



# Genomic Insights into the Distribution of Peptidases and Proteolytic Capacity among *Prevotella* and *Paraprevotella* Species

Amlan Kumar Patra,<sup>a</sup>  Zhongtang Yu<sup>b</sup>

<sup>a</sup>Department of Animal Nutrition, West Bengal University of Animal and Fishery Sciences, Belgachia, Kolkata, India

<sup>b</sup>Department of Animal Sciences, Ohio State University, Columbus, Ohio, USA

**ABSTRACT** Bacterial peptidases play important roles in health and nutrient digestion in both humans and animals, especially ruminant animals. In this study, we examined and compared the various peptidases (both total and secretory) among species of *Prevotella* (44 in total) and *Paraprevotella* (2) revealed in their sequenced genomes that were archived in the MEROPS database. The phylogenetic relationships were also compared among the species based on 16S rRNA gene sequences and the occurrence of peptidases. A rich repertoire of peptidases was found that represents six catalytic types of peptidases (aspartic, cysteine, glutamic, metallo, mixed, and serine), together with some with unknown catalytic mechanisms, and 78 families. Metallopeptidases were the most predominant, followed by serine and cysteine peptidases. Considerable variations in peptidase occurrence and distribution were noted among the species and the different families of peptidases. A total of 48 different families of secretory peptidases were found in the genomes of the *Prevotella* and *Paraprevotella* species. Secretory peptidases in the families of S41 and M13 were ubiquitous, and S9, M16, C1, S13, and C69 were found in more than 95% of the species. Multivariate analysis of the peptidases indicated that species were mostly clustered except for a few species. Analysis using a bipartite association network showed that the majority of peptidase families were shared among the species. The relatedness of peptidase distributions among the species did not significantly correlate with their phylogenetic relationship based on the 16S rRNA genes. The genomic overview on the peptidases of *Prevotella* and *Paraprevotella* species provided new insights into their potential capacity to degrade proteins.

**IMPORTANCE** Species of *Prevotella* are prevalent and predominant bacteria residing in animals and humans, and their proteolytic capacity and activity play important roles in nutrient utilization in animals (especially ruminants) and some anaerobic infections of the intestinal, respiratory, and urinary tracts in humans. This study reveals the large repertoire and wide distribution of metallo, serine, and cysteine peptidases, especially secretory peptidases, among the *Prevotella* species. The information presented here could aid in the identification of the *Prevotella* species and the peptidases to target to decrease the excessive protein degradation in the rumen and improve dietary nitrogen utilization by ruminant animals.

**KEYWORDS** *Prevotella*, protease, protein degradation, ruminants, secretory peptidase, phylogenetic relationship, genome analysis, peptidases

**B**acterial degradation of both dietary and host-derived protein includes proteolysis, by which protein is depolymerized into smaller polypeptides or amino acids (AAs) and further catabolism of the proteolytic products. Bacterial protein degradation in the gastrointestinal (GI) tract is an important process in nutrient utilization and some

**Editor** Paul A. Jensen, University of Illinois at Urbana Champaign

**Copyright** © 2022 Patra and Yu. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Zhongtang Yu, yu.226@osu.edu.

The authors declare no conflict of interest.

**Received** 8 November 2021

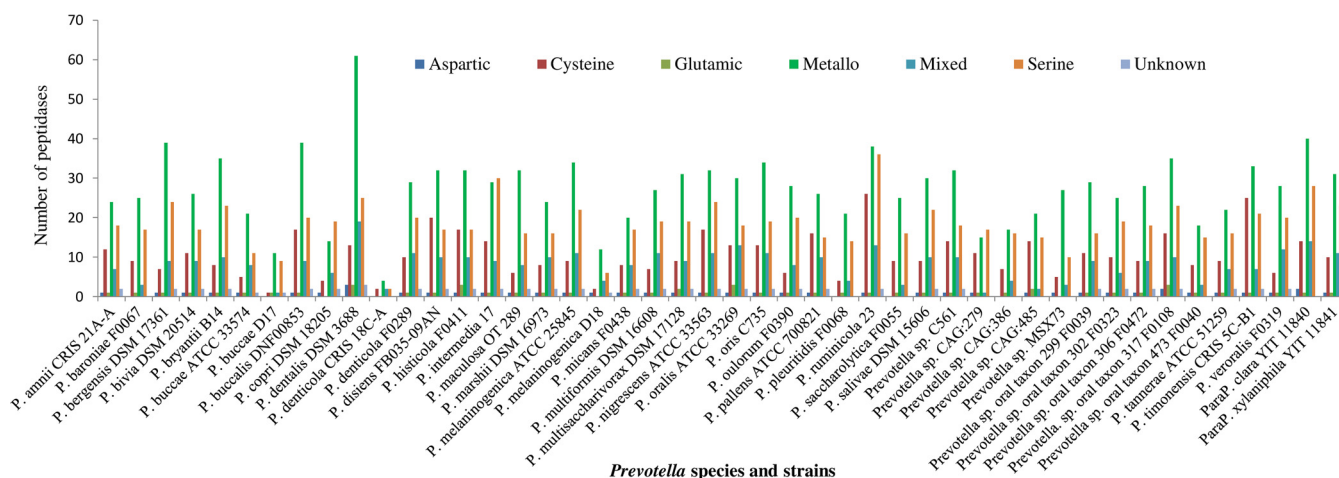
**Accepted** 16 March 2022

**Published** 4 April 2022

diseases in humans and animals (1). In humans and nonruminant animals, bacterial degradation of dietary and host-derived protein occurs primarily in the large intestines, where most of the gut microbes reside (2), producing oligopeptides, AAs, and short-chain fatty acids (SCFAs), including branched-chain fatty acids (BCFA). Both AAs and SCFAs can be absorbed by enterocytes as nutrients, even though only a small amount of AAs can be absorbed by the large intestine (1). However, many of the oligopeptides and AAs are fermented by GI bacteria producing SCFAs and several toxic metabolites, including phenols, indoles, ammonia, sulfide, and amines, which are genotoxic and can increase the risk of colorectal cancer (3, 4). Proteolytic degradation mediated by bacterial peptidases is also involved in some other diseases (5). For example, proteolytic degradation of the extracellular matrix of the GI tract by bacteria is implicated in inflammatory bowel diseases (IBDs) in humans (5–7). Bacterial proteolytic degradation of host protein is implicated in periodontitis (8) and urinary tract infections (9). In ruminants, bacterial proteolytic degradation of dietary protein in the rumen plays a central role in converting dietary protein to microbial protein, which represents 60 to 85% of the AAs reaching the animal's small intestine (10). Furthermore, microbial protein provides balanced AA profiles for milk and meat production (11–13). Therefore, bacterial proteolytic degradation of dietary protein is an essential process for microbial protein synthesis in the rumen. However, a significant portion of the proteolytic products, both oligopeptides and AAs, is fermented to SCFAs and ammonia or decarboxylated to amines due to the excessive degradation of dietary and microbial proteins (14). Although SCFAs are used as nutrients, much of the ammonia is converted to urea and excreted by the animals (15, 16), lowering nitrogen utilization efficiency (only about 30% in dairy cows). Furthermore, amines cannot be used by animals or humans, and they are also toxic. Species of *Prevotella* play an important role in protein degradation, especially in the rumen (17).

*Prevotella* species are anaerobic Gram-negative bacteria, and they inhabit the oral cavity, GI tract, and urinary tract of animals and humans (18–22). *Prevotella* is particularly predominant in the GI microbiome of humans with a *Prevotella* enterotype (23) and the rumen microbiome (21). Species of *Prevotella* are commensal, but several *Prevotella* species are implicated in some anaerobic infections of the GI tract, respiratory tract, and urinary tract (19, 22). For example, *Prevotella dentalis*, *Prevotella melaninogenica*, *Prevotella denticola*, *Prevotella enoeca*, *Prevotella fusca*, *Prevotella intermedia*, and *Prevotella* sp. oral taxon 299 are involved in periodontal diseases, such as gingivitis and periodontitis in the oral cavity of humans (20). In the rumen of ruminants, species of *Prevotella* are among the most predominant and proteolytic isolates (24, 25), and most of them are classified as *Prevotella ruminicola*, followed by *Prevotella brevis*, *Prevotella bryantii*, and *Prevotella albensis*. Based on quantitative PCR (qPCR) analysis, the genus *Prevotella* can account for 20 to 60% of the bacterial abundance and represents a substantial functional diversity of the rumen microbiome (18, 21, 26). However, *P. ruminicola* and *P. bryantii* represent only 2 to 4.4% of the total bacterial rRNA gene copies (18, 21), while *P. brevis* and *P. albensis* have a negligible abundance (21). Many *Prevotella* species evidently remain to be cultured and characterized. Indeed, most (88%) of the sequenced 16S rRNA gene clones prepared from rumen samples shared less than 97% sequence similarity with a known ruminal *Prevotella* species (18). In a phylogenetic census, there were 673 operational taxonomic units (OTUs, defined at 97% sequence similarity), but only 20 of them contained sequences from cultured/characterized *Prevotella* strains (27). Known rumen species of *Prevotella* produce various extracellular digestive enzymes, including amylases, xylanases, peptidases, and deaminase (28, 29), allowing them to be nutritionally versatile and grow on small peptides, many AAs, and sugars (21). Therefore, *Prevotella* species play significant roles in the digestion and fermentation of nutrients, especially dietary protein and noncellulosic polysaccharides (18, 21, 25). Peptidases are essential for the proteolytic activities of *Prevotella* species (30).

Peptidases (or proteases or proteinases) are hydrolytic enzymes ubiquitous among microbes (31–34). Peptidases are classified into different catalytic types based on the



**FIG 1** Catalytic types of peptidases found in the genomes of *Prevotella* (*P.*) and *Paraprevotella* (*ParaP.*) species and strains. Metallo, metallopeptidase.

catalytic residues: aspartic peptidases (A), cysteine peptidases (C), glutamic peptidases (G), metallopeptidases (M), asparagine peptidases (N), mixed peptidases (P), serine peptidases (S), threonine peptidases (T), and unknown peptidases (U). Each catalytic type of peptidases is classified into different clans, each of which has arisen from a single evolutionary origin of peptidases (35). Each catalytic type is then classified into one or more families based on AA sequence homology. Chemical characterization of individual peptidases from individual bacteria is time-consuming and laborious. The genome sequences of bacteria provide an opportunity to examine the catalytic types and families, distribution, and abundance of peptidases that can be produced by individual bacteria. In the study presented here, we analyzed and compared the catalytic types and distribution of the peptidases identified in the sequenced genomes of *Prevotella* species. We also compared the phylogenetic relationships among individual *Prevotella* bacteria using both peptidase sequences and 16S rRNA gene sequences. Species of *Paraprevotella*, which is a genus taxonomically established in 2009, are phylogenetically and phenotypically most similar to *Prevotella* species (36), and they were also included.

## RESULTS

### Abundance and distribution of different catalytic types and families of peptidases.

The MEROPS database (<https://www.ebi.ac.uk/merops/>) is a specialty peptidase database of manually curated proteolytic enzymes (35). It has archived peptidases of 34 strains of *Prevotella* within 31 known species, 10 strains that have not been assigned to any known species, and 2 strains of 2 *Paraprevotella* species (Fig. 1). A total of 3,131 peptidases and 550 nonpeptidase homologs were found in the genomes of the 44 strains of *Prevotella* and two strains of *Paraprevotella*. The nonpeptidase homologs were not further analyzed. All the known catalytic types of peptidases, except asparagine and threonine peptidase, were found in the sequenced genomes of the *Prevotella* and *Paraprevotella* species (Fig. S1). On average, each genome had ( $\pm$  standard deviation) 68.1 ( $\pm$ 22.1) peptidases. The number of peptidases per genome varied considerably among the species, with the largest in *P. dentalis* DSM 3688 ( $n = 127$ ) followed by *P. ruminicola* 23 ( $n = 117$ ), and the smallest were in *P. denticola* CRIS 18C-A ( $n = 10$ ) and *P. buccae* D17 ( $n = 24$ ) (Fig. 1). The catalytic type of metallopeptidase was the most predominant in all the species except for *P. intermedia* 17 and *Prevotella* sp. CAG:279, in which serine peptidase was the most predominant ( $n = 29$  versus 30 and  $n = 15$  versus 17, respectively). Serine peptidase was the second most predominant in all the species except for *Prevotella disiens* FB035-09AN, *Prevotella pallens* ATCC 700821, and *Prevotella timonensis* CRIS 5C-B1, in which cysteine peptidase was the second most predominant. Aspartic peptidase and glutamic peptidase were the least

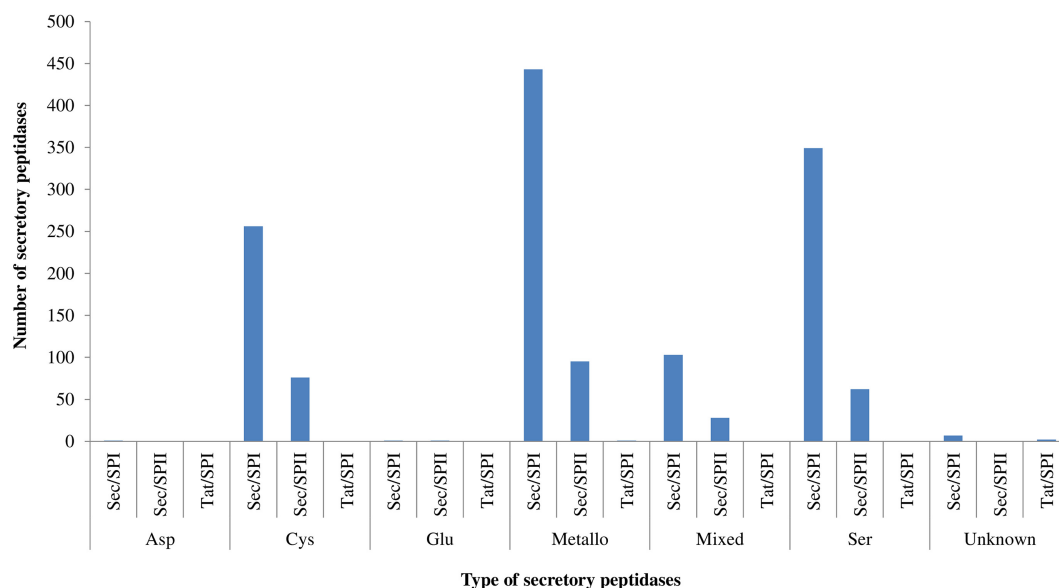
**TABLE 1** Different peptidase families present in *Prevotella* and *Paraprevotella* species/strains<sup>a</sup>

Peptidase family	Total no.	Percentage of total peptidases (%)	Total carrying species/strains	Percentage of total carrying species/strain (%)
M16	186	5.94	46	100
S41	139	4.44	46	100
C1	122	3.90	46	100
S9	219	6.99	45	97.8
M3	77	2.46	45	97.8
M13	74	2.36	45	97.8
M20	196	6.26	44	95.7
S13	57	1.82	44	95.7
S49	48	1.53	44	95.7
C69	65	2.08	43	93.5
S16	64	2.04	43	93.5
C25	47	1.50	43	93.5
G5	51	1.63	42	91.3
M49	45	1.44	42	91.3
M41	43	1.37	42	91.3
S8	118	3.77	41	89.1
M6	102	3.26	41	89.1
S1	50	1.60	41	89.1
S54	47	1.50	41	89.1
S26	46	1.47	41	89.1
A8	43	1.37	41	89.1
C10	187	5.97	40	87.0
M19	44	1.41	40	87.0
S46	41	1.31	40	87.0
C26	179	5.72	39	84.8
S14	47	1.50	39	84.8
M50	39	1.25	39	84.8
M24	129	4.12	37	80.4
M23	90	2.87	37	80.4
U32	70	2.24	36	78.3
M48	51	1.63	34	73.9
M15	33	1.05	31	67.4
C39	72	2.30	26	56.5
M93	38	1.21	25	54.3

<sup>a</sup>The peptidase families (of 78 in total) present in more than 50% of the bacterial species/strains are shown here.

prevalent and predominant (ranging from 0 to 3 per genome). The distribution of the catalytic types of peptidases also varied between some strains of the same species, such as *P. buccae* D17 versus *P. buccae* ATCC 33574 and *P. denticola* CRIS 18C-A versus *P. denticola* F0289 (Fig. 1).

A total of 78 distinct families of peptidases were found in the sequenced genomes of the *Prevotella* and *Paraprevotella* species, with 34 families being found in more than 50% of the species/strains (Table 1) and the remaining 44 being found in less than 50% of the species/strains (Table S1). Among these families, S9 (serine peptidase), M20 and M16 (metallopeptidase), and C10 and C26 (cysteine peptidase) were the most predominant families, with each representing more than 5% of the total peptidases found in all the species/strains. The families M16, S41, and C1 were the most prevalent in all the species (100%), and S9, M3, M13, M20, S13, and S49 were found in more than 95% of the species/strains. The families S13 and S49 were at low abundance, but they were found in 95.7% of the species/strains. Several families were found only in some of the species/strains, including 36 peptidase families that were found in less than 20% of the species/strains. The number of peptidases in a family varied considerably among the genomes, ranging from 0 to 17, with the largest count in the family C10 ( $n = 17$ ) in *P. timonensis* CRIS 5C-B1 followed by S9 ( $n = 16$ ) in *P. ruminicola* 23, M20 ( $n = 15$ ) in *P. dentalis* DSM 3688, S8 ( $n = 14$ ) in *P. intermedia* 17, C26 ( $n = 12$ ) in *P. dentalis* DSM 3688, and C10 ( $n = 12$ ) in *P. disiens* FB035-09AN (Table S2).



**FIG 2** Secretory types of peptidases of different catalytic types found in the genomes of *Prevotella* and *Paraprevotella* species/strains. Asp, aspartate; Cys, cysteine; Glu, glutamate; Ser, serine; Sec/SPI, peptidases Sec/SPI, secretory peptidases with a signal peptide transported by secretion translocation and cleaved by signal peptidase I; Sec/SPII, secretory peptidases with a lipoprotein signal peptide transported by secretion translocation and cleaved by signal peptidase II; Tat/SPI, secretory peptidases with a Tat signal peptide transported by twin-arginine translocation and cleaved by signal peptidase I.

**Abundance and distribution of different types of secretory peptidases.** A total of 1,425 secretory peptidases, which represented 45.5% of the total peptidases, were found in the sequenced genomes of *Prevotella* and *Paraprevotella* species/strains. On average ( $\pm$  standard error), 33.1 ( $\pm$  7.39) secretory peptidases were found per genome. Similar to the occurrence of total peptidases, secretory peptidases were mostly metallo (37.8%), serine (28.8%), and cysteine (23.3%) peptidase, the three of which together represented 90% of the total secretory peptidases (Fig. 2). Glutamic peptidase (0.07%) and aspartic peptidase (0.14%) were the least predominant. Among the secretory peptidases, the Sec/SPI secretory type was the most predominant (81.3%), followed by lipoprotein signal peptides transported by secretion translocation and cleaved by signal peptidase II (Sec/SPII) (18.4%), whereas the Tat signal peptides transported by twin-arginine translocation and cleaved by signal peptidase I (Tat/SPI) secretory type was sparse among the *Prevotella* and *Paraprevotella* species/strains (Fig. 2).

The secretory peptidases represented a total of 48 different families, including 18 prevalent with each having a prevalence of more than 50% (Table 2) and 30 families with each having a prevalence of less than 50% (Table S3). The secretory peptidases in the families S41 and M13 were ubiquitous (present in all species), and S9, M16, C1, S13, and C69 were found in more than 95% of the species/strain (Table 2). Those in families S9 (11.5%), C10 (10.2%), and M16 (9.26%) were most predominant, while many families were minor (less than 0.5% for 24 families combined). The number of secretory peptidases in a family varied among the species/strains, ranging from 0 to 12 (Table S4). Among all the species, the top 5 peptidase family with highest counts was the C10 family, which was present in *P. timonensis* CRIS 5C-B1 ( $n = 12$ ), *P. disiens* FB035-09AN ( $n = 9$ ), *Prevotella histicola* F0411 ( $n = 9$ ), *Prevotella buccalis* DNF00853 ( $n = 8$ ), and *P. ruminicola* 23 ( $n = 8$ ) (Table S4).

Of all the secretory peptidases, Sec/SPI was more prevalent in all species/strains than other secretory types, which was more predominant in *Prevotella nigrescens* ( $n = 38$ ), *P. melaninogenica* ( $n = 37$ ), and *P. disiens* ( $n = 36$ ) than in other *Prevotella* and *Paraprevotella* species/strains (Fig. 3). Among the species/strains of *Prevotella* and *Paraprevotella*, *P. melaninogenica* had the most secretory peptidases ( $n = 53$ ), followed

**TABLE 2** Different secretory peptidase families found in *Prevotella* and *Paraprevotella* species/strains<sup>a</sup>

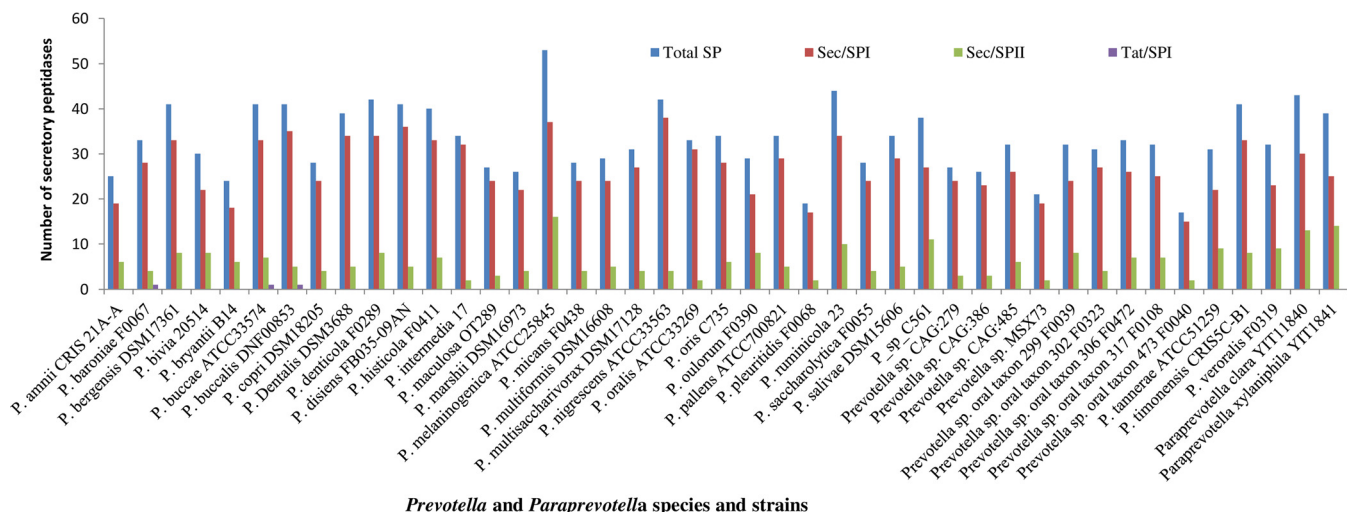
Peptidase family	Total no.	Percentage of total secretory peptidases (%)	Total carrying species/strains	Percentage of total carrying species/strains (%)
S41	84	5.89	43	100.0
M13	65	4.56	43	100.0
S9	164	11.51	42	97.7
M16	132	9.26	42	97.7
C1	107	7.51	42	97.7
S13	52	3.65	41	95.3
C69	43	3.02	41	95.3
M6	88	6.18	40	93.0
C10	145	10.18	39	90.7
S1	45	3.16	39	90.7
S8	83	5.82	36	83.7
S46	36	2.53	36	83.7
M23	75	5.26	35	81.4
C25	38	2.67	35	81.4
M15	27	1.89	27	62.8
M93	40	2.81	26	60.5
M48	30	2.11	25	58.1
M3	26	1.82	24	55.8

<sup>a</sup>The secretory peptidase families (of 48 in total) present in more than 50% of the bacterial species/strains are shown here.

by *P. ruminicola* (n = 44), *Paraprevotella clara* (n = 43), *P. denticola* (n = 42), and *P. nigrescens* (n = 42), while *Prevotella* sp. oral taxon 473 F0040 (n = 17) and *Prevotella pleuritidis* (n = 19) had the fewest secretory peptidases. Metallo, serine, and cysteine peptidases had more secretory peptidases than the other catalytic types (Fig. 2). Sec/SPII was more predominant in *P. melaninogenica* (n = 16), *Paraprevotella xylaniphila* (n = 14), and *P. clara* (n = 13) than in other *Prevotella* and *Paraprevotella* species/strains.

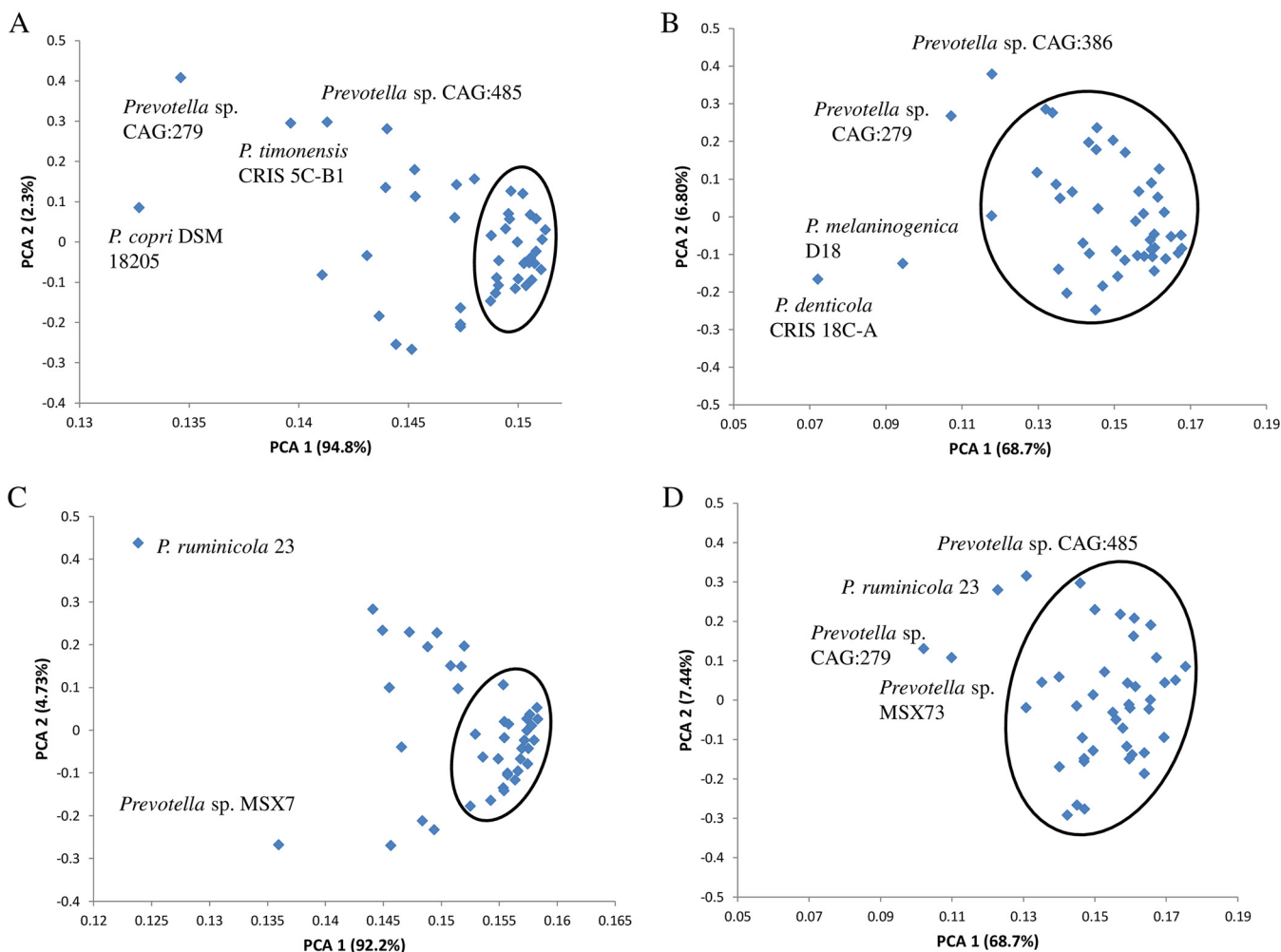
**Comparison of the species/strains for their peptidase occurrence patterns.**

Principal-component analysis based on total peptidases and their occurrence at the catalytic type and family levels showed that PCA1 accounted for most of the variations (94.8 and 68.7% at the catalytic type and family levels, respectively) (Fig. 4A). Based on the occurrence of catalytic types, more than 50% of the species formed a tight cluster, and the remaining species scattered along both the PC1 and PC2 axes, with *Prevotella copri* DSM 18205, *Prevotella* sp. CAG:279, *P. timonensis* CRIS 5C-B1, and *Prevotella* sp. CAG:485 scattered further away. However, all but *P. denticola* CRIS 18C-A, *P. melaninogenica* D18, *Prevotella* sp.



**FIG 3** Secretory types of peptidases (peptidases with a signal peptide) found in the genomes of *Prevotella* and *Paraprevotella* species and strains. Total SP, total secretory peptidases; Sec/SPI, secretory peptidases with a signal peptide transported by secretion translocation and cleaved by signal peptidase I; Sec/SPII, secretory peptidases with a lipoprotein signal peptide transported by secretion translocation and cleaved by signal peptidase II; Tat/SPI, secretory peptidases with a Tat signal peptide transported by twin-arginine translocation and cleaved by signal peptidase I.





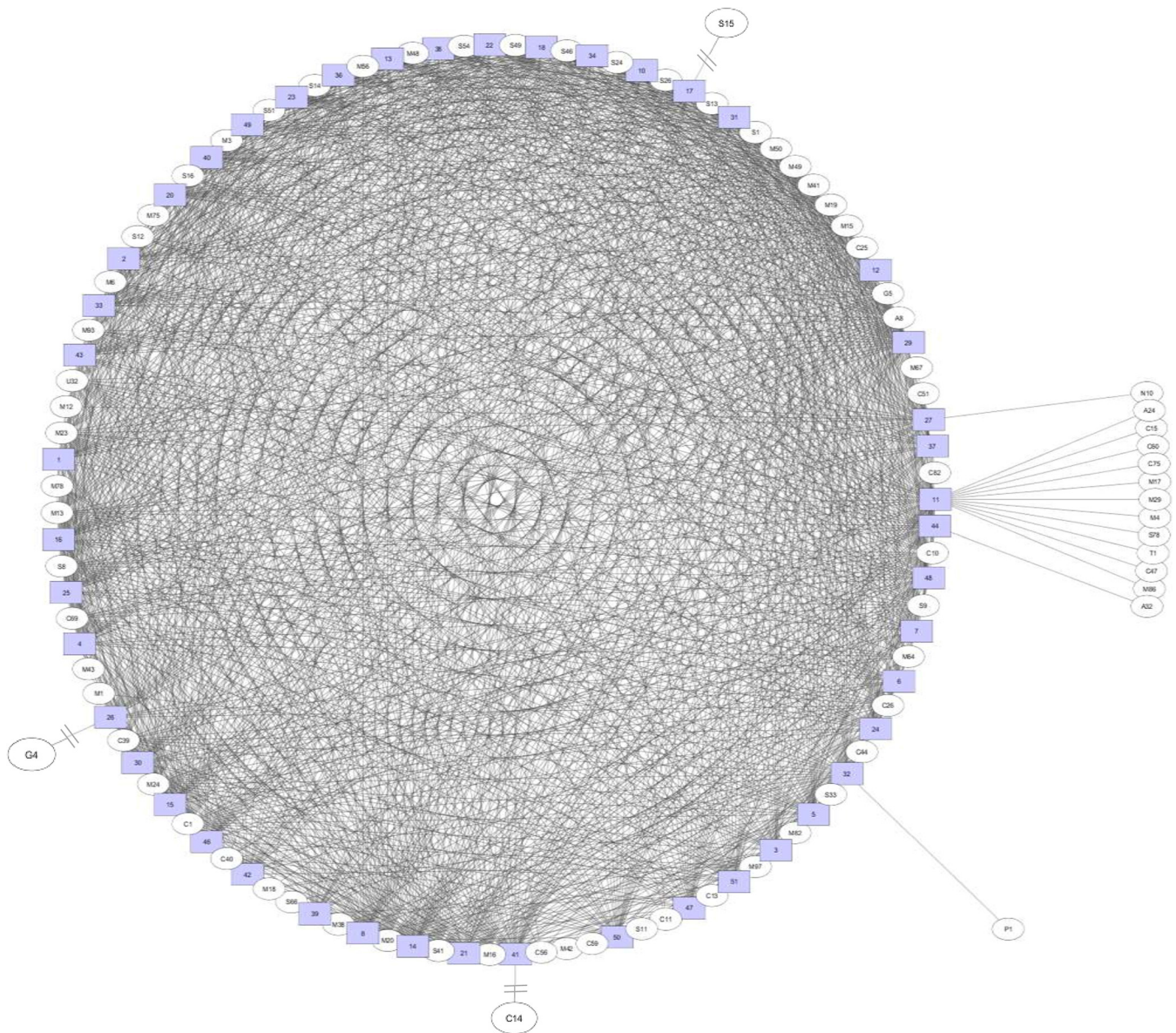
**FIG 4** Principal-component analysis (PCA) of peptidase occurrence based on the distribution of catalytic types of all the peptidases (A) and secretory peptidases (C) and the distribution of peptidase families of all peptidases (B) and secretory peptidases (D).

CAG:279, and *Prevotella* sp. CAG:386 formed one cluster when analyzed based on the occurrence of peptidase families (Fig. 4B).

For secretory peptidases, PCA1 also accounted for most of the variation (94.8 and 68.7% at the catalytic type and family levels, respectively) (Fig. 4C). Similar to all the peptidases, most species formed one tight cluster, and the remaining species scattered along both the PC1 and PC2 axes, with *P. ruminicola* 23 and *Prevotella* sp. MSX73 scattering further away. At the family level of secretory peptidases, all but *P. ruminicola* 23, *Prevotella* sp. CAG:279, *Prevotella* sp. CAG:485, and *Prevotella* sp. MSX73 formed one large cluster (Fig. 4D).

The bipartite association network for total peptidase families revealed that most of the peptidase families were shared among *Prevotella* species (Fig. 5). Only 16 peptidase families were unique to some *Prevotella* species, with S15 for *Prevotella maculosa* OT 289, G4 for *Prevotella oralis* ATCC 33269, P1 (mixed) for *P. ruminicola* 23, A32 for *Prevotella* sp. oral taxon 317 F0108, and N10 for *Prevotella oris* C735 (Fig. 5). *P. dentalis* DSM 3688 had many unique peptidase families (i.e., A24, C15, C47, C60, C75, M4, M17, M29, M86, S78, and T1) that were not present in the other *Prevotella* species (Fig. 5).

For secretory peptidases, 13 peptidase families were exclusively found in some *Prevotella* species: G4 in *P. oralis* ATCC 33269, G5 in *Prevotella marshii* DSM 16973, M19 in *Prevotella baroniae* F0067, S15 in *P. maculosa* OT 289, S66 in *P. ruminicola* 23, S16 in *P. buccalis* DNF00853, S26 in *P. xylaniphila* YIT 11841, S33 in *P. buccae* D17, C39 in *P. bivia* DSM 20514, and A32 in *Prevotella* sp. oral taxon 317 F0108 (Fig. 6). Same as for



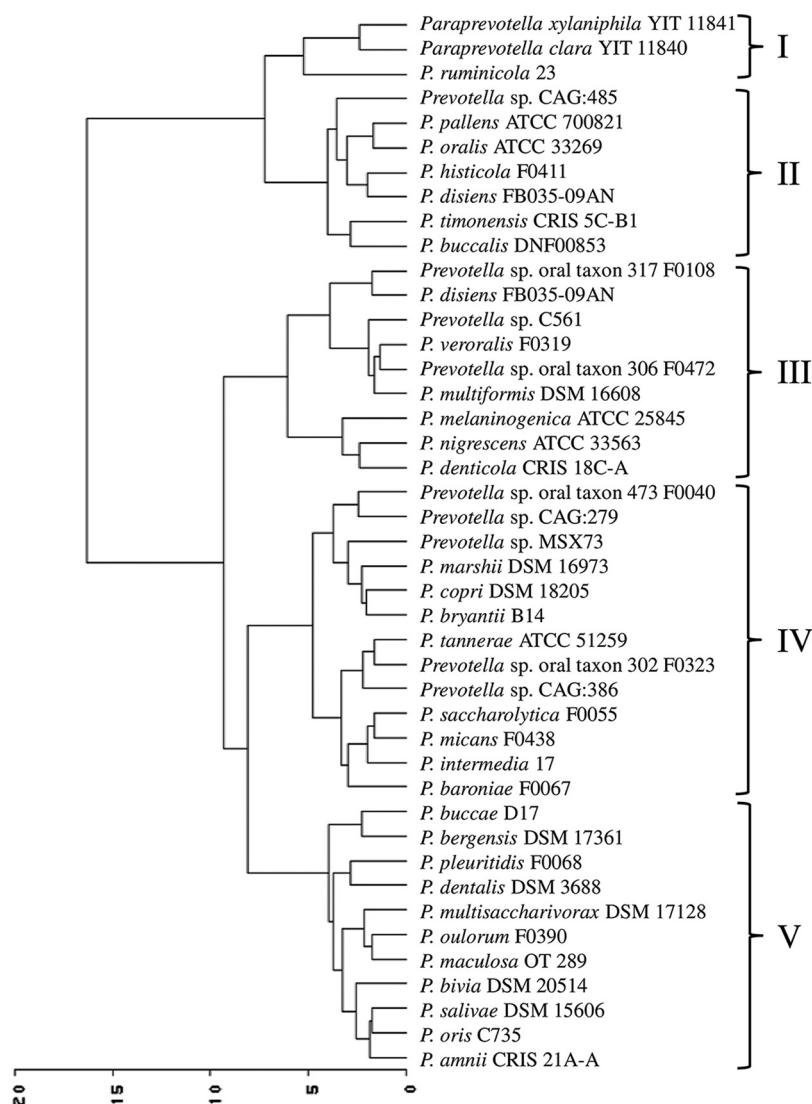
**FIG 5** A bipartite association network comparing the occurrence of peptidase families. The peptidase families are shown as ovals, while bacterial species names are shown as numbers in square boxes. 1, *P. amnii* CRIS 21A-A; 2, *P. baroniae* F0067; 3, *P. bergensis* DSM 17361; 4, *P. bivia* DSM 20514; 5, *P. bryantii* B14; 6, *P. buccae* ATCC 33574; 7, *P. buccae* D17; 8, *P. buccalis* DNF00853; 10, *P. copri* DSM 18205; 11, *P. dentalis* DSM 3688; 12, *P. denticola* CRIS 18C-A; 13, *P. denticola* F0289; 14, *P. disiens* FB035-09AN; 15, *P. histicola* F0411; 16, *P. intermedia* 17; 17, *P. maculosa* OT289; 18, *P. marshii* DSM 16973; 20, *P. melaninogenica* ATCC 25845; 21, *P. melaninogenica* D18; 22, *P. micans* F0438; 23, *P. multififormis* DSM 16608; 24, *P. multisaccharivorax* DSM 17128; 25, *P. nigrescens* ATCC 33563; 26, *P. oralis* ATCC 33269; 27, *P. oris* C735; 29, *P. oulorum* F0390; 30, *P. pallens* ATCC 700821; 31, *P. pleuritidis* F0068; 32, *P. ruminicola* 23; 33, *P. saccharolytica* F0055; 34, *P. salivae* DSM 15606; 36, *Prevotella* sp. C561; 37, *Prevotella* sp. CAG:279; 38, *Prevotella* sp. CAG:386; 39, *Prevotella* sp. CAG:485; 40, *Prevotella* sp. MSX73; 41, *Prevotella* sp. oral taxon 299 F0039; 42, *Prevotella* sp. oral taxon 302 F0323; 43, *Prevotella* sp. oral taxon 306 F0472; 44, *Prevotella* sp. oral taxon 317 F0108; 46, *Prevotella* sp. oral taxon 473 F004; 47, *P. tanneriae* ATCC 51259; 48, *P. timonensis* CRIS 5C-B1; 49, *P. veroralis* F0319; 50, *P. clara* YIT 11840; 51, *P. xylaniphila* YIT 11841.

total peptidases, *P. dentalis* DSM 3688 had three unique secretory peptidase families (i.e., M4, M86, and C47 families) that were not present in the other *Prevotella* species (Fig. 6). Cluster analysis of the secretory peptidase families revealed five clusters, with *P. ruminicola* and the two *Paraprevotella* species (*P. clara* and *P. xylaniphila*) forming one cluster and *Prevotella* sp. CAG:485, *P. pallens*, *P. oralis*, *P. histicola*, *P. disiens*, *P. timonensis*, and *P. buccalis* forming another cluster, and both clusters were separated from the clusters containing the other species (Fig. 7).

The phylogenetic tree constructed based on the 16S rRNA gene sequences of the *Prevotella* and *Paraprevotella* species along with the distribution of catalytic types of all





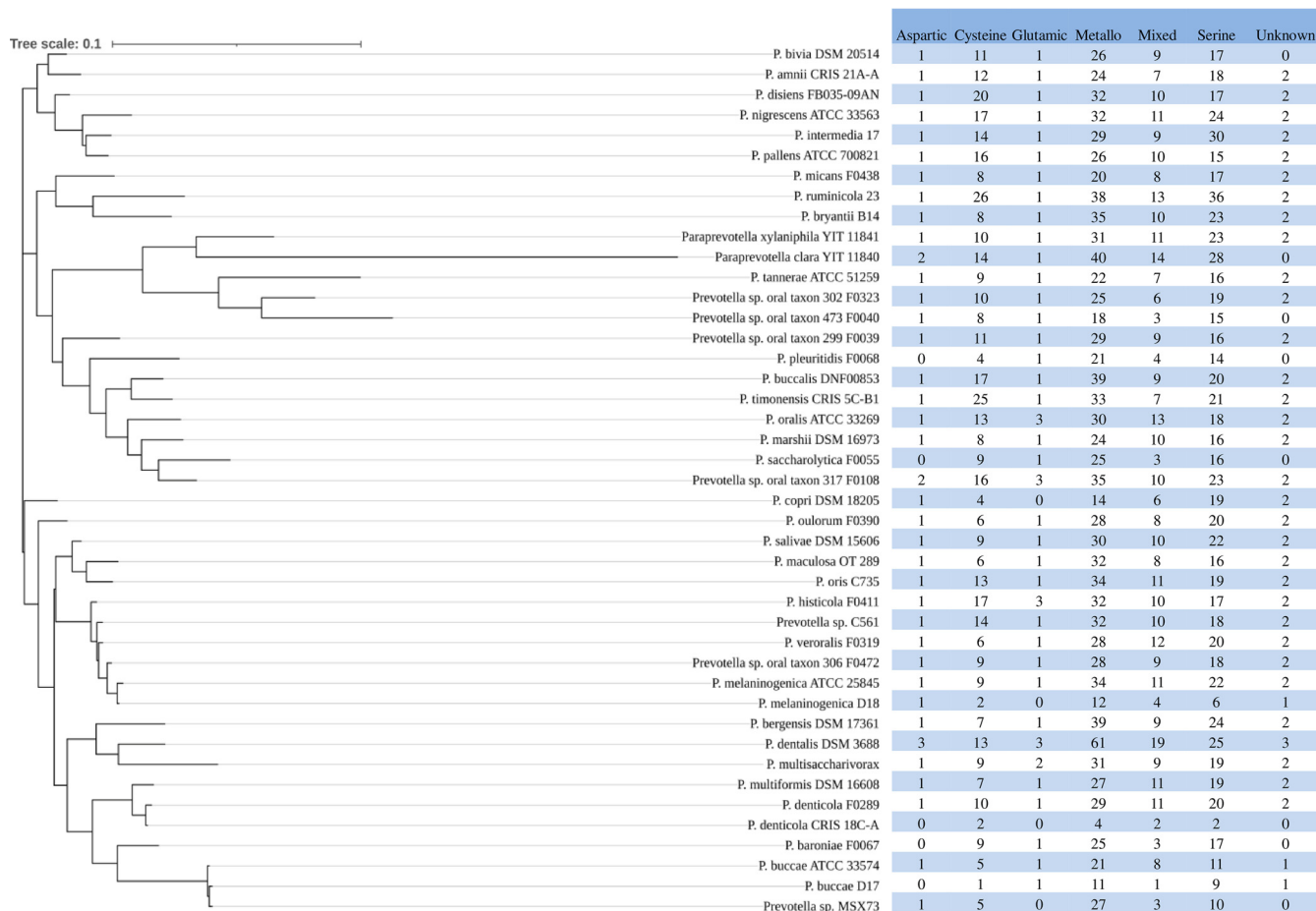


**FIG 7** Ward cluster analysis of secretory peptidase families in different species of *Prevotella* and *Paraprevotella*.

distribution distance matrix of secretory peptidase families and the 16S rRNA phylogenetic distance matrix was also not significant ( $r_{\text{Mantel}} = -0.051$ , simulated  $P$  value = 0.61 to 0.63, based on more than 1,000 replicates) (correlation tree not shown).

## DISCUSSION

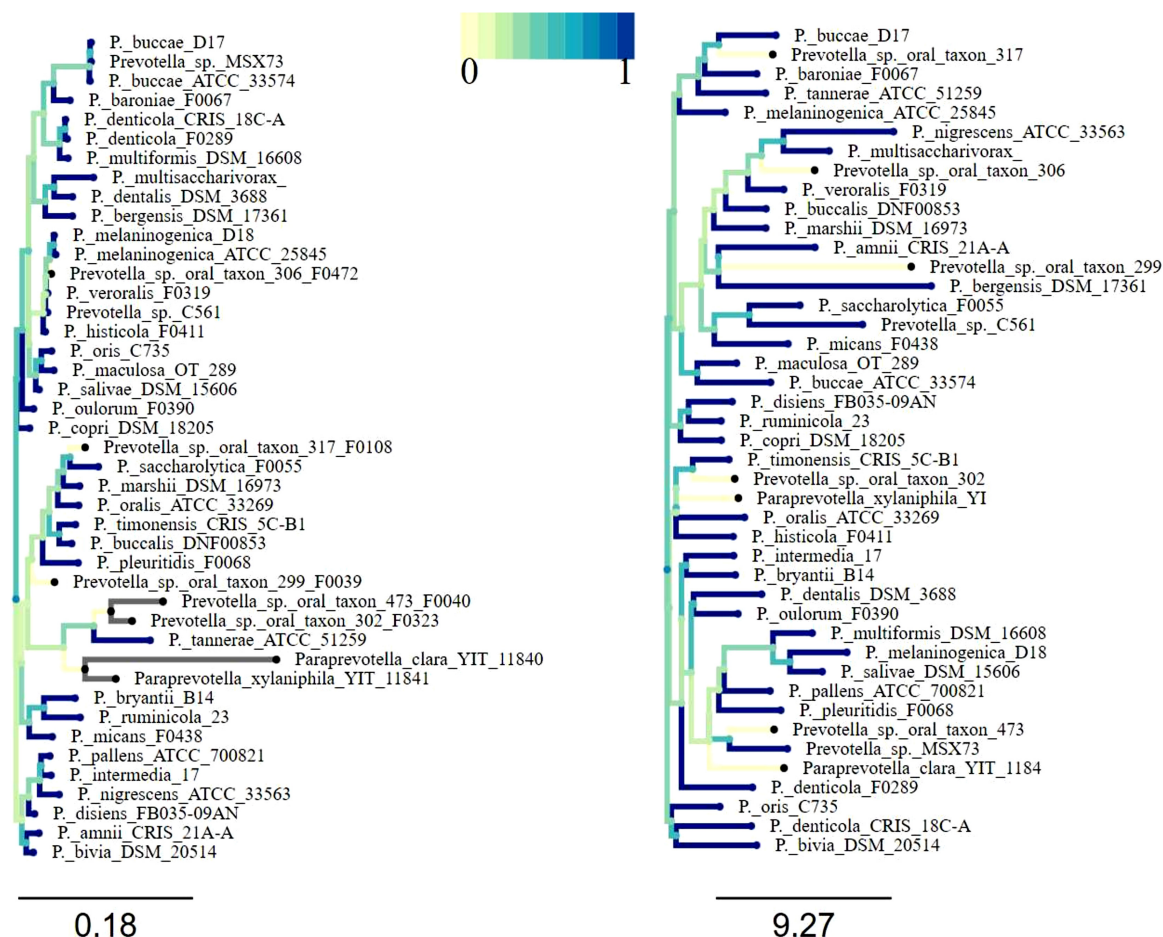
The MEROPS database classifies the peptidases into eight distinct catalytic types, and *Prevotella* and *Paraprevotella* genomes encode six of them (i.e., asparagine and threonine peptidases were absent). One peptidase analysis of 128 genomes of *Bacteria*, *Archaea*, and *Eukarya* revealed that serine (41%), metallo (29%), and cysteine (22%) peptidases were most prevalent (37). In the study presented here, metallo (40.4%), serine (26.9%), and cysteine (15.4%) peptidases were most prevalent, indicating that *Prevotella* and *Paraprevotella* species have different peptidase occurrence patterns compared to the global peptidase patterns. In three of the *Prevotella* species (i.e., *P. copri* DSM 18205, *P. intermedia* 17, and *Prevotella* sp. CAG:279), serine peptidases were, however, more abundant than metallopeptidases. At the family level, the families S9 (prolyl oligopeptidase), M20 (glutamate carboxypeptidase), M16 (pitrilysin), C10 (streptopain), and C26 ( $\gamma$ -glutamyl hydrolase) were most abundant, and the families



**FIG 8** Phylogenetic tree of the *Prevotella* and *Paraprevotella* species showing the distribution of different catalytic types of peptidases.

M16, S41 (C-terminal processing peptidase-1), and C1 (papain) were ubiquitous. Most members of M16 and S41 are involved in protein processing (38, 39). Various *Prevotella* species such as *P. oralis*, *P. melaninogenica*, *P. intermedia*, *P. nigrescens*, *P. buccae*, and *P. denticola* are implicated in dentoalveolar (purulent) infection (40). The peptidases of these *Prevotella* species may be involved in the degradation of connective tissue in gingiva, breakdown of proteinaceous compounds involved in the defense mechanisms in hosts, and virulence in the pathogenicity of oral infections (40, 41). In one study with different *Prevotella* species (one strain for each species), *P. intermedia* and *P. nigrescens* were shown to have much higher peptidase activities than *P. buccae*, *P. denticola*, *P. melaninogenica*, and *P. oralis* (21.1 to 23.5 versus 8.5 to 10.5 × 10<sup>-8</sup> A-units) (40). In the present study, we found that the two “high-protease-activity” *Prevotella* species (*P. intermedia* and *P. nigrescens*) contained more peptidases (86 and 88 peptidases, respectively) than the four “low-protease-activity” species (*P. buccae*, *P. denticola*, *P. melaninogenica*, and *P. oralis*; 36, 42, 53, and 80 peptidases, respectively), and this suggests that *Prevotella* species that carry a large number of peptidases may be more proteolytic.

In the large intestines of humans and animals and the rumen of ruminants, protein degradation is primarily driven by proteolytic bacteria, including species of *Prevotella*. In the rumen, excessive proteolysis leads to decreased dietary nitrogen utilization efficiency (42). Inhibition of some of the secretory peptidases of *Prevotella* species may help reduce their proteolytic activities to achieve improved nitrogen utilization. Indeed, inhibitors of dipeptidyl peptidase decreased ammonia concentration in cultures of *P. albensis*, *P. ruminicola*, *P. bryantii*, and *P. brevis* and in mixed cultures of ruminal bacteria (43). The decreased ammonia concentration indicates reduced protein degradation and production of ammonia, which cannot be assimilated by ruminants



**FIG 9** A phylogenetic tree based on 16S rRNA gene sequences (left) and a cladogram based on a distance matrix of the distribution of peptidase families (right). The deviation/distance from the blue to the light green lines corresponds to the dissimilarity at the branches of the phylogenetic tree and cladogram.

for protein synthesis, and increased dipeptides, which can be utilized by ruminants. In the present study, 59 peptidases (prolyl oligopeptidase, dipeptidyl peptidase IV, and acylaminoacyl peptidase) were found, and they may be targets for inhibition to improve dietary nitrogen utilization efficiency in ruminants.

Most nascent proteins destined for secretion into the extracellular milieu contain a signal peptide at the N terminus (44). After signal peptides are cleaved off during the secretion process, the mature secretory protein is exported to extracellular environments through different secretion systems (44, 45). Secretory proteins can be predicted based on the presence of a signal peptide. We used SignalP to predict the signal peptides of the peptidases. In one study, SignalP and Phobius (another software tool to predict signal peptides) were shown to produce nearly identical predictions (32). It has been predicted that 7.8 to 10.5% of the proteins of various mycobacterial proteomes (46) and one-third of the entire eukaryotic proteome (45) are putative secretory proteins. Although total proteins in proteomes were not analyzed in the present study, secretory peptidases accounted for 45.5% of the total peptidases present in the *Prevotella* and *Paraprevotella* species. In *Bacteria* and *Archaea*, serine, metallo, and cysteine secretory peptidases represent more than 80% of the secretory peptidase with serine peptidases accounting for more than 50% (32). In the present study, we found that nearly 90% of the secretory peptidases of the *Prevotella* and *Paraprevotella* species, similar to total peptidases, were metallo (37.8%), serine (28.8%), or cysteine (23.3%) peptidases. Compared with the distributions of secretory peptidases (about



48% serine, 31% metallo, and 17% cysteine peptidases; estimated from Fig. 1 of reference [32]) in the *Bacteria* kingdom, the secretory peptidases among the *Prevotella* species were enriched with more metallopeptidase and cysteine peptidase at the expense of serine peptidase. Future research can help determine the actual contribution of these three types of secretory peptidases to the overall protein degradation in the rumen.

The abundance of secretory peptidases varies widely between the *Archaea* and *Bacteria* kingdoms (32), and the present study revealed extensive variations even among the species of *Prevotella* and *Paraprevotella*. Large variations in the genetic diversity and proteolytic activities among *Prevotella* species have been reported earlier (28). Differences in the number of peptidase families, including secretory peptidases, may reflect their niche adaptation, such as to temperature, pH, substrate, and competitors (19, 32, 47, 48). Among the secretory peptidases, standard secretory signal peptides transported by secretion translocation and cleaved by signal peptidase I (Sec/SPI) was predominant (81.3%), followed by Sec/SPII (18.4%) in the *Prevotella* species, and these patterns of Sec/SPI and Sec/SPII peptidases were noted in other organisms (45, 46).

Intracellular peptidases are usually associated with regulatory functions and protein turnover, whereas secretory or extracellular peptidases (both free and membrane-bound) are involved in the enzymatic degradation of proteinaceous materials for the acquisition of carbon and nitrogen from the environment (32, 49). As revealed in the present study, the number of secretory peptidases varied considerably among the species; *P. melaninogenica* ATCC 25845 had the greatest number of secretory peptidases, followed by *P. ruminicola* 23 and *P. clara* YIT 11840. Different from most bacterial species, *P. melaninogenica* ATCC 25845 possesses two chromosomes: the primary chromosome carrying 1,318 coding genes and the secondary chromosome carrying 978 coding genes (19). *P. melaninogenica* is an important human pathogen of the upper respiratory tract, and it is the most prevalent and abundant anaerobic bacterium in cystic fibrosis lungs (50). This species has the most secretory peptidases (53 in total and 36 Sec/SPI type), and these secretory peptidases may make it a prevalent and abundant pathogen. Secretory peptidases vary between *Bacteria* and *Archaea* and between Gram-positive and Gram-negative bacteria (32). A study on several rumen *Prevotella* species also found great variability in their proteolytic activities (28). The greater number of secretory peptidases may correlate to a broader range of proteolytic activity of the species.

It will be interesting to compare the repertoires of the peptidases and the actual proteolytic activities between *Prevotella* species, but few studies have compared the proteolytic activities of multiple *Prevotella* species. *P. ruminicola* 23 had more total secretory peptidases than *P. bryantii* B14 (44 versus 24). *P. ruminicola* had the greatest range and activity of dipeptidyl peptidases (30, 51), which release dipeptides from the N terminus of peptides. The families C40, M49, S9, S15, and S46 contain dipeptidyl peptidases. *P. ruminicola* 23 had more secretory dipeptidyl peptidases (one C40, eight S9, and one S46) than *P. bryantii* B14 (two S9 and one S46). The differences in the abundance of secretory peptidases corroborate the greater extracellular peptidase activities observed in *P. ruminicola* D31d than in *P. bryantii* B14 (28). *P. ruminicola* 23 also had more secretory cysteine and serine peptidases than *P. bryantii* B14 (20 versus 5 and 15 versus 8, respectively). These two types of secretory peptidases may play a greater role in proteolysis by *P. ruminicola* than by *P. bryantii*. Indeed, specific inhibitors of cysteine peptidase (iodoacetate) inhibited the extracellular peptidase activity of *P. ruminicola* to a greater extent than that of *P. bryantii* (28). Furthermore, *P. ruminicola* 23 and *P. bryantii* B14 had similar numbers of secretory metallopeptidases (11 versus 9), and specific inhibition of metallopeptidases by 1,10-phenanthroline decreased the extracellular peptidase activities of these two strains to similar degrees (28). The sequencing of the genome of *P. ruminicola* D31d has not been completed, and thus its repertoire of peptidases is not known and cannot be directly compared to that of *P. ruminicola* 23. However, the difference in the repertoire of secretory peptidases between *P. ruminicola* 23 and *P. bryantii* B14 parallels the differences in the extracellular peptidase activities and

in their response to metallopeptidase inhibition between *P. ruminicola* D31d and *P. bryantii* B14. It is reasonable to postulate that the type and abundance of peptidases identified from the sequenced genomes of *Prevotella* and *Paraprevotella* species can serve as a proxy of their proteolytic capacity and activities. However, although *P. ruminicola* 23 had more secretory serine peptidase than *P. bryantii* B14, phenylmethylsulfonyl fluoride, a specific inhibitor of serine peptidases, inhibited the extracellular peptidase activity of the latter to a greater extent than that of the former (28). The reason is unclear, but differences between strains 23 and D31d (52) may also determine the differences in their proteolytic activities along with protein substrates. Once the sequencing of the genome of *P. ruminicola* D31d is completed, the repertoire of the secretory peptidases and their activities can be compared between *P. ruminicola* D31d and *P. bryantii* B14.

Bipartite network analysis revealed that most peptidase families were shared among the *Prevotella* and *Paraprevotella* species, but *P. dentalis* had a few peptidase families that were not rare in the other species. *P. dentalis* inhabits the periodontal area of oral cavities, and it is predominantly responsible for periodontal diseases (20). It is not clear whether these unique peptidases are the results of the adaptation to this habitat by *P. dentalis*. Principal-component analysis based on total peptidase families showed that most of the bacterial species formed one cluster except for a few exceptions. Especially, *Prevotella* sp. CAG:279, *Prevotella* sp. CAG:485, and *Prevotella* sp. CAG:386 were separated from the other species. The 16S rRNA gene sequences of these three strains are not available in Silva or other databases, so we could not determine their phylogenetic relatedness with the other species. Based on the secretory peptidases, most of the bacterial species also formed one cluster except *P. ruminicola* 23, *Prevotella* sp. MSX73, *Prevotella* sp. CAG:279, and *Prevotella* sp. CAG:485. Different strains of *P. ruminicola* can occupy different ecological niches, but *P. ruminicola* 23 is solely present in the rumen along with *P. brevis* GA33 (53, 54). Although *P. bryantii* B14 is frequently observed in the rumen, it is not closely related to *P. ruminicola* 23 as shown in the left panel of Fig. 9 (only 89.7% 16S rRNA gene sequence similarity). It has also been demonstrated from 16S rRNA analysis that *P. bryantii* B14 clustered with a nonruminant *Prevotella* supercluster that has additional ecological niches such as the oral cavity, whereas *P. ruminicola* 23 clustered with a ruminal *Prevotella* supercluster (54). Cluster analysis based on the secretory peptidase families indicated that *P. ruminicola*, *P. clara*, and *P. xylaniphila* formed a distinct cluster. Although frequently reported in the rumen (55–57), the role of the genus *Paraprevotella* therein is not well defined, *P. clara* and *P. xylaniphila* may be involved in proteolysis as the well-characterized ruminal *P. ruminicola* owing to their similar secretory peptidases.

The cladogram based on total peptidase families or the secretory peptidases and the phylogenetic tree based on the 16S rRNA gene sequences had different topologies across the *Prevotella* and *Paraprevotella* species. This contradicts the report that the distribution of secretory peptidases correlated significantly with the 16S rRNA-based phylogeny within the *Archaea* or *Bacteria* kingdom, and such correlation was most significant between closely related taxa for both *Archaea* and *Bacteria* though the correlations were weak (32). Among the species analyzed in the present study, the distribution of peptidases may not reflect their phylogenetic relationship, and recruitment of peptidases and speciation might not be closely evolutionarily linked.

In conclusion, *Prevotella* and *Paraprevotella* species possess 6 distinct catalytic types and 78 families (48 families of secretory peptidases) of peptidases, with metallopeptidases being present in the highest proportion, followed by serine and cysteine peptidases. The number and families of peptidases varied considerably among the species, but a few families of peptidases were ubiquitous to all species. Based on the peptidase distribution patterns, the majority of the species clustered together with a few exceptions. The bipartite association network based on the peptidase families revealed that most of the peptidase families were shared among the species. This study did not reveal any significant correlation between the phylogenetic relationship and the distribution of peptidase families, suggesting no evolutionary linkage of peptidases among the species. The results of this study could be used to identify some secretory

peptidases, especially those that are abundant and ubiquitous in the *Prevotella* species, which can be inhibited to reduce excessive dietary protein degradation in the rumen, improving protein utilization efficiency in ruminant animals.

## MATERIALS AND METHODS

**Collection of peptidase sequences and analyses.** Sequences and related data of the peptidases identified in the genomes of *Prevotella* and *Paraprevotella* species were downloaded from the MEROPS database (<https://www.ebi.ac.uk/merops/>) (35). In this database, peptidases are manually curated and hierarchically classified into eight catalytic types based on the catalytic residue serving at the active site of the enzymes, clans based on evolutionary origins, and families and then subfamilies based on AA sequence homology (35, 47, 58). When the database (release 12.3) was accessed (April 15, 2021), it had 16, 97, 6, 74, 10, 2, 54, 6, and 8 families (total 265) of peptidases for the aspartic, cysteine, glutamic, metallo, asparagine, mixed, serine, threonine, and unknown peptidases, respectively. Data relating to catalytic types, clan, family, MEROPS ID, or MERNUM (the MEROPS sequence identifier) of peptidases or non-peptidase homologs (proteins of known sequences that can be placed in a peptidase family but lack one or more of the expected catalytic residues) of strains of *Prevotella* ( $n = 44$ ) and *Paraprevotella* ( $n = 2$ ), whose whole genomes had been completely sequenced, were extracted based on the Index of Organisms. The details of species and the information of the peptidases are shown in Table S5. The peptidase sequences of all the species of *Prevotella* ( $n = 41$ ) and *Paraprevotella* ( $n = 2$ ) were also downloaded. All the peptidases involved in the proteolytic degradation of dietary protein are extracellular enzymes, and they have a signal peptide (SP) to allow for their excretion. Therefore, we predicted the presence of SP (short AA sequences at the N terminus), with the “Gram-negative” option selected, among all the downloaded peptidase sequences (3,752 in total) using the SignalP-5.0 server (<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>), which predicts the presence of SP and secretory types based on the deep neural network-based method (59). As a suitable method to predict prokaryotic secretory peptidases (32), the SignalP algorithm predicts proteins as standard secretory signal peptides transported by secretion translocation and cleaved by signal peptidase I (Sec/SPI), Tat signal peptides transported by twin-arginine translocation and cleaved by signal peptidase I (Tat/SPI), lipoprotein signal peptides transported by secretion translocation and cleaved by signal peptidase II (Sec/SPII), or no signal peptides (other). The prediction summary for each peptidase was downloaded from the SignalP-5.0 server. The MEROPS database does not have the AA sequences of peptidases for the individual strains of *Prevotella* and *Paraprevotella*. Therefore, species differences, but not strain variations, in the secretory peptidases of *Prevotella* and *Paraprevotella* species were extracted.

**Phylogenetic analysis and comparisons.** The taxonomy database identifiers of the individual species of *Prevotella* and *Paraprevotella* recorded in the MEROPS database were used to search for the corresponding 16S rRNA gene sequences in the SILVA database (release 138.1) using “NCBI tx ID” (60). The taxonomic identifiers present in both databases were used to identify the genomes of interest in this study. Only the completely annotated genomes with available 16S rRNA gene sequences in the SILVA database (release 138.1) were included in this study. The 16S rRNA gene sequences of *Prevotella* ( $n = 41$ ) and *Paraprevotella* ( $n = 2$ ) species were downloaded from the SILVA database as a Fasta file (60). The 16S rRNA gene sequences of *Prevotella* sp. CAG:279, *Prevotella* sp. CAG:386, and *Prevotella* sp. CAG:485 were not available in any databases. The 16S rRNA gene sequences were then aligned using the SILVA Incremental Aligner (SINA) (61), and the tree (as a Newick file) and FASTA files were downloaded. A phylogenetic distance matrix was obtained from the Newick file using the T-REX web-based server (62), which was then imported into R. A distance matrix of the bacterial species based on the distribution of peptidase (total and secretory peptidases) families (counts of individual families) was also constructed using the *ade4* package in R (63). Then, a correlation between the phylogenetic distance matrix of 16S rRNA genes and the peptidase distribution distance matrix was evaluated using the Mantel test of *ade4* in R (63). The phylogenetic distance matrix and the distance matrix of the peptidase family distribution in the Newick format were uploaded to the Plyo.io web server (<http://phylo.io/>) to visualize and compare them (64). The phylogenetic tree of *Prevotella* and *Paraprevotella* species and the distributions of different catalytic types of peptidases were visualized using the Interactive Tree Of Life (iTOL; version 6.1.1) online tool (65).

Principal-component analysis (PCA) was performed using “Proc Princomp” of SAS package (66) to explore the similarity of peptidase distributions among different *Prevotella* species and strains. The PCA was based on the correlation matrix of the numbers of peptidases, including secretory peptidases in various catalytic types and families in each *Prevotella* species.

The secretory peptidases were summarized in a matrix containing the counts of secretory peptidases assigned to peptidase families (rows) across all *Prevotella* species (columns). Indices between the secretory peptidases of individual species were calculated from this matrix using the Ward’s minimum-variance method and were used to generate a secretory peptidase distance matrix using the “Proc cluster” followed by the “Proc tree” procedures of the SAS package (66). A bipartite association network of unique and shared families of peptidases in *Prevotella* species was generated and visualized using Cytoscape release 3.6.1 (67).

**Data availability.** All the data are presented in the tables or in the supplementary files.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, XLSX file, 0.04 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.03 MB.

**SUPPLEMENTAL FILE 3**, XLSX file, 0.6 MB.

**SUPPLEMENTAL FILE 4**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

This study was partially supported by Agriculture and Food Research Initiative Competitive grant 2019-67030-29003 from the U.S. Department of Agriculture, National Institute of Food and Agriculture.

We declare no conflict of interest.

## REFERENCES

- Gurina TS, Mohiuddin SS. 2021. Biochemistry, protein catabolism. StatPearls Publishing, Treasure Island, FL.
- Macfarlane GT, Cummings JH, Allison C. 1986. Protein degradation by human intestinal bacteria. *J Gen Microbiol* 132:1647–1656. <https://doi.org/10.1099/00221287-132-6-1647>.
- Hughes R, Magee EAM, Bingham S. 2000. Protein degradation in the large intestine: relevance to colorectal cancer. *Curr Issues Intest Microbiol* 1: 51–58.
- Macfarlane GT, Macfarlane S. 1997. Human colonic microbiota: ecology, physiology and metabolic potential of intestinal bacteria. *Scand J Gastroenterol* 32:3–9. <https://doi.org/10.1080/00365521.1997.11720708>.
- Carroll IM, Maharshak N. 2013. Enteric bacterial proteases in inflammatory bowel disease-pathophysiology and clinical implications. *World J Gastroenterol* 19:7531–7543. <https://doi.org/10.3748/wjg.v19.i43.7531>.
- Pruteanu M, Hyland NP, Clarke DJ, Kiely B, Shanahan F. 2011. Degradation of the extracellular matrix components by bacterial-derived metalloproteases: implications for inflammatory bowel diseases. *Inflamm Bowel Dis* 17:1189–1200. <https://doi.org/10.1002/ibd.21475>.
- Steck N, Mueller K, Schemann M, Haller D. 2012. Bacterial proteases in IBD and IBS. *Gut* 61:1610–1618. <https://doi.org/10.1136/gutjnl-2011-300775>.
- O'Brien-Simpson NM, Veith PD, Dashper SG, Reynolds EC. 2003. *Porphyromonas gingivalis* gingipains: the molecular teeth of a microbial vampire. *Curr Protein Pept Sci* 4:409–426. <https://doi.org/10.2174/1389203033487009>.
- Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol* 13:269–284. <https://doi.org/10.1038/nrmicro3432>.
- Storm E, Ørskov ER, Smart R. 1983. The nutritive value of rumen microorganisms in ruminants: 2. The apparent digestibility and net utilization of microbial N for growing lambs. *Br J Nutr* 50:471–478. <https://doi.org/10.1079/bjn19830115>.
- Mabjeesh SJ, Arieli A, Bruckental I, Zamwell S, Tagari H. 1997. Effect of ruminal degradability of crude protein and nonstructural carbohydrates on the efficiency of bacterial crude protein synthesis and amino acid flow to the abomasum of dairy cows. *J Dairy Sci* 80:2939–2949. [https://doi.org/10.3168/jds.S0022-0302\(97\)76260-X](https://doi.org/10.3168/jds.S0022-0302(97)76260-X).
- Schingoethe DJ. 1996. Balancing the amino acid needs of the dairy cow. *Anim Feed Sci Technol* 60:153–160. [https://doi.org/10.1016/0377-8401\(96\)00976-5](https://doi.org/10.1016/0377-8401(96)00976-5).
- Schwab CG, Broderick GA. 2017. A 100-year review: protein and amino acid nutrition in dairy cows. *J Dairy Sci* 100:10094–10112. <https://doi.org/10.3168/jds.2017-13320>.
- Kung L, Jr, Rode LM. 1996. Amino acid metabolism in ruminants. *Anim Feed Sci Technol* 59:167–172. [https://doi.org/10.1016/0377-8401\(95\)00897-7](https://doi.org/10.1016/0377-8401(95)00897-7).
- Patra AK. 2015. Urea/ammonia metabolism in the rumen and toxicity in ruminants, p 329–341. *In* *Rumen microbiology: from evolution to revolution*. Springer, New York, NY.
- Tamminga S. 1979. Protein degradation in the forestomachs of ruminants. *J Anim Sci* 49:1615–1630. <https://doi.org/10.2527/jas1979.4961615x>.
- Wallace RJ, Onodera R, Cotta MA. 1997. Metabolism of nitrogen-containing compounds, p 283–328. *In* *The rumen microbial ecosystem*. Springer, New York, NY.
- Bekele AZ, Koike S, Kobayashi Y. 2010. Genetic diversity and diet specificity of ruminal *Prevotella* revealed by 16S rRNA gene-based analysis. *FEMS Microbiol Lett* 305:49–57. <https://doi.org/10.1111/j.1574-6968.2010.01911.x>.
- Gupta VK, Chaudhari NM, Iskepalli S, Dutta C. 2015. Divergences in gene repertoire among the reference *Prevotella* genomes derived from distinct body sites of humans. *BMC Genomics* 16:153. <https://doi.org/10.1186/s12864-015-1350-6>.
- Ibrahim M, Subramanian A, Anishetty S. 2017. Comparative pan-genome analysis of oral *Prevotella* species implicated in periodontitis. *Funct Integr Genomics* 17:513–536. <https://doi.org/10.1007/s10142-017-0550-3>.
- Stevenson DM, Weimer PJ. 2007. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* 75: 165–174. <https://doi.org/10.1007/s00253-006-0802-y>.
- Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, Armanini F, Manghi P, Bonham K, Zolfo M, De Filippis F, Magnabosco C, Bonneau R, Lusingu J, Amuasi J, Reinhard K, Rattei T, Boulund F, Engstrand L, Zink A, Collado MC, Littman DR, Eibach D, Ercolini D, Rota-Stabelli O, Huttenhower C, Maixner F, Segata N. 2019. The *Prevotella copri* complex comprises four distinct clades underrepresented in Westernized populations. *Cell Host Microbe* 26:666–679.e7. <https://doi.org/10.1016/j.chom.2019.08.018>.
- Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334:105–108. <https://doi.org/10.1126/science.1208344>.
- Attwood GT, Reilly K. 1995. Identification of proteolytic rumen bacteria isolated from New Zealand cattle. *J Appl Bacteriol* 79:22–29. <https://doi.org/10.1111/j.1365-2672.1995.tb03119.x>.
- Wallace RJ. 1996. Ruminal microbial metabolism of peptides and amino acids. *J Nutr* 126:1326S–1334S. [https://doi.org/10.1093/jn/126.suppl\\_4.1326S](https://doi.org/10.1093/jn/126.suppl_4.1326S).
- Wood J, Scott KP, Avguštin G, Newbold CJ, Flint HJ. 1998. Estimation of the relative abundance of different *Bacteroides* and *Prevotella* ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. *Appl Environ Microbiol* 64:3683–3689. <https://doi.org/10.1128/AEM.64.10.3683-3689.1998>.
- Kim M, Morrison M, Yu Z. 2011. Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* 76:49–63. <https://doi.org/10.1111/j.1574-6941.2010.01029.x>.
- Griswold KE, White BA, Mackie RI. 1999. Diversity of extracellular proteolytic activities among *Prevotella* species from the rumen. *Curr Microbiol* 39:187–194. <https://doi.org/10.1007/s002849900443>.
- Shen J, Yu Z, Zhu W. 2018. Insights into the populations of proteolytic and amino acid-fermenting bacteria from microbiota analysis using *in vitro* enrichment cultures. *Curr Microbiol* 75:1543–1550. <https://doi.org/10.1007/s00284-018-1558-1>.
- Wallace RJ, McKain N, Broderick GA, Rode LM, Walker ND, Newbold CJ, Kopecny J. 1997. Peptidases of the rumen bacterium, *Prevotella ruminicola*. *Aerobio* 3:35–42. <https://doi.org/10.1006/anae.1996.0065>.
- Kalisz HM. 1988. Microbial proteinases. *Adv Biochem Eng Biotechnol* 36: 1–65.
- Nguyen TT, Myrold DD, Mueller RS. 2019. Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. *Front Microbiol* 10:413. <https://doi.org/10.3389/fmicb.2019.00413>.
- Potempa J, Pike RN. 2005. Bacterial peptidases. *Contrib Microbiol* 12: 132–180. <https://doi.org/10.1159/000081693>.
- Rawlings ND, Bateman A. 2019. Origins of peptidases. *Biochimie* 166: 4–18. <https://doi.org/10.1016/j.biochi.2019.07.026>.
- Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* 46:D624–D632. <https://doi.org/10.1093/nar/gkx1134>.



36. Morotomi M, Nagai F, Sakon H, Tanaka R. 2009. *Paraprevotella clara* gen. nov., sp. nov. and *Paraprevotella xylaniphila* sp. nov., members of the family 'Prevotellaceae' isolated from human faeces. *Int J Syst Evol Microbiol* 59:1895–1900. <https://doi.org/10.1099/ijms.0.008169-0>.
37. Page MJ, Di Cera E. 2008. Evolution of peptidase diversity. *J Biol Chem* 283:30010–30014. <https://doi.org/10.1074/jbc.M804650200>.
38. Carroll RK, Rivera FE, Cavaco CK, Johnson GM, Martin D, Shaw LN. 2014. The lone S41 family C-terminal processing protease in *Staphylococcus aureus* is localized to the cell wall and contributes to virulence. *Microbiology* 160:1737–1748. <https://doi.org/10.1099/mic.0.079798-0>.
39. Nägler DK, Ménard R. 2003. Family C1 cysteine proteases: biological diversity or redundancy? *Biol Chem* 384:837–843.
40. Yanagisawa M, Kuriyama T, Williams DW, Nakagawa K, Karasawa T. 2006. Proteinase activity of *Prevotella* species associated with oral purulent infection. *Curr Microbiol* 52:375–378. <https://doi.org/10.1007/s00284-005-0261-1>.
41. Eley BM, Cox SW. 2003. Proteolytic and hydrolytic enzymes from putative periodontal pathogens: characterization, molecular genetics, effects on host defenses and tissues and detection in gingival crevice fluid. *Periodontol* 2000 31:105–124. <https://doi.org/10.1034/j.1600-0757.2003.03107.x>.
42. Newbold CJ, Ramos-Morales E. 2020. Review: ruminal microbiome and microbial metabolome: effects of diet and ruminant host. *Animal* 14: s78–s86. <https://doi.org/10.1017/S1751731119003252>.
43. Wang H, McKain N, Walker ND, Wallace RJ. 2004. Influence of dipeptidyl peptidase inhibitors on growth, peptidase activity, and ammonia production by ruminal microorganisms. *Curr Microbiol* 49:115–122. <https://doi.org/10.1007/s00284-004-4295-6>.
44. Chen Y, Zhang Y, Yin Y, Gao G, Li S, Jiang Y, Gu X, Luo J. 2005. SPD—a web-based secreted protein database. *Nucleic Acids Res* 33:D169–D173. <https://doi.org/10.1093/nar/gki093>.
45. Yan S, Wu G. 2014. Signal peptide of cellulase. *Appl Microbiol Biotechnol* 98:5329–5362. <https://doi.org/10.1007/s00253-014-5742-3>.
46. Leversen NA, de Souza GA, Mälen H, Prasad S, Jonassen I, Wiker HG. 2009. Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods. *Microbiology* 155:2375–2383. <https://doi.org/10.1099/mic.0.025270-0>.
47. Rawlings ND. 2016. Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation. *Biochimie* 122:5–30. <https://doi.org/10.1016/j.biochi.2015.10.003>.
48. Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. 1998. Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* 62:597–635. <https://doi.org/10.1128/MMBR.62.3.597-635.1998>.
49. Geisseler D, Horwath WR. 2008. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. *Soil Biol Biochem* 40:3040–3048. <https://doi.org/10.1016/j.soilbio.2008.09.001>.
50. Lamoureux C, Guilloux C-A, Beauvueille C, Gouriou S, Ramel S, Dirou A, Le Bihan J, Revert K, Ropars T, Lagrèfeuille R. 2021. An observational study of anaerobic bacteria in cystic fibrosis lung using culture dependant and independent approaches. *Sci Rep* 11:6845. <https://doi.org/10.1038/s41598-021-85592-w>.
51. Wallace RJ, McKain N. 1991. A survey of peptidase activity in rumen bacteria. *J Gen Microbiol* 137:2259–2259. <https://doi.org/10.1099/00221287-137-9-2259>.
52. Madeira HM, Morrison M. 1997. Growth inhibition of *Prevotella ruminicola* by protamine. *FEMS Microbiol Lett* 150:81–88. <https://doi.org/10.1111/j.1574-6968.1997.tb10353.x>.
53. Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM, Henrissat B, Nelson KE, North American Consortium for Rumen Bacteria. 2010. Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microb Ecol* 60: 721–729. <https://doi.org/10.1007/s00248-010-9692-8>.
54. Ramšak A, Peterka M, Tajima K, Martin JC, Wood J, Johnston MEA, Aminov RI, Flint HJ, Avguštin G. 2000. Unraveling the genetic diversity of ruminal bacteria belonging to the CFB phylum. *FEMS Microbiol Ecol* 33:69–79. <https://doi.org/10.1111/j.1574-6941.2000.tb00728.x>.
55. Dias J, Marcondes MI, Motta de Souza S, Cardoso da Mata e Silva B, Fontes Noronha M, Tassinari Resende R, Machado FS, Cuquetto Mantovani H, Dill-McFarland KA, Suen G. 2018. Bacterial community dynamics across the gastrointestinal tracts of dairy calves during preweaning development. *Appl Environ Microbiol* 84:e02675-17. <https://doi.org/10.1128/AEM.02675-17>.
56. Patra AK, Park T, Braun H-S, Geiger S, Pieper R, Yu Z, Aschenbach JR. 2019. Dietary bioactive lipid compounds rich in menthol alter interactions among members of ruminal microbiota in sheep. *Front Microbiol* 10: 2038. <https://doi.org/10.3389/fmicb.2019.02038>.
57. Xie X, Yang C, Guan LL, Wang J, Xue M, Liu JX. 2018. Persistence of cellulolytic bacteria *Fibrobacter* and *Treponema* after short-term corn stover-based dietary intervention reveals the potential to improve rumen fibrolytic function. *Front Microbiol* 9:1363. <https://doi.org/10.3389/fmicb.2018.01363>.
58. Rawlings ND, Barrett AJ. 1993. Evolutionary families of peptidases. *Biochem J* 290:205–218. <https://doi.org/10.1042/bj2900205>.
59. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
60. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
61. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28: 1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>.
62. Boc A, Diallo AB, Makarenkov V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res* 40:W573–W579. <https://doi.org/10.1093/nar/gks485>.
63. Dray S, Dufour A-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22:1–20.
64. Robinson O, Dylus D, Dessimoz C. 2016. Phylo. io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol Biol Evol* 33: 2163–2166. <https://doi.org/10.1093/molbev/msw080>.
65. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49: W293–W296. <https://doi.org/10.1093/nar/gkab301>.
66. SAS. 2001. Statistical analysis system. User's guide: statistics, version 8.2. SAS Institute Inc., Cary, NC.
67. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504. <https://doi.org/10.1101/gr.1239303>.