BMC Genomics

Genomic analysis of *Helicobacter himalayensis* sp. nov. isolated from *Marmota himalayana*

Shoukui Hu[1†], Lina Niu[2†], Lei Wu[1†], Xiaoxue Zhu[1†], Yu Cai[1], Dong Jin[3], Linlin Yan[1*] and Fan Zhao[1*]

Abstract

Background: *Helicobacter himalayensis* was isolated from *Marmota himalayana* in the Qinghai-Tibet Plateau, China, and is a new non-*H. pylori* species, with unclear taxonomy, phylogeny, and pathogenicity.

Results: A comparative genomic analysis was performed between the *H. himalayensis* type strain 80(YS1)[T] and other the genomes of *Helicobacter* species present in the National Center for Biotechnology Information (NCBI) database to explore the molecular evolution and potential pathogenicity of *H. himalayensis*. *H. himalayensis* 80(YS1)[T] formed a clade with *H. cinaedi* and *H. hepaticus* that was phylogenetically distant from *H. pylori*. The *H. himalayensis* genome showed extensive collinearity with *H. hepaticus* and *H. cinaedi*. However, it also revealed a low degree of genome collinearity with *H. pylori*. The genome of 80(YS1)[T] comprised 1,829,936 bp, with a 39.89% GC content, a predicted genomic island, and 1769 genes. Comparatively, *H. himalayensis* has more genes for functions in "cell wall/membrane/envelope biogenesis" and "coenzyme transport and metabolism" sub-branches than the other compared helicobacters, and its genome contained 42 virulence factors genes, including that encoding cytolethal distending toxin (CDT).

Conclusions: We characterized the *H. himalayensis* 80(YS1)[T] genome, its phylogenetic position, and its potential pathogenicity. However, further understanding of the pathogenesis of this potentially pathogenic bacterium is required, which might help to manage *H. himalayensis*-induced diseases.

Keywords: Helicobacter, Comparative genomics, *Helicobacter himalayensis*, Virulence factor

Background

The *Helicobacter* genus comprises gram-negative bacteria with characteristic spiral shapes [1]. The type species, *Helicobacter pylori*, was the first cultivated species of the genus. *H. pylori* was first isolated from in the stomachs of patients with gastritis and peptic ulceration in 1984 [2]. Currently, over 30 species have been isolated, including many candidate or unclassified species [3].

Except for *H. pylori*, most of the other *Helicobacter* have been isolated and identified from wild animals, including rodents, cats, dogs, rabbits, chickens, sheep, cattle, swine, cheetahs, ferrets, dolphins, whales, and non-human primates [4–8]. In addition, certain Helicobacters were isolated from laboratory animals [7, 9–11]. Convincing evidence has identified *H. pylori* as a causative bacterium of gastric adenocarcinoma, mucosa-associated lymphoid tissue lymphoma, and peptic ulcer in humans [12]. However, studies of the association between diseases and non-*H. pylori* species are lacking. There are only a few reports about non-*H. pylori* infection in humans, mostly in immunocompromised patients [13–16]. Thus, the

* Correspondence: ylinlin3@163.com; fannyhere@163.com
†Shoukui Hu, Lina Niu, Lei Wu and Xiaoxue Zhu contributed equally to this work.
[1]Department of Clinical Laboratory, Peking University Shougang Hospital, Beijing 100144, China
Full list of author information is available at the end of the article

Hu *et al. BMC Genomics*   (2020) 21:826

Page 2 of 10

pathogenicity and genetic characteristics of non-*H. pylori* remain to be explored.

In 2015, we reported the isolation of *Helicobacter himalayensis sp.* nov. strain 80(YS1)[T] from *Marmota himalayana* in the Qinghai-Tibet Plateau, China. This was the first report that a new species of helicobacter could be isolated and cultured directly in the gastrointestinal mucosa of a *Marmota himalayana*. *M. himalayana* live in the Qinghai-Tibet Plateau, a relatively isolated environment with an average altitude over 4000 m, which induces great differences in animal and plant biodiversity compared with plain regions. *H. himalayensis*, a new species of *Helicobacter* isolated from this plateau, may have unique evolutionary, pathogenic, and drug resistance characteristics because of its unique natural environment and host.

With development of next-generation sequencing technologies, many whole-genome sequences for non-*H. pylori* species have been generated in the past five years [17–20]. The deposition of this genomic data in the National Center for Biotechnology Information (NCBI) database has encouraged research seeking a deeper understanding of *Helicobacter* species at the genomic level. Therefore, in the present study we aimed to sequence the whole genome of *H. himalayensis*, and to explore the molecular evolution and potential pathogenicity of *H. himalayensis* using genome-wide comparative analyses based on the existing genomes of *Helicobacter* species.

Results

General features of *H. himalayensis* genome

The nucleotide sequence of the genome of *H. himalayensis* strain 80(YS1)[T] was deposited in the NCBI Databases with accession No. CP014991. The final genome assembly of *H. himalayensis* 80(YS1)[T] contained 103 contigs with a draft genome size of 1,829,936 bp and a genomic GC content of 39.89% (Table 1). The general annotation was performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). It predicted 1769 genes in total, among which 1664 belongs to predicted proteins, 1290 (77.52%) could be assigned to a function with a high level of confidence, and 374 (22.48%) were assigned as hypothetical proteins. The genome contained 39 tRNAs, a pair of 16S rRNAs, a pair of 23S rRNAs, a pair of 5S rRNAs, 3 ncRNAs and one phage intergrase, however, no plasmids or insertion sequence (IS) elements were found. Moreover, a genomic island predicted by a online tool IslandViewer 4, was named HhiG1 in the genome. The genomic island contained multiple genes encoding four restriction endonuclease subunit S, a restriction endonuclease subunit R, a type I restriction-modification system subunit M, a type II toxin-antitoxin system RelE/ParE family toxin, a type II toxin-antitoxin system RelB/DinJ family antitoxin, and site-specific integrase, as well as some hypothetical proteins (See Supplementary Table 1, Additional File 1). HhiG1 comprised 52,677 bp at position 1,738,356 to 1,791,033 in genome .

Table 1 General features of the *H. himalayensis*, *H. cinaedi*, *H. hepaticus* and *H. pylori*

Species	H. himalayensis	H. cinaedi	H. hepaticus	H. pylori
Strain	80 (YS1)[T]	ATCC BAA 847	ATCC 51449	26,695
Host	Marmota	Human	Rodent	Human
Genome size (bp)	1,829,936	2,240,130	1,799,146	1,667,867
GC content (%)	39.89	38.34	35.93	38.87
Predicted genes	1769	2365	1845	1583
Predicted proteins	1664	2233	1769	1447
Coding area (%)	91	91	93	91
Average gene length (bp)	942	865	881	913
plasmid	None	None	None	None
phages and phage like elements	1 phage intergrase	3 phage like element	3 phage genes	None
IS elements	None	Unknown	None	IS605, IS606
Genomic islands	HHiGI1 (predicted by IslandViewer 4)	Unknown	HHGI1	cag PAI
Regions of deviating GC content	DNA restriction/ modification systerm	Unknown	Genomic islets & islands, DNA restriction/ modification system, translation machinery	Genomic islets & islands, DNA-restriction/ modification system, translation machinery

Hu *et al. BMC Genomics*   (2020) 21:826

Page 3 of 10

Phylogenetic relationship and genomic collinearity features of *H. himalayensis* with other *Helicobacter* species

Seventeen whole-genome sequences of *Helicobacter* species, including *H. himalayensis*, and two whole-genome sequences of *Campylobacter jejuni* and *Acetobacter pasteurianus* as outgroups, were selected to construct the phylogenetic tree. *H. himalayensis* was in a single clade that was adjacent to *H. cinaedi*, *H. bilis* and *H. hepaticus*. These three *Helicobacter* spp. were in a node that was far away from *H. pylori* (Fig. 1a). Moreover, the phylogenetic relationship between *H. himalayensis* and the selected sixteen *Helicobacter* species, based on the core-genome, were also analyzed (Fig. 1b). The results were consistent with the phylogenetic map based on the *Helicobacter* genome sequences, which indicated that *H. himalayensis* is evolutionarily close to *H. hepaticus* and *H. cinaedi* but not to *H. pylori*. The general genomic features of *H. himalayensis*, *H. cinaedi*, *H. hepaticus*, and *H. pylori* are presented for comparison in Table 1. *H. himalayensis* has a larger genome than that of *H. hepaticus* (1,799,146 bp) and smaller genome than that of *H. cinaedi* (2,240,130 bp) [21]. The genomes of all three *Helicobacter* spp. were larger than that of *H. pylori*. All three *Helicobacter* spp. have phages or phage like elements in their genome, but *H. pylori* has none. By contrast, *H. pylori* has IS elements (IS605 and IS606), whereas *H. himalayensis*, *H. cinaedi,* and *H. hepaticus* do not have.

Figure 2 shows a comparison of the gene sequences of *H. himalayensis* with those of *H. cinaedi*, *H. hepaticus*, and *H. pylori* at the whole-genome scale. The results revealed a very high degree of genome collinearity between *H. himalayensis* and *H. hepaticus* (Fig. 2a), as well as between *H. himalayensis* and *H. cinaedi* (Fig. 2b). However, it also revealed a low degree of genome collinearity between *H. himalayensis* and *H. pylori* (Fig. 2c).

Function analysis of *H. himalayensis*

Of the 1664 predicted proteins, 1184 were clearly assigned to a functional classification with evidence and 202 were poorly functionally characterized using the Clusters of Orthologous Groups (COG) database (Fig. 3). Moreover, there were 105 proteins without any annotations in the COG database and 173 were not in the database. The numbers of genes identified through functional classification in the COG database for "information storage and processing", "metabolism", and "cellular processes and signaling" were 284, 546, and 426 respectively. Compared with *H. cinaedi*, *H. hepaticus*, *H. bilis*, and *H. pylori*, *H. himalayensis* has more genes for functions in the "cell wall/membrane/envelope biogenesis" and "coenzyme transport and metabolism" sub-branches, but fewer genes for functions in the "cell cycle control/ cell division/ chromosome partitioning",

"intracellular trafficking/ secretion/ vesicular transport", "lipid transport and metabolism" and "carbohydrate transport and metabolism" sub-branches (Table 2). The circular genome atlas of *H. himalayensis* integrating different kinds of information is shown in Fig. 4, that including the protein-coding genes, tRNA and rRNA genes, GC content and GC skew information.

In addition, Forty-two genes were predicted to match virulence factors genes in the VFDB from *Helicobacter* genus, and 83.3% (35/42) of them were associated with flagella motility and bacterial invasion (See Supplementary Table 2, Additional File 2). Other virulence factors included cytolethal distending toxin (*cdtA*, *cdtB*, *cdtC*), lipopolysaccharide Lewis antigens (*futA*), neutrophil activating protein (*napA*), autoinducer-2 production protein (*luxS*), and catalase (*katA*). The *H. himalayensis* genome does not contain a urease and the Cag pathogenic island, like *H. pylori*. However, the genome contained other predicted virulence factors associated with migration, invasion, colonization, and carcinogenesis, which might be pathogenic to the host.

Discussion

In the present study, the whole genome sequence data of *H. himalayensis*, a new species of *Helicobacter*, were obtained. The *H. himalayensis* genome was compared with those of other *Helicobacter* species deposited in the NCBI database. The results of the comparative genomic analyses not only confirmed phylogenetic position of *H. himalayensis* in relation to other *Helicobacter* species, but also predicted its potential pathogenicity by matching the virulence factor genes in VFDB.

Several smaller stretches of anomalous GC content were identified using the GC mol% content analysis around the genome (Fig. 4). In pathogenic bacteria, this type of anomalous GC content is usually associated with genome islands [22]. For example, the *H. pylori* cag pathogenicity island is intimately associated with increased potential to cause more severe pathology, including cancer and gastric ulcers [23]. There were five fragments with anomalous GC content in the *H. himalayensis* genome, and the predicted genome island HhiG1 was covered by one of the five fragments, which suggested that HhiG1 might be related with the pathogenicity of *H. himalayensis*.

H. himalayensis is phylogenetically close to *H. cinaedi* and *H. hepaticus*, based on the whole genome sequence, which is consistent with the previously reported phylogenetic relationship based on the analysis of housekeeping gene analysis [24]. Moreover, the predicted genes of *H. himalayensis* had higher identity to those of *H. cinaedi* and *H. hepaticus* than to those of *H. pylori*, which confirmed the previous result of evolutionary analyses from another perspective. *H. cinaedi* was first
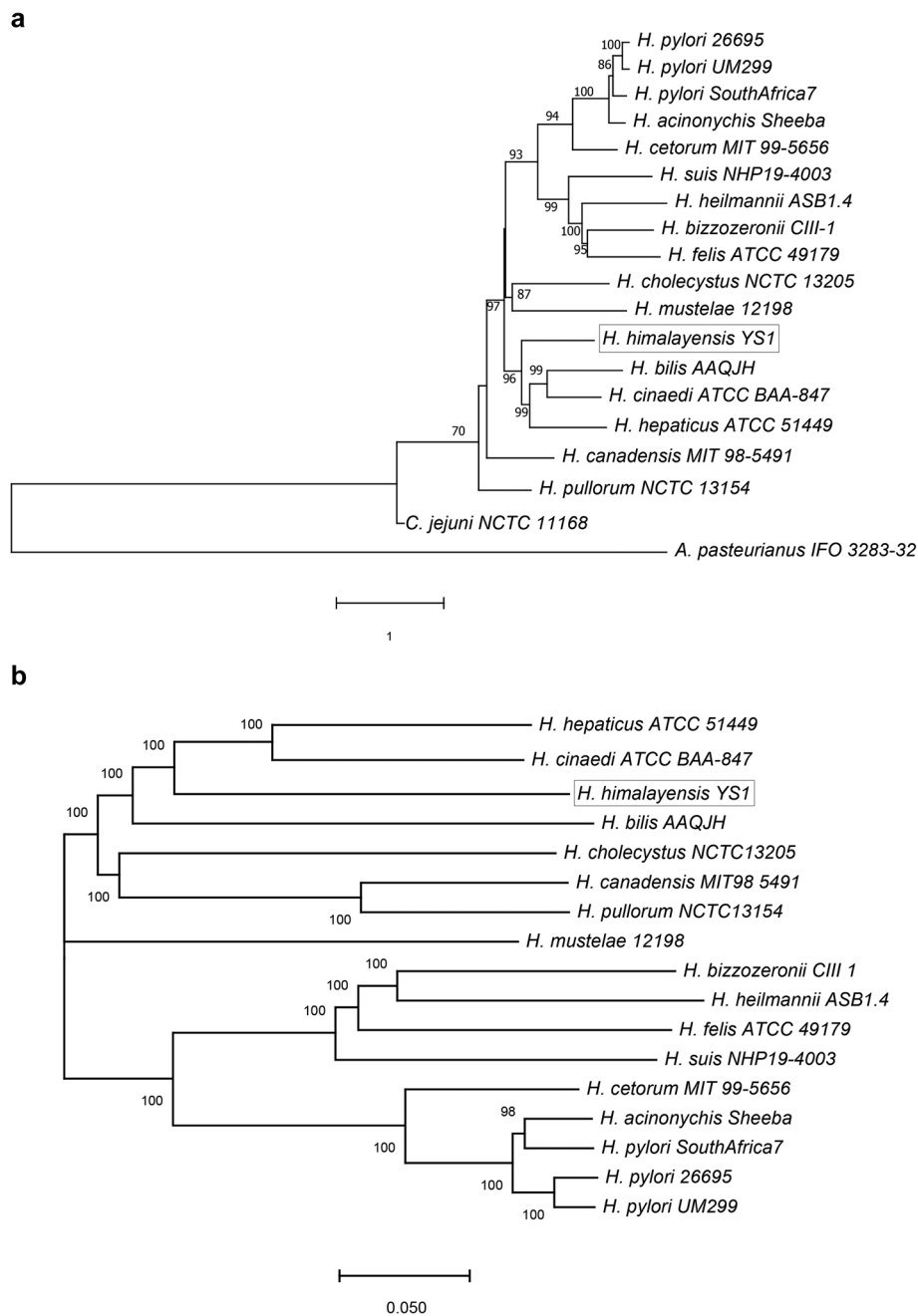
Fig. 1 Phylogenetic tree produced by the comparison of the genome sequences of selected bacteria using MEGA X. **a** Phylogenetic tree based on the pan-genomes of the seventeen *Helicobacter* species and two outgroups. *Campylobacter jejuni* NCTC 11168 and *Acetobacter pasteurianus* IFO 3283–32 were used as outgroups. **b** Phylogenetic tree based on the core-genomes of the seventeen *Helicobacter* species without outgroup. The scale indicates the divergence (substitutions/site). The labeled numbers represent the posterior mean of rates on the corresponding branches

isolated from a rectal swab from a homosexual men, while *H. hepaticus* was first isolated from the livers of mice. Both of them are unable to colonize the gastric mucosa, because they do not express a urease to overcome the acidic environment in the stomach. These helicobacters with the term "enterohepatic helicobacters (EHH)" colonize the intestinal mucosa or the liver, and are associated with chronic liver or intestinal inflammation [25]. Recently, many cases of EHH infection in humans were reported, mainly in immunocompromised patients [3]. *H. himalayensis* is closely related to *H. cinaedi* and *H. hepaticus*; therefore, *H. himalayensis* is predicted to have similar pathogenic effects.

Fig. 2 Map of genome collinearity between *H. himalayensis* with *H. hepaticus*, *H. cinaedi*, and *H. pylori*. **a** *H. hepaticus*. **b** *H. cinaedi*. **c** *H. pylori*. A red line indicates that the gene sequence is transcribed in the forward direction, and a blue line indicates that the gene is transcribed in the reverse orientation. The homologous sequences was identified as > 100 bp in length and at least 70% homology

Based on the genomic functional analysis of *H. himalayensis*, we found that in addition to the 35 virulence factor genes related to flagellar motility, *H. himalayensis* also has genes encoding CDT, lipopolysaccharide Lewis antigens, neutrophil activating protein, autoinducer-2 production protein, and catalase. These virulence factor genes from the VFDB have been identified in *H. hepaticus*, *H. pylori*, or other bacteria in experiments [26]. CDT is a typical virulence factor, having DNase I activity that induces DNA double-strand breaks, apoptosis, and

G2/M cell cycle arrest in cultured mammalian cells. In vivo, CDT is also associated with carcinogenesis [27, 28]. The Lewis antigen (encoded by *futA*) suppresses the immune response to the bacteria permits it to adhere to the gastric mucosa, or has a role in certain aspects of autoimmunity [29]. Neutrophil activating protein (encoded by *napA*) could promote the production of reactive oxygen radicals in *H. pylori* or the adhesion of human neutrophils to endothelial cells [30, 31]. Autoinducer-2, encoded by *luxS*, is produced and

Hu *et al. BMC Genomics* (2020) 21:826

Page 6 of 10

Fig. 3 COG functional classification of genes belonging to *H. himalayensis*

Legend:
- CELLULAR PROCESSES AND SIGNALING
- INFORMATION STORAGE AND PROCESSING
- METABOLISM
- POORLY CHARACTERIZED

C: Energy production and conversion 7.36 %
D: Cell cycle control, cell division, chromosome partitioning 1.23 %
E: Amino acid transport and metabolism 8.59 %
F: Nucleotide transport and metabolism 4.47 %
G: Carbohydrate transport and metabolism 2.74 %
H: Coenzyme transport and metabolism 7.14 %
I: Lipid transport and metabolism 2.96 %
J: Translation, ribosomal structure and biogenesis 10.82 %
K: Transcription 2.6 %
L: Replication, recombination and repair 7.07 %
M: Cell wall/membrane/envelope biogenesis 11.69 %
N: Cell motility 4.69 %
O: Posttranslational modification, protein turnover, chaperones 4.83 %
P: Inorganic ion transport and metabolism 5.19 %
Q: Secondary metabolites biosynthesis, transport and catabolism 0.94 %
S: Function unknown 15.3 %
T: Signal transduction mechanisms 3.39 %
U: Intracellular trafficking, secretion, and vesicular transport 2.96 %
V: Defense mechanisms 1.95 %

detected by a wide variety of bacteria, and is presumed to facilitate interspecies communications [32]. *katA* is a gene coding catalase, which protects against reactive oxygen species damage [33]. It is hypothesized that these orthologous genes in the *H. himalayensis* genome have similar effects and may also confer potential pathogenicity on the host. Till now, no animal experiment has been done with *H. himalayensis* infection, but the result of histopathological examination for the gastrointestinal mucosa of *M. himalayana* where *H. himalayensis* isolated showed obvious lesions, while the gastrointestinal mucosa of *M. himalayana* without *H. himalayensis* isolated had no lesions (See Supplementary Figure 1, Additional File 3). Therefore, whether *H. himalayensis* containing these virulence factor genes is a pathogen to host, should be confirmed by further experimental studies.

Conclusions

The present study described the characteristic of the *H. himalayensis* genome, analyzed its phylogenetic relationship with other helicobacters, and highlighted the potential pathogenicity of 80(YS1)^T to the host. Further research into the pathogenesis of this potentially pathogenic bacterium is necessary and might help to manage diseases caused by *H. himalayensis*.

Methods

Bacterial strain

H. himalayensis sp. nov strain 80(YS1)^T (identical to CGMCC1.12864 and DSM28742) was used in this study. It was cultured as previously described, with some modifications [24]. Briefly, the stored strain was revived and cultured on brain heart infusion agar plates containing 8% (v/v) horse blood, 2.5 μg/mL trimethoprim, 10 μg/mL vancomycin, and 1.25 U/mL polymyxin B, under a hypoxic atmosphere (5% CO_2, 5% H_2, 0.5% O_2, and 90% N_2) at 35 °C for 3–5 days.

Genome DNA extraction and whole genome sequencing

A Wizard® Genomic DNA Purification kit (Promega, Madison, WI, USA) was used to isolate the genomic DNA from collected cells of strain 80(YS1)^T, according to the manufacturer's instructions. Sequencing and

Hu *et al. BMC Genomics* (2020) 21:826

Page 7 of 10

Table 2 Proportion of genes in COG function clssification for helicobacters

COG function classification	H.himalayensis	H.cinaedi	H.hepaticus	H.bilis	H.pylori
J: Translation, ribosomal structure and biogenesis	10.82%	8.60%	9.95%	9.53%	11.18%
A: RNA processing and modification	0.00%	0.00%	0.00%	0.06%	0.08%
K: Transcription	2.60%	2.81%	3.16%	3.07%	2.64%
L: Replication, recombination and repair	7.07%	7.75%	5.77%	7.04%	9.32%
D: Cell cycle control, cell division, chromosome partitioning	1.23%	1.57%	1.30%	1.66%	2.17%
V: Defense mechanisms	1.95%	1.85%	1.37%	1.98%	2.56%
T: Signal transduction mechanisms	3.39%	3.88%	3.29%	3.07%	2.41%
M: Cell wall/membrane/envelope biogenesis	11.69%	10.39%	8.99%	10.04%	9.47%
N: Cell motility	4.69%	4.27%	4.46%	4.16%	6.60%
Z: Cytoskeleton	0.00%	0.00%	0.07%	0.00%	0.00%
U: Intracellular trafficking, secretion, and vesicular transport	2.96%	3.76%	3.16%	3.26%	7.14%
O: Posttranslational modification, protein turnover, chaperones	4.83%	4.55%	4.94%	4.61%	4.58%
C: Energy production and conversion	7.36%	7.19%	7.89%	6.33%	6.52%
G: Carbohydrate transport and metabolism	2.74%	3.48%	3.71%	2.88%	2.87%
E: Amino acid transport and metabolism	8.59%	7.75%	8.51%	8.64%	6.52%
F: Nucleotide transport and metabolism	4.47%	4.04%	4.60%	4.29%	5.12%
H: Coenzyme transport and metabolism	7.14%	6.24%	6.31%	5.63%	6.37%
I: Lipid transport and metabolism	2.96%	3.76%	3.64%	3.45%	3.26%
P: Inorganic ion transport and metabolism	5.19%	5.39%	5.28%	5.57%	5.12%
Q: Secondary metabolites biosynthesis, transport and catabolism	0.94%	2.70%	1.72%	1.73%	0.93%
S: Function unknown	15.30%	17.47%	18.53%	19.45%	13.12%

library construction were carried out using the Hiseq2000 90 bp paired end sequencing platform (Illumina, San Diego, CA, USA). In total, 273,256 polymerase reads (242 bp/reads) were generated with 33.1× theoretical coverage. After filtering out all low quality reads, 266,891 reads (97.7%) were assembled into 103 contigs using 454/Roche Newbler (Roche Molecular Systems Inc., Branchburg, NJ, USA). All the contigs were sorted using the genomes of *H. pylori* 26695 and *H. cinaedi* ATCC BAA847, and verified by sequencing using an ABI BigDye Terminator V3.1 Kit and ABI 3730 sequencer (Applied Biosystems, Foster City, CA, USA). Conventional Sanger sequencing of polymerase chain reaction (PCR) fragments, based on brute force PCR, was used to for gap filling among the contigs and scaffolds to complete the genome.

Comparative genomics analysis

The assembled genome of strain 80(YS1)^T was compared with the available *Helicobacter spp.* genomes deposited in the NCBI Database. Sixteen *Helicobacter spp.* and two outgroup stains (*Campylobacter jejuni* and *Acetobacter pasteurianus*) were selected, and the whole genomes were download from NCBI Database. For pan-genome analysis, SNP sequence from the selected genomes were generated using KSNP3 (version 3.0) [34], and analysed using MEGA X [35] to construct the phylogenetic trees by Neighbor-Joining algorithm with 1000 bootstraps. For core-genome analysis, genes present in each selected Helicobacter genome were classified as core. The core-genome compared files were generated using roary (version 3.12.0) with a BLAST cutoff of 70% identity of gene (−i 70). MEGA X was used to construct the phylogenetic trees by Neighbor-Joining algorithm with 1000 bootstraps based on the core-genome. Collinearity analysis of *H. hemalayensis* with other three helicobacter species were performed using BLASTn. To identify homologous sequences (> 100 bp in length and at least 70% homology), 80(YS1)^T genomic contigs were aligned with those of three other *Helicobacter* species, using BLASTn as implemented in NCBI BLAST+ v. 2.2.8 [36]. Progressive Mauve, with match seed weight = 15, minimum Locally Collinear Block weight = 45, minimum island size = 50, maximum backbone gap size = 50, and a minimum backbone size = 50 [37] was used to compare the genomes of the *Helicobacter* spp. using multiple sequence alignment. CIRCOS v. 0.64 [38] was used to construct a visual representation of the links between homologous regions of *H. himalayensis sp.* nov strain 80(YS1)^T and those of three other *Helicobacter* species.
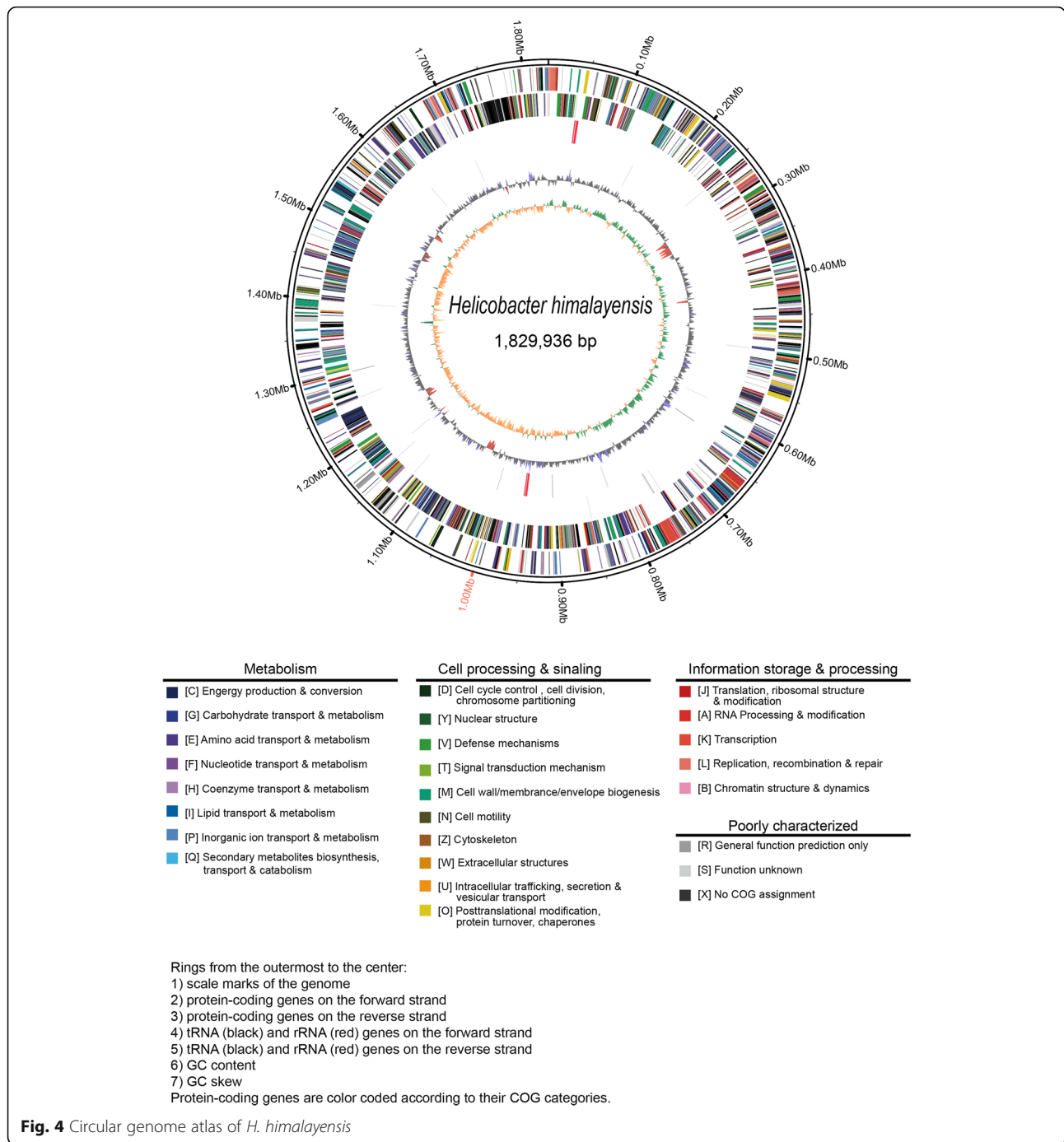
Metabolism

- [C] Engergy production & conversion
- [G] Carbohydrate transport & metabolism
- [E] Amino acid transport & metabolism
- [F] Nucleotide transport & metabolism
- [H] Coenzyme transport & metabolism
- [I] Lipid transport & metabolism
- [P] Inorganic ion transport & metabolism
- [Q] Secondary metabolites biosynthesis, transport & catabolism

Cell processing & sinaling

- [D] Cell cycle control , cell division, chromosome partitioning
- [Y] Nuclear structure
- [V] Defense mechanisms
- [T] Signal transduction mechanism
- [M] Cell wall/membrane/envelope biogenesis
- [N] Cell motility
- [Z] Cytoskeleton
- [W] Extracellular structures
- [U] Intracellular trafficking, secretion & vesicular transport
- [O] Posttranslational modification, protein turnover, chaperones

Information storage & processing

- [J] Translation, ribosomal structure & modification
- [A] RNA Processing & modification
- [K] Transcription
- [L] Replication, recombination & repair
- [B] Chromatin structure & dynamics

Poorly characterized

- [R] General function prediction only
- [S] Function unknown
- [X] No COG assignment

Rings from the outermost to the center:
1) scale marks of the genome
2) protein-coding genes on the forward strand
3) protein-coding genes on the reverse strand
4) tRNA (black) and rRNA (red) genes on the forward strand
5) tRNA (black) and rRNA (red) genes on the reverse strand
6) GC content
7) GC skew
Protein-coding genes are color coded according to their COG categories.

Fig. 4 Circular genome atlas of *H. himalayensis*

Function analysis of *H. himalayensis*

The identified *H. himalayensis* proteins were assigned and annotated to the database of COG by eggnog-mapper (version 2) [39]. COG classification was conducted by using DIAMOND v0.8.37.99 (−evalue 1e-5) [40]. The results of the COG classification were visualized using R (version 3.6.2). The genomic island of *H.* himalayensis was predicted by an online tool Island-Viewer 4 [41] (http://www.pathogenomics.sfu.ca/island-viewer/upload/). The virulence factors of *H. himalayensis* were searched in the virulence factors database (VFDB) using the integrated and automatic pipeline, Vfanalyzer [42] (http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi?JobID=May_28-1427107863).

Hu *et al. BMC Genomics* (2020) 21:826

Page 9 of 10

Supplementary Information

Additional file 1: Supplementary Table 1. Results of genomic island predicted by Island Viewer 4.

Additional file 2: Supplementary Table 2. Virulence factors genes present in genome of *H.himalayensis*.

Additional file 3: Supplementary Figure 1. Samples for histological examination. **a** Image and **c** histological examination of the intestinal mucosa of *Marmota himalayana* without *H. himalayensis* isolated (hematoxylin-eosin, original magnification × 100). A black arrow indicates normal histoarchitecture of the intestinal mucosa. **b** Image and **d** pathological examination of the intestinal mucosa of *Marmota himalayana* where *H. himalayensis* was isolated (hematoxylin-eosin, original magnification × 100). A black arrow indicates necrosis area and/or lesion area of the intestinal mucosa.

Abbreviations

NCBI: National Center for Biotechnology Information; CDT: Cytolethal Distending Toxin; IS: Insertion Sequence; COG: Clusters of Orthologous Groups; VFDB: Virulence factors database; EHH: Enterohepatic helicobacters

Acknowledgements

Not applicable.

Authors' contributions

All authors read and approved the final manuscript. HS and ZF contributed to the conception and design of the work; HS, NL, ZX and WL contributed to the acquisition of data; ZX, CY, JD, WL and YL participated in the analysis and interpretation of data; NL and HS drafted the manuscript; HS, YL and ZF revised the manuscript and provided important suggestions.

Funding

Availability of data and materials

The datasets analysed during the current study are available in the NCBI repository, accession numbers: NZ_CP014991 for *H. himalayensis* strain 80 (YS1)^T, complete genome; NC_008229 for *H. acinonychis* Sheeba, complete genome; NZ_CP019645 for *H. bilis* strain AAQJH, complete genome; NC_015674 for *H. bizzozeronii* CIII-1, complete genome; NZ_CM000776 for *H. canadensis* MIT 98–5491; NC_017735 for *H. cetorum* MIT 99–5656, complete genome; NZ_LR134518 for *H. cholecystus* strain NCTC13205, complete genome; AP012492 for *H. cinaedi* ATCC BAA-847, complete genome; NC_014810 for *H. felis* ATCC 49179, complete genome; NC_019674 for *H. heilmannii* ASB1.4, complete genome; AE017125 for *H. hepaticus* ATCC 51449, complete genome; NC_013949 for *H. mustelae* 12198, complete genome; NZ_LR134509 for *H. pullorum* NCTC13154, complete genome; CP003904 for *H. pylori* 26695, complete genome; NC_017361 for *H. pylori* SouthAfrica7, complete genome; CP005491 for *H. pylori* UM299, complete genome; AP023039 for *H. suis* NHP19–4003, complete genome; NC_017111 for *Acetobacter pasteurianus* IFO 3283–32, complete genome; NC_002163 for *Campylobacter jejuni* NCTC 11168, complete genome.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

We declare no competing interests.

Author details

^{1}Department of Clinical Laboratory, Peking University Shougang Hospital, Beijing 100144, China. ^{2}Department of Pathogen Biology, School of Basic Medicine and Lifescience, Hainan Medical University, Haikou 571101, China. ^{3}State Key Laboratory of Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing 102206, China.

References

1. Sycuro LK, Pincus Z, Gutierrez KD, Biboy J, Stern CA, Vollmer W, et al. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. Cell. 2010;141(5):822–33.
2. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet. 1984;1(8390):1311–5.
3. Péré-Védrenne C, Flahou B, Loke MF, Ménard A, Vadivelu J. Other Helicobacters, gastric and gut microbiota. Helicobacter. 2017;22(Suppl 1): e12407.
4. Van den Bulck K, Decostere A, Baele M, Driessen A, Debongnie JC, Burette A, et al. Identification of non-*Helicobacter pylori* spiral organisms in gastric samples from humans, dogs, and cats. J Clin Microbiol. 2005;43(5):2256–60.
5. Neiger R, Dieterich C, Burnens A, Waldvogel A, Corthésy-Theulaz I, Halter F, et al. Detection and prevalence of *Helicobacter* infection in pet cats. J Clin Microbiol. 1998;36(3):634–7.
6. Hwang CY, Han HR, Youn HY. Prevalence and clinical characterization of gastric *Helicobacter* species infection of dogs and cats in Korea. J Vet Sci. 2002;3(2):123–33.
7. O'Rourke JL, Solnick JV, Neilan BA, Seidel K, Hayter R, Hansen LM, et al. Description of 'Candidatus Helicobacter heilmannii' based on DNA sequence analysis of 16S rRNA and urease genes. Int J Syst Evol Microbiol. 2004;54(Pt 6):2203–11.
8. Collado L, Jara R, González S. Description of *Helicobacter valdiviensis* sp. nov., an *Epsilonproteobacteria* isolated from wild bird faecal samples. Int J Syst Evol Microbiol. 2014;64:1913-9.
9. Franklin CL, Beckwith CS, Livingston RS, Riley LK, Gibson SV, Besch-Williford CL, et al. Isolation of a novel *Helicobacter* species, *Helicobacter cholecystus* sp. nov., from the gallbladders of Syrian hamsters with cholangiofibrosis and centrilobular pancreatitis. J Clin Microbiol. 1996;34(12):2952–8.
10. Coldham T, Rose K, O'Rourke J, Neilan BA, Dalton H, Lee A, et al. Detection, isolation, and characterization of *Helicobacter* species from the gastrointestinal tract of the brushtail possum. Appl Environ Microbiol. 2011; 77(5):1581–7.
11. Shen Z, Fox JG, Dewhirst FE, Paster BJ, Foltz CJ, Yan L, et al. *Helicobacter rodentium* sp. nov., a urease-negative *Helicobacter* species isolated from laboratory mice. Int J Syst Bacteriol. 1997;47(3):627–34.
12. Sonnenberg A. Review article: historic changes of *Helicobacter pylori*-associated diseases. Aliment Pharmacol Ther. 2013;38(4):329–42.
13. Saito S, Tsukahara M, Ohkusu K, Kurai H. *Helicobacter fennelliae* bacteremia: three case reports and literature review. Medicine. 2016;95(18):e3556.
14. Shimizu Y, Gomi H, Ishioka H, Isono M. Refractory to treat *Helicobacter cinaedi* bacteremia with bilateral lower extremities cellulitis in an immunocompetent patient. IDCases. 2016;5:9–11.
15. Hayashi T, Tomida J, Kawamura Y, Yoshida M, Yokozawa I, Kaneko S. Unusual manifestation of *Helicobacter cinaedi* infection: a case report of intracranial subdural empyema and bacteremia. BMC Infect Dis. 2017; 17(1):40.
16. Katsuma A, Yamamoto I, Tsuchiya Y, Kawabe M, Yamakawa T, Katsumata H, et al. *Helicobacter cinaedi* bacteremia with cellulitis in a living-donor kidney transplant recipient identified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a case report. BMC Res Notes. 2017;10(1):87.
17. Behrens W, Bönig T, Suerbaum S, Josenhans C. Genome sequence of *Helicobacter pylori* hpEurope strain N6. J Bacteriol. 2012;194(14):3725–6.
18. Choo SW, Ang MY, Fouladi H, Tan SY, Siow CC, Mutha NV, et al. HelicoBase: a *Helicobacter* genomic resource and analysis platform. BMC Genomics. 2014;15:600.

Hu *et al. BMC Genomics*   (2020) 21:826

Page 10 of 10

19. Clancy CD, Forde BM, Moore SA, O'Toole PW. Draft genome sequences of *Helicobacter pylori* strains 17874 and P79. J Bacteriol. 2012;194(9):2402.

20. Duncan SS, Bertoli MT, Kersulyte D, Valk PL, Tamma S, Segal I, et al. Genome Sequences of Three hpAfrica2 Strains of *Helicobacter pylori*. Genome Announc. 2013;1(5):e00729-13.

21. Miyoshi-Akiyama T, Takeshita N, Ohmagari N, Kirikae T. Complete genome sequence of *Helicobacter cinaedi* type strain ATCC BAA-847. J Bacteriol. 2012; 194(20):5692.

22. Novick RP, Ram G. The floating (pathogenicity) island: a genomic dessert. Trends Genet. 2016;32(2):114–26.

23. Raei N, Latifi-Navid S, Zahri S. *Helicobacter pylori cag* Pathogenicity Island *cagL* and *orf17* genotypes predict risk of peptic ulcerations but not gastric Cancer in Iran. Asian Pac J Cancer Prev. 2015;16(15):6645–50.

24. Hu S, Jin D, Lu S, Liu S, Zhang J, Wang Y, et al. *Helicobacter himalayensis* sp. nov. isolated from gastric mucosa of *Marmota himalayana*. Int J Syst Evol Microbiol. 2015;65(Pt 6):1719–25.

25. Fresia P, Jara R, Sierra R, Ferrés I, Greif G, Iraola G, et al. Genomic and clinical evidence uncovers the enterohepatic species *Helicobacter valdiviensis* as a potential human intestinal pathogen. Helicobacter. 2017;22(5):e12425.

26. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33(Database issue):D325–8.

27. Ge Z, Feng Y, Ge L, Parry N, Muthupalani S, Fox JG. *Helicobacter hepaticus* cytolethal distending toxin promotes intestinal carcinogenesis in 129*Rag2*-deficient mice. Cell Microbiol. 2017;19(7):e12728.

28. Ge Z, Rogers AB, Feng Y, Lee A, Xu S, Taylor NS, et al. Bacterial cytolethal distending toxin promotes the development of dysplasia in a model of microbially induced hepatocarcinogenesis. Cell Microbiol. 2007;9(8):2070–80.

29. Moran AP, Prendergast MM. Molecular mimicry in *Campylobacter jejuni* and *Helicobacter pylori* lipopolysaccharides: contribution of gastrointestinal infections to autoimmunity. J Autoimmun. 2001;16(3):241–56.

30. Satin B, Del Giudice G, Della Bianca V, Dusi S, Laudanna C, Tonello F, et al. The neutrophil-activating protein (HP-NAP) of *Helicobacter pylori* is a protective antigen and a major virulence factor. J Exp Med. 2000;191(9):1467–76.

31. Nishioka H, Baesso I, Semenzato G, Trentin L, Rappuoli R, Del Giudice G, et al. The neutrophil-activating protein of *Helicobacter pylori* (HP-NAP) activates the MAPK pathway in human neutrophils. Eur J Immunol. 2003;33(4):840–9.

32. Chen X, Schauder S, Potier N, Van Dorsselaer A, Pelczer I, Bassler BL, et al. Structural identification of a bacterial quorum-sensing signal containing boron. Nature. 2002;415(6871):545–9.

33. Wu H, Soler-Garcia AA, Jerse AE. A strain-specific catalase mutation and mutation of the metal-binding transporter gene *mntC* attenuate *Neisseria gonorrhoeae* in vivo but not by increasing susceptibility to oxidative killing by phagocytes. Infect Immun. 2009;77(3):1091–102.

34. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31(17):2877–8.

35. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.

36. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

37. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5(6):e11147.

38. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.

39. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through Orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34(8):2115–22.

40. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.

41. Bertelli C, Laird MR, Williams KP. Simon Fraser University research computing G, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45(W1):W30–W5.

42. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019. https://doi.org/10.1093/nar/gky1080.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.