ELSEVIER

FEMS Microbiology Ecology 38 (2001) 141–151

FEMS MICROBIOLOGY Ecology

www.fems-microbiology.org

Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes

Kristian Daly [a,*], Colin S. Stewart [b], Harry J. Flint [b], Soraya P. Shirazi-Beechey [a]

[a] *Epithelial Function and Development Group, Department of Veterinary Preclinical Sciences, University of Liverpool, Liverpool L69 7ZJ, UK*
[b] *Microbial and Intestinal Division, Rowett Research Institute, Bucksburn, Aberdeen AB21 9SB, UK*

Received 13 July 2001; received in revised form 24 September 2001; accepted 25 September 2001

First published online 29 October 2001

Abstract

The molecular diversity of the microflora present within the equine large intestine was investigated through the analysis of PCR-amplified 16S ribosomal RNA gene sequences. Total genomic DNA, recovered from samples of large intestinal wall tissue and lumen contents, was used to generate 272 random clones that were subjected to comparative phylogenetic analysis. The 272 sequences were classified into 168 operational taxonomic units/molecular species (at least 97% similarity), with 92% of recovered sequences being placed within two major phyla: low %G+C Gram-positive bacteria (72%) and *Cytophaga–Flexibacter–Bacteroides* (20%). Over one-third (37%) of all sequences were affiliated with one clostridial group, cluster XIVa. The remaining sequences were associated with *Spirochaetaceae* (3%), *Verrucomicrobiales* (3%), high %G+C Gram-positive bacteria ($<1\%$) and *Proteobacteria* ($<1\%$). Within the recovered equine clone population only 5% of the sequences corresponded to known organisms whose sequences are available in public databases. A further 6% corresponded to unidentified sequences retrieved in similar 16S rDNA PCR-based studies. The vast majority of sequences recovered (89%) did not correspond to any recorded sequences suggesting that the anaerobic microflora of the equine large intestine is severely underrepresented in the public domain and the lack of recognised sequences in many branches of the phylogenetic tree suggests the equine flora may contain many novel bacterial species.

Keywords: Equine large intestine; 16S rDNA; Bacterial diversity

1. Introduction

In horse, a non-ruminant herbivore, the large intestine is an immensely enlarged fermentative chamber which contains an extremely abundant and highly complex community of anaerobic microorganisms. The composition and activity of this obligatory anaerobic microflora, which includes bacteria, fungi, protozoa and archaea, have a profound effect on the health, growth, development and the performance of the animal [1–3]. The large intestine (caecum and colon) accounts for two-thirds of the volume of the digestive tract, with a combined capacity of over 200 l [4]. The microbial hydrolysis of dietary plant fibre within the large intestine leads to the release of soluble sugars which are fermented to short chain fatty acids (SCFA), most notably, acetate, propionate and butyrate. SCFA are absorbed across the large intestinal epithelium and provide 60–70% of horse's body energy [5,6].

Despite its importance, the microbial community of the equine hindgut has received relatively little attention. Molecular analysis suggests that *Ruminococcus flavefaciens* and *Fibrobacter* spp. are among the cellulolytic bacteria present in horses [7,8]. Knowledge of these and other groups of bacteria is however very limited [9,10], compared for example with the bacteria of the rumen [11–13] and of the colon of human and pigs [14–16]. Furthermore, information on the predominant cellulolytic microbial populations of equine gut is scarce [7–9]. There is an urgent need therefore for information on the equine gut microflora and their interspecies interactions especially in relation to their physiological contribution to the health and nutrition of the host. Understanding and predicting the effects of dietary change, stress, exercise, age, drug treatment or disease upon the complex microbial ecosystem of the gut requires both a basic knowledge of the composition of the microflora and a rapid and convenient technique for their study. Culture-based methods are often

* Corresponding author. Tel.: +44 (151) 794-4276; Fax: +44 (151) 794-4243.
E-mail address: nkd@liv.ac.uk (K. Daly).

0168-6496/01/$20.00
PII: S0168-6496(01)00178-7

laborious, time consuming and may recover only a small fraction of the total microbial diversity present within the gut [17]. Molecular methods based on PCR have the advantage of producing direct information on community structure and provide an efficient strategy for describing microbial biodiversity within environmental samples [13,18–20].

In the present study, we aimed to provide information on bacterial community members within the equine gut by analysis of 16S ribosomal RNA (rRNA) gene sequences retrieved directly from samples of wall tissue and lumen contents taken from the large intestines of several grass-fed horses. Comparative phylogenetic analysis has shown the majority of recovered sequences to be affiliated with two distinct phyla, the low %G+C Gram-positive bacteria (LGCGP) and the *Cytophaga–Flexibacter–Bacteroides* (CFB) assemblage. Furthermore, we report that the majority of sequences recovered here are only distantly related to known sequences from cultured bacteria and as such may represent many novel bacterial species.

2. Materials and methods

2.1. Collection and processing of samples

Samples of colonic wall tissue (caecum, CW; proximal colon, PW; mid colon, MW; distal colon, DW) and colonic lumen contents (caecum, CL; proximal colon, PL; mid colon, ML; distal colon, DL) were taken from five freshly slaughtered grass-fed animals, not suffering from any known intestinal diseases, obtained from the local abattoir. Lumen contents were placed in marked aluminium foils and immediately frozen in liquid nitrogen, while tissue samples were washed thoroughly in sterile 0.9% saline, pH 7.0, to remove luminal material prior to being frozen in liquid nitrogen. Frozen samples were then stored at −80°C until required.

2.2. DNA extraction

DNA was extracted from frozen tissue and luminal samples by a modification of the methods of Stahl et al. [21] and Pryde et al. [15]. Frozen samples were thawed on ice and 0.5-g aliquots were transferred to capped tubes containing phenol–chloroform–isoamylalcohol (25:24:1) and sterile glass beads (0.1-mm diameter). Sterile 10 mM Tris buffer (pH 8.5) was then added to completely fill the tubes. The samples were beaten for 30 s followed by chilling on ice for 1 min using a mini beadbeater (Biospec Corporation, Stratech Scientific, Luton, UK); this procedure was repeated six times. The aqueous supernatant was extracted twice more with phenol–chloroform–isoamylalcohol (25:24:1) and precipitated with isopropanol and ethanol. The DNA pellets were resuspended in sterile 10 mM Tris buffer (pH 8.5). To minimise variation between animals, DNA from each gut region was pooled before final purification with 13% polyethylene glycol (8000); 1.6 M NaCl. Purified DNA was resuspended in sterile 10 mM Tris buffer (pH 8.5) and stored at −20°C.

2.3. PCR

Total genomic DNA (2 μg from wall tissue samples; 50 ng from lumen content samples) was used as template for PCR amplification of approximately 720 bp of 16S rDNA with the eubacterial primers P3-Mod (5′-CGCGCCGC-ATTAGATACCCTDGTAGTCC-3′ [*Escherichia coli* positions 787–806]) and PC5 (5′-GCGGCCGC-TACCTT-GTTACGACTT-3′ [*E. coli* positions 1507–1492]) [15,16]. Each reaction contained 2.5 U Extensor Hi-Fidelity Enzyme Mix (ABGene) and PCR cycling was performed as follows: initial denaturation at 94°C for 5 min, denaturation at 94°C for 20 s, annealing at 55°C for 20 s and extension at 72°C for 1 min and 30 s. Twenty cycles of PCR were used to minimise the risk of preferential amplification of certain 16S rDNA sequence types.

2.4. Random cloning of bacterial DNA

Amplified PCR products were excised from agarose gels and purified using a commercial gel extraction kit (Qiagen). PCR products were then cloned into a pGEM-T Easy vector plasmid (Promega). Ligation was performed at 4°C overnight followed by transformation into competent *E. coli* JM109 cells. The clones were screened for successful transformation using X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) and IPTG (isopropyl-β-D-thiogalactopyranoside).

Transformed clones were selected and grown overnight in Luria broth with ampicillin selection. Plasmid DNA was isolated from 72 selected transformed clones from each gut region (24 from wall tissue; 48 from lumen contents) using a commercial plasmid miniprep kit (Eppendorf). Insert DNA was sequenced at the University of Liverpool sequencing facility. Only clones yielding clear and unambiguous sequence data were included in subsequent analyses.

2.5. Phylogenetic analysis

The partial 16S rDNA sequences corresponding to *E. coli* positions 807–1491 were compared directly to the EMBL and GenBank non-redundant nucleotide databases using the BLAST search facility [22]. Sequences derived from previously cultured and described organisms of known phylogeny which corresponded to major subdivisions of the domain *Bacteria* were included in the phylogenetic analysis. The sequence data approximating to *E. coli* positions 807–1491 were aligned with ALIGN X, within Vector NTi suite 6.0 (Informax Inc.) and phylogenetic trees were generated using the PHYLIP package [23].

Table 1
Summary of microbial diversity recovered from equine gut

Gut region	Number of sequences (%) affiliated with:															
	I[a]	III	IV	IX	XI	XIVa	XIVb	UCA[b]	UCB[c]	UCC[d]	B–L[e]	M[f]	CFB[g]	HGCGP[h]	V[i]	P[j]	S[k]
Caecum	0	5(2)	8(3)	2(<1)	1(<1)	26(9)	1(<1)	0	5(2)	1(<1)	2(<1)	0	16(6)	0	0	0	1(<1)
Proximal colon	2(<1)	8(3)	3(1)	1(<1)	0	21(8)	0	0	6(2)	1(<1)	2(<1)	1(<1)	15(5)	1(<1)	4(2)	1(<1)	3(1)
Mid colon	0	8(3)	5(2)	0	1(<1)	29(11)	0	0	2(<1)	2(<1)	2(<1)	2(<1)	14(5)	1(<1)	0	1(<1)	3(1)
Distal colon	0	9(3)	5(2)	0	0	24(9)	0	3(1)	0	6(2)	0	0	11(4)	0	4(2)	0	2(>1)
Total	2(<1)	30(11)	21(8)	3(1)	2(<1)	100(37)	1(<1)	3(1)	13(5)	10(4)	6(2)	3(1)	56(20)	2(<1)	8(3)	2(<1)	9(3)

[a] Roman numerals indicate phylogenetic cluster of *Clostridiaceae* [30].
[b] 'Unknown cluster A'.
[c] 'Unknown cluster B'.
[d] 'Unknown cluster C'.
[e] *Bacillales–Lactobacillales*.
[g] *Mycoplasmatales*.
[g] *Cytophaga–Flexibacter–Bacteroides*.
[h] High G+C Gram-positive bacteria.
[i] *Verrucomicrobiales*.
[j] *Proteobacteria*.
[k] *Spirochaetaceae*.

The DNADIST program analysed distances with the Kimura–Nei correction [24] and trees were generated from distance matrices by the neighbour-joining method [25]. Bootstrap analysis was performed by resampling the data 1000 times using the program SEQBOOT [26].

2.6. Nucleotide sequence accession numbers

The 16S rDNA sequences of clones used in phylogenetic analysis have been deposited in the EMBL data library under accession numbers AJ408036–AJ408266. The reference strains used in phylogenetic analysis were also from the EMBL database.

3. Results

3.1. 16S rDNA sequence diversity of the equine colonic microflora

Sequence analysis showed all sequences to contain conserved regions typical of eubacterial 16S rDNA. Six possible chimeric sequences found with the program CHIMERA_CHECK [27] and by comparison of trees derived from 200 bases at the 5′ and 3′ ends of the sequences were disregarded.

In total, 272 clones yielding unambiguous sequence data were analysed both by comparative analysis to public databases and by phylogenetic inference. The 272 sequences were classified into 168 operational taxonomic units (OTUs)/molecular species based on at least 97% sequence similarity. This value has been reported as discriminating between bacterial species previously defined on the basis of DNA–DNA reassociation values [28], albeit for full length 16S rDNA sequences.

The vast majority of sequences recovered from the equine gut in this study did not correspond to any recorded entries in the EMBL–GenBank databases. 77% of the total number of sequences showed similarity values in the range 90–97%, while 32 sequences (12%) showed less than 90% homology to their nearest database entry. Only 29 (11%) of the sequences recovered from the equine large intestine showed >97% homology to any recorded entries. Of these, 13 (5%) were related to a previously described organism, all but one belonging to the LGCGP. *Streptococcus bovis* was represented by five sequences comprising almost 2% of the total isolated population, with *R. flavefaciens* being represented by three sequences (1%). *Clostridium barati* was represented by two sequences, while *Butyrivibrio fibrisolvens*, *Lactobacillus salivarius* and the proteobacterium *Campylobacter lanienae* were all represented by one sequence each. The remaining 16 sequences (6%) were related to unidentified sequences retrieved in other 16S rDNA clone libraries [11,12,29].

3.2. Phylogenetic analysis

Table 1 shows the distribution of sequences recovered from each region of the equine large intestine. Representative clones encompassing all of the sequences recovered were used in the construction of phylogenetic trees by the methods previously described, as shown in Figs. 1–4. Reference sequences shown include those identified as being the closest relatives of the cloned sequences in the database comparisons.

The LGCGP phylum dominates the recovered equine sequences comprising 72% of the total number analysed (Figs. 1 and 2). Many isolates belonged to established clusters within the *Clostridiaceae* as defined by Collins et al. [30]. However, there are also three separate, distinct and novel clusters within the LGCGP phylum containing both equine and rumen sequences that do not correspond to any of the known clostridial clusters. Clones DL23, DL28, and DL29, all isolated from distal colon lumen,

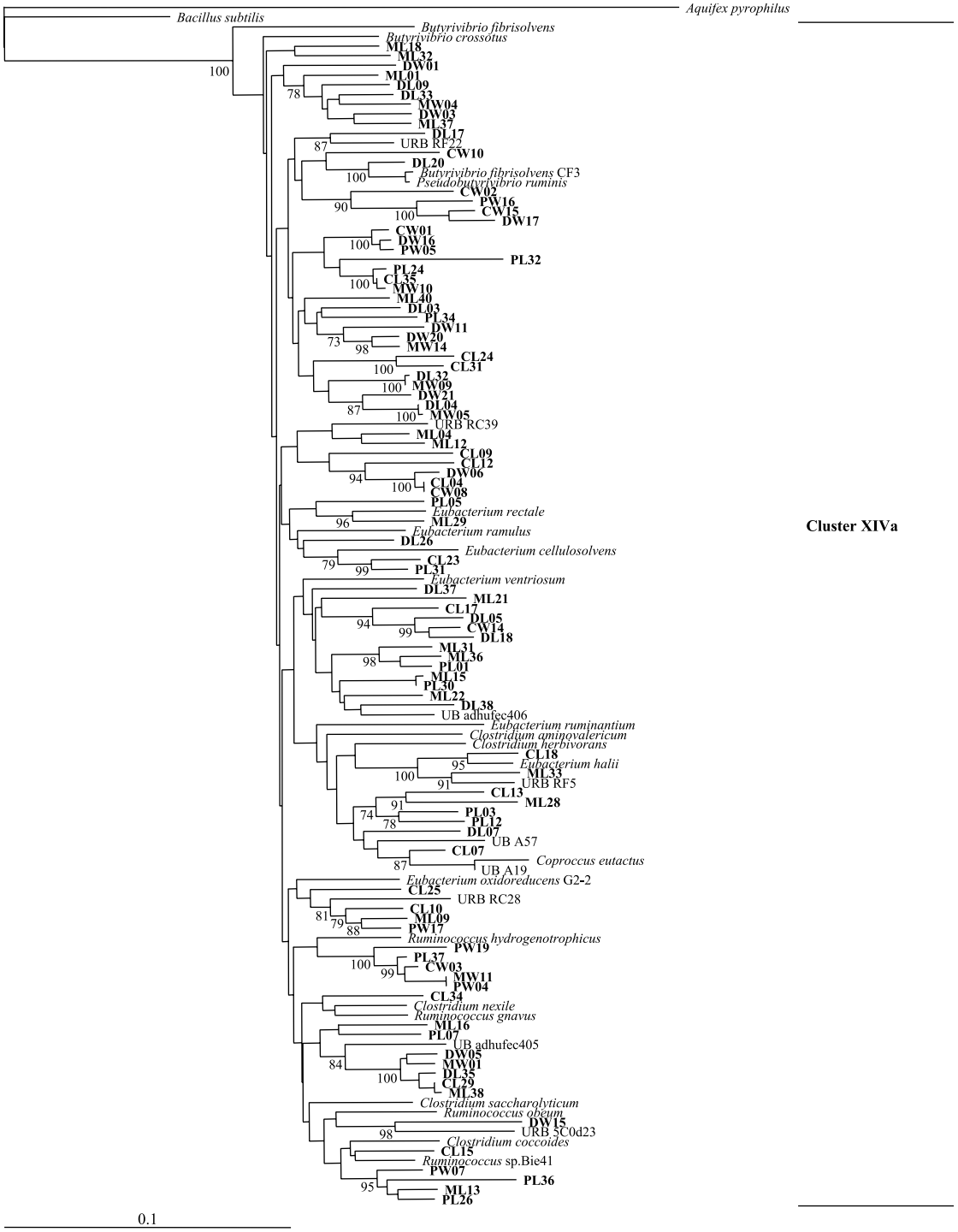

Fig. 1. Phylogenetic tree derived from partial 16S rDNA sequence data recovered from equine large intestine for members of cluster XIVa of the LGCGP. The tree was constructed using neighbour-joining analysis of a distance matrix obtained from a multiple sequence alignment. Bootstrap values are expressed as a percentage of 1000 trees; values below 70% are not shown. Sequences obtained from known species are shown in italics. *Aquifex pyrophilus* is used as the outgroup sequence. Scale bar represents the number of substitutions per nucleotide position.

formed part of 'unknown cluster A' (Fig. 2). Twelve clones (CL16, CL26, CL28, CL36, PW11, PW15, PL15, PL20, PL27, PL29, MW13, and ML27), representing 5% of the total clone population, were identified as members of 'unknown cluster B', while 'unknown cluster C' comprised 10 equine sequences (CW06, PL04, MW12, ML26, DW07, DW18, DL19, and DL22) isolated from all four regions of equine large intestine sampled, representing 4% of the total number. The robustness of all three clusters was supported by strong bootstrap confidence (99–100%).

Cluster XIVa of the *Clostridiaceae* [30] (Fig. 1), which contains many of the cellulolytic *Clostridium* spp., together

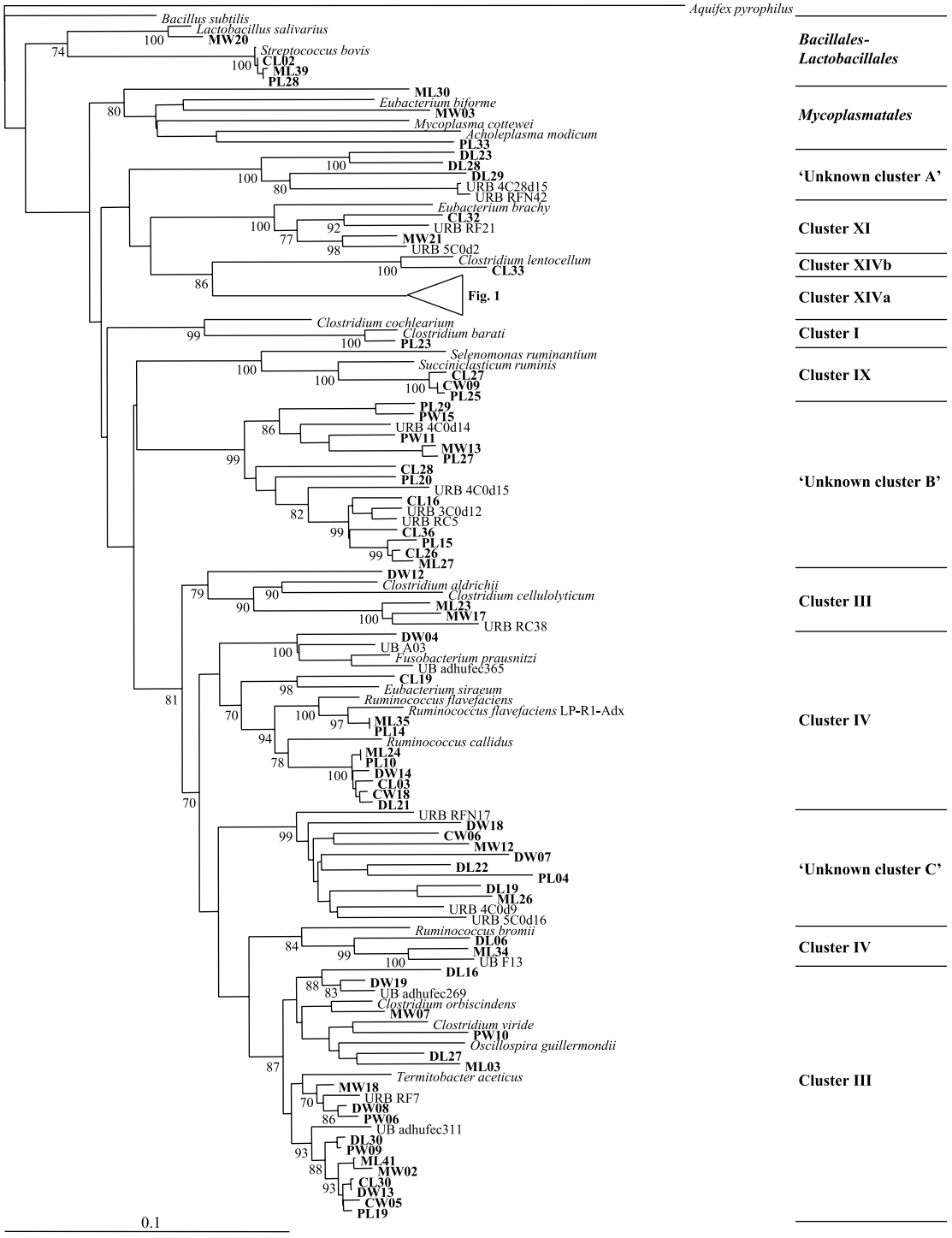

Fig. 2. Phylogenetic tree derived from partial 16S rDNA sequence data recovered from equine large intestine for members of the LGCGP. The tree was constructed using neighbour-joining analysis of a distance matrix obtained from a multiple sequence alignment. Bootstrap values are expressed as a percentage of 1000 trees; values below 70% are not shown. Sequences obtained from known species are shown in italics. *A. pyrophilus* is used as the outgroup sequence. Scale bar represents the number of substitutions per nucleotide position.

with *Butyrivibrio* spp., *Ruminococcus* spp., and *Eubacterium* spp., is represented by the largest number of recovered sequences, accounting for 37% of the total number (Fig. 1). The equine sequences contained within cluster XIVa are extremely diverse (Fig. 1), with 100 sequences classified into 68 OTUs (>97% sequence similarity). The majority of isolated clones within cluster XIVa occupy branches within the cluster that have no specific affinity to any cultured strains. A number of these novel subclusters contain both equine and other unidentified sequences (rumen and human), however, there are also subclusters that contain only equine sequences recovered in this study.

Fig. 3. Phylogenetic tree derived from partial 16S rDNA sequence data recovered from equine large intestine for members of the CFB assemblage. The tree was constructed using neighbour-joining analysis of a distance matrix obtained from a multiple sequence alignment. Bootstrap values are expressed as a percentage of 1000 trees; values below 70% are not shown. Sequences obtained from known species are shown in italics. *A. pyrophilus* is used as the outgroup sequence. Scale bar represents the number of substitutions per nucleotide position.

Only one cluster XIVa sequence showed greater than 97% sequence homology to any recorded database entry (clone DL20). Four of the clones within this cluster (CW15, PW16, ML28, and DW17) were less than 90% similar to any recorded sequences.

Twenty clones comprising 11% of the total number of sequences were phylogenetically positioned with representatives of cluster III of the *Clostridiaceae* (Fig. 2). Eight of these clones (CW05, CL30, PW09, PL19, MW02, ML41, DW08, and DL30), representing all four regions of the horse gut sampled including both wall tissue and lumen contents, formed the single most predominant OTU recovered (93% bootstrap confidence), represented 6% of the total clone population and occupied a branch that did not contain any known sequences. However, two of the clones within the OTU (PW09 and DL30) did correspond

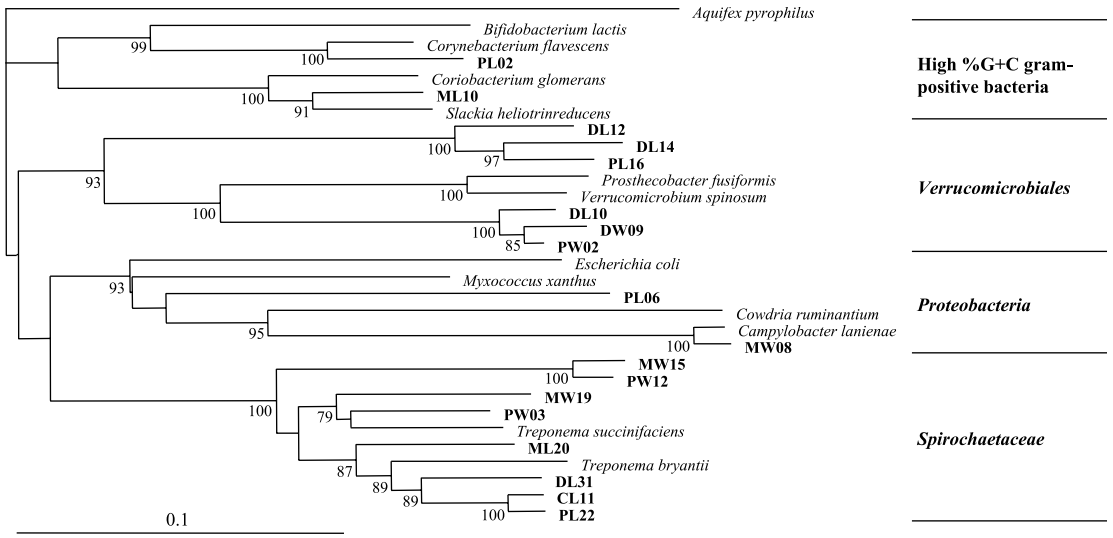

Fig. 4. Phylogenetic tree derived from partial 16S rDNA sequence data recovered from equine large intestine. The tree was constructed using neighbour-joining analysis of a distance matrix obtained from a multiple sequence alignment. Bootstrap values are expressed as a percentage of 1000 trees; values below 70% are not shown. Sequences obtained from known species are shown in italics. *A. pyrophilus* is used as the outgroup sequence. Scale bar represents the number of substitutions per nucleotide position.

to an unidentified sequence isolated from a human colonic sample [29]. Clones DW19 and DL16 also formed a separate cluster within this group (bootstrap value 88%) related to a human colonic sequence [29].

Cluster IV was represented by 21 sequences, 8% of the total clone population. Clones CW18, CL03, PL10, ML24, DW14, and DL21, encompassing all four gut regions sampled, formed a single OTU (5% of the total number of sequences) associated with *Ruminococcus callidus* (78% bootstrap confidence), while *R. flavefaciens* formed a single OTU with clones PL14 and ML35. Clone DW04 was associated with the *Fusobacterium prausnitzi* subcluster (100% bootstrap confidence).

Other LGCGP groups represented include clusters I, IX, XI, XIVb, the *Bacillales–Lactobacillales* group and the *Mycoplasmatales* (Fig. 2). Cluster IX of the *Clostridiaceae* was represented by three sequences (CW09, CL27, and PL25) that formed a single OTU, with *Succiniclasticum ruminis* being the nearest relative. Cluster I was represented by one clone, PL23, which formed a single OTU with *C. barati*. There were two clones placed within cluster XI, CL32 and MW21 (bootstrap value 100%), which were related to two unidentified rumen sequences [11,12], while clone CL33 was affiliated to cluster XIVb (bootstrap value 100%).

The *Bacillales–Lactobacillales* cluster contained four clones, three of which formed a single OTU with *S. bovis*, representing just under 2% of the total recovered clone population. The remaining clone, MW20, formed a single OTU with *L. salivarius*. Clones PL33, MW03, and ML30 were phylogenetically positioned with members of the *Mycoplasmatales* (bootstrap value 80%).

The second most represented phylum within the recovered equine sequences is the CFB assemblage (Fig. 3). This phylum comprises 20% of the total recovered sequences, with the majority (18% of total) falling within the *Bacteroides–Porphyromonas–Prevotella* (BPP) group. However, only one clone, PL11, showed more than 97% sequence homology to any recorded database entry, this being another unidentified equine sequence (accession number

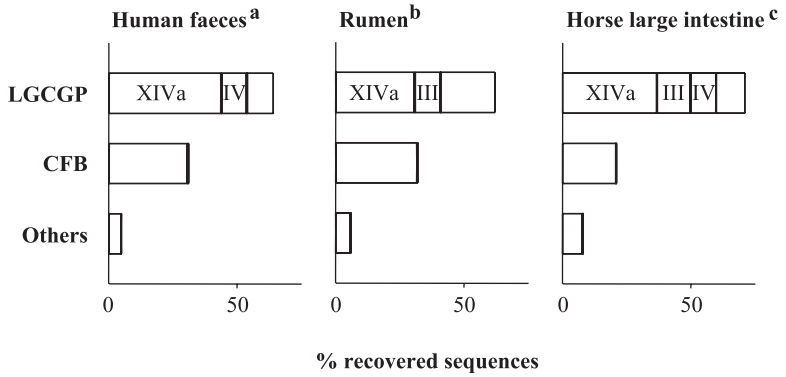

Fig. 5. Gut bacterial diversity (from 16S rRNA gene sequences). [a]References [16,29], [b]references [11,13], [c]this study.

AB045748). A large proportion of the clones did not correspond to previously assigned groups. Clones CW11, CL08, DW02, DL15, and DL24 formed a separate cluster (bootstrap value 100%) related to *Bacteroides* cluster II [27]. A further eight clones (CW16, CL01, PW01, PL38, ML05, ML25, DW10, and DL08), representing all four regions of the horse gut sampled, formed a novel, distinct cluster comprising 3% of the total number of sequences, and also included an unidentified equine sequence recovered in another study (accession number AB045750). The robustness of this cluster was confirmed by 100% bootstrap confidence, and all of the clones contained signature nucleotides relating to the BPP group [31]. Within the *Bacteroides* group, there were a number of equine clones that clustered with known subgroups. Clones PL40 and MW16 clustered with the ruminal sequences RF17 and RFN47 (bootstrap value 99%), members of *Bacteroides* cluster I [32], with clones CL14, PW08, ML14, Ml19, DL13, and DL24 related to ruminal sequences assigned to *Bacteroides* cluster II [32] (bootstrap value 92%).

Twenty four sequences within the BPP group are affiliated with the *Prevotella* subgroup (9% of the total number), although the two main clusters, the non-ruminal *Prevotella* supercluster I, and the ruminal *Prevotella* supercluster II, as defined by Ramsak et al. [32], are not represented by the recovered equine sequences. However, seven clones (CW04, CL05, PL13, and CL22, PW14, PL11, ML07) forming two separate clusters (94% and 81% bootstrap confidence respectively) were distantly related to *Prevotella albensis* and *Prevotella bryantii*, members of the non-ruminal *Prevotella* supercluster I. A further eight clones (PW13, PL17, ML11, DL01, and CL06, PL21, ML08, DL02) formed two novel clusters within the *Prevotella* subgroup not affiliated to any known sequences supported by bootstrap values of 87% and 100% respectively.

The remaining phyla represented in the equine microflora comprise only 8% of the total number of sequences recovered (Fig. 4). Nine sequences (3%) are affiliated with the *Spirochaetaceae*, with a bootstrap confidence value of 100%. Clones CL11, PL22, ML20, DL31, and PW03, MW19 clustered with the known spirochaetes *Treponema bryantii* and *Treponema succinifaciens* respectively. The *Verrucomicrobiales* group also makes up 3% of the total equine clone population. Six clones are affiliated with this group. Three of the clones (PW02, DW09, and DL10) formed a distinct separate cluster, but showed a definite (100% bootstrap confidence), albeit distant, relationship to *Prosthecobacter fusiformis* and *Verrucomicrobium spinosum*. The high %G+C Gram-positive bacteria (HGCGP) and the *Proteobacteria* each comprised $< 1\%$ of the total recovered population. Both clones within the HGCGP phylum were associated with known bacteria. Clone ML10 was related to *Slackia heliotrinreducens*, a member of the *Coriobacteriaceae*, while clone PL02 was closely related to the *Corynebacteriales*. Clone MW08, within the *Proteobacteria* assemblage, formed a single OTU with *C. lanienae*, a member of the *epsilon* subdivision, while clone PL06 formed a single, deep branch that did not correspond to any of the known subdivisions of the *Proteobacteria*. The phylogenetic position of this clone within the *Proteobacteria* assemblage, however, was supported by a strong bootstrap value (93%).

4. Discussion

Horses are trickle feeders, possessing a voluminous and elaborate intestinal tract. They are hindgut fermenters with complex microbial digestion, uniquely adapted to grazing on high fibre, low energy fodder. Domestication of the horse has led to this natural pattern of feeding and digestion being disturbed. Consequently, gastrointestinal disease is the single most important cause of mortality in the domestic horse and a major cause of morbidity and financial loss to the equine industry. An understanding of the biology and microbiology of the intestinal tract is essential for prevention and treatment of intestinal diseases.

The aim of this study was to provide information on the microbial diversity within the large intestine of horses maintained on a conventional grass-based diet, and to provide the comparative framework that is required for the further understanding of host–microflora relationships within the horse. To our knowledge, there have been no published data on the composition of the equine intestinal microflora using PCR-based 16S rDNA sequence analysis, although there are limited unidentified 16S rDNA sequences recovered from equine intestine in the public database.

One of the major findings of this study is the overwhelming representation of the LGCGP, comprising 72% of the total recovered clone population, with one group, cluster XIVa, accounting for over one-third (37%) of all sequences. The CFB group was represented by 20% of all sequences, while the remaining 8% were split between several other phyla. Although it is acknowledged that this study is limited by the use of only one primer set and by the analysis of partial sequences, the distribution of sequences seen here (at the phylum level) correlates well with results obtained in similar studies for other species, e.g. ruminants, humans [11,13,16,29] (Fig. 5), where the majority of sequences also correspond to the LGCGP and CFB. The numerical prevalence of the LGCGP bacteria, as shown in these similar 16S rDNA-based studies, would suggest that this is the most important functional group within intestinal ecosystems containing, as it does, the majority of cellulolytic and fibrolytic organisms such as *Clostridium* spp., *Ruminococcus* spp., *Butyrivibrio* spp., and *Eubacterium* spp. It is, however, very difficult to assign phenotypic characteristics to many of the recovered equine bacterial sequences here, as most do not correspond to any previously characterised species.

The most remarkable aspect of the data presented here

is the extremely high degree of genetic diversity within the bacterial populations of the equine large caecum and colon. Furthermore, the data highlight the scarcity of similar sequences in the public domain. The vast majority of sequences recovered in this study (89%) do not correspond to any previously recorded 16S rDNA sequences and the lack of recognised sequences in many branches of the phylogenetic trees indicates that new and, as yet, uncharacterised phylogenetic groups could be represented among the recovered equine sequences. Only 11% of the equine sequences recovered were >97% homologous to any recorded sequences, with only 5% corresponding to a recognised species.

There are three novel clusters within the LGCGP phylum (unknown clusters A, B, and C; Fig. 2), containing both equine and rumen sequences, that do not correspond to any of the clostridial clusters as defined by Collins et al. [30], which may represent new and functionally important groups of anaerobic bacteria that have yet to be characterised.

Even within known groups, there is a high level of divergence between the equine sequences recovered here and sequences available in the public databases. Within cluster XIVa and the BPP assemblage, there is a remarkable degree of diversity with a number of novel clusters occupying branches that do not contain any known sequences (Figs. 1 and 3). These include clusters that are only represented by equine sequences, and as such may be regarded as possible 'equine only' groups. Furthermore, many other branches in the phylogenetic trees that do not contain any recognised sequences are occupied by unidentified sequences recovered from horses, ruminants, mice and humans, which suggests that these clusters may be new groups of anaerobic bacteria that can occur in many host species.

There have been remarkably few studies concerning the composition of the equine microflora using modern molecular methods, with ruminants and humans seemingly the favoured subjects for study. Studies that do concern horses have also had a tendency to be focused towards specific components of the microflora, rather than the microflora as a whole. Lin and Stahl [8] focused their efforts on the quantification of the cellulolytic *Fibrobacter* spp. in the equine caecum and colon by oligonucleotide probing and concluded that *Fibrobacter succinogenes* accounted for up to 12% of total rRNA extracted from the caecum and 4% from the colon. Similarly, Julliand et al. [7] detected *F. succinogenes*, *R. flavefaciens* and *Ruminococcus albus* in equine caecum by oligonucleotide probing, but concluded that *R. flavefaciens* was the predominant cellulolytic species in the caecum (up to 9% of total rRNA). Interestingly, in our study we did not recover any sequences at all relating to *F. succinogenes* or *R. albus*, although we did recover three sequences that formed a single OTU with *R. flavefaciens* (Fig. 2). The apparent differences between our study and those mentioned above [7,8] could be as a result of an unknown PCR bias or perhaps differences in diet resulting in low abundance of *F. succinogenes* and *R. flavefaciens*. Tajima et al. [11,12] also failed to detect *F. succinogenes* in two separate 16S rRNA studies on the rumen. However, a recent publication by these authors [33] found that DNA extracted from *F. succinogenes* was amplified less efficiently by PCR in comparison to other gut bacteria. Although different primers were used in this study, it may be the case that this template is poorly amplified by any set of primers leading to its underrepresentation in PCR-generated libraries and giving rise to the apparent contradiction between PCR and hybridisation data [7,8,33].

In order to ascertain any potential differences in microbial community structure between different intestinal sites, samples from both the wall tissue and lumen contents from different regions of the equine caecum and colon were taken. Differences in colonic wall and lumen populations have already been shown in pigs [15]. However, our data do not indicate any significant differences in community structure between any of the different sample sites in the equine gut, and therefore, sequences from each separate library were analysed and are presented as one library.

The major impact of the activity of the intestinal microflora in the horse is in the production of fermentation products, mainly three monocarboxylates commonly referred to as SCFA, acetate, propionate and butyrate. The absorption of these fermentation products is of critical importance to the horse as 60–70% of its body energy requirements are met by SCFA [5,6]. It is, therefore, not surprising that there are specialised membrane transport systems in the caecum and colon known to transport SCFA, expression of which is tightly regulated [34]. Acetate is metabolised by many tissues such as heart, muscle and brain whilst propionate is one of the major precursors of gluconeogenesis [1]. Butyrate, in addition to being the preferred energy source for colonocytes [35], has been shown to have a much greater impact upon colonic epithelia. The identification of butyrate response elements upstream of some genes [36,37] suggests a much wider role for butyrate in the control of gene expression and it is thought that butyrate regulates proliferation and differentiation in gut epithelia [38].

The production of SCFA, or alternatively other monocarboxylates, in the large intestine is dependent on two interrelated factors: the composition of the resident microflora and the nature of the diet. In our study, we have shown that the vast majority of recovered sequences analysed are affiliated with members of the LGCGP, with over one-third belonging to cluster XIVa. In a study of colonic bacteria isolated from human colon, it was observed that 100% of isolates producing >2 mM butyrate in vitro were members of the LGCGP, and that over 80% of butyrate-producing isolates fell within cluster XIVa [14]. Changes in diet, though, can radically alter both

the community structure of the intestinal microflora [12] and thus the relative proportions of SCFA/monocarboxylates produced.

High grain diets containing soluble carbohydrates, if given over and above the capacity of the small intestine to digest and absorb them, can lead to elevated levels of soluble carbohydrates entering the caecum and colon. This has been implicated in the pathogenesis of equine acidosis [39,40], a similar condition to that which occurs in the rumen following a carbohydrate overload [41]. One particular organism, *S. bovis*, which ferments soluble carbohydrates, and produces lactate as its main metabolic product, has been associated as a major culprit in the onset of lactic acidosis [41,42]. Increased levels of lactic acid have not only been shown to inhibit the absorption of SCFA [43] but also can cause a rapid decline in the intestinal luminal pH, and in severe cases can lead to the death of the animal. Interestingly, in the present study, we have identified sequences relating to *S. bovis* (Fig. 2), and the presence of such bacteria in the equine large intestine may prove to be significant during periods of dietary change. Also recovered in this study were sequences relating to bacteria capable of utilising lactate, most notably members of cluster IX of the LGCGP (Fig. 2). These bacteria, which are able to tolerate low pH conditions, have been shown to be effective in preventing lactic acid accumulation in the rumen [41,44] and have been observed to increase in number from 4% to 46% during transition from high fibre to high grain diet in ruminant animals [12].

Information on the basic structure of the equine intestinal flora, under steady-state conditions, is of vital importance in determining which components of the microflora respond to dietary variations and are responsible for changes in the gut environment. The microbial sequences recovered in this study were amplified using primers that were identified in previous studies as universal eubacterial primers. It is possible, however, that certain bacterial groups, for example *Fibrobacter*, *Bifidobacteria*, which were not identified in this study, may not be efficiently amplified by these primers. In this case, further studies with additional primer pairings (including archaeal) plus the sequencing of full length 16S rDNA sequences would be required for an exhaustive survey of microbial diversity in the equine large intestine. Nevertheless, the data presented here provide an important first step in indicating the considerable uncharacterised microbial diversity that is present within the equine gut. The remarkable diversity observed with the equine sequences and the lack of recognised sequences in many branches of the phylogenetic trees suggest there may be significant new groups of intestinal anaerobic bacteria which may be 'equine only' or occur in many host species. Furthermore, the sequences identified here allow for the design of specific oligonucleotide probes that can be used to screen for the potential presence/absence of various phylogenetic groups in the equine intestinal tract. The use of specific probes will also allow for the quantitative determination of the predominant bacterial populations in the equine gut under any given circumstance. This will provide a rapid and convenient method for monitoring the effects of various known modulators such as dietary change, stress, exercise, age, drug treatment or disease on the complex microflora of the equine large intestine. This will allow a better understanding of the interrelationship between microbial composition and the host physiology and pathophysiology.

Acknowledgements

This study was supported by a grant from the Horserace Betting Levy Board under the Advancement of Veterinary Science and Education funding scheme. K.D. is a postdoctoral fellow funded by the Horserace Betting Levy Board.

References

[1] Cummings, J.H. and Macfarlane, G.T. (1997) Role of intestinal bacteria in nutrient metabolism. J. Parenter. Enter. Nutr. 21, 357–365.
[2] Cummings, J.H. and Macfarlane, G.T. (1997) Colonic microflora: nutrition and health. Nutrition 13, 476–478.
[3] Bergman, E.N. (1990) Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol. Rev. 70, 567–590.
[4] Jones, S.L., Snyder, J.R. and Spier, S.J. (1998) Physiology of the large intestine. In: Equine Internal Medicine (Reed, S.M. and Bayly, W.M., Eds.), pp. 651–655. W.B. Saunders, Philadelphia, PA.
[5] Argenzio, R.A., Southworth, M. and Stevens, C.E. (1974) Sites of organic acid production and absorption in the equine gastrointestinal tract. Am. J. Physiol. 226, 1043–1050.
[6] Argenzio, R.A. (1975) Functions of the equine large intestine and their interrelationship in disease. Cornell Vet. 65, 303–327.
[7] Julliand, V., de Vaux, A., Millet, L. and Fonty, G. (1999) Identification of *Ruminococcus flavefaciens* as the predominant cellulolytic bacterial species of the equine cecum. Appl. Environ. Microbiol. 65, 3738–3741.
[8] Lin, C. and Stahl, D.A. (1995) Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. Appl. Environ. Microbiol. 61, 1348–1351.
[9] Kern, D.L., Slyter, L.L., Leffel, E.C., Weaver, J.M. and Oltjen, R.R. (1974) Ponies vs. steers: microbial and chemical characteristics of intestinal ingesta. J. Anim. Sci. 38, 559–564.
[10] Mackie, R.I. and Wilkins, C.A. (1988) Enumeration of anaerobic bacterial microflora of the equine gastrointestinal tract. Appl. Environ. Microbiol. 54, 2155–2160.
[11] Tajima, K., Aminov, R.I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H. and Benno, Y. (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. FEMS Microbiol. Ecol. 29, 159–169.
[12] Tajima, K., Arai, S., Ogata, K., Nagamine, T., Matsui, H., Nakamura, M., Aminov, R.I. and Benno, Y. (2000) Rumen bacterial community transition during adaption to high-grain diet. Anaerobe 6, 273–284.
[13] Whitford, M.F., Forster, R.J., Beard, C.E., Gong, J. and Teather, R.M. (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. Anaerobe 4, 153–163.

[14] Barcenilla, A., Pryde, S.E., Martin, J.C., Duncan, S.H., Stewart, C.S., Henderson, C. and Flint, H.J. (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl. Environ. Microbiol. 66, 1654–1661.
[15] Pryde, S.E., Richardson, A.J., Stewart, C.S. and Flint, H.J. (1999) Molecular analysis of the microbial diversity present in the colonic wall, colonic lumen, and cecal lumen of a pig. Appl. Environ. Microbiol. 65, 5372–5377.
[16] Wilson, K.H. and Blitchington, R.B. (1996) Human colonic biota studied by ribosomal DNA sequence analysis. Appl. Environ. Microbiol. 62, 2273–2278.
[17] Langendijk, P.S., Schut, F., Jansen, G.J., Raangs, G.C., Kamphuis, G.R., Wilkinson, M.H. and Welling, G.W. (1995) Quantitative fluorescence in situ hybridization of *Bifidobacterium* spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. Appl. Environ. Microbiol. 61, 3069–3075.
[18] Godon, J.J., Zumstein, E., Dabert, P., Habouzit, F. and Moletta, R. (1997) Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. Appl. Environ. Microbiol. 63, 2802–2813.
[19] Gray, J.P. and Herwig, R.P. (1996) Phylogenetic analysis of the bacterial communities in marine sediments. Appl. Environ. Microbiol. 62, 4049–4059.
[20] Wise, M.G., McArthur, J.V. and Shimkets, L.J. (1997) Bacterial diversity of a Carolina bay as determined by 16S rRNA gene analysis: confirmation of novel taxa. Appl. Environ. Microbiol. 63, 1505–1514.
[21] Stahl, D.A., Flesher, B., Mansfield, H.R. and Montgomery, L. (1988) Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. Appl. Environ. Microbiol. 54, 1079–1084.
[22] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[23] Felsenstein, J. (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.
[24] Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
[25] Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
[26] Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
[27] Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker Jr., C.T., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M. and Tiedje, J.M. (2001) The RDP-II (Ribosomal database project). Nucleic Acids Res. 29, 173–174.
[28] Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic notice: a place for DNA–DNA reassociation and 16S rDNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849.
[29] Suau, A., Bonnet, R., Sutren, M., Godon, J.J., Gibson, G.R., Collins, M.D. and Dore, J. (1999) Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. 65, 4799–4807.
[30] Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. and Farrow, J.A. (1994) The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. 44, 812–826.
[31] Paster, B.J., Dewhirst, F.E., Olsen, I. and Fraser, G.J. (1994) Phylogeny of *Bacteroides*, *Prevotella*, and *Porphyromonas* spp. and related bacteria. J. Bacteriol. 176, 725–732.
[32] Ramsak, A., Peterka, M., Tajima, K., Martin, J.C., Wood, J., Johnston, M.E., Aminov, R.I., Flint, H.J. and Avgustin, G. (2000) Unravelling the genetic diversity of ruminal bacteria belonging to the CFB phylum. FEMS Microbiol. Ecol. 33, 69–79.
[33] Tajima, K., Aminov, R.I., Nagamine, T., Matsui, H., Nakamura, M. and Benno, Y. (2001) Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. Appl. Environ. Microbiol. 67, 2766–2774.
[34] Ritzhaupt, A., Wood, I.S., Ellis, A., Hosie, K.B. and Shirazi-Beechey, S.P. (1998) Identification and characterization of a monocarboxylate transporter (MCT1) in pig and human colon: its potential to transport L-lactate as well as butyrate. J. Physiol. 513, 719–732.
[35] Sheppach, W., Bartram, H.P. and Richter, F. (1998) Role of short chain fatty acids in the prevention of colorectal cancer. Eur. J. Cancer 31A, 1070–1080.
[36] Csordas, A. (1995) Toxicology of butyrate and short chain fatty acids. In: Role of Gut Bacteria in Human Toxicology and Pharmacology (M. Hill, Ed.), pp. 105–125. Taylor and Francis, London.
[37] Tran, C.P., Familari, M., Parker, L.M., Whitehead, R.H. and Giraud, A.S. (1998) Short-chain fatty acids inhibit intestinal trefoil factor gene expression in colon cancer cells. Am. J. Physiol. 275, G85–G94.
[38] Treem, W.R., Ahsan, N., Shoup, M. and Hyams, J.S. (1994) Fecal short-chain fatty acids in children with inflammatory bowel disease. J. Pediatr. Gastroenterol. Nutr. 18, 159–164.
[39] Garner, H.E., Hutcheson, D.P., Coffman, J.R., Hahn, A.W. and Salem, C. (1977) Lactic acidosis: a factor associated with equine laminitis. J. Anim. Sci. 45, 1037–1041.
[40] Goodson, J., Tyznik, W.J., Cline, J.H. and Dehority, B.A. (1988) Effects of an abrupt diet change from hay to concentrate on microbial numbers and physical environment in the cecum of the pony. Appl. Environ. Microbiol. 54, 1946–1950.
[41] Owens, F.N., Secrist, D.S., Hill, W.J. and Gill, D.R. (1998) Acidosis in cattle: a review. J. Anim. Sci. 76, 275–286.
[42] Goad, D.W., Goad, C.L. and Nagaraja, T.G. (1998) Ruminal microbial and fermentative changes associated with experimentally induced subacute acidosis in steers. J. Anim. Sci. 76, 234–241.
[43] Ritzhaupt, A., Ellis, A., Hosie, K.B. and Shirazi-Beechey, S.P. (1998) The characterization of butyrate transport across pig and human colonic luminal membrane. J. Physiol. 507, 819–830.
[44] Nocek, J.E. (1997) Bovine acidosis: implications on laminitis. J. Dairy Sci. 80, 1005–1028.