production of a 0.9 Mb assemblage of DNA representing one of the novel Bacteroidales clades numerically dominant in this community. Further in silico analysis revealed the presence of polysaccharide utilization loci-like systems (PULs) that resemble multiprotein starch utilization systems (Sus) [11] as well as the presence of carbohydrate-active enzymes targeting a broad spectrum of polysaccharides, in both this clade and the unassembled metagenome. Additional functional screens of fosmid libraries and data-mining of existing metagenome datasets revealed cellulose-degrading loci and PUL-like systems at an exceptionally high frequency in the Svalbard reindeer rumen as well as other herbivore gut environments [7,12].

Results and Discussion

Metagenomic Sequence Generation

Total DNA was extracted from rumen samples collected from two Svalbard reindeer (SR1 and SR2) feeding on natural winter pasture, when the percentage of fibre degrading bacteria is believed to be at its maximum [8,9]. A 454 pyrosequencing scheme with four technical replicate PCR reactions (each with a unique barcode) for each animal (2 animals × 4 replicates; 8 samples in total), was used to obtain ~80,000 sequence reads from PCR-amplified V1–V3 regions of bacterial 16S rRNA genes (average read length ~490 nt). In order to eliminate noise introduced during PCR and sequencing of 16S rRNA genes, operational taxonomic units (OTUs: defined using a 97% sequence identity threshold) were only included in the community analysis if their representatives were found in at least four of the eight samples.

To facilitate access to the metagenome and subsequent descriptions of microbial community function, approximately 503 Mb of raw single- and paired-end shotgun reads were generated from pooled rumen microbiome community DNA (2 animals). A proportion of the reads could be assembled into contigs greater than 500 nt (a total of 32,073 contigs, 26 Mb in total) with the largest contig being 17,446 nt. Paired-end sequences were used to construct 1364 scaffolds (scaffolds represent one or more contigs ordered and oriented using paired-end reads) for taxonomic binning using PhyloPythiaS (avg 3986 nt, largest 46,178 nt; 5.44 Mb in total). Another 1.2 Mb of metagenomic DNA sequence was obtained by assembling and manual editing of the sequences of selected fosmids.

Microbial Community Composition and the Dominance of the Bacteroidetes

The microbial community structure was determined using both 16S rRNA gene amplicon pyrosequencing and sequence-composition binning (PhyloPythiaS) of scaffolds assembled from metagenomic sequences (Figure 1). Using 16S rRNA gene analysis we determined there was little variation between the two animals used for this study: 90.4% and 91.5% of OTUs found in the rumen communities of animals SR1 and SR2, respectively, were shared between the two samples (Figure 1, Table S1). There were only a few instances of large differences in OTU relative abundance between the two samples (Table S1).

Comparison of OTUs against the Ribosomal Database Project [13] revealed that the Bacteroidetes and Firmicutes were largely predominant constituting 61% and 27% of the gene amplicon sequences respectively (Figure 1a). Dominance of these phyla is commonly observed in gut microbiomes. However, at an OTU-level as many as 63% of the OTUs could not be classified to a Genus-rank (Table S1), whilst the 20 most abundant OTUs demonstrated low sequence identities with cultured representatives (Figure S1). This suggests that the OTUs generally are only distantly related to any of the cultivated species from other gut environments. Furthermore, OTU-level comparisons using network maps showed a limited degree of shared OTUs between the Svalbard reindeer and other foregut-digesting herbivores, including the Norwegian reindeer (Rangifer tarandus tarandus) fed a commercial pelleted feed (Figure 1b). Community-level comparisons with Unifrac, which analyze phylogenetic lineages and not just shared OTUs, showed that the microbiome of the Svalbard reindeer rumen was more similar to that of the Norwegian reindeer than to those from other herbivores (Figure 1c). A small core-set of six OTUs was present in all ruminants (both reindeer species and cow), as well as two more that were present in all foregut digesters (macropods included). Interestingly these eight OTUs were all affiliated to either Ruminococcaceae or Lachnospiraceae lineages (Firmicutes) (Figure 1b, Table S1). Recent broad-scale phylogenetic analysis of the human gastrointestinal microbe has identified highly prevalent core phylogroups belonging to the Lachnospiraceae [14], and evidence presented here suggests that such groups exist for foregut digesters as well. Most OTUs in the Svalbard reindeer rumen were present at low abundance, however several deeply branched unique members of the Bacteroidales (Bacteroidetes) were found in high abundance (Table S1). One OTU in particular, hereafter referred to as SRM-1, comprised 11% of the total 16S rRNA gene dataset. The SRM-1 OTU was found in both reindeer species only (Figure 1b) and has never before been reported. It demonstrates only 91% sequence identity to its closest cultured relative, Bacteroidales genosp. P1.

Taxonomic analysis of the metagenome sequences by sequence-composition binning of assembled scaffolds with PhyloPythiaS revealed a similar community structure to that observed when using 16S rRNA gene amplicon pyrosequencing analysis, with ~75% of scaffolds assigned to the Bacteroidetes and Firmicutes (Figure 1d, Table S2). The dominance of SRM-1 was confirmed with over 880 kb (~16%) of scaffold sequence being assigned to this OTU. Noteworthy contrasts between the two approaches included: (1) the absence of metagenomic scaffolds binned to the phylum Chloroflexi, which had several numerically abundant OTUs in the 16S rRNA gene amplicon pyrosequencing analysis (Figure 1a, Figure S1), and (2) the higher percentage of
metagenomic scaffolds that were assigned as “other”. We hypothesize that due to the deep-branching nature of the Chloroflexi OTUs (the closest cultured relative has 90–91% 16S sequence identity) and an under-representation of Chloroflexi genetic information available for PhyloPythiaS training (Table S3), metagenomic scaffolds originating from Chloroflexi species were not detected and instead assigned as “others” (assignments not extending deeper than Domain-rank). This could explain both inconsistencies, at least in part.

Overall, both analysis yielded a considerable proportion of shallow assignments, with 60% of the PhyloPythiaS assignments (Table S2) and 53% of the OTU lineages (Table S1) not extending deeper than an Order-rank of classification. This confirms that the Svalbard rumen microbiome consists of unique bacterial lineages, with limited similarity to published sequence data from other organisms and/or environments.

Plant Polysaccharide Degradative enzymes in the Svalbard Reindeer Rumen Microbiome and the Dominant, Novel and Saccharolytic Bacteroidales Clade SRM-1

In order to characterise the biomass-degrading capabilities of the Svalbard reindeer rumen microbiome gene-centric metagenomic datasets were constructed and annotation efforts focused on...
enzymes were identified using pfamHMMs and grouped accord-
and scaffolds (for SRM-1; Table S4) were subjected to automated
Both filtered unassembled reads (Table 1) and assembled contigs
profiles of hemicellulases and non-cellulase polysaccharide degrad-
Some complexes occur in the microbiome of ruminants [16],
Predicted abundant enzyme activities include hydrolysis of
galacturonans (GH28) and rhamnoses (GH76) commonly found
Having observed the abundance of Sus-like proteins in the
suggested on the basis of the metagenome described for the
Interestingly, Hess et al. showed that both these cellulase genes
Interestingly, Bacteroides thetaiotaomicron [21]. SusC-
Their involve-
functional Screens and Gene-mining Linking Cellulose
Degradation to the Bacteroidales and PULs
The frequency of fosmid clones testing positive for CMCase
activity was the highest reported so far (48 positive from ~3,000
screened, corresponding to ~1% of the clones and a hit rate of 1
per 3.5 Mb screened; [20]). The majority of the sequenced fosmid
clones were predicted to be derived from Bacteroidetes-affiliated
lineages (Table 2). More than 70% of the scaffolds constructed
carry one or more genes encoding a presumptive GH5
endoglucanase. However, two scaffolds (Sc00001 & ScSc00021)
did not possess any GH families representing known endoglucan-
ases, suggesting members of the Bacteroidiales might also possess
novel cellulose-degrading mechanisms. Further studies, e.g.
employing random transposon mutagenesis, are needed to identify
exactly which genes are responsible for the enzymatic activity
encoded by these fonsmids. The presumptive GH5 endoglucanase
genes were most often located in close proximity to multi-gene
PULs which were all found on fosmids assigned to Bacteroidetes,
including one fosmid assigned to SRM-1 (see Figure 2, Table 2).
Their gene-organisation included Sus-like outer membrane
proteins homologous to SusC and SusD, which are essential for
the import and degradation of starch by the Sus initially described
in the human gut bacterium Bacteroides thetaiotaomicron [21]. SusC-
and SusD-like genes which typify a Sus-like PUL, function
together with other hypothetical outer-membrane proteins (Sus/E-
F-like) and linked carbohydrate-active enzymes [11]. Whilst
originally characterised on starch, Sus-like PULs encoded in
human gut Bacteroidetes have since been described with
capabilities to degrade other plant polysaccharides that include
pectins and hemicellulosic substrates [22]. Predictions of polysac-
charide degradation by Sus-like PULs in other environments have
previously been made [11,23], but so far there are only few studies
addressing their importance in the herbivore gut. Their involve-
ment in xylan degradation by the rumen bacterium Prevotella
bryantii has been illustrated by gene transcriptome studies [23],
whereas their potential role in cellulose degradation has been
suggested on the basis of the metagenome described for the
anaerobic microbiome of the Tammar wallaby foregut [3].
Having observed the abundance of Sus-like proteins in the
Svalbard reindeer microbiome, we re-evaluated existing herbivore
metagenomic data. Cellulosytic screens using fosmids constructed
from the metagenome of the buffalo rumen microbiome have
previously identified GH5 endoglucanases affiliated to the
Bacteroidetes [12]; closer examination of flanking genetic regions
revealed the presence of Sus-like PULs (Figure 2). Gene-mining
within the rumen metagenome constructed by Hess et al. [7],
revealed a hitherto non-detected abundance of Sus-like genes
(Table 1) as well as a cellulase-linked PUL encoded within the
partial genome assembled for the as-yet uncultured Bacteroidetes
phylotype AC2a (Figure 2, Table S5). The cellulases (GH3 and
GH9) and a GH94 cellulose phosphorylase encoded within the
AC2a PUL suggest activity against cellulose substrates (Figure 2).
Interestingly, Hess et al. showed that both these cellulase genes
expressed proteins with enzymatic activity against either CMC or
avice and pretreated Miscanthus substrates [7] (Table S5). Overall
these findings suggest that PULs including those linked with
cellulases are widespread in herbivore gut environments and may
play a major role with respect to plant biomass degradation,
similar to what has been demonstrated in the human distal gut.
Taken together the accumulated data also suggests involvement
of Sus-like proteins in cellulose degradation, a process which is not
predominant in the human gut but essential in herbivores.
| Predominant activity                      | Macropod | Termite | Bovine | Bovine | Reindeer |
|------------------------------------------|----------|---------|--------|--------|----------|
| **Cellulases**                           |          |         |        |        |          |
| GH5 cellulases                           | 10       | 56      | 8      | 1451   | 287      |
| GH6 endoglucanases                       | 0        | 0       | 0      | 0      | 0        |
| GH7 endoglucanases                       | 0        | 0       | 0      | 1      | 0        |
| GH9 endoglucanases                       | 0        | 9       | 6      | 795    | 109      |
| GH44 endoglucanases                      | 0        | 6       | 0      | 99     | 5        |
| GH45 endoglucanases                      | 0        | 4       | 0      | 115    | 0        |
| GH48 cellobiohydrolases                  | 0        | 0       | 0      | 3      | 5        |
| **Total**                                | 10 (4)   | 75 (20) | 14 (2) | 2464 (13) | 406 (8) |
| **Endohemicellulases**                   |          |         |        |        |          |
| GH8 endoxylanases                        | 1        | 5       | 4      | 329    | 35       |
| GH10 endo-1,4-β-xylanases                | 11       | 46      | 7      | 1025   | 190      |
| GH11 xylanases                           | 0        | 0       | 0      | 0      | 0        |
| GH26 β-mannanase & xylanases             | 5        | 15      | 5      | 369    | 153      |
| GH28 galacturonases                      | 2        | 6       | 5      | 472    | 120      |
| GH53 endo-1,4-β-galactanases             | 9        | 12      | 17     | 483    | 125      |
| **Total**                                | 28 (10)  | 98 (26) | 39 (6) | 2843 (15) | 631 (12) |
| **Xyloglucanases**                       |          |         |        |        |          |
| GH16 xyloglucanases                      | 4        | 1       | 1      | 483    | 116      |
| GH74 xyloglucanases                      | 1        | 7       | 0      | 385    | 44       |
| **Total**                                | 5 (2)    | 8 (2)   | 1 (0)  | 868    | 160 (3)  |
| **Debranching enzymes**                  |          |         |        |        |          |
| GH51 α-L-arabinofuranosidases            | 12       | 18      | 64     | 1249   | 488      |
| GH54 α-L-arabinofuranosidases            | 0        | 0       | 1      | 76     | 23       |
| GH62 α-L-arabinofuranosidases            | 0        | 0       | 0      | 1      | 0        |
| GH67 α-glucuronidases                    | 5        | 10      | 0      | 120    | 74       |
| GH78 α-L-rhamnosidases                   | 25       | 0       | 34     | 1260   | 313      |
| **Total**                                | 42 (15)  | 18 (5)  | 99 (14) | 2706 (15) | 898 (17) |
| **Oligosaccharide-degrading enzymes**    |          |         |        |        |          |
| GH1 β-glucosidases                       | 61       | 22      | 10     | 253    | 122      |
| GH2 β-galactosidases                     | 24       | 23      | 186    | 1436   | 716      |
| GH3 β-glucosidases                       | 72       | 69      | 176    | 2844   | 844      |
| GH29 α-L-fucosidases                     | 2        | 0       | 74     | 939    | 268      |
| GH35 β-galactosidases                    | 3        | 3       | 12     | 158    | 39       |
| GH38 α-mannosidases                      | 3        | 11      | 17     | 272    | 116      |
| GH39 β-xylidoses                         | 1        | 3       | 2      | 315    | 76       |
| GH42 β-galactosidases                    | 8        | 24      | 11     | 374    | 95       |
| GH43 arabinobio/xylidoses                | 10       | 16      | 61     | 2932   | 787      |
| GH52 β-xylidoses                         | 0        | 3       | 0      | 1      | 2        |
| **Total**                                | 184 (70) | 174 (47) | 549 (78) | 9524 (52) | 3065 (60) |
| **Other domains associated with GHs**     |          |         |        |        |          |
| Cohesin                                  | 0        | 0       | 0      | 80     | 52       |
| Dockerin                                 | 41       | 0       | 8      | 188    | 92       |
| SusC                                     | 36       | 0       | 9      | 3110   | 1122     |
| SusD                                     | 42       | 0       | 11     | 1889   | 685      |
| **Metagenome size**                      | 0.054 Gb | 0.062 Gb | 0.026 Gb | 268 Gb | 0.30 Gb |

*GHs are grouped according to their major functional role in the degradation of plant fiber. The numbers in parentheses represent the percentages of these groups relative to the total number of GHs presented in this table. In addition to GHs, the table shows data for four other selected proteins possibly involved in biomass turnover; see main text for details. Note that some of the GH profiles were derived from contigs rather than (filtered) unassembled reads, explaining part of the differences in absolute gene numbers.

Data calculated in this study using searches against the rumen metagenome dataset [7] using SusC (PF00593) and SusD (PF07980) HMMs. Pfam HMMs not available for CAZy families; data generated by dbCAN (http://csbl.bmb.uga.edu/dbCANdev/index.php).

doi:10.1371/journal.pone.0038571.t001
| Scaffold | GH2 | GH3 | GH5 | GH9 | GH10 | GH16 | GH20 | GH23 | GH26 | GH28 | GH31 | GH33 | GH35 | GH43 | GH51 | GH53 | GH55 | GH74 | GH76 | GH78 | GH94 | GH109 | GH117 | PL1 | PL22 | CE1 | CE7 | CE11 | susD | susEF | BinningSb. | Substrate. | SusC |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Sc00001 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00002 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00003 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00004 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00005 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00006 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00007 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00008 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00009 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00010 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00011 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00012 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00013 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00014 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00015 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00016 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00017 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00018 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00019 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00020 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00021 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00022 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00023 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00024 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00025 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00026 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00027 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00028 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sc00029 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

**Notes:** Carbohydrate binding modules (CBM or BACON) are indicated in superscript text. Note that Sc00001 and Sc00021 do not contain any gene(s) coding for proteins known to be involved in the degradation of cellulose. Partial scaffolds not encoding documented cellulose active genes are excluded (n = 13).

**doi:** 10.1371/journal.pone.0038571.t002
Figure 2. Comparative gene organization of presumptive cellulolytic PULs identified in the Svalbard reindeer rumen and other gut environments.

Gene clusters were recovered from the cow rumen metagenome (partial genome of an as-yet uncultured Bacteroidales phylotype AC2a) and sequenced cellulolytic fosmids constructed from environmental DNA originating from the gut microbiomes of the wallaby foregut and the reindeer and buffalo rumen (upper four clusters). All fosmid sequences originate from Bacteroidetes, according to PhyloPythiaS binning, and for all cellulolytic activity has been detected in functional screens. Green genes represent SusE/SusF-like genes predicted to encode outer-membrane proteins whose function is currently unknown. Black genes encode putative response-regulators. BACON: indicates a carbohydrate binding domain linked PULs in the cow and buffalo rumen and macropod foregut included one linked to SRM-1. The finding of similar cellulase-affiliated to the abundant SRM-1 clade that lacked known endoglucanases, inferring the possibility of novel cellulolytic mechanisms that are yet to be characterised. This approach also identified Bacteroidetes-affiliated GH5-linked PULs, which included one linked to SRM-1. The finding of similar cellulase-linked PULs in the cow and buffalo rumen and macropod foregut microbiomes adds further weight to the hypothesis that these structures perhaps represent a key adaptation to growth on cellulose by Bacteroidetes species. It would seem that these PULs and their constituent genes present important targets for further research, possibly representing completely novel mechanisms for enzymatic cellulose conversion in anaerobic gut environments.

Concluding Remarks

The high-Arctic Svalbard reindeer survive under austere nutritional conditions relying on the ability of the rumen microbiome to digest poor quality moses and fibrous plants available in winter. The metagenomic analyses in this study illustrate that the Svalbard reindeer rumen is host to novel and numerically dominant bacterial lineages. The deeply branched and numerically dominant SRM-1 lineage is predicted to play a key role in the deconstruction of plant biomass by producing an array of glycoside hydrolases targeting cellulose, hemicelluloses, pectin and other oligosaccharides. Functional screening and sequencing of fosmid libraries revealed CMCase positive clones affiliated to the abundant SRM-1 clade that lacked known endoglucanases, inferring the possibility of novel cellulolytic mechanisms that are yet to be characterised. This approach also identified Bacteroidetes-affiliated GH5-linked PULs, which included one linked to SRM-1. The finding of similar cellulase-linked PULs in the cow and buffalo rumen and macropod foregut microbiomes adds further weight to the hypothesis that these structures perhaps represent a key adaptation to growth on cellulose by Bacteroidetes species. It would seem that these PULs and their constituent genes present important targets for further research, possibly representing completely novel mechanisms for enzymatic cellulose conversion in anaerobic gut environments.

Materials and Methods

Reindeer Sampling and Ethics Statement

Rumen contents were sampled from two adult, female Svalbard reindeer (Rangifer tarandus platyrhynchus) (SR1 and SR2) aged between 1.5 and 4.5 years and grazing on their natural winter pastures in Bjørndalen near Longyearbyen, Svalbard (Norway) on January 31st 2010. The Svalbard Environmental Protection Fund “Svalbard miljøvernfond,” Acting Environment Manager Per Kyrre Reymert and Conservation Advisor Tor Punsvik, approved the use of Svalbard Reindeer in this study (Permission reference number: 2009/00420-4). The Svalbard Environmental Protection Fund is the competent authority for sampling from wild animal populations. The reindeer were euthanized by an experienced hunter by shooting followed by exsanguination, which is an approved procedure under the Norwegian Animal Welfare Act. Since the reindeer were euthanized for the post mortem collection of samples, the procedure did not require ethical approval from the Norwegian Animal Research Authority (NARA).

In addition, for comparative rs analysis, two adult Norwegian reindeer (Rangifer tarandus tarandus) (NR1 and NR2) fed a commercially available pelleted concentrate feed for reindeer (RF-80; Felleskjøpet, Norway), were sampled from the domestic herd at the University of Tromso, Norway (December 2009). NR1 and NR2 were purchased from local reindeer herders and maintained in the animal research facility at the University of Tromso. This facility (and staff) has been inspected and approved by NARA, and therefore fulfills current animal welfare criteria. The reindeer were euthanized by trained personnel by stunning followed by exsanguination, which is an approved procedure under the Norwegian Animal Welfare Act.

Rumen contents were transferred immediately after slaughter to sterile containers and frozen at −80°C.

Cell Dissociation and DNA Extraction

Cell dissociation from plant material and DNA extraction were performed on individual samples (SR1, SR2, NR1 and NR2) and pooled Svalbard reindeer samples that contained equal amounts of SR1 and SR2 material. To desorb and recover microbes adhered to plant biomass 5–10 g of the pooled samples were centrifuged at 14 000 rpm for 2 minutes, and the pellets were resuspended in
dissociation buffer and subjected to a dissociation and DNA extraction procedure described by Rosewarne et al. [24].

16S rRNA Gene Amplicon Sequencing

Bacterial 16S genes were amplified from the individual NR1, NR2, SR1 and SR2 metagenomic DNA samples using the forward primer (5’- CCA TGG TGG CAG TCT CAG CAA CAG CTA GAG TTT GAT CCT GG -3’), which contained the 454 Life Sciences primer B sequence and the broad- conserved bacterial primer 27F, and the reverse primer (5’-CCA TCT CAT CCC TGC GTG TCT CGG ACT CAG NNN NNN NNT TAG CGC GGC TGG T -3’), which contained the 454 Life Sciences primer A sequence, the broadly-conserved bacterial primer 515R and a unique 8-nt multiplex identifier (MID) used to tag each amplicon (designated by NNNNNNNN) [25]. Four technical replicates PCR reactions (each with a unique barcode) were performed for each DNA sample. Pyrosequencing of m gene amplicons was performed on the 454 Genome Sequencer FLX-Titanium system according to manufacturer’s instructions (454 Life Sciences). Signal processing and base calling were performed using the bundled 454 Data Analysis Software version 2.3.

Phylogenetic Analysis of 16S rRNA Gene Sequences

Rrs gene sequences were processed using the QIIME software package [26] and removed from the analysis if they were <350 or >550 nt in length, contained ambiguous bases, had a mean quality score <25, contained a homopolymer run exceeding 6 nt, or did not contain a primer or barcode sequence. Similar sequences were clustered into operational taxonomic units (OTUs) using UCLUST software [27] and a 97% sequence identity threshold. To eliminate noise and possible artifacts introduced during PCR and sequencing, OTUs were filtered so that only those that contained representatives from a minimum of 4 samples were used, and the most abundant sequence in each OTU was chosen as the representative sequence. As an added precaution chimeras were removed from the representative set using Chimera Slayer as previous work suggests that chimera formation is reproducible across technical replicates [28]. Representative sequences (accession numbers JN802705 - JN803885, SRM-1: JN802985) were aligned against the Greengenes core set [29] using PyNAST software [30] with a minimum alignment length of 150 and a minimum identity of 75%. Taxonomy was assigned to each OTU using the Ribosomal Database Project (RDP) classifier [31] with a minimum support threshold of 80% and using the RDP taxonomic nomenclature. The alignments were then filtered to remove gaps and hypervariable regions using a Lane mask [26], and an maximum-likelihood tree was constructed from the filtered alignment using FastTree [32]. Prior to comparison of reinder rrs gene sequences with wallaby, rumen and termite samples, each rrs dataset was randomly “subsampled” using QIME to normalize each dataset and remove sample heterogeneity. An unweighted UniFrac distance matrix [32] was constructed from the phylogenetic tree and visualised using principal coordinates analysis. The OTU network maps were generated using QIIME and visualised with Cytoscape [33].

Metagene Processing: Shotgun Library Preparation, Sequencing and Assembly

Shotgun sequencing runs were performed on libraries prepared from pooled Svalbard reindeer rumen community DNA using the 454 Genome Sequencer FLX-Titanium single- and paired-end protocols (total 1,453,100 reads, 503 Mb). Sequencing reads were assembled using Newbler (GSassembler v. 2.3) resulting in 32,073 contigs ≥500 nt, totalling 26 Mb and 1,459 scaffolds totalling 5.44 Mb. Due to low assembly (334,500 out of 1,453,100: -23%), unassembled single-end reads were used for community GH profile analysis. Importantly this approach also ensured that differences in species abundance distribution were incorporated (i.e. dominant populations producing multiple hits to the same gene will be weighted in the analysis). Unassembled sequencing reads with degenerate bases (“Ns”) were removed along with all replicate sequences that were detected using the following parameters: 0.9 (90% ID), length difference requirement = 0 and 3 beginning bases checked [34]. A total of 605,636 (300 Mb) reads passed this quality filtering. The raw sequencing reads and the assembled metagenome dataset have been deposited at the NCBI Short Read Archive under BioProject ID PRJNA73677 and accession number SRA046345.1.

Fosmid Construction, Screening, Sequencing and Assembly

A 36 kb insert fosmid library was cloned in pCC1Fos (Epiprocite Corp.), using previously described methods [24]. Fosmid clones bearing endoglucanase and/or xylanase activity were detected by plating the E. coli library on LB-chloramphenicol agar plates containing either 0.2% (w/v) carboxymethylcellulose or 0.2% (w/v) birchwood xylan (Sigma). Recombinant strains were plated and incubated overnight at 37°C. The plates were then stained with Congo red dye and de-stained with 1M NaCl to reveal zones of hydrolysis. Positive colonies were isolated and reexamined to confirm activity. This approach yielded 48 fosmid clones positive for carboxymethylcellulase hydrolysis (~5,000 screened) and three positive for xylan hydrolysis (~1,000 screened). These 51 clones were subsequently sequenced and assembled. Fosmid copy numbers were enhanced using Epicentre protocols, and the fosmid DNA was purified using Qiagen MiniPrep columns. Equimolar amounts of the fosmids were pooled together (~20 µg total DNA) and both a 3 kb paired-end library and a 454 standard shotgun library were constructed. Both libraries were directly sequenced with the 454 Life Sciences Genome Sequencer GS FLX and assembled into 40 scaffolds using Newbler. Redundancy was observed within the assembly with several fosmids overlapping at least one other fosmid, however no instances of fosmid replicates were observed. For nine fosmids, insert sequences were completely assembled, with no gaps. In total, 1.2 Mb of metagenomic DNA sequence was assembled and manually edited.

Gene Annotation and Phylogenetic Analysis

Three metagenomic datasets ( unassembled single-end reads, assembled contigs ≥500 nt, and assembled fosmids) from the Svalbard reindeer rumen microbiome were annotated via the IMG/M-ER annotation pipeline and loaded as independent data sets into IMG/M-ER [15] (http://img.jgi-doe.gov/cgi-bin/m/main.cgi), a data-management and analysis platform for genomic and metagenomic data based on IMG [33]. Complete annotated data for the fosmid scaffolds, assembled contigs and unassembled reads can also be accessed through the IMG/M-ER (http://img.jgi-doe.gov) under Taxon Object ID 2199532020, 2081372005 and 2088090000 respectively. Putative genes were called with a combination of GeneMark.hmm for Prokaryotes (v. 2.4) [36], MetaGene [37], Prodigal (v2.00) [38] and multiBLASTx. Datasets from the Hess et al., rumen metagenome [7] were downloaded from ftp://ftp.jgi-psf.org/pub/rnd2/Cow_Rumen/. This included genome bins for as-yet uncultured bacteria (ftp://ftp.jgi-psf.org/pub/rnd2/Cow_Rumen/cow_rumen_genome_bins.tar.gz) and putative genes for the total dataset that included genome bins
as well as assembled metagenome contigs (ftp://ftp.jgi-psf.org/pub/rnd2/Cow_Rumen/metagenemark_predictions.faa.gz).

Searches for glycoside hydrolases (GHs) of selected functional classes (e.g. cellulases, hemicellulases, debranching enzymes, "other") were performed with pfam HMMs (Pfam version 24.0 and HMMER v3.0), named in accordance with the CAZy nomenclature scheme [1]. The specific cut-off was set to Gathering Threshold (HMMER). For those GH families for which there is currently no representation in Pfam, HMMs were generated using hmmbuild (HMMER). For those GH families for which there is currently no representation in Pfam, HMMs were generated using hmmbuild (HMMER) and multiple sequence alignments of representative sequences selected from the CAZy database.

**Binning**

Assembled metagenomic contigs, scaffolds and fosmids were binned (classified) using Phylotyping [10], a kmer-based taxonomic classifier. The classifier was trained to include clades at the taxonomic ranks of domain, phylum, class, order and family and the clade "uncultured Bacteroidales bacterium" (SRM-1). The models include all clades covered by two or more species at the corresponding ranks among 2193 sequenced microbial isolates and clade SRM-1 (Table S3). The classifier consists of an ensemble of six structural support vector machines (SVMs) models, created by using fragments of 1, 3, 5, 10, 15 and 50 kb in length, respectively, for training (see [10] for details). For SRM-1, thirteen assembled scaffolds (a total of 205,517 bp) were used for training, which were assigned unambiguously through a combination of high read coverage, consistent GC% and affiliation of selected phylogenetic marker genes [39] to the order Bacteroidales. Sample specific sequences for other clades were obtained by similarity searches and application of the lowest common ancestor algorithm on taxonomic identifiers of the best-scoring hits. Input fragments of a particular length were generated by using a sliding window with a step size of one-tenth of the generated fragment size (for example 5 kb for 50-kb fragments) on sample-derived sequences and a step size of generated fragment length on the isolate sequences. The classifier was then used to assign all assembled scaffolds and contigs larger than 500 bp. Results of this binning process were loaded into IMG/M-ER to allow independent analysis of the component populations.

**Supporting Information**

Table S1 Operational taxonomic units (OTU) representatives of 16S rRNA gene sequences obtained from the rumen microbiome of the Svalbard reindeer. * Hierarchical taxonomic assignment for each OTU calculated using the RDP naive Bayesian Classifier [13]. Lineages are displayed only where OTUs could be assigned with an 80% bootstrap confidence estimate. SR1 and SR2 indicate animal number and a-d indicate PCR replicates used for OTU filtering (see Materials and Methods). Rows highlighted in yellow indicate OTUs shared with PCR replicates used for OTU filtering (see Materials and Methods). Rows highlighted in yellow indicate OTUs shared with PCR replicates used for OTU filtering (see Materials and Methods).

**References**

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: 233–238.
2. Bayer EA, Lamed R, White BA, Flint HJ (2008) From cellulosomes to cellulosomics. *Chem. Rev.* 108: 364–377.
3. Suze G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, et al. (2011) The Complete Genome Sequence of *Fibrobacter succinogenes* S395 Reveals a Cellulolytic and Metabolic Specialist. *PLoS ONE* 6: e18014.
4. Warnecke F, Lugtenbel P, Ivanova N, Ghasemian M, Richardson TH, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450: 360–365.
5. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different to other herbivores. *Proc. Natl Acad. Sci. USA* 107: 14793–14798.
6. Bude JM, Antonopoulous DA, Berg Miller ME, Wilson MR, Yuanzall AG, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl Acad. Sci. USA* 106: 1948–1953.
7. Hess M, Scevola A, Egan R, Kim TW, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331: 463–467.
8. Orpin CG, Mathiesen SD, Greenwood Y, Blix AS (1985) Seasonal changes in the ruminal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Appl. Environ. Microbiol.* 50: 144–151.

**Table S2 Phylogenetic profile of the Svalbard reindeer rumen metagenome sequence dataset, based on sequence composition-based binning of assembled scaffolds using Phylotyping.** Values described in non-bold text represent sub-category counts for order/family lineages within the Bacteroidetes and Firmicutes.

**Table S3 Input clades, sample specific data and genomes/Whole Genome Shotgun submissions (WGS) used for Phylotyping training models.** * temporary ncbid for SRM-1.

**Table S4 Glycoside hydrolases and related proteins recovered from the putative partial SRM-1 genome.** * indicates best match; † indicates contig is linked to an uncultured Bacteroidales bacterium scaffold, based on Phylotyping analysis; +SP indicates signal peptide detected. Taxonomic assignment was predicted using GC%, high coverage (greater than 6x) and Phylotyping binning.

**Table S5 A cellulase-linked PUL encoded within the as-yet uncultured Bacteroidales phylotype AC2a genome bin, reconstructed from the rumen metagenome.** Gene ID's are in the following format: NODE_ORF (see [7]). All data downloaded from ftp://ftp.jgi-psf.org/pub/rnd2/Cow_Rumen/† Sample ID and Substrate are as in Figure 3 and Table S6 from Hess et al. [7].

**Figure S1 Relative abundance of the 20 most dominant bacterial taxa in the Svalbard reindeer rumen microbiome.** Percentages are calculated against the total number of 16S rRNA gene sequences recovered. The closest cultured relative of each OTU and the sequence similarity % ID is indicated in parentheses. The lineage of each OTU is indicated by colour of text: Bacteroidetes maroon, Chloroflexi blue and Firmicutes green.

**Acknowledgments**

We are especially grateful to the support from Karl-Arne Stokkan (University of Tromsø) and Tommy Frantzen (Svalbard) who assisted in sample collection, in addition to Sebastian Konietzny and Aaron Weimann (Heinrich-Heine University) who assisted in data annotation.

**Author Contributions**

Conceived and designed the experiments: PBP VGHE. Performed the experiments: PBP AKM IG WS. Analyzed the data: PBP IG. Contributed reagents/materials/analysis tools: MAS ACM MM VGHE. Wrote the paper: PBP MM VGHE.

**Table S2**

| Gene | Order | Family | Genus | Species | Bin Cluster |
|------|-------|--------|-------|---------|-------------|
| ... | ... | ... | ... | ... | ... |

**Table S3**

| Sample ID | Substrate | Clade |
|-----------|-----------|-------|
| ... | ... | ... |

**Table S4**

| Glycoside Hydrolase | Class | Function | Clade |
|---------------------|-------|----------|-------|
| ... | ... | ... | ... |

**Table S5**

| Genome Bin | Description |
|------------|-------------|
| ... | ... |

**Figure S1**

Relative abundance of the 20 most dominant bacterial taxa in the Svalbard reindeer rumen microbiome. Percentages are calculated against the total number of 16S rRNA gene sequences recovered. The closest cultured relative of each OTU and the sequence similarity % ID is indicated in parentheses. The lineage of each OTU is indicated by colour of text: Bacteroidetes maroon, Chloroflexi blue and Firmicutes green. (TIF)
9. Sundet MA, Praesteng KE, Cann IK, Mathiesen SD, Mackie RI (2007) Novel rumen bacterial diversity in two geographically separated sub-species of reindeer. *Microb. Ecol.* 54: 424–438.

10. Paul KK, Haider F, Pope PB, Turnbaugh PJ, Morrison M, et al. (2011) Taxonomic metagenome sequence assignment with structured output models. *Nat. Meth.* 8: 191–192.

11. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota. The bacteroidetes Sue-like paradigm. *J. Biol. Chem.* 284: 24673–24677.

12. Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, et al. (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* 107: 245–256.

13. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for tRNA analysis. *Nucleic Acids Res.* 37: D141–D145.

14. Sekelja M, Berget I, Næs T, Rudi K (2011) Unveiling an abundant core microbiota in the human adult colon by a phylogenetic-independent searching approach. *ISME J.* 5: 519–531.

15. Markowitz VM, Ivanova NN, Sato E, Palaniappan K, Chu K, et al. (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* 36: 534–538.

16. Ding SY, Rincon MT, Lamed R, Martin JC, McCrae SI, et al. (2001) Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* 183: 1945–1953.

17. Sørmo W, Haga ØE, Gaare E, Langvatn R, Mathiesen SD (1999) Forage chemistry and fermentation chambers in Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *J. Zool., Lond.* 247: 301–307.

18. Bååth RB, Selten J, Brul S (2006) Bioinformatic analysis of the microbial biofilms adherent to plant biomass in rumen bacterial diversity in two geographically separated sub-species of reindeer. *Microb. Ecol.* 54: 424–438.

19. Bobolowski M, Mills R, Beesmer J, Lomsadze A (2003) Prokaryotic gene prediction using GeneMark and GeneMark.mhm. *Curr. Protoc. Bioinformatics Chapter 4.*

20. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71: 8228–8235.

21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498–2504.

22. Noguchi H, Park J, Takagi T, Tsuchiya T, Watanabe K, et al. (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* 34: 344–348.

23. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21: 494–504.

24. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72: 5202–5207.

25. Hamady M, Walker J, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Meth.* 5: 235–237.

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

27. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows integration and analysis of high-throughput community sequencing data. *Nat. Meth.* 7: 335–336.

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

29. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. *PloS one* 5: e9490.

30. Microb. Ecol.* 61: 448–454.

31. Gebauer J, Pfeifer T, Helmisaari K, Nyman J, Peschke M, et al. (2006) Metagenomics of the Reindeer Rumen Microbiome. *Nat. Meth.* 3: 1314–1317.

32. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71: 8228–8235.

33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498–2504.

34. Salmivirta J, Välikangas L, Tenkanen V, Arvilommi H, Kere J, et al. (2003) Development and validation of an immune gene expression index. *Nat. Meth.* 5: 235–237.

35. Hamady M, Walker J, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Meth.* 5: 235–237.

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

37. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72: 5202–5207.

38. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. *PloS one* 5: e9490.

39. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71: 8228–8235.

40. Microb. Ecol.* 61: 448–454.

41. Gebauer J, Pfeifer T, Helmisaari K, Nyman J, Peschke M, et al. (2006) Metagenomics of the Reindeer Rumen Microbiome. *Nat. Meth.* 3: 1314–1317.

42. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71: 8228–8235.