Ormerod et al. Microbiome (2016) 4:36
DOI 10.1186/s40168-016-0181-2

Microbiome

RESEARCH Open Access

Genomic characterization of the uncultured *Bacteroidales* family *S24-7* inhabiting the guts of homeothermic animals

Kate L. Ormerod[1], David L. A. Wood[1], Nancy Lachner[1], Shaan L. Gellatly[2], Joshua N. Daly[1], Jeremy D. Parsons[3], Cristiana G. O. Dal'Molin[4], Robin W. Palfreyman[4], Lars K. Nielsen[4], Matthew A. Cooper[5], Mark Morrison[6], Philip M. Hansbro[2] and Philip Hugenholtz[1*]

Abstract

Background: Our view of host-associated microbiota remains incomplete due to the presence of as yet uncultured constituents. The *Bacteroidales* family *S24-7* is a prominent example of one of these groups. Marker gene surveys indicate that members of this family are highly localized to the gastrointestinal tracts of homeothermic animals and are increasingly being recognized as a numerically predominant member of the gut microbiota; however, little is known about the nature of their interactions with the host.

Results: Here, we provide the first whole genome exploration of this family, for which we propose the name "*Candidatus* Homeothermaceae," using 30 population genomes extracted from fecal samples of four different animal hosts: human, mouse, koala, and guinea pig. We infer the core metabolism of "*Ca.* Homeothermaceae" to be that of fermentative or nanaerobic bacteria, resembling that of related *Bacteroidales* families. In addition, we describe three trophic guilds within the family, plant glycan (hemicellulose and pectin), host glycan, and α-glucan, each broadly defined by increased abundance of enzymes involved in the degradation of particular carbohydrates.

Conclusions: "*Ca.* Homeothermaceae" representatives constitute a substantial component of the murine gut microbiota, as well as being present within the human gut, and this study provides important first insights into the nature of their residency. The presence of trophic guilds within the family indicates the potential for niche partitioning and specific roles for each guild in gut health and dysbiosis.

Keywords: Gut microbiome, S24-7, *Homeothermaceae*, Population genomes, Metagenomics, Comparative genomics

Background

The host microbiome has been firmly established as critical to host physiology. Evidence now supports the microbiome as influential in diverse processes ranging from infection susceptibility [1] to behavior [2]. Unique anatomical sites are occupied by microbiota of distinct composition [3], supporting alternative functions being carried out at each site. Of clear significance is the gut microbiome, as metabolic capacity is a product of the capabilities encoded within both the host and the microbiome. The typical vertebrate gut microbiome is dominated by the *Firmicutes* and *Bacteroidetes*, and the divergent nature of gut-associated genera in comparison to other phylum members not associated with this environment indicates host selection and evolution occurring over a long period [4]. The relationship is also dynamic, evidenced by shifts in the composition of the gut microbiota encountered with perturbations to a person's diet, as well as with many acute and chronic, non-communicable diseases, such as inflammatory bowel diseases, or with their treatment (reviewed in [5, 6]). Despite our advances in describing these fluctuations, many members of the communities have yet to be cultured and characterized. As such, it remains difficult to ascribe their contributions to gut and systemic function, and thereby, host health and well-being.

* Correspondence: p.hugenholtz@uq.edu.au
[1]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia
Full list of author information is available at the end of the article

BioMed Central

© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

One such uncharacterized inhabitant of the gastrointestinal tract is a novel branch of the "*Bacteroides* group" first recognized in 2002 [7]. This branch was subsequently classified by Greengenes [8] and Silva [9] as an uncultured family of the order *Bacteroidales*, named after one of the earliest environmental clones belonging to the lineage, S24-7 (acc. AJ400263, [7]). Multiple studies have since reported the altered abundance of *S24-7* family members in association with different environmental conditions, e.g., *S24-7* is more abundant in diabetes-sensitive mice fed a high-fat diet, in particular when chow is supplemented with gluco-oligosaccharides [10] and following treatment-induced remission of colitis in mice [11]. Members of the *S24-7* family are also differentiated by their degree of IgA-labeling [12, 13] suggesting at least some members of the group are targeted by the innate immune system. While these observations are currently limited to murine-based studies, they do suggest that *S24-7* is involved in host-microbe interactions that impact on gut function and health.

To further our understanding of the *S24-7* family, we obtained 30 population genomes from four different hosts (human, mouse, koala, and guinea pig) and performed a comparative genomics analysis. The recovered genomes define a family that is most closely related to, but distinct from, the genera *Barnesiella* and *Coprobacter*. Analysis of 16S rRNA gene databases suggests a strong habitat preference for the homeothermic gut. Metabolically, we infer *S24-7* are fermentative or nanaerobic [14] species, consistent with their environmental niche. We describe three trophic guilds within the family focusing on α-glucan, host glycan, or plant glycan-based carbohydrates suggesting the capacity for niche partitioning and/or divergent spatial organization of its members.

Results

"*Candidatus* Homeothermaceae" (*S24-7*) members are found almost exclusively in the guts of homeothermic animals

The Silva database (release 119, [9]) contains over 3000 sequences designated as being within the *S24-7* family. Notably, 98 % of these sequences originate from homeothermic animals, 99 % of which are associated with the gastrointestinal system (Fig. 1a). There is a single sequence obtained from the intestine of the anchovy *Coilia mystus*. We therefore propose the name "*Candidatus* Homeothermaceae" in reference to the homeothermic preference of the family. Similar, although not as extreme, trends are observed in other *Bacteroidales* families, both in terms of homeothermic hosts and gastrointestinal preference (Fig. 1a). The association of *Bacteroidales* with feces is well documented and has led to the establishment of host-specific detection of fecal contamination based on this order [15, 16]. Of the 57 unique animal species identified as hosting "*Ca.* Homeothermaceae," the majority (96 %) are herbivores or omnivores, with many of the omnivores likely to consume a mostly herbivorous diet (e.g., chimpanzee, gorilla). The only two carnivorous hosts within the database are the dhole (*Cuon alpinus*), one of only four canids classified as hypercarnivores [17], and the sea lion. The substantial dominance of plant-based diets amongst "*Ca.* Homeothermaceae" hosts potentially reflects the metabolic capacity of the family.

"*Ca.* Homeothermaceae" population genomes

We obtained 30 near complete "*Ca.* Homeothermaceae" draft genomes from fecal metagenomic datasets: 10 from human (genomes H1 to H10), 14 from mouse (*Mus musculus*, order: *Rodentia*, family: *Muridae*; genomes M1 to M14), four from guinea pig (*Cavia porcellus*, order: *Rodentia*, family: *Caviidae*; genomes GP1 to GP4), and two from koala (*Phascolarctos cereus*, order: *Diprotodontia*, family: *Phascolarctidae*; K1 and K10). Average genome size was 2.69 Mb, with a notable outlier from the koala gut of 4.46 Mb (Additional file 1: Table S1). Abundance of each population bin within their respective metagenomic dataset varied from 0.4 to 14.8 % indicating that members of this family represent large fractions of the gut community in some animal hosts (Additional file 1: Table S1). We selected the most complete genome with the lowest inferred contamination, M4 (99.4 % complete; 0.4 % contamination), as a representative of the "*Ca.* Homeothermaceae" family, for which we propose the name "*Candidatus* Homeothermus arabinoxylanisolvens." Identification of protein orthologs between each of the genomes revealed a core of 503 proteins present in at least 28 of the 30 assembled "*Ca.* Homeothermaceae" genomes with an average of 14 % unique genes (minimum 6 % (H9), maximum 28 % (K1), Additional file 2: Table S2). Unique genes were distributed throughout each genome, including the large genome obtained from koala, with only a small number clustered in apparent genomic islands (Additional file 3: Figure S1).

Genome-based phylogenetic classification of the "*Ca.* Homeothermaceae"

The assembled "*Ca.* Homeothermaceae" population genomes were phylogenetically placed within the *Bacteriodales* order by generating a genome tree based on 120 single-copy marker genes within 300 *Bacteroidales* reference genomes obtained from NCBI (Fig. 1b). The closest characterized relatives of "*Ca.* Homeothermaceae" are members of the bacterial genera *Barnesiella* and *Coprobacter*. These genera are currently classified as members of the family *Porphyromonadaceae* but according to our genome-based inference, we propose that they

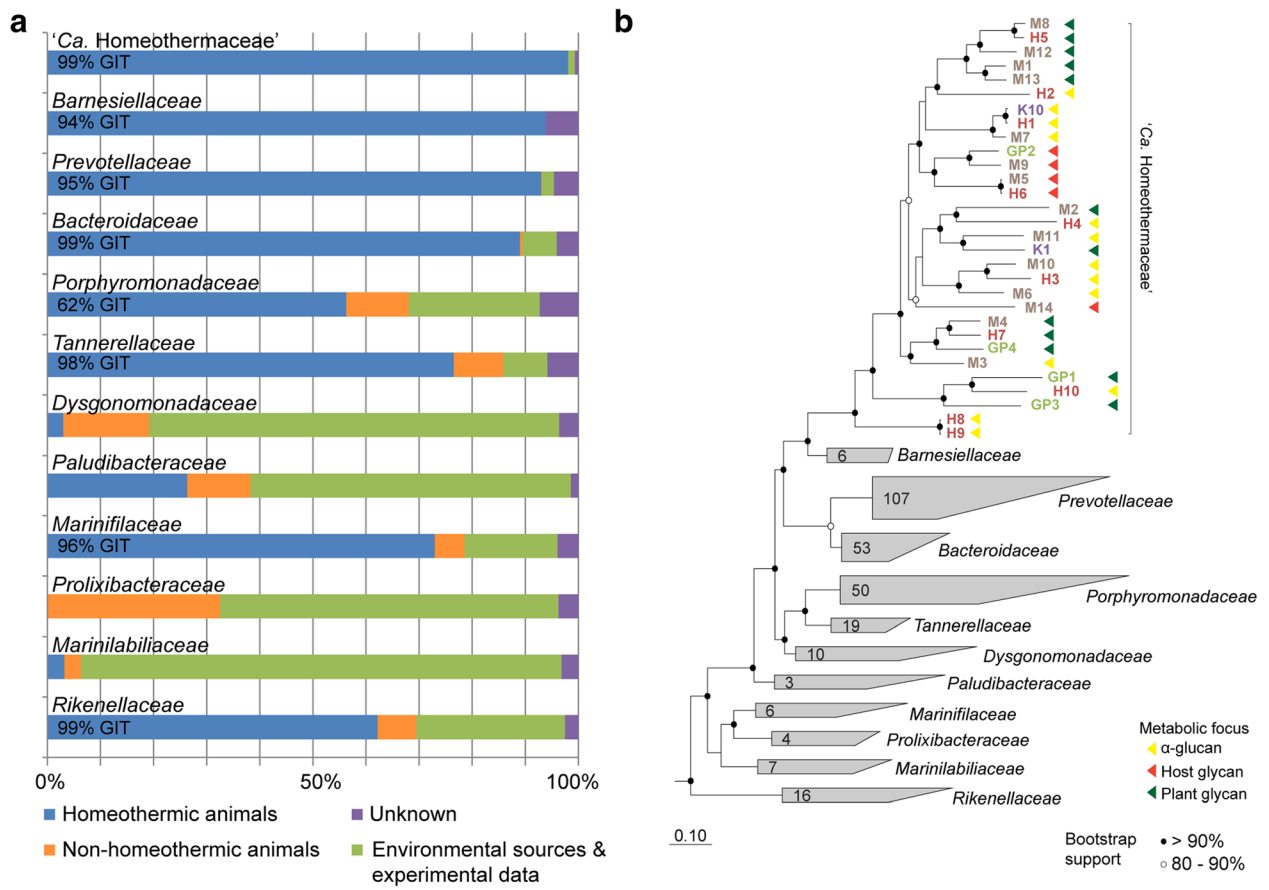

Fig. 1 Habitat and phylogenetic positioning of "*Ca.* Homeothermaceae" (*S24-7*). **a** Habitat source of 16S rRNA gene sequences described in the Silva database [9] belonging to members of the *Bacteroidales* order. Written percentages indicate the proportion of homeothermic animal samples which were derived from the gastrointestinal tract (GIT) for families dominated by homeothermic samples. Note that families in this panel have been reclassified according to the genome tree (panel **b**). **b** A maximum-likelihood tree of "*Ca.* Homeothermaceae" population genomes based on alignment of 120 concatenated marker genes. Bootstrap support derived from 100 replicates. Genome identifiers indicate host from which genome was recovered: human (*H*), mouse (*M*), guinea pig (*GP*), or koala (*K*). The metabolic focus, as determined by carbohydrate active gene annotations, is indicated by *colored triangles* following each population genome

be reclassified as a separate family, *Barnesiellaceae* fam. nov. (Fig. 1b and Additional file 4: Figure S2). Calculation of average nucleotide sequence identity (ANI) between "*Ca.* Homeothermaceae" genomes supports the dataset representing 27 different species, with only three examples of members of the same species having been sampled on two independent occasions (ANI >95 %, [18]). Two of these genome pairs were recovered from different hosts: M5 and H6, originating from mouse and human, and H1 and K10, originating from human and koala (Additional file 5: Figure S3). The third pair, H8 and H9, shares a human origin. Thus, "*Ca.* Homeothermaceae" species are not restricted to specific hosts. Above the species level, there are also five clades within the dataset sharing increased average amino acid identity that may represent distinct genera (Fig. 1b and Additional file 5: Figure S3). Nine population genomes fall outside of both inferred species and genera indicating further diversity exists within the family.

Shared features of "*Ca.* Homeothermaceae" genomes

Annotation of each genome supports the family as being composed of primary fermenters capable of producing acetate, propionate, and succinate (Fig. 2). The "*Ca.* Homeothermaceae" cell envelope is that of a diderm (Gram-negative), as demonstrated by the absence of typical monoderm protein domains and the presence of the majority of typical diderm domains, consistent with the *Bacteroidetes* phylum (Additional file 6: Table S3, [19]). A number of the genomes encode putative capsular polysaccharide synthesis loci defined by the presence of the regulatory homolog UpxY in association with a number of glycosyltransferase genes, as seen in *Bacteroides* (Additional file 7: Figure S4, [20]). Predicted fimbrial

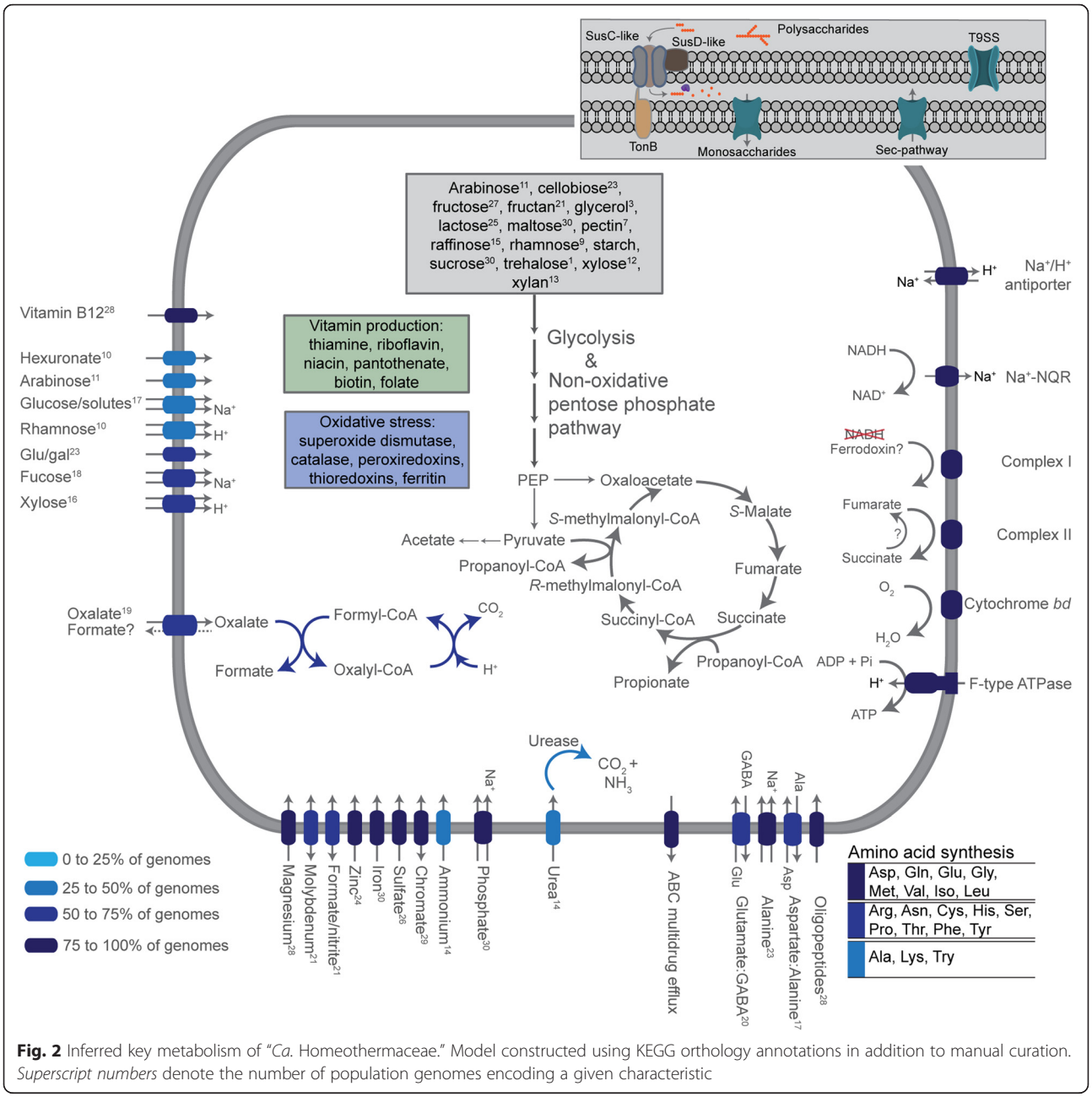

Fig. 2 Inferred key metabolism of "*Ca.* Homeothermaceae." Model constructed using KEGG orthology annotations in addition to manual curation. *Superscript numbers* denote the number of population genomes encoding a given characteristic

genes are present in 80 % of the genomes and carry the FimA domain (pfam06321), and/or the associated fimbrillin C domain (pfam15495), or, more commonly, the Mfa-like-1 (pfam13149) or Mfa2 (pfam08842) domain indicating "*Ca.* Homeothermaceae," may produce fimbriae resembling that of the periodontal pathogen *Porphyromonas gingivalis* where they have been shown to bind both host proteins and those of other bacteria, as well as promote inflammatory responses (reviewed in [21]). All "*Ca.* Homeothermaceae" contain at least one putative antibiotic efflux pump, and 60 % also encode a class A β-lactamase (Additional file 7: Figure S4), a profile that is less extensive than that of their close relatives *Bacteroides* (reviewed in [22]).

All genomes encode the capacity for the production of vitamins B_1 (thiamine), B_2 (riboflavin), B_3 (niacin), B_5 (pantothenate), B_7 (biotin), and B_9 (folate), a range which is consistent with other *Bacteroidetes* [23]. No homolog of PdxH, the final enzyme required for active vitamin B_6 production, was identified in any of the genomes, despite the remainder of the pathway being present. The lack of PdxH has also been noted in some *Bacteroides* species [23]. The complete vitamin B_{12} (cobalamin) production pathway is absent from all "*Ca.* Homeothermaceae,"

however, a subset encode a partial pathway originating from adenosyl cobyrinic acid. Vitamin B_{12} transporters were identified within 28 of the 30 genomes (Fig. 2). Therefore, members of this family are predicted to rely on neighboring populations for the production of B_{12}, an important cofactor, as is seen in other *Bacteroidetes* [24].

In addition to fermentation capacity, "*Ca.* Homeothermaceae" also encode elements of an electron transport chain indicating possible alternative modes of energy production. Complex I is found in the majority of "*Ca.* Homeothermaceae" genomes (25 of 30, Additional file 7: Figure S4) and comprises 11 of the 14 canonical subunits [25], lacking NuoEFG, the NADH dehydrogenase module. Such complexes are found in multiple phyla; however, their redox function is often unclear [26]. Of the remaining five genomes, four (GP3, GP4, M10, and M14) have loci adjacent to contig boundaries suggesting incomplete assembly as the reason for the missing elements. The remaining genome, H5, entirely lacks all components of complex I. While this genome has a relatively low level of completeness (85 %), other genomes with similar completion levels contain complex I, suggesting true absence in H5. The closest relative of this population, M8, contains a complete (11/14 subunit) complex, indicating this would be a recent loss from H5 (Fig. 1b). An F-type ATP synthase was identified in 26 "*Ca.* Homeothermaceae" genomes and is integrated within the genomic locus of complex I. Consequently, several genomes with incomplete complex I also harbor incomplete ATP synthases, including H5 in which all subunits are absent.

In addition to complex I, complex II (fumarate reductase/succinate dehydrogenase) is present in 28 genomes and is composed of three subunits: a flavoprotein, an iron-sulfur protein, and a single transmembrane protein, indicating a type B structure [27]. Both genomes with incomplete complex II gene sets, GP3 (missing two genes) and GP4 (missing all genes), were also missing elements of complex I, consistent with their lower completeness and increased fragmentation (Additional file 7: Figure S4 and Additional file 1: Table S1). The flavoprotein catalytic subunit of the complex resembles that of other related anaerobic *Bacteroidales* species suggesting fumarate as the terminal electron acceptor in an anaerobic respiratory chain, as is described in *Bacteroides* (Additional file 8: Figure S5, [28]). However, also like *Bacteroides* and most other members of the *Bacteroidales* order, the majority of "*Ca.* Homeothermaceae" genomes contain a vertically inherited aerobic reductase operon, *cydAB* (Additional file 9: Figure S6), a high-affinity bd-type oxidase induced under low oxygen conditions permitting growth in nanomolar concentrations of oxygen [29, 30]. Thus, "*Ca.* Homeothermaceae" are likely nanaerobes, able to inhabit both anoxic and marginally oxic environments [14]. Four genomes lack the *cydAB* operon (H4, H9, M9, and GP4), one of which is 97 % complete, suggesting that not all "*Ca.* Homeothermaceae" have this capacity. We were unable to confirm the operons' absence through genomic context due to a lack of synteny in the surrounding regions despite close relatives in some instances (H8-H9, M9-GP2). If truly absent, as with complex I in H5, this would indicate relatively recent independent loss of this operon from multiple "*Ca.* Homeothermaceae" lineages. The variable presence of respiratory complexes in "*Ca.* Homeothermaceae" suggests a level of energetic flexibility within the family and potentially relatively recent purging of non-essential respiratory elements in the gastrointestinal environment.

In addition to the typical electron transport chain elements, the Na^+ translocating NADH:ubiquinone oxidoreductase Nqr complex was identified in 27 "*Ca.* Homeothermaceae" genomes (Additional file 7: Figure S4). Nqr permits the use of a Na^+ gradient for energy production and suggests NADH may be oxidized by this complex in "*Ca.* Homeothermaceae" rather than by complex I, which is missing the NADH dehydrogenase module. All six Nqr subunits are present within the 27 genomes, which include H5. Three genomes, H2, H9, and M11, are missing all six components. At least one H^+/Na^+ antiporter is found within all "*Ca.* Homeothermaceae" genomes supporting the use of this system.

To support a prediction of "*Ca.* Homeothermaceae" as nanaerobes, we looked for proteins within the genomes that would provide oxidative stress protection. Superoxide dismutase (O_2^- to O_2 or H_2O_2) was identified in 24 of the genomes, and eight of these also encode a catalase (H_2O_2 to $H_2O + O_2$) protein (Additional file 7: Figure S4). All eight catalase-positive genomes were obtained from mice. Peroxide reduction may also be achieved via several peroxiredoxins that are present within the "*Ca.* Homeothermaceae" genomes. Firstly, the alkyl hydroperoxide reductase, AhpC, and associated disulfide reductase, AhpF, were identified in nine genomes. A further six contained AhpC only, representing a separate cluster within generated gene trees (Additional file 7: Figure S4 and Additional file 10: Figure S7). There are also between one and four copies of rubrerythrin in all "*Ca.* Homeothermaceae" genomes. Finally, 25 genomes encode the thiol peroxidase bacterioferritin comigratory protein. Protein stability and reduced state regeneration during oxidative stress is supported by the presence of between two and six TRX family (group I) thioredoxins within each genome and a single copy of the thioredoxin reductase TrxB in all but three genomes: GP1 and M13 lack TrxB, while M12 contains two copies. In addition, 26 genomes contain a non-heme ferritin protein permitting the storage of excess iron. Overall, "*Ca.* Homeothermaceae" members appear well equipped to

deal with oxidative stress, supporting potential microaerobic growth.

The "*Ca.* Homeothermaceae" secretome

Proteins secreted by gut-inhabiting bacteria can influence interactions with both the host and other microbes. Approximately 15 % of "*Ca.* Homeothermaceae" proteins carry the general secretory pathway signal peptide, as predicted by SignalP [31], which is at the lower end of the range predicted for Gram-negative species (13 to 42 %) [32]. Within this secretome, ~15 % of the proteins were annotated with carbohydrate-based activity and thus potentially provide nutrients for "*Ca.* Homeothermaceae" or neighboring populations. Several immune-related peptidases are also secreted: 20 of the 30 population genomes contain a metalloprotease belonging to peptidase family M6 (immune inhibitor A family) (Additional file 7: Figure S4). Members of this family have been demonstrated to degrade antimicrobial peptides [33] and components of the extracellular matrix [34] and may therefore play a role in invasiveness or persistence within the host (reviewed in [35]). In addition, 11 genomes contain an IgA degrading peptidase (peptidase family M64) that may assist with immune evasion by these populations [36].

Working in concert with the general secretory pathway, a type IX secretion system (T9SS) was also identified in the majority of the "*Ca.* Homeothermaceae" genomes. All 10 components of the system (PorK, PorL, PorM, PorN, PorP, PorT, PorU, PorV, PorW, and Sov) were present in 22 genomes (other genomes contained incomplete gene sets), in addition to the regulatory two-component sensor system, PorX (response regulator) and PorY (histidine kinase), responsible for the co-regulation of a subset of T9SS genes (Additional file 7: Figure S4, [37]). No other secretion system was identified within "*Ca.* Homeothermaceae". Within the T9SS, PorU acts as a peptidase for proteins containing a conserved C-terminal domain (TIGR04183) that dictates the use of the system and is cleaved during translocation [38, 39]. We identified 161 proteins containing this domain within the "*Ca.* Homeothermaceae" genomes, ~75 % of which also carried a general secretory pathway signal peptide, supporting their movement to the periplasm and subsequent secretion by the T9SS. The majority of proteins within this group are annotated as hypothetical (60 %); however, there is a homolog of a characterized immune-related peptidase, streptopain (SpeB). SpeB, encoded by *Streptococcus pyogenes*, contains the peptidase C10 domain and is capable of degrading multiple components of the immune system (reviewed in [40]). Streptopain homologs are present in 26 "*Ca.* Homeothermaceae" genomes.

Potential metabolic guilds within "*Ca.* Homeothermaceae"

Carbohydrate-active enzymes constitute ~6 % of "*Ca.* Homeothermaceae" coding sequences, a level similar to that of other *Bacteroidales* families (Additional file 11: Table S4). GH13 is the most abundant glycoside hydrolase family and largely comprises α-amylases, suggesting starch is a key resource of the family. In support of this, GH13 was found to be significantly ($P = 1.52E-08$) enriched in "*Ca.* Homeothermaceae" in comparison to related *Bacteroidales* families, as was the starch binding module CBM26 ($P = 1.37E-06$, Additional file 12: Table S5). The next most abundant glycoside hydrolase family is GH43 and is dominated by xylosidase and arabinosidase enzymes capable of degrading xylan and arabinan, respectively. However, this ability is not ubiquitous amongst "*Ca.* Homeothermaceae" as 12 genomes contain no genes in this category (Fig. 3). Differential abundance of such enzymes indicates potential niche partitioning, and comparative analysis across the population genomes suggests three trophic guilds with differential capacity for carbohydrate degradation: α-glucans, complex plant cell wall glycans (hemicellulose and pectin), or host-derived glycans (Figs. 1b and 3).

The α-glucan and plant glycan guilds constitute the majority of the "*Ca.* Homeothermaceae" genomes, comprising 13 and 12 genomes respectively, with the host glycan group composed of five members. We used a combination of indicator species identification and pairwise differential abundance analysis to confirm the enriched enzymes within each group, retaining those enzymes identified by both methods as defining the guild (Additional files 13 and 14: Tables S6 and S7). The α-glucan guild is the most highly selective group, with only two significantly enriched enzyme categories, both starch related (Additional file 13: Table S6). The plant guild is equipped for the degradation of arabinan, xylan, and pectin, all plant cell wall constituents. Finally, the host glycan guild is enriched in β-hexosaminidases, capable of cleaving glucosamine and galactosamine residues, α-fucosidases, capable of cleaving fucose residues and comprises the only genomes to contain the sialic acid cleaving sialidase, supporting a capacity for host glycan degradation (other enzymes carrying the GH33 domain do not display homology to known sialidase enzymes, Additional file 15: Figure S8). Integration of trophic guild membership with phylogeny reveals the presence of some clades with a shared substrate focus, while others are mixed (Fig. 1b). Each guild is also mixed in terms of host distribution, with no foci found in only one host. There are, however, dominant guilds within both guinea pig and human samples: 70 % of human origin genomes are α-glucan focused and 75 % of guinea pig origin genomes are plant focused. This may reflect diet preference of these hosts; however, more genomes

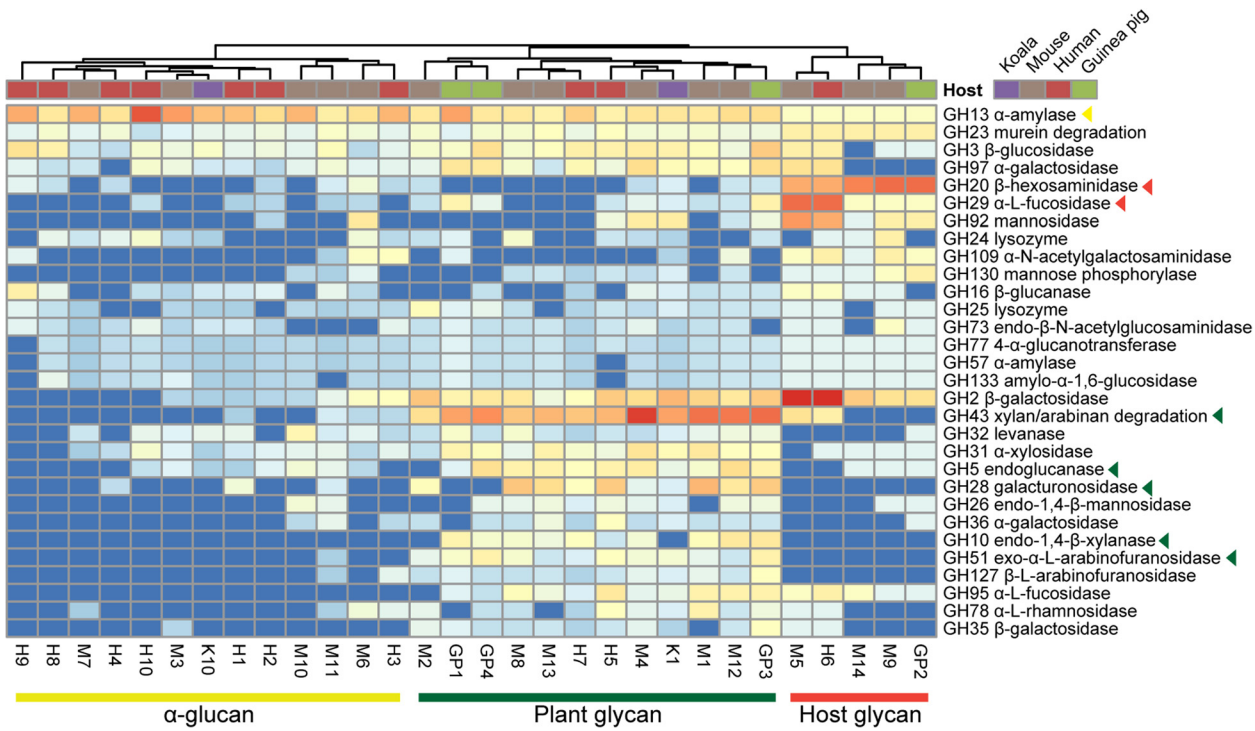

Fig. 3 Carbohydrate use differs amongst "*Ca.* Homeothermaceae" population genomes. Heatmap displaying the 30 most abundant glycoside hydrolase categories within "*Ca.* Homeothermaceae" population genomes. Colored lines denote predicted carbohydrate focus. Matching colored triangles indicate categories significantly enriched within a particular guild in comparison to both alternative guilds (Additional file 13: Table S6)

are required to provide support for this theory. We also predict that these guilds may occupy distinct spatial niches within the gut; the host glycan group primarily associating with the mucus layer, as has been demonstrated for the known mucin degrader *Akkermansia muciniphila* [41], and the plant and starch groups associating primarily with the digesta.

To provide contextual data, we generated carbohydrate-active enzyme (CAZy) profiles for a selection of microbial isolates from the gut and compared their glycoside hydrolase distribution with that of the "*Ca.* Homeothermaceae" population genomes (Additional file 15: Figure S8). The trophic guild division between "*Ca.* Homeothermaceae" genomes remained largely intact with this extended analysis; however, two genomes from the α-glucan guild, M6 and M11, became associated with members of an alternate guild, M6 with host glycan and M11 with plant glycan. Both these populations are positioned on the boundary between their two alternative guilds when broader analysis was conducted based on Clusters of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology annotation (Fig. 4a, b). In support of the predicted metabolic guilds, *A. muciniphila*, a known mucin degrader [42], was found to associate with the host glycan guild (Additional file 15: Figure S8). The plant guild associates with a single species, *Prevotella copri*, which has been demonstrated to increase in abundance following a controlled diet containing barley kernel-based bread [43]. Functional metagenomic analysis of fecal samples following this controlled diet revealed an increase in genes necessary for the degradation of polysaccharides within the bread, including xylan-degrading enzymes which were found to be enriched within the plant guild (Additional file 13: Table S6). Finally, the α-glucan guild associates with the acetogen *Blautia hydrogenotrophica*, which is known to harbor few glycoside hydrolases and consequently has been proposed to inhabit a syntrophic niche within the gut [28]. Therefore, the α-glucan guild may actually represent populations utilizing the metabolism of their neighbors and providing an alternative function in return.

Polysaccharide utilization loci

Carbohydrate degrading enzymes may be clustered together in polysaccharide utilization loci (PUL), first described in *Bacteroides thetaiotaomicron* [20] and now known to be typical of *Bacteroidetes* [44]. Such loci are defined by the presence of orthologs of the starch utilization system (Sus) components responsible for starch and maltooligosaccharide binding (SusD) and transfer to the periplasmic space (SusC) [45, 46]. Multiple *susCD*-like gene pairs are found within all "*Ca.*

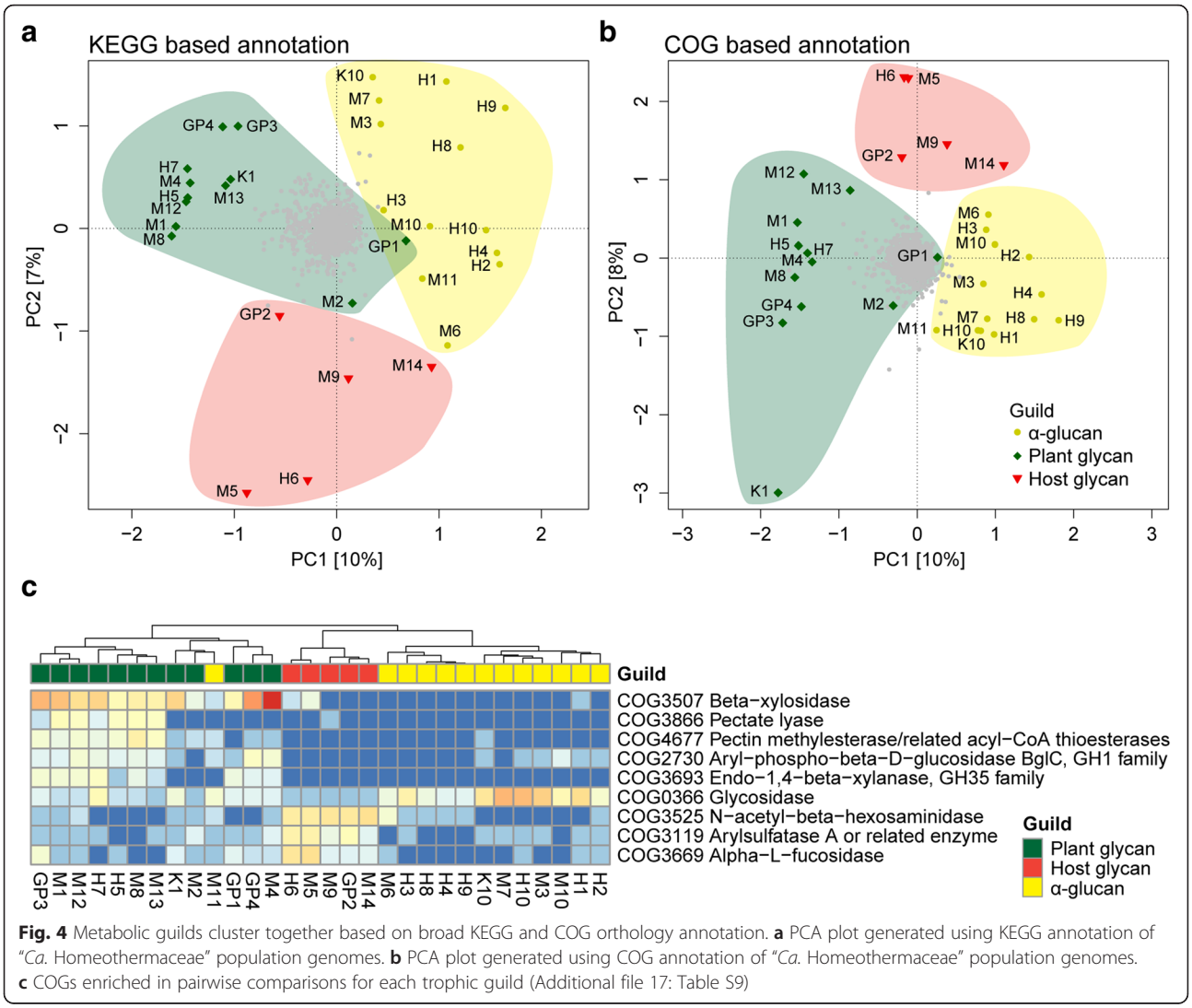

Fig. 4 Metabolic guilds cluster together based on broad KEGG and COG orthology annotation. **a** PCA plot generated using KEGG annotation of "*Ca.* Homeothermaceae" population genomes. **b** PCA plot generated using COG annotation of "*Ca.* Homeothermaceae" population genomes. **c** COGs enriched in pairwise comparisons for each trophic guild (Additional file 17: Table S9)

Homeothermaceae" genomes, and most also contain the gene pair in association with carbohydrate-active enzymes (Additional file 16: Table S8). In addition, ~10 % of "*Ca.* Homeothermaceae" PULs are located in close proximity to a hybrid two-component system protein, which have been demonstrated to regulate PUL expression (Additional file 16: Table S8) [47, 48]. While homologs of extracytoplasmic function σ-factors, also linked to PUL expression [49], were identified in many "*Ca.* Homeothermaceae" genomes, they were not located near PULs. The average number of both *susCD* pairs and pairs with associated carbohydrate enzymes is highest in the plant guild, suggesting these as requiring the most varied carbohydrate degradative machinery (Additional file 16: Table S8). Both *susC* and *susD* are present in large numbers throughout the *Bacteroidetes* with variable sequence identity existing both within and between species [20, 50, 51]. While initially described as a starch binding system, the association of *susCD* pairs with enzymes targeting a variety of substrates supports the broader use of this system, and the differential regulation of distinct *susCD* pairs in response to dietary changes has been demonstrated in *B. thetaiotaomicron* [52]. We constructed gene trees based on both genes (data not shown) to determine whether there was greater variability within a particular "*Ca.* Homeothermaceae" guild, which could reflect an ability to bind and transport a wider variety of substrates. However, average phylogenetic diversity scores [53] for all three guilds were similar for both *susC* (α-glucan:1.6, host glycan: 1.7, plant:1.5) and *susD* (α-glucan:1.8, host glycan:1.8, plant:1.7) suggesting that equivalent sequence diversity exists within each group and therefore potentially a similar, although not necessarily overlapping, diversity of substrates available to this system.

Broader functional comparative analysis of "*Ca.* Homeothermaceae"

To determine whether the metabolic guilds extend to broader genome content, we annotated each genome using both KEGG and COG orthology detection. Ordination plots generated from both annotation systems clustered the metabolic guilds discretely, although with low levels of separation (Fig. 4a, b). Only a single KEGG orthology group was found to be significantly enriched; K12373 (hexosaminidase), which was increased within the host glycan guild, consistent with the previous CAZy-based observations. COG annotation yielded several enriched protein families within each guild (Fig. 4c and Additional file 17: Table S9), all of which are associated with carbohydrate-active enzymes, indicating these as the key differentiating characteristic of each group. In addition to functions noted previously, the host glycan guild is also enriched for arylsulfatase-related enzymes (COG3119), which plays a role in the degradation of host glycans and is therefore consistent with the guild focus.

Using COG-based annotations, we then extended the analysis to compare "*Ca.* Homeothermaceae" to other *Bacteroidales* families. "*Ca.* Homeothermaceae," *Bacteroidaceae*, and *Prevotellaceae* separated clearly based on COG annotations, while *Porphyromonadaceae* members were intermingled with other families, potentially reflecting the phenotypic diversity of this family (Additional file 18: Figure S9, [54]). The abundance of over 450 individual COGs was found to be significantly different within "*Ca.* Homeothermaceae" compared to the other families; ~75 % of which were decreased (Additional file 19: Table S10). Out of those with increased abundance, several are of interest. Multiple urease-associated COGs were significantly enriched, and the urease gene cluster was subsequently identified in twelve genomes, suggesting a role in nitrogen recycling and a source of ammonia (Additional file 7: Figure S4). Gene trees produced for each of the urease subunits confirm that the presence of urease is unusual within the *Bacteroidales* and suggest lateral transfer of the gene cluster to the common ancestor of "*Ca.* Homeothermaceae" from an *Alistipes*- or *Odoribacter*-like ancestor and subsequent loss from a subset of the "*Ca.* Homeothermaceae" genomes (Additional file 20: Figure S10, data not shown). Putative formyl-CoA transferases (COG1804) were also enriched in "*Ca.* Homeothermaceae," revealing the presence of the oxalate degrading gene pair formyl-CoA transferase and oxalyl-CoA decarboxylase in 19 of the 30 genomes (Additional file 7: Figure S4). Oxalate is likely transported into the cell via a permease located adjacent to oxalyl-CoA decarboxylase in all 19 oxalate degraders (and which is not found in "*Ca.* Homeothermaceae" lacking the oxalate degrading gene pair) as suggested by genomic analysis of other oxalate degrading species [55]. Oxalate degradation appears to be linked to metabolic guilds, that is, 10 of 12 plant, and four of five host glycan-focused "*Ca.* Homeothermaceae" have oxalate degrading genes, whereas only five of the 13 α-glucan-focused guild have them. As with urease, gene trees confirm that this function is rare within the *Bacteroidales* (Additional file 21: Figure S11, data not shown).

Relative abundance and prevalence within the sampled mammalian hosts

A large portion of "*Ca.* Homeothermaceae" sequences within 16S rRNA databases originate from mice suggesting high prevalence of the family in the murine host. To determine whether this reflects true prevalence or database bias, we searched for evidence of "*Ca.* Homeothermaceae" in available gut metagenome datasets from both mice and humans. We found "*Ca.* Homeothermaceae," as represented by the analyzed population genomes, present in ~50 % of mice samples, with an average relative abundance of 6 % of the gut community as estimated by read mapping (from ~100 metagenomes, Additional file 22: Table S11). The prevalence in humans was lower, ~20 %, with an average abundance of ~2 % of the community (from ~300 metagenomes). "*Ca.* Homeothermaceae" is therefore more common in mice than humans but is nonetheless likely to be present in a sizable fraction of the human population.

The most prevalent "*Ca.* Homeothermaceae" species was H8/H9 in 20 % of human datasets, and M6 in 35 % of murine datasets (Additional file 22: Table S11), both members of the α-glucan guild. We were interested to see if this prevalence was consistent across different dietary backgrounds as a potential source of support for our proposed guild structure. To do this, we subdivided the analyzed datasets according to available diet-related metadata. We found mice fed a high-fat diet carried a narrower range of "*Ca.* Homeothermaceae" populations, and these were present at a lower abundance than those fed a standard chow diet (Additional file 22: Table S11). However, no particular guild showed dominance, and M6 remained the most prevalent population overall. Within the public human datasets sampled, very few non-α-glucan guild representatives were detected (Additional file 22: Table S11) in agreement with the dominance of this guild within the "*Ca.* Homeothermaceae" genomes recovered from humans and suggestive of the human diet as most supportive of members of the α-glucan guild. We found a higher prevalence of "*Ca.* Homeothermaceae" in obese individuals (23 %) than in lean (10 %), suggesting that a higher energy diet may better support the family, although there was no dietary composition information available for this dataset. We also identified a substantial increase in the prevalence of

"*Ca.* Homeothermaceae" within the Hadza hunter-gatherer population; 70 % of individuals were found to carry at least one population. Only two species were identified within this group of individuals: the typically prevalent H8/H9 and species H4, also a member of the α-glucan guild. The Hadza diet is heavily plant based, composed of tubers, meat, honey, foliage, and berries, with tubers being particularly important due to their constant availability [56]. Tubers consumed by the Hadza contain a large portion of indigestible fibers that are expectorated after chewing. As such, tubers provide a source of moisture, simple sugars, starch, and soluble fiber [56]. The increased prevalence of "*Ca.* Homeothermaceae" within the Hadza suggests this diet is particularly amenable to the maintenance of the identified species, although does not result in any increase in their overall abundance.

Discussion

The *Bacteroidales* family *S24-7* is encountered frequently in culture-independent studies and is gaining recognition due to both its prevalence, particularly in murine-based datasets, and its fluctuating abundance in cross-sectional and intervention type studies. Increased abundance has been described in mice fed a low-fat diet and, in association with increased exercise [57], in diabetes-sensitive mice fed a high-fat diet [10] and following remission of colitis in a mouse model [11]. *S24-7* is also the dominant family during hibernation of arctic ground squirrels [58]. The consequence of these fluctuations in the abundance of *S24-7* is currently unknown, as they remain uncultured and no genomic studies have been undertaken. Here, we recover a set of *S24-7* population genomes from metagenomic samples, enabling inference of their core metabolism, and propose the name "*Ca.* Homeothermaceae" reflecting an ecological distribution limited to the guts of homeothermic animals (Fig. 1a).

Carbohydrate composition and availability is known to be a primary driver of microbial community structure in gut ecosystems [48, 59–61]. We identified three trophic guilds within the "*Ca.* Homeothermaceae" based on their encoded carbohydrate-active enzymes (Fig. 3) that suggest the family has the capacity to occupy multiple niches within the gut, which may include spatial partitioning. Similar metabolic differentiation is observed in other gut-inhabiting genera such as *Bacteroides* [48, 50, 62] and *Bifidobacterium* [61] where different species encode alternative carbohydrate utilization machinery. Some gut inhabitants may be able to occupy multiple carbohydrate-based trophic niches: *B. thetaiotaomicron* displays a preference for diet-derived polysaccharides, such as xylan-, pectin-, and arabinose-based compounds, over host glycans; however, when such polysaccharides are scarce, host glycan degradation activity is upregulated [52]. "*Ca.* Homeothermaceae" guild representatives may also utilize this strategy; all analyzed population genomes encode starch-degrading α-amylases and as such, while they may have a preference for their specialist carbohydrate source, starch can serve as a foundation carbohydrate for all family members.

We identified two characteristics within a subset of "*Ca.* Homeothermaceae" populations that were relatively unusual for members of the *Bacteriodales*: the capacity for oxalate degradation and the production of urease (Additional files 20 and 21: Figure S10 and S11). Plants are the primary source of dietary oxalate, which in excess contributes to the formation of renal stones by complexing calcium (reviewed in [63]). In agreement with this dietary source, the majority of "*Ca.* Homeothermaceae" populations within the plant-focused guild encode the necessary components for oxalate degradation. Oxalate degrading potential is also encoded within four of the five host glycan guild population genomes while only 40 % of α-glucan guild members contain the necessary genes (Additional file 7: Figure S4). *Oxalobacter formigenes* is the best characterized oxalate degrading gut bacterium, and colonization is associated with the decreased incidence of calcium oxalate stone formation [64]. While *O. formigenes* is dependent on the presence of oxalate and uses it as a sole carbon source [65], other oxalate degraders, such as lactic acid bacteria, require the presence of an additional carbon source [55, 66]. "*Ca.* Homeothermaceae" oxalate degraders are likely in the latter category due to sporadic distribution of the trait across the family. This distribution appears to be the result of lineage-specific loss rather than multiple independent lateral acquisitions (Additional file 20: Figure S10). Urease releases ammonia from urea, which can then be incorporated into microbial amino acids [67] and contributes to nitrogen level stability of the host particularly when protein consumption is low [68]. Urease activity can therefore be advantageous for both host and microbe. However, urease-positive microbes can be detrimental in combination with elevated circulating ammonia levels associated with liver disease [69]. Urease is also a recognized virulence factor in both bacterial and fungal infection (reviewed in [70]). The abundance of urease genes within the gut microbiota in humans differs with age and geography and is potentially reflective of diet [71]. We identified urease positive "*Ca.* Homeothermaceae" populations in all four hosts (Additional file 7: Figure S4), and presence did not correlate with a particular trophic guild making it difficult to predict a role for urease within the group; however, a metabolic role for the enzyme is supported by the presence of glutamine synthetase in the majority of genomes (Fig. 2).

A key question relating to newly characterized members of the microbiota is whether they are friend or foe. "*Ca.* Homeothermaceae" representatives have been shown to be IgA coated [12]. IgA production is induced by both commensal and pathogenic intestinal inhabitants, and both are believed to be able to induce the production of highly specific IgA leading to microbial cell coating [72]. "*Ca.* Homeothermaceae" are found in both IgA+ and IgA− fractions of fecal microbiota [12, 13], however, on the whole, are not highly coated with IgA in contrast to families known to be inflammatory in murine models of colitis, such as *Prevotellaceae* [12]. The presence within some "*Ca.* Homeothermaceae" genomes of an IgA protease (Additional file 7: Figure S4) may contribute to their identification within both IgA+ and IgA− community fractions, although the significance of IgA proteases in vivo remains unclear (reviewed in [73]). The majority of "*Ca.* Homeothermaceae" genomes also contain a homolog of SpeB (Additional file 7: Figure S4), a peptidase capable of degrading multiple immunologically relevant proteins (reviewed in [40]). SpeB homologs are found in other gut bacteria including *Bacteroides fragilis* and *B. thetaiotaomicron* [74, 75] and in the periodontal pathogen *Prevotella intermedia* where the homolog interpain A is involved in the inhibition of the immune response via complement degradation [76]. The presence of multiple potential immune evasion peptidases within members of "*Ca.* Homeothermaceae" does not preclude a typically commensal relationship with the host; however, they may provide the capacity for opportunistic infection under the appropriate conditions [75].

Conclusions

Overall, this study provides the first genomic insights into the uncultured gut-inhabiting "*Ca.* Homeothermaceae" family through the comparative analysis of 30 draft genomes obtained from metagenomic datasets. We describe varied carbohydrate utilization mechanisms existing within the family in agreement with other genera occupying the same environmental niche. As a group that is particularly prevalent within a key experimental environment, the mouse gut (and also present in the human gut and potentially relevant to human health), further reports confirming the roles of "*Ca.* Homeothermaceae" in vivo are likely to appear in the future.

Description of "*Candidatus* Homeothermus"

Homeothermus (Ho.me.o.ther'mus Gr. adj. *homoios*, similar, Gr. n. *thermē*, heat. *Homeothermus* of homeothermic origin). Inferred to be Gram-negative, nonmotile, nanaerobic, and able to ferment a wide range of carbohydrates including arabinose, cellobiose, fructan, fructose, glycerol, lactose, maltose, raffinose, sucrose, xylan, and xylose, with a focus on arabinan and xylan based on enzyme abundance.

Description of "*Ca.* Homeothermus arabinoxylanisolvens"

Homeothermus arabinoxylanisolvens (Ho.me.o.ther'mus Gr. adj. *homoios*, similar, Gr. n. *thermē*, heat. a.rab.in.o.xy.lan.i.sol'vens n. *arabin*, a carbohydrate derived from gum arabic, M.L. n. *xylan*, xylan, M. L. part. adj. *solvere*, to loosen, untie, free up). Description is the same as that for genus "*Ca.* Homeothermus." Represented by population genome M4 (acc. no. LUJO00000000) obtained from metagenomic sequencing of *Mus musculus* fecal sample.

Description of "*Ca.* Homeothermaceae"

The description is the same as for the genus "*Ca.* Homeothermus" with the following additions; *-aceae* ending to denote a family. Additional fermentation substrates include pectin, rhamnose, and trehalose. Three trophic guilds are proposed within the family based on the relative abundance of carbohydrate-active enzymes with different substrates: α-glucans, complex plant cell wall components, and host glycans. Type genus: "*Ca.* Homeothermus." Order: *Bacteroidales*.

Emended description of the family *Porphyromonadaceae* Krieg, Staley et al. 2011

The description is the same as that given by Krieg et al. [54] with the following amendment. The genera, *Barnesiella*, *Butyricimonas*, *Coprobacter*, *Dysgonomonas*, *Odoribacter*, *Paludibacter*, *Parabacteroides*, *Proteiniphilum*, "*Ca.* Sanguibacteroides," and *Tannerella* have been removed as they do not form a monophyletic group with the type genus, *Porphyromonas.*

Description of *Barnesiellaceae* fam. nov.

Includes the genera *Barnesiella* (type genus) and *Coprobacter*. Description is drawn from that of *Barnesiella* given by Sakamoto et al. [77] and *Coprobacter* given by Shkoporov et al. [78]: *-aceae* ending to denote a family. Cells are Gram-negative, obligately anaerobic, non-spore-forming, and non-motile. Saccharolytic. Type genus: *Barnesiella*. Order: *Bacteroidales*.

Description of *Dysgonamonadaceae* fam. nov.

Includes the genera *Dysgonomonas* (type genus) and *Proteiniphilum*. Description is drawn from that of *Dysgonomonas* given by Hofstad et al. [79] and *Proteiniphilum* given by Chen et al. [80]: *-aceae* ending to denote a family. Cells are Gram-negative, fermentative, and facultatively (*Dysgonomonas*) or obligately (*Proteiniphilum*) anaerobic. Type genus: *Dysgonomonas*. Order: *Bacteroidales*.

Emended description of the family *Marinifilaceae* Iino, Mori et al. 2014

The description is drawn from that of *Marinifilaceae* given by Iino et al. [81] with the following amendment. The family *Marinifilaceae* contains the genera *Butyricimonas*, *Marinifilum* (type genus), *Odoribacter*, and "*Ca.* Sanguibacteroides." Cells are Gram-negative, non-spore-forming, non-motile, and facultatively (*Marinifilum*) or obligately (*Butyricimonas*, *Odoribacter*) anaerobic.

Description of *Paludibacteraceae* fam. nov.

Includes the genus *Paludibacter* (type genus). Description is the same as for the genus *Paludibacter* given by Ueki et al. [82]: *-aceae* ending to denote a family. Type genus: *Paludibacter*. Order: *Bacteroidales*.

Description of *Tannerellaceae* fam. nov.

Includes the genera *Parabacteroides* and *Tannerella* (type genus). Description is drawn from that of *Tannerella* given by Sakamoto et al. [83] and *Parabacteroides* given by Sakamoto et al. [84]: *-aceae* ending to denote a family. Cells are Gram-negative, non-motile, and obligately anaerobic. Type genus: *Tannerella*. Order: *Bacteroidales*.

Methods

Sample collection and sequencing

Fecal samples were obtained from six 12-week-old female C57BL/6 mice housed in accordance with the University of Newcastle Animal Care and Ethics Committee; reference number A-2013-303. DNA was extracted from feces using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, CA, USA) according to the manufacturer's instructions. Library preparation was performed using the Nextera DNA Library Preparation Kit (Illumina, CA, USA). Libraries were sequenced at the Diamantina Institute, The University of Queensland, using the Illumina HiSeq 2500 platform generating ~9 Gbp of 100 bp paired-end reads per sample.

Koala fecal samples originated from a 12-year-old male as previously described [85]. Public metagenomic datasets generated from fecal samples of both human and guinea pig were downloaded from the NCBI sequence read archive (SRA); human samples were obtained from multiple projects, specifically runs (ERR209459, ERR209707, ERR525737, ERR688517, ERR688528, ERR710427, ERR710429, SRR413598, SRR413599); guinea pig samples were obtained from a single project (BioProject: SRP012966).

Sequence assembly and population genome recovery

Reads from mouse fecal samples were adapter trimmed and merged using SeqPrep (https://github.com/jstjohn/SeqPrep) then remaining pairs were quality trimmed using Nesoni v0.128 (https://github.com/Victorian-Bioinformatics-Consortium/nesoni) with a minimum Phred quality threshold of 20. Assembly of pooled reads was performed using CLC Genomics Workbench v7.0.4 (QIAGEN, Aarhus, Denmark) using a word size of 30 and bubble size of 1500. Scaffolding was performed during assembly, and reads were mapped back to contigs with default settings. Minimum contig length was 300. Mapping of reads to final assemblies was performed using BWA v0.7.10 [86] with default settings.

SRA data from guinea pig and human datasets was quality and adapter trimmed using Trimmomatic v0.3.2 [87] with default settings plus a head crop of 10 and minimum read length of 30. Trimmed reads were merged using BBMerge (https://sourceforge.net/projects/bbmap/). Assembly was performed either per individual run (human) or pooled (guinea pig) using CLC Genomics Workbench v8.5.1 (QIAGEN, Aarhus, Denmark) either using default settings (human) or using a word size of 30 and a bubble size of 1000 (guinea pig). Minimum contig length was 500. Scaffolding was performed during assembly, and reads were mapped back to contigs with default settings. Gap filling was performed on assemblies from human datasets using Abyss-sealer [88] with default settings. Mapping of reads to final assemblies was performed using BamM v1.5.1 (http://ecogenomics.github.io/BamM/) with default settings.

Population genomes were obtained either from previous studies [85] or de novo from metagenomic datasets using GroopM v0.2 ([89], mouse) or MetaBAT v0.25.4 ([90], human, guinea pig). Phylogenetic analysis and estimation of contamination and completeness of recovered genomes was performed using CheckM v1.0.3 [91], which utilizes a set of single-copy marker genes. "*Ca.* Homeothermaceae" genomes were refined by removing contigs with incongruent coverage profiles as identified by RefineM v0.0.3 (https://github.com/dparks1134/RefineM). Gap filling was performed on these assemblies using FinishM v0.0.6 (https://github.com/wwood/finishm). Reads mapping to each genome were extracted from a complete assembly mapping (incorporating refined genomes and additional unrefined genomes as the reference) using BamM v1.5.1 (http://ecogenomics.github.io/BamM/) then remapped to each individual genome using CLC Genomics Workbench v8.5.1 (QIAGEN, Aarhus, Denmark) with default settings. These mappings were used for manual investigation of specific genomic features.

Phylogenetic resolution

A maximum-likelihood tree of "*Ca.* Homeothermaceae" population genomes in the context of the order *Bacteroidales* was generated based on an alignment of 120 concatenated single-copy, bacterial-specific, marker genes implemented within an in-house pipeline. Marker genes from 300 *Bacteroidales* and 37 *Sphingobacteriales*

(outgroup) genomes were obtained from publicly available genomes within the NCBI database and aligned using Mafft v7.221 [92]. A maximum-likelihood tree was inferred using FastTree v2.1.7 [93] under the WAG + GAMMA model based on alignment positions containing a residue within ≥90 % of sequences (36,713 positions). Bootstrap support was derived from 100 replicates.

Individual gene trees were generated using FastTree v2.1.7 [93] implemented within Mingle v0.0.15 (https://github.com/Ecogenomics/mingle) utilizing the BLAST [94] workflow, identifying homologs within the IMG database [95]. Default Mingle settings used for all genes except for oxalyl-CoA decarboxylase where percent identity was increased to 40 % due to an alignment length of <100 amino acids with default settings. Bootstrap support was derived from 100 replicates, also implemented within Mingle. Phylogenetic diversity scores for *susC* and *susD* gene trees were calculated using the picante R package [96]. All trees were visually inspected using ARB v6.0.2 [97].

Genome annotation and core metabolic analysis

Genomes were annotated using Prokka v1.11 [98]. All subsequent gene-based analysis was performed using the output of this annotation process. Orthologous proteins within each genome were identified using Proteinortho v5.11 [99]. Average nucleotide identity was calculated using the Goris method [18] implemented in calculate_ani.py (available at: https://github.com/widdowquinn/scripts/blob/master/bioinformatics/calculate_ani.py). Average amino acid identity between each genome pair was calculated using CompareM v0.0.5 (https://github.com/dparks1134/CompareM). Protein families were identified using pfam_scan.pl against the Pfam database release 28 [100] employing HMMER v3.1b2 [101]. Archetypal cell envelope families as per Albertsen et al. [19] were used for predicting cell structure. TIGRFAMs were identified using hmmscan from HMMER v3.1b2 [101] against database release 15.0 (downloaded from ftp://ftp.jcvi.org/pub/data/TIGRFAMs/). Antibiotic resistance genes were predicted using the Resfams database [102].

Kyoto Encyclopedia of Genes and Genomes (KEGG) term annotation was performed using KAAS [103] and KEGG maps in combination with RAST [104], KBase (http://kbase.us/), and Pathway Tools v19.0 [105], and curated gene lists [106] were used to elucidate the general metabolic profile of "*Ca.* Homeothermaceae." Clusters of Orthologous Groups (COG) profiles were identified using BLASTP v2.2.30+ [94] against 2014 update of the 2003 COG database downloaded from NCBI (http://www.ncbi.nlm.nih.gov/COG/, last accessed July 2015) with e-value cutoff of 1e-6.

Carbohydrate-active enzymes (CAZy) were identified using hmmscan from HMMER v3.1b2 [101] against the dbCAN database v4 [107]. Percentage abundance of each CAZy category was assessed against the total number of genes with CAZy annotation within each genome. Signal peptide sequences within CAZy annotated genes were predicted using SignalP v4.1 [31] and LipoP v1.0 [108] with a margin cutoff of 4 for LipoP. Prediction of membrane positioning was based on the presence of a transmembrane domain (SignalP) or a type II signal peptide (LipoP).

Differential abundance comparisons

Differential abundance of annotated features (COG, KO, CAZy) was analyzed using DESeq2 R package [109] based on count data. Heatmaps were generated using pheatmap [110] following variance stabilizing normalization of data by DESeq2. Indicator annotations were identified using the labdsv R package [111]. PCA plots were created using the vegan R package [112].

Prevalence and relative abundance

Prevalence of each population genome and overall abundance of "*Ca.* Homeothermaceae" were assessed by mapping public human and mouse gut metagenome datasets downloaded from the SRA against all genomes using BWA v0.7.12 [86]. Genome coverage amounting to ≥0.5 % of total reads was used as the minimum cutoff for the presence of a given population in a particular sample. Cutoff was based on minimum coverage within read mappings of datasets used to produce "*Ca.* Homeothermaceae" population bins and thus represents a conservative estimate. Relative abundance of each population was based on the percentage of total reads attributed to each genome exceeding the minimum cutoff percentage, normalized for genome size. Diet-related datasets were PRJEB7759 (mouse), PRJNA278393 (Hadza), and PRJEB2054 (lean vs. obese). Additional human samples originated from projects PRJEB1220, PRJEB4410, PRJEB6456, and PRJEB7774

Additional files

Additional file 1: Table S1. "*Ca.* Homeothermaceae" population genome properties. (DOCX 19 kb)

Additional file 2: Table S2. Unique genes within each "*Ca.* Homeothermaceae" population genome. (DOCX 14 kb)

Additional file 3: Figure S1. Distribution of unique genes in selected "*Ca.* Homeothermaceae" population genomes. Coding sequences within each genome are denoted by vertical yellow lines, unique genes are denoted by vertical blue lines. Classification of unique genes is based on failure to find an ortholog to a given gene in any of the other "*Ca.* Homeothermaceae" genomes using Proteinortho [99]. (TIF 249 kb)

Additional file 4: Figure S2. Phylogenetic tree of the order *Bacteroidales*. Maximum-likelihood tree of the *Bacteroidales* based on the concatenated alignment of 120 concatenated marker genes (36,713 amino acids) using genomes available within the NCBI database. Bootstrap support derived from 100 replicates. (TIF 281 kb)

Additional file 5: Figure S3. Average nucleotide and amino acid identity between "*Ca.* Homeothermaceae" population genomes. ANI calculated using Goris method [18] implemented in calculate_ani.py (https://github.com/widdowquinn/scripts/blob/master/bioinformatics/calculate_ani.py). AAI calculated using CompareM v0.0.5 (https://github.com/dparks1134/CompareM). Values over 95 % indicate population genomes originating from the same species. (TIF 2541 kb)

Additional file 6: Table S3. Cell wall associated PFAMs identified in "*Ca.* Homeothermaceae" genomes. (XLSX 13 kb)

Additional file 7: Figure S4. Presence of described characteristics within "*Ca.* Homeothermaceae" population genomes. Presence determined via BLAST [94] search of annotated proteins within each population genome. AR: antibiotic resistance. (TIF 724 kb)

Additional file 8: Figure S5. Gene tree of *frdA*, catalytic subunit of complex II. Maximum-likelihood gene tree was inferred using FastTree 2 [93] based on a 640 amino acid alignment of sequences, implemented within the in-house script Mingle (https://github.com/Ecogenomics/mingle). Bootstrap values represent result of 100 replicates. "*Ca.* Homeothermaceae" *frdA* genes shown in blue. Where shown, tips display IMG genome ID, species name and gene annotation. (TIF 475 kb)

Additional file 9: Figure S6. Gene tree of *cydA*, cytochrome *bd* subunit. Maximum-likelihood gene tree was inferred using FastTree 2 [93] based on a 500 amino acid alignment of sequences, implemented within the in-house script Mingle (https://github.com/Ecogenomics/mingle). Bootstrap values represent result of 100 replicates. "*Ca.* Homeothermaceae" *cydA* genes shown in blue. Where shown, tips display IMG genome ID, species name and gene annotation. (TIF 347 kb)

Additional file 10: Figure S7. Gene tree of *ahpC*, alkyl hydroperoxide. Maximum-likelihood gene tree was inferred using FastTree 2 [93] based on a 180 amino acid alignment of sequences implemented, within the in-house script Mingle (https://github.com/Ecogenomics/mingle). Bootstrap values represent result of 100 replicates. "*Ca.* Homeothermaceae" *ahpC* genes shown in blue. Where shown, tips display IMG genome ID, species name and gene annotation. (TIF 787 kb)

Additional file 11: Table S4. Top 20 most abundant CAZy categories within "*Ca.* Homeothermaceae" as a percentage of total genes with CAZy annotation in each genome. (DOCX 17 kb)

Additional file 12: Table S5. CAZy categories with significantly different abundance in "*Ca.* Homeothermaceae" in comparison with collated counts from *Bacteroidacea*e, *Prevotellaceae*, and *Porphyromonadaceae. (DOCX 17 kb)*

Additional file 13: Table S6. Significantly enriched carbohydrate-active enzymes within each trophic guild. (DOCX 18 kb)

Additional file 14: Table S7. Indicator carbohydrate-active enzymes identified within each trophic guild. (DOCX 23 kb)

Additional file 15: Figure S8. 50 most abundant glycoside hydrolases in "*Ca.* Homeothermaceae" and selected gut-associated species. Heatmap displaying the top 50 most abundant GH enzymes within "Ca. Homeothermaceae" in addition to a selection of reference genomes available on NCBI also originating from fecal samples. (TIF 1691 kb)

Additional file 16: Table S8. *SusCD*-like genes present in "*Ca.* Homeothermaceae" genomes. (DOCX 17 kb)

Additional file 17: Table S9. Significantly enriched COGs within each trophic guild. (DOCX 15 kb)

Additional file 18: Figure S9. COG category abundance between "*Ca.* Homeothermaceae" and related *Bacteroidales* families. PCA plots generated using annotated COGs within "*Ca.* Homeothermaceae" and IMG genomes [95] from other families. Family designation of each genome is based on IMG phylogenetic annotation except for *Barnesiellaceae*, which includes *Barnesiella* and *Coprobacter* (previously *Porphyromonadaceae*). (TIF 550 kb)

Additional file 19: Table S10. COGs with significantly altered abundance in "*Ca.* Homeothermaceae" versus related *Bacteroidales* families. (XLSX 57 kb)

Additional file 20: Figure S10. Gene tree of *ureA*, urease subunit. Maximum-likelihood gene tree was inferred using FastTree 2 [93] based on a 100 amino acid alignment of sequences, implemented within the in-house script Mingle (https://github.com/Ecogenomics/mingle). Bootstrap values represent result of 100 replicates. "*Ca.* Homeothermaceae" *ureA* genes shown in blue. Where shown, tips display IMG genome ID, species name and gene annotation. (TIF 470 kb)

Additional file 21: Figure S11. Gene tree of *oxa*, oxalyl-CoA decarboxylase. Maximum-likelihood gene tree was inferred using FastTree 2 [93] based on a 580 amino acid alignment of sequences, implemented within the in-house script Mingle (https://github.com/Ecogenomics/mingle). Bootstrap values represent result of 100 replicates. "*Ca.* Homeothermaceae" *oxa* genes shown in blue. Where shown, tips display IMG genome ID, species name and gene annotation. (TIF 296 kb)

Additional file 22: Table S11. Prevalence and relative abundance of "*Ca.* Homeothermaceae" populations within public metagenomics datasets from human and mouse gut. (DOCX 25 kb)

Abbreviations

ANI, average nucleotide identity; AR, antibiotic resistance; CAZy, carbohydrate-active enzyme; COG, Clusters of Orthologous Groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; PUL, polysaccharide utilization loci; Sus, starch utilization system

Acknowledgements

We thank Nicola Angel and Serene Low for the library preparation and sequencing; Donovan Parks, Adam Skarshewski, and Pierre-Alain Chaumeil for providing early access to the Genome Tree Database; and Jennifer Hosmer, Emily Hoedt, Gene Tyson, and Rick Webb for their fruitful discussions.

Funding

KLO was supported by the NHMRC Program Grant APP1071822, SLG was supported by a Lung Foundation Australia Fellowship, and the mouse work was supported by the NHMRC Project Grant APP1059239.

Availability of data and materials

The population genomes supporting the conclusions of this article have been uploaded to the NCBI BioProject accession number PRJNA313232.

Authors' contributions

KLO, JND, and JP generated the population genomes. KLO and DLAW performed the bioinformatic analysis. NL, SLG, and PMH performed the sample collection and generation of murine data. KLO, CGOD, RWP, and LKN performed the metabolic analysis and interpretation of the results. LKN, MAC, MM, PMH, and PH assisted in study design and data interpretation. KLO and PH wrote the manuscript. All authors read, edited, and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

Fecal samples were obtained in accordance with the University of Newcastle Animal Care and Ethics Committee; reference number A-2013-303.

Author details

[1]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia. [2]Priority Research Centre for Healthy Lungs, The University of Newcastle and Hunter Medical Research Institute, Newcastle, Australia. [3]QFAB Bioinformatics, The University of Queensland, Brisbane, Australia. [4]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Australia. [5]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. [6]Microbial Biology and Metagenomics, The University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia.

Received: 31 March 2016 Accepted: 23 June 2016
Published online: 07 July 2016

References

1. Sekirov I, Tam NM, Jogova M, Robertson ML, Li YL, Lupp C, et al. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. Infect Immun. 2008;76(10):4726–36. doi:10.1128/iai.00319-08.
2. Cryan JF, O'Mahony SM. The microbiome-gut-brain axis: from bowel to behavior. Neurogastroenterol Motil. 2011;23(3):187–92. doi:10.1111/j.1365-2982.2010.01664.x.
3. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13(4):260–70. doi:10.1038/nrg3182.
4. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915–20. doi:10.2307/3841877.
5. Wlodarska M, Kostic Aleksandar D, Xavier Ramnik J. An integrative view of microbiome-host interactions in inflammatory bowel diseases. Cell Host Microbe. 2015;17(5):577–91. doi:10.1016/j.chom.2015.04.008.
6. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011;9(4):279–90. doi:10.1038/nrmicro2540.
7. Salzman NH, de Jong H, Paterson Y, Harmsen HJ, Welling GW, Bos NA. Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. Microbiology. 2002;148(Pt 11):3651–60.
8. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012;6(3):610–8. doi:10.1038/ismej.2011.139.
9. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):D590–6. doi:10.1093/nar/gks1219.
10. Serino M, Luche E, Gres S, Baylac A, Bergé M, Cenac C, et al. Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota. Gut. 2012;61(4):543–53. doi:10.1136/gutjnl-2011-301012.
11. Rooks MG, Veiga P, Wardwell-Scott LH, Tickle T, Segata N, Michaud M, et al. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. ISME J. 2014;8(7):1403–17. doi:10.1038/ismej.2014.3.
12. Palm Noah W, de Zoete Marcel R, Cullen Thomas W, Barry Natasha A, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014;158(5):1000–10. doi:10.1016/j.cell.2014.08.006.
13. Bunker Jeffrey J, Flynn Theodore M, Koval Jason C, Shaw Dustin G, Meisel M, McDonald Benjamin D, et al. Innate and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A. Immunity. 2015;43(3):541–53. doi:10.1016/j.immuni.2015.08.007.
14. Morris RL, Schmidt TM. Shallow breathing: bacterial life at low O_2. Nat Rev Microbiol. 2013;11(3):205–12. doi:10.1038/nrmicro2970.
15. Dick LK, Bernhard AE, Brodeur TJ, Domingo JWS, Simpson JM, Walters SP, et al. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. Appl Environ Microbiol. 2005;71(6):3184–91. doi:10.1128/aem.71.6.3184-3191.2005.
16. Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol. 2006;72(6):4214–24. doi:10.1128/aem.01036-05.
17. Van Valkenburgh B. Iterative evolution of hypercarnivory in canids (*Mammalia*: *Carnivora*): evolutionary interactions among sympatric predators. Paleobiology. 1991;17(4):340–62.
18. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57(1):81–91. doi:10.1099/ijs.0.64483-0.
19. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol. 2013;31(6):533–8. doi:10.1038/nbt.2579.
20. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, et al. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science. 2003;299(5615):2074–6. doi:10.1126/science.1080029.
21. Yoshimura F, Murakami Y, Nishikawa K, Hasegawa Y, Kawaminami S. Surface components of *Porphyromonas gingivalis*. J Periodontal Res. 2009;44(1):1–12. doi:10.1111/j.1600-0765.2008.01135.x.
22. Wexler HM. *Bacteroides*: the good, the bad, and the nitty-gritty. Clin Microbiol Rev. 2007;20(4):593–621. doi:10.1128/cmr.00008-07.
23. Magnúsdóttir S, Ravcheev DA, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. Front Genet. 2015;6(148). doi:10.3389/fgene.2015.00148.
24. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe. 2009;6(3):279–89. doi:10.1016/j.chom.2009.08.003.
25. Friedrich T, Scheide D. The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. FEBS Lett. 2000;479(1–2):1–5. doi:10.1016/S0014-5793(00)01867-6.
26. Moparthi VK, Hagerhall C. The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. J Mol Evol. 2011;72(5-6):484–97. doi:10.1007/s00239-011-9447-2.
27. Lemos RS, Fernandes AS, Pereira MM, Gomes CM, Teixeira M. Quinol: fumarate oxidoreductases and succinate:quinone oxidoreductases: phylogenetic relationships, metal centres and membrane attachment. Biochim Biophys Acta. 2002;1553(1–2):158–70. doi:10.1016/S0005-2728(01)00239-0.
28. Fischbach MA, Sonnenburg JL. Eating for two: how metabolism establishes interspecies interactions in the gut. Cell Host Microbe. 2011;10(4):336–47. doi:10.1016/j.chom.2011.10.002.
29. Baughn AD, Malamy MH. The strict anaerobe *Bacteroides fragilis* grows in and benefits from nanomolar concentrations of oxygen. Nature. 2004;427(6973):441–4. doi:10.1038/nature02285.
30. Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. The cytochrome *bd* respiratory oxygen reductases. Biochim Biophys Acta. 2011;1807(11):1398–413. doi:10.1016/j.bbabio.2011.06.016.
31. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785–6. doi:10.1038/nmeth.1701.
32. Song C, Kumar A, Saleh M. Bioinformatic comparison of bacterial secretomes. Genomics Proteomics Bioinformatics. 2009;7(1-2):37–46. doi:10.1016/s1672-0229(08)60031-5.
33. Dalhammar G, Steiner H. Characterization of inhibitor A, a protease from *Bacillus thuringiensis* which degrades attacins and cecropins, two classes of antibacterial proteins in insects. Eur J Biochem. 1984;139(2):247–52.
34. Vaitkevicius K, Rompikuntal PK, Lindmark B, Vaitkevicius R, Song T, Wai SN. The metalloprotease PrtV from *Vibrio cholerae*: purification and properties. FEBS J. 2008;275(12):3167–77. doi:10.1111/j.1742-4658.2008.06470.x.
35. Singh B, Fleury C, Jalalvand F, Riesbeck K. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol Rev. 2012;36(6):1122–80. doi:10.1111/j.1574-6976.2012.00340.x.
36. Kosowska K, Reinholdt J, Rasmussen LK, Sabat A, Potempa J, Kilian M, et al. The *Clostridium ramosum* IgA proteinase represents a novel type of metalloendopeptidase. J Biol Chem. 2002;277(14):11987–94. doi:10.1074/jbc.M110883200.
37. Nakayama K. *Porphyromonas gingivalis* and related bacteria: from colonial pigmentation to the type IX secretion system and gliding motility. J Periodontal Res. 2015;50(1):1–8. doi:10.1111/jre.12255.
38. Seers CA, Slakeski N, Veith PD, Nikolof T, Chen YY, Dashper SG, et al. The RgpB C-terminal domain has a role in attachment of RgpB to the outer membrane and belongs to a novel C-terminal-domain family found in *Porphyromonas gingivalis*. J Bacteriol. 2006;188(17):6376–86. doi:10.1128/jb.00731-06.
39. Glew MD, Veith PD, Peng B, Chen YY, Gorasia DG, Yang Q, et al. PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. J Biol Chem. 2012;287(29):24605–17. doi:10.1074/jbc.M112.369223.
40. Nelson DC, Garbe J, Collin M. Cysteine proteinase SpeB from *Streptococcus pyogenes* - a potent modifier of immunologically important host and bacterial proteins. Biol Chem. 2011;392(12):1077–88. doi:10.1515/bc-2011-208.
41. Berry D, Stecher B, Schintlmeister A, Reichert J, Brugiroux S, Wild B, et al. Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. Proc Natl Acad Sci. 2013;110(12):4720–5. doi:10.1073/pnas.1219247110.
42. Derrien M, Vaughan EE, Plugge CM, de Vos WM. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol. 2004;54(5):1469–76. doi:10.1099/ijs.0.02873-0.
43. Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee Ying S, De Vadder F, Arora T, et al. Dietary fiber-induced improvement in glucose metabolism is

associated with increased abundance of *Prevotella*. Cell Metab. 2015;22(6): 971–82. doi:10.1016/j.cmet.2015.10.001.
44. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem. 2009;284(37):24673–7. doi:10.1074/jbc.R109.022848.
45. Reeves AR, D'Elia JN, Frias J, Salyers AA. A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. J Bacteriol. 1996;178(3):823–30.
46. Reeves AR, Wang GR, Salyers AA. Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. J Bacteriol. 1997;179(3):643–9.
47. Sonnenburg ED, Sonnenburg JL, Manchester JK, Hansen EE, Chiang HC, Gordon JI. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. Proc Natl Acad Sci. 2006;103(23):8834–9. doi:10.1073/pnas.0603249103.
48. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, et al. Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell. 2010;141(7):1241–52. doi:10.1016/j.cell.2010.05.005.
49. Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe. 2008;4(5):447–57. doi:10.1016/j.chom.2008.09.007.
50. Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, et al. Evolution of symbiotic bacteria in the distal human intestine. PLoS Biol. 2007;5:e156. 2007/06/21.
51. Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. Genome Biol. 2015;16(1):82. doi:10.1186/s13059-015-0646-9.
52. Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, et al. Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science. 2005;307(5717):1955–9. doi:10.1126/science.1109051.
53. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61(1):1–10. doi:10.1016/0006-3207(92)91201-3.
54. Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, et al., editors. Bergey's manual of systematic bacteriology, 2 edn. New York: Springer; 2011.
55. Turroni S, Bendazzoli C, Dipalo SCF, Candela M, Vitali B, Gotti R, et al. Oxalate-degrading activity in *Bifidobacterium animalis* subsp. *lactis*: impact of acidic conditions on the transcriptional levels of the oxalyl coenzyme A (CoA) decarboxylase and formyl-CoA transferase genes. Appl Environ Microbiol. 2010;76(16):5609–20. doi:10.1128/aem.00844-10.
56. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G et al. Gut microbiome of the Hadza hunter-gatherers. Nat Commun. 2014;5. doi:10.1038/ncomms4654.
57. Evans CC, LePard KJ, Kwak JW, Stancukas MC, Laskowski S, Dougherty J, et al. Exercise prevents weight gain and alters the gut microbiota in a mouse model of high fat diet-induced obesity. PLoS One. 2014;9:e92193. Public Library of Science.
58. Stevenson TJ, Duddleston KN, Buck CL. Effects of season and host physiological state on the diversity, density, and activity of the arctic ground squirrel cecal microbiota. Appl Environ Microbiol. 2014;80(18):5611–22. doi:10.1128/aem.01537-14.
59. Kolida S, Meyer D, Gibson GR. A double-blind placebo-controlled study to establish the bifidogenic dose of inulin in healthy humans. Eur J Clin Nutr. 2007;61(10):1189–95.
60. Cantarel BL, Lombard V, Henrissat B. Complex carbohydrate utilization by the healthy human microbiome. PLoS One. 2012;7:e28742. 2012/06/22.
61. Milani C, Andrea Lugli G, Duranti S, Turroni F, Mancabelli L, Ferrario C, et al. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. Sci Rep. 2015;5:15782. Macmillan Publishers Limited.
62. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, et al. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. Nature. 2014;506(7489):498–502. doi:10.1038/nature12907.
63. Whiteside SA, Razvi H, Dave S, Reid G, Burton JP. The microbiome of the urinary tract: a role beyond infection. Nat Rev Urol. 2015;12(2):81–90. doi:10.1038/nrurol.2014.361.
64. Troxel SA, Sidhu H, Kaul P, Low RK. Intestinal *Oxalobacter formigenes* colonization in calcium oxalate stone formers and its relation to urinary oxalate. J Endourol. 2003;17(3):173–6. doi:10.1089/089277903321618743.
65. Dawson KA, Allison MJ, Hartman PA. Isolation and some characteristics of anaerobic oxalate-degrading bacteria from the rumen. Appl Environ Microbiol. 1980;40(4):833–9.
66. Campieri C, Campieri M, Bertuzzi V, Swennen E, Matteuzzi D, Stefoni S, et al. Reduction of oxaluria after an oral course of lactic acid bacteria at high concentration. Kidney Int. 2001;60(3):1097–105.
67. Metges CC, Petzke KJ, El-Khoury AE, Henneman L, Grant I, Bedri S, et al. Incorporation of urea and ammonia nitrogen into ileal and fecal microbial proteins and plasma free amino acids in normal men and ileostomates. Am J Clin Nutr. 1999;70(6):1046–58.
68. Meakins TS, Jackson AA. Salvage of exogenous urea nitrogen enhances nitrogen balance in normal men consuming marginally inadequate protein diets. Clin Sci (London). 1996;90(3):215–25.
69. Shen T-CD, Albenberg L, Bittinger K, Chehoud C, Chen Y-Y, Judge CA, et al. Engineering the gut microbiota to treat hyperammonemia. J Clin Invest. 2015;125(7):2841–50. doi:10.1172/JCI79214.
70. Mora D, Arioli S. Microbial urease in health and disease. PLoS Pathog. 2014; 10:e1004472. Public Library of Science.
71. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222–7. doi:10.1038/nature11053.
72. Pabst O. New concepts in the generation and functions of IgA. Nat Rev Immunol. 2012;12(12):821–32. doi:10.1038/nri3322.
73. Mistry D, Stockley RA. IgA1 protease. Int J Biochem Cell Biol. 2006;38(8): 1244–8. doi:10.1016/j.biocel.2005.10.005.
74. Thornton RF, Kagawa TF, O'Toole PW, Cooney JC. The dissemination of C10 cysteine protease genes in *Bacteroides fragilis* by mobile genetic elements. BMC Microbiol. 2010;10(1):1–15. doi:10.1186/1471-2180-10-122.
75. Thornton RF, Murphy EC, Kagawa TF, O'Toole PW, Cooney JC. The effect of environmental conditions on expression of *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* C10 protease genes. BMC Microbiol. 2012;12(1): 1–11. doi:10.1186/1471-2180-12-190.
76. Potempa M, Potempa J, Kantyka T, Nguyen K-A, Wawrzonek K, Manandhar SP, et al. Interpain A, a cysteine proteinase from *Prevotella intermedia*, inhibits complement by degrading complement factor C3. PLoS Pathog. 2009;5:e1000316.
77. Sakamoto M, Lan PT, Benno Y. *Barnesiella viscericola* gen. nov., sp. nov., a novel member of the family *Porphyromonadaceae* isolated from chicken caecum. Int J Syst Evol Microbiol. 2007;57(2):342–6. doi:10.1099/ijs.0.64709-0.
78. Shkoporov AN, Khokhlova EV, Chaplin AV, Kafarskaia LI, Nikolin AA, Polyakov VY, et al. *Coprobacter fastidiosus* gen. nov., sp. nov., a novel member of the family *Porphyromonadaceae* isolated from infant faeces. Int J Syst Evol Microbiol. 2013;63(11):4181–8. doi:10.1099/ijs.0.052126-0.
79. Hofstad T, Olsen I, Eribe ER, Falsen E, Collins MD, Lawson PA. *Dysgonomonas* gen. nov. to accommodate *Dysgonomonas gadei* sp. nov., an organism isolated from a human gall bladder, and *Dysgonomonas capnocytophagoides* (formerly CDC group DF-3). Int J Syst Evol Microbiol. 2000;50(6):2189–95. doi:10.1099/00207713-50-6-2189.
80. Chen S, Dong X. *Proteiniphilum acetatigenes* gen. nov., sp. nov., from a UASB reactor treating brewery wastewater. Int J Syst Evol Microbiol. 2005;55(6): 2257–61. doi:10.1099/ijs.0.63807-0.
81. Iino T, Mori K, Itoh T, Kudo T, Suzuki K-i, Ohkuma M. Description of *Mariniphaga anaerophila* gen. nov., sp. nov., a facultatively aerobic marine bacterium isolated from tidal flat sediment, reclassification of the *Draconibacteriaceae* as a later heterotypic synonym of the *Prolixibacteraceae* and description of the family *Marinifilaceae* fam. nov. Int J Syst Evol Microbiol. 2014;64(11):3660–7. doi:10.1099/ijs.0.066274-0.
82. Ueki A, Akasaka H, Suzuki D, Ueki K. *Paludibacter propionicigenes* gen. nov., sp. nov., a novel strictly anaerobic, Gram-negative, propionate-producing bacterium isolated from plant residue in irrigated rice-field soil in Japan. Int J Syst Evol Microbiol. 2006;56(1):39–44. doi:10.1099/ijs.0.63896-0.
83. Sakamoto M, Suzuki M, Umeda M, Ishikawa I, Benno Y. Reclassification of *Bacteroides forsythus* (Tanner *et al.* 1986) as *Tannerella forsythensis* corrig., gen. nov., comb. nov. Int J Syst Evol Microbiol. 2002;52(3):841–9. doi:10.1099/00207713-52-3-841.
84. Sakamoto M, Benno Y. Reclassification of *Bacteroides distasonis*, *Bacteroides goldsteinii* and *Bacteroides merdae* as *Parabacteroides distasonis* gen. nov., comb. nov., *Parabacteroides goldsteinii* comb. nov. and *Parabacteroides merdae* comb. nov. Int J Syst Evol Microbiol. 2006;56(7):1599–605. doi:10.1099/ijs.0.64192-0.
85. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An expanded genomic representation of the phylum cyanobacteria. Genome Biol Evol. 2014;6(5):1031–45. doi:10.1093/gbe/evu073.

86. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324.
87. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. doi:10.1093/bioinformatics/btu170.
88. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinformatics. 2015;16(1):1–8. doi:10.1186/s12859-015-0663-4.
89. Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. GroopM: an automated tool for the recovery of population genomes from related metagenomes. In: Peer J, vol. 2. San Francisco: PeerJ Inc; 2014. p. e603.
90. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. In: PeerJ. Edited by Rahmann S, vol. 3; 2015: e1165.
91. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55. doi:10.1101/gr.186072.114.
92. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80. doi:10.1093/molbev/mst010.
93. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490. 2010/03/13.
94. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
95. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42(D1):D560–7. doi:10.1093/nar/gkt963.
96. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26(11):1463–4. doi:10.1093/bioinformatics/btq166.
97. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. Nucleic Acids Res. 2004;32(4):1363–71. doi:10.1093/nar/gkh293.
98. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9. doi:10.1093/bioinformatics/btu153.
99. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12(1):1–9. doi:10.1186/1471-2105-12-124.
100. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–30. doi:10.1093/nar/gkt1223.
101. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009;23(1):205–11.
102. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 2015;9(1):207–16. doi:10.1038/ismej.2014.106.
103. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(Web Server issue):W182–5. doi:10.1093/nar/gkm321.
104. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75. doi:10.1186/1471-2164-9-75.
105. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, et al. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform. 2010;11(1):40–79. doi:10.1093/bib/bbp043.
106. Wu M, McNulty NP, Rodionov DA, Khoroshkin MS, Griffin NW, Cheng J et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut *Bacteroides*. Science. 2015;350(6256). doi:10.1126/science.aac5992.
107. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51. Oxford University Press.
108. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 2003;12(8):1652–62. doi:10.1110/ps.0303703.
109. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8.
110. Kolde R. pheatmap: Pretty Heatmaps. 2015. R package version 107 from http://CRAN.R-project.org/package=pheatmap.
111. Roberts D. Labdsv: ordination and multivariate analysis for ecology. 2007. R package version 18-0 from https://cran.r-project.org/web/packages/labdsv/index.html.
112. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB et al. Vegan: community ecology package. 2015. R package version 23-1 from http://CRAN.R-project.org/package=vegan.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit