

GENOME REPORT

Open Access



# The complete genome and methylome of *Helicobacter pylori* hpNEAfrica strain HP14039

Binit Lamichhane<sup>1</sup>, Eng-Guan Chua<sup>1</sup>, Michael J. Wise<sup>1,2</sup>, Connor Laming<sup>1</sup>, Barry J. Marshall<sup>1,3</sup> and Chin-Yen Tay<sup>1,3\*</sup> 

## Abstract

**Background:** *Helicobacter pylori* is a Gram-negative bacterium which mainly causes peptic ulcer disease in human, but is also the predominant cause of stomach cancer. It has been coevolving with human since 120,000 years and, according to Multi-locus sequence typing (MLST), *H. pylori* can be classified into seven major population types, namely, hpAfrica1, hpAfrica2, hpNEAfrica, hpEastAsia, hpAsia2, hpEurope and hpSahul. *Helicobacter pylori* harbours a large number of restriction-modification (R-M) systems. The methyltransferase (MTase) unit plays a significant role in gene regulation and also possibly modulates pathogenicity. The diversity in MTase can act as geomarkers to correlate strains with the phylogeographic origins. This paper describes the complete genome sequence and methylome of gastric pathogen *H. pylori* belonging to the population hpNEAfrica.

**Results:** In this paper, we present the complete genome sequence and the methylome profile of *H. pylori* hpNEAfrica strain HP14039, isolated from a patient who was born in Somalia and likely to be infected locally during early childhood prior to migration. The genome of HP14039 consists of 1,678,260 bp with 1574 coding genes and 38.7% GC content. The sequence analysis showed that this strain lacks the *cag* pathogenicity island. The *vacA* gene is of S2M2 type. We have also identified 15 methylation motifs, including WCANHNNNTG and CTANNNNNNTAYG that were not previously described.

**Conclusions:** We have described the complete genome of *H. pylori* strain HP14039. The information regarding phylogeography, methylome and associated metadata would help scientific community to study more about hpNEAfrica population type.

**Keywords:** Complete genome, Methylome, *Helicobacter pylori*, hpNEAfrica

## Background

*Helicobacter pylori* is a Gram-negative bacterium that colonises human stomach, resulting in various gastric diseases including gastritis, peptic ulcer disease and gastric cancer. About half of the world population is infected with *H. pylori* with prevalence in developing countries reaching up to 90% [1, 2]. *H. pylori* is well-known for its genome's ability to mirror the history of human migration history due to its very long association with humans and transmission being within families [2–5]. This

coevolution has led to the emergence of seven different MLST population types of *H. pylori*, based on the geographical origins, namely hpAfrica1, hpAfrica2, hpNEAfrica, hpEastAsia, hpAsia2, hpEurope and hpSahul [2, 3, 5].

*Helicobacter pylori* harbours large number of type II restriction-modification (R-M) systems [6–9]. The type II systems have separate DNA methyltransferase (MTase) and restriction endonuclease proteins (REase) that act on the same DNA sequence motif. Apart from protecting host DNA from foreign DNA, DNA methylation has been implicated in the regulation of bacterial gene expression [10]. The Type II RM systems in *H. pylori* are substantially diverse among strains and therefore can be

\*Correspondence: alfred.tay@uwa.edu.au

<sup>1</sup> Marshall Centre for Infectious Diseases Research and Training, School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia

Full list of author information is available at the end of the article



used as a biomarker to trace geographical association of *H. pylori* [11, 12].

*Helicobacter pylori* strains of hpNEAfrica population type are mainly found circulating in Ethiopia, Somalia, Sudan and Nilo-Saharan speakers in northern Nigeria [2, 13]. No complete genome of a *H. pylori* strain originated from this region is available in public databases. Our complete genome and methylome data of *H. pylori* HP14039 hence will provide further indicators on the evolution and genetic diversity of this human gastric pathogen.

## Methods

### Bacterial culture and genomic DNA extraction

*Helicobacter pylori* strain HP14039 was isolated from a patient gastric biopsy sample onto selective agar plates. The selective plates used were Columbia blood agar plates (CBA) containing 5% horse blood (PathWest Laboratory Medicine WA Media, Australia) with Dent supplement (Oxoid, UK). The plates were incubated for 3–4 days at 37 °C in a 10% CO<sub>2</sub> environment. The genomic DNA extraction was performed on 48 h bacterial culture using phenol–chloroform method [14]. Cells were harvested from culture plates and washed with PBS (pH 8) followed by centrifugation at 14,000 rpm for 1 min. Following the removal of supernatant, the pellet was resuspended in 50 µl of 0.5 M EDTA and 200 µl of sodium dodecyl sulphate and incubated at 50 °C for 2 h. Resultant lysate was thoroughly mixed with one volume 25:24:1 phenol:chloroform:isoamyl alcohol solution in a phase separating gel tube and spun at 14,000 rpm for 5 min; repeated once, then again subsequently with 24:1 chloroform:isoamyl alcohol. The aqueous layer was transferred to two volumes ice cold ethanol and gently mixed immediately. Precipitated DNA was then washed with 70% ethanol and solubilised in TE buffer. DNA quality and quantity were assessed using both Nanodrop (ThermoFisher, USA) and Qubit (ThermoFisher, USA).

### PacBio and Illumina MiSeq genome sequencing

The genomic DNA was sequenced using two sequencing platforms, PacBio RSII and Illumina MiSeq. The PacBio sequencing was conducted by Novogene (HK) Co., Ltd, Hong Kong. For Illumina MiSeq sequencing, the genomic library was prepared using Nextera XT kit (Illumina, San Diego, USA) according to manufacturer's protocol and sequenced using 2 × 300 paired-end protocol on an Illumina MiSeq instrument.

### Genome assembly and annotation

The PacBio raw reads were assembled into a single contig using Canu assembler v1.7 [15], after which the assembly was circularized using Circlator v1.5.5 [16]. The

circularised contig was subjected to further correction by mapping of Illumina MiSeq-generated paired-end reads using CLC Genomics Workbench 11 with default parameters (QIAGEN). Genome annotation was performed using Prokka v1.12 [17]. The annotation features are available in Additional file 1. Genome completeness and contamination of HP14039 genome was assessed using the taxonomy\_wf workflow at species level available in CheckM v1.0.13.

### Processing of PacBio methylome data

All raw data in bax.h5 format were converted and merged into a bam file using bax2bam v0.0.8 prior to alignment to *H. pylori* HP14039 complete genome sequence using blasr v5.3.2 with default parameters. The aligned bam output file was then subjected to ipdSummary v2.3 to detect kinetic variations that were predictive of DNA modification events. Finally, the methylated DNA motifs were deduced using MotifMaker v0.3.1 [18, 19]. The density of methylated sites was plotted using Circos v0.69-6 with a window of 5000 bp [20].

### Phylogenetic analysis

The complete genome of HP14039, and 47 publicly available *H. pylori* complete genomes from NCBI database and 12 draft genomes of *H. pylori* strains isolated from our patients who were born in Northeast Africa, were used for core genome phylogeny analysis. The accession numbers of all *H. pylori* genomes used in this study are provided in Additional file 2: Table S1. For consistency, all genomes were annotated by Prokka v1.12 prior to using Roary v3.12.0 [21] for core genome analysis. In the Roary pipeline, sequence alignment of multiple core genes was performed using MAFFT v7.271 [22] and we specified that a gene must be present in all *H. pylori* strains to be considered as a core gene with the percentage identity cut-off of 95. The core alignment was then used to construct a neighbour joining tree using Mega v7.0.2 [23] and the output phylogenetic tree was visualized using Figtree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree>).

### Quality assurance

Species confirmation was performed by using biochemical tests (urease, catalase and oxidase positives) and PCR with seven species-specific housekeeping genes (*atpA*, *esp*, *mutY*, *ppa*, *trpC*, *ureI* and *yphC*). Bacterial culture of pure single colony was used for genomic DNA extraction.

## Results and discussion

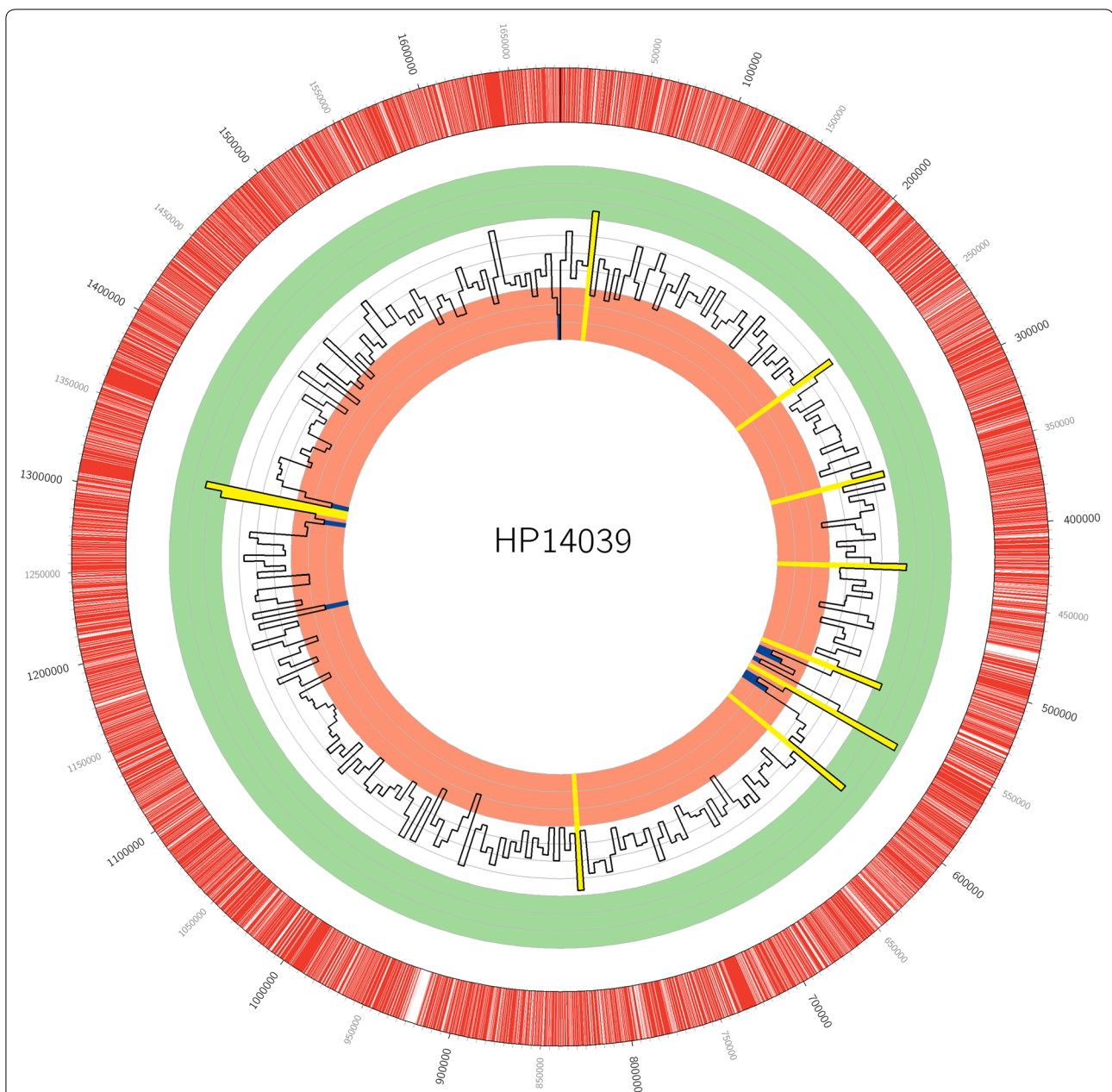
### Strain metadata and genomic characteristics

*Helicobacter pylori* strain HP14039 was isolated from the gastric biopsy of an Australian resident who was born in Somalia, located in the Northeast Africa region.

**Table 1 Presence and absence of major virulence factors in HP14039 complete genome and other hpNEAfrica draft genomes (“+” means presence and “-” means absence)**

| Virulence factors   | HP14039 | HP01234 | HP07036 | HP08058 | HP08061        | HP08074        | HP11049 | HP13005        | HP13050        | HP13068        | HP15005 | HP98490 | HP99255 |
|---------------------|---------|---------|---------|---------|----------------|----------------|---------|----------------|----------------|----------------|---------|---------|---------|
| <i>cag</i> -PAI     | -       | -       | +       | -       | +              | -              | -       | + <sup>a</sup> | -              | +              | -       | +       | -       |
| <i>vacA</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>dupA</i>         | +       | +       | -       | -       | -              | + <sup>b</sup> | -       | -              | -              | -              | -       | -       | -       |
| <i>iceA</i>         | +       | +       | +       | -       | -              | +              | +       | -              | -              | +              | +       | +       | +       |
| <i>babA</i>         | +       | +       | +       | +       | -              | + <sup>b</sup> | +       | + <sup>b</sup> | + <sup>b</sup> | + <sup>b</sup> | -       | +       | -       |
| <i>babB</i>         | +       | -       | +       | +       | + <sup>b</sup> | + <sup>b</sup> | -       | -              | -              | + <sup>b</sup> | +       | -       | -       |
| <i>babC</i>         | -       | +       | -       | -       | + <sup>b</sup> | -              | -       | -              | -              | -              | -       | -       | +       |
| <i>sabA</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>oipA</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>alpA</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>alpB</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>hopZ</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>napA</i>         | +       | +       | +       | +       | +              | +              | +       | +              | +              | +              | +       | +       | +       |
| <i>tieA</i> /hp0986 | -       | -       | -       | -       | -              | -              | -       | -              | -              | -              | -       | -       | -       |
| <i>ctkA</i> /hp0940 | -       | -       | -       | -       | -              | -              | -       | -              | -              | -              | -       | -       | -       |

<sup>a</sup> All *cag*-PAI elements are present, except *cagA*<sup>b</sup> Truncated gene



**Fig. 1** Circos plot displaying the density of methylated sites in HP14039 complete genome (5000 bp window). Open reading frames are highlighted in red in the outer ring. The inner histogram displays the abundance of methylated bases within every 5000 bp region over the chromosome, with an average of  $367.9 \pm 49.7$ . The hypermethylated and hypomethylated regions with methylated sites exceeding two standard deviations ( $> 466.7$  and  $< 267.9$ , respectively) are highlighted in yellow and dark blue, respectively. The three outer green tracks within the histogram represent density values ranging from 600 to 481 (from outer to inner). The three inner orange tracks within the histogram represent density values ranging from 320 to 201 (from outer to inner)

The genome of *H. pylori* HP14039 was sequenced using PacBio and Illumina technologies at 257 $\times$  and 159 $\times$  genome coverages, respectively. The final assembled genome is 1,678,260 bp in length with 1574 coding sequences, 36 tRNA genes and 38.7% G+C content. Genome assessment using CheckM revealed

no contamination and 99.13% genome completeness. We found that it completely lacks the *cag* pathogenicity island, which is one of the major virulence factors and is thought to be associated with the development of gastric cancer [24]. Other major *H. pylori* virulence factors present in *H. pylori* HP14039 are listed in Table 1.

**Table 2 Methylated motifs detected in HP14039**

| Recognition site <sup>a</sup> | Modification type | No. detected | No. in genome | Detected (%) | Restriction component(s)   |                      | Methylation component(s) |                            | Refs        |            |            |
|-------------------------------|-------------------|--------------|---------------|--------------|----------------------------|----------------------|--------------------------|----------------------------|-------------|------------|------------|
|                               |                   |              |               |              | Locus tag                  | Ortholog             | % identity               | Locus tag                  |             | Ortholog   | % identity |
| 5'-ACNGT                      | m4c               | 976          | 1112          | 87.8         | HP14039_01407-01406        | K747_11000           | 94.1, 93.9               | HP14039_01408              | K747_10995  | 95.5       | [8]        |
| 5'-ATTAAT                     | m6a               | 949          | 952           | 99.7         | HP14039_01410              | jhp0431              | 94.1                     | HP14039_01411              | jhp0430     | 94.9       | [26, 29]   |
| 5'-CATG                       | m6a               | 14,872       | 14,874        | 100          | HP14039_00716              | hp1209               | 93.3                     | HP14039_00717              | hp1208      | 93.8       | [29]       |
| 5'-CCATC                      | m6a               | 2190         | 2199          | 99.6         | HP14039_00375-00376        | K747_12645           | 95.7, 92.7               | HP14039_00372-00373        | K747_03690  | 96.2, 97.7 | [8]        |
| 5'-CCGG                       | m4c               | 3535         | 3542          | 99.8         | HP14039_01604              | hp0262               | 96.1                     | HP14039_01603              | hp0263      | 94.2       | [29]       |
| 5'-CTNAG                      | m4c               | 6164         | 6172          | 99.9         | HP14039_01555              | HpyHI                | 96                       | HP14039_01554              | Mi.HpyHI    | 94.9       | [30]       |
| 5'-GAGG                       | m6a               | 4672         | 4718          | 99           | HP14039_00203 <sup>b</sup> | -                    | -                        | HP14039_00204              | hp0050      | 95.2       | [26, 29]   |
| 5'-GATC                       | m6a               | 10,541       | 10,548        | 99.9         | HP14039_00158              | hp0091               | 93.7                     | HP14039_00157              | hp0092      | 92.8       | [29]       |
| 5'-GTAC                       | m6a               | 300          | 304           | 98.7         | HP14039_01382-01381        | jhp0455              | 93.4, 93.7               | HP14039_01383              | jhp0454     | 95.8       | [26, 31]   |
| 5'-GTNNAC                     | m6a               | 820          | 826           | 99.3         | HP14039_00987              | hp0909               | 91.7                     | HP14039_00986              | hp0910      | 96.1       | [29]       |
| 5'-TCGA                       | m6a               | 610          | 612           | 99.7         | -                          | -                    | -                        | HP14039_01606-01607        | hp0260      | 96.1, 97.6 | [32]       |
| 5'-TCNNGA                     | m6a               | 3795         | 3808          | 99.7         | HP14039_00847-00846        | jhp1013 <sup>c</sup> | 96.8, 97.5               | HP14039_00848              | jhp1012     | 96.1       | [26]       |
| 5'-TGCA                       | m6a               | 11,060       | 11,072        | 99.9         | HP14039_00305              | HpyCH4 V             | 96.6                     | HP14039_00306              | Mi.HpyCH4 V | 93.4       | [12, 30]   |
| 5'-CTANNNNNITAYG              | m6a               | 191          | 192           | 99.5         | -                          | -                    | -                        | HP14039_00325 <sup>d</sup> | K747_03505  | 93         | [8]        |
| 5'-WCANHHNNNTG                | m6a               | 1732         | 4440          | 39           | -                          | -                    | -                        | HP14039_01429 <sup>e</sup> | K747_10905  | 92         | [8]        |

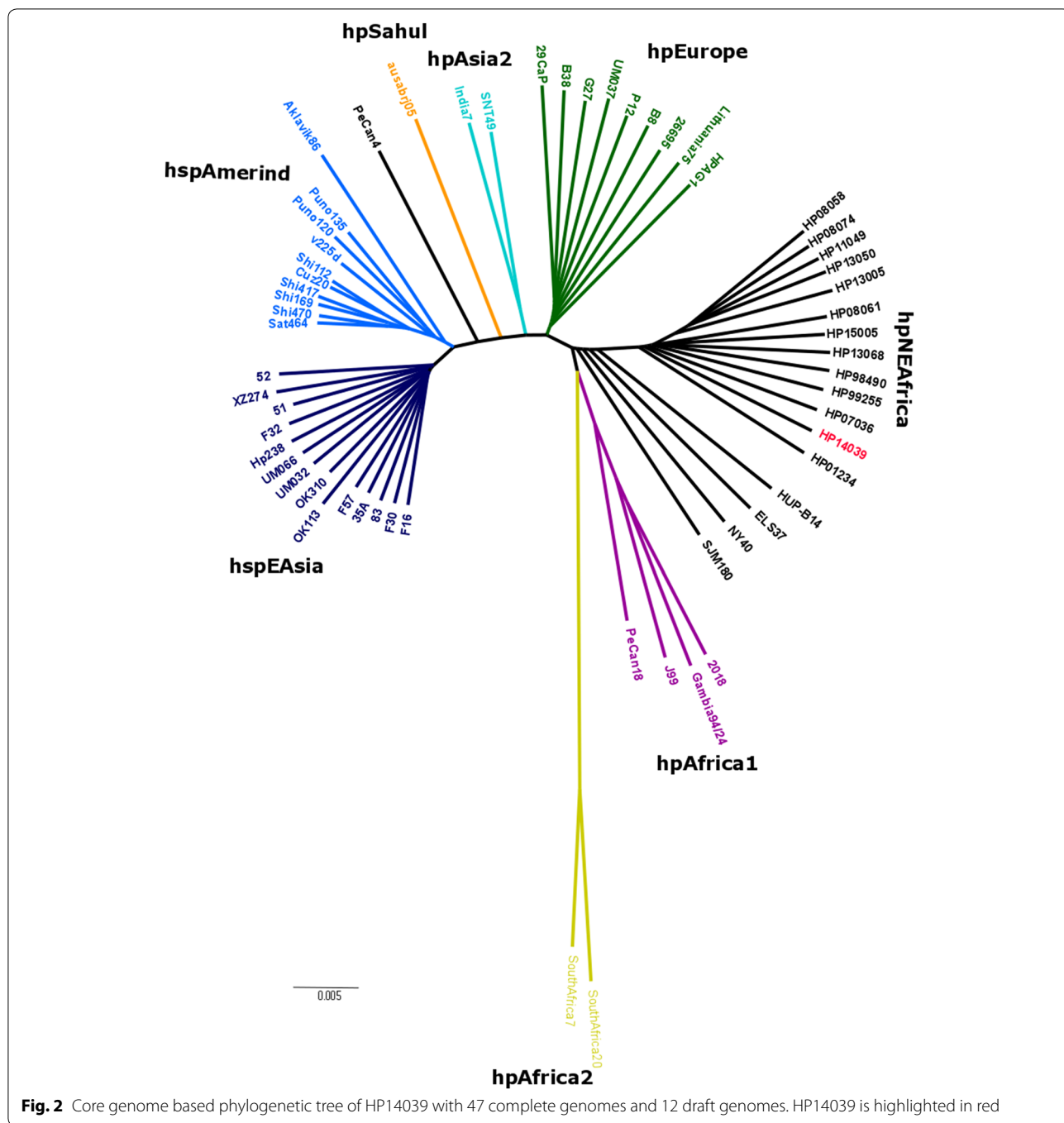
<sup>a</sup> The methylated base within the motif is underlined while the modified base in the complementary strand is highlighted in red

<sup>b</sup> The 1–563 bp and 550–918 bp regions of HP14039\_00203 demonstrated significant nucleotide sequence homology to the N-terminus of hp0052 and the C-terminus of hp0051, respectively

<sup>c</sup> Putative type II restriction enzyme probably recognising TCNNGA

<sup>d</sup> Predicted type I DNA methylase probably recognising CTANNNNNITAYG based on high sequence similarity with K747\_03505 that recognises a closely related sequence motif, which is GANNNNNNITAYG

<sup>e</sup> Predicted type I DNA methylase probably recognising WCANHHNNNTG. This gene is highly similar to K747\_10905, which encodes a type I DNA methylase that recognises CCANNNNNITC sequence motif



**Methylome of HP14039**

Pacbio SMRT sequencing technology has the advantage of being able to detect the epigenetic state of sequenced DNA, and allow identification of modified nucleotides and methylated motifs. In HP14039 genome, a total of 62,407 methylated genomic positions were detected (m6A and m4C). The distribution of methylated bases over the HP14039 chromosome is presented in Fig. 1.

Fifteen functional MTases were identified of which thirteen were assigned to their MTase genes based on previous studies [8, 25, 26]. Two methylated motifs, WCANHHNNNTG and CTANNNNNNTAYG detected in this study were not described in earlier studies. All recognition sequence motifs and their corresponding MTases are summarised in Table 2.



## Phylogeny

The neighbour joining tree was constructed using core genome alignment derived from 48 complete *H. pylori* genomes including HP14039, and additionally 12 draft genomes of *H. pylori* strains isolated from patients originated from similar African region as HP14039. Among the 12 clinical strains that were included, two were from Somalia, identical to that of HP14039; four each from Sudan and Ethiopia, respectively; and the remaining two were from Eritrea. As *H. pylori* infection is common in early childhood [27], it is therefore highly likely that the patients have acquired these individual strains locally when young prior to their migration to Australia. The phylogenetic tree showed clear separation of *H. pylori* population types (Fig. 2). As expected, HP14039, along with other 12 clinical strains with similar geographical origins, were found clustered together. Importantly, HUP-B14, ELS37 and SJM180, which were isolated from Spain, El Salvador and Peru, respectively, were found to be closely related to hpNEAfrica and hpAfrica1 populations despite previous reports that these strains belong to the hpEurope population [28]. This indicates that the birthplace of the patient plays a more important and accurate role in determining the population type of a *H. pylori* isolate, than the geographical origin where the clinical isolate was acquired, as countless individuals are constantly migrating and moving in today's globalised world.

## Additional files

**Additional file 1.** The original genbank file that shows annotation features of HP14039.

**Additional file 2: Table S1.** Strains used in this study to construct core genome phylogenetic tree.

## Authors' contributions

Preparation of draft manuscript: BL. Manuscript revision: EGC, MJW and CYT. Data analysis: BL, EGC and CYT. Endoscopy: BJM. Bacterial culture and Illumina MiSeq sequencing: BL and CL. Study design and conceptualisation: CYT and MJW. Funding and resources: BJM. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Marshall Centre for Infectious Diseases Research and Training, School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia. <sup>2</sup> Department of Computer Science and Software Engineering, University of Western Australia, Perth, WA, Australia. <sup>3</sup> Shenzhen Dapeng New District Kuichong People Hospital, Shenzhen City 518119, Guangdong Province, China.

## Acknowledgements

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The complete genome sequence of *H. pylori* HP14039 has been deposited in DDBJ/ENA/Genbank under the accession number CP034147.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

This study was approved by the Sir Charles Gairdner and Osborne Park Health Care Group Human Research Ethics Committee (HREC No: 2013-007). The gastric biopsy sample used for culturing in this study was obtained with informed and written consent from a patient who presented for endoscopy at Sir Charles Gairdner Hospital.

## Funding

This project was supported by Shenzhen's Sanming Project (Grant No: SZSM201510050), National Health and Medical Research Council (Grant No. 572723) and the Vice Chancellor of the University of Western Australia.

## Publisher's Note

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

Received: 14 December 2018 Accepted: 1 February 2019

Published online: 20 February 2019

## References

- Covacci A, Telford JL, Del Giudice G, Parsonnet J, Rappuoli R. *Helicobacter pylori* virulence and genetic geography. *Science*. 1999;284(5418):1328–33.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*. 2007;445(7130):915–8.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, et al. Traces of human migrations in *Helicobacter pylori* populations. *Science*. 2003;299(5612):1582–5.
- Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, et al. Age of the association between *Helicobacter pylori* and man. *PLoS Pathog*. 2012;8(5):e1002693.
- Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, et al. The peopling of the Pacific from a bacterial perspective. *Science*. 2009;323(5913):527–30.
- Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*. 1999;397(6715):176–80.
- Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, et al. The complete genome sequence of *Helicobacter pylori* strain G27. *J Bacteriol*. 2009;191(1):447–8.
- Lee WC, Anton BP, Wang S, Baybayan P, Singh S, Ashby M, et al. The complete methylome of *Helicobacter pylori* UM032. *BMC Genomics*. 2015;16:424.
- Oh JD, Kling-Backhed H, Giannakis M, Xu J, Fulton RS, Fulton LA, et al. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc Natl Acad Sci U S A*. 2006;103(26):9999–10004.
- Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev*. 2013;77(1):53–72.
- Vale FF, Megraud F, Vitor JM. Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration. *BMC Microbiol*. 2009;9:193.
- Xu Q, Morgan RD, Roberts RJ, Blaser MJ. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc Natl Acad Sci U S A*. 2000;97(17):9671–6.
- Nell S, Eibach D, Montano V, Maady A, Nkwescheu A, Siri J, et al. Recent acquisition of *Helicobacter pylori* by Baka pygmies. *PLoS Genet*. 2013;9(9):e1003775.
- Marmur J. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J Mol Biol*. 1961;3(2):208–IN1.

15. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36.
16. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 2015;16:294.
17. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9.
18. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 2012;40(4):e29.
19. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010;7(6):461–5.
20. 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.
21. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31(22):3691–3.
22. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
23. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
24. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, et al. Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.* 1995;55(10):2111–5.
25. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* 2014;10(4):e1004272.
26. Krebs J, Morgan RD, Bunk B, Sproer C, Luong K, Parusel R, et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* 2014;42(4):2415–32.
27. Weyermann M, Rothenbacher D, Brenner H. Acquisition of *Helicobacter pylori* infection in early childhood: independent contributions of infected mothers, fathers, and siblings. *Am J Gastroenterol.* 2009;104(1):182–9.
28. Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet.* 2017;13(2):e1006546.
29. Lin LF, Posfai J, Roberts RJ, Kong H. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc Natl Acad Sci U S A.* 2001;98(5):2740–5.
30. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 2015;43(Database issue):D298–9.
31. Humbert O, Salama NR. The *Helicobacter pylori* HpyAXII restriction-modification system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric conservation of the DNA methyltransferase and restriction endonuclease components. *Nucleic Acids Res.* 2008;36(21):6893–906.
32. Vitkute J, Stankevicius K, Tamulaitiene G, Maneliene Z, Timinskas A, Berg DE, et al. Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. *J Bacteriol.* 2001;183(2):443–50.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

