PCY Woo 胡釗逸
KY Yuen 袁國勇
SKP Lau 劉嘉珮
H Tse 謝勰文

Random exploration of the *Laribacter hongkongensis* genome

Key Messages

1. Random exploration of the *Laribacter hongkongensis* genome was performed. A total of 1957 random sequence tags were sequenced and analysed, which represents about 26% of the *L hongkongensis* genome.
2. Random exploration of the genome can be completed at a fraction of the complete genome costing. The limited partial sequence obtained in this study revealed ample evidence about the genetic composition and potential virulence factors of *L hongkongensis*.
3. Different classes of potential virulence factors were identified. Among these, a complete urease cassette, which is a major virulence factor in another gastrointestinal pathogen, *Helicobacter pylori*, was observed.
4. Random exploration also provided a foundation for the complete genome sequencing of *L hongkongensis*.

Hong Kong Med J 2010;16(Suppl 4):S15-7

Department of Microbiology, The University of Hong Kong
PCY Woo, KY Yuen, SKP Lau, H Tse

RFCID project number: 05050012

Principal applicant:
Prof Patrick CY Woo
Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China
Tel: (852) 2855 4892
Fax: (852) 2855 1241
Email: pcywoo@hkucc.hku.hk

Corresponding author:
Prof Kwok-yung Yuen
Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China
Tel: (852) 2855 4892
Fax: (852) 2855 1241
Email: hkumicro@hkucc.hku.hk

Introduction

Laribacter hongkongensis, a novel genus and species, was first discovered in Hong Kong in 2001.[1] Phenotypically, it is a facultative anaerobic, motile, non-sporulating, urease-positive, Gram-negative, S-shaped bacillus. Phylogenetic analysis using 16S rRNA gene sequences revealed that *L hongkongensis* belongs to the *Neisseriaceae* family of the β-subclass of Proteobacteria. During a period of 2 months, *L hongkongensis* was discovered, on charcoal cefoperazone deoxycholate agar, in three of our patients with community-acquired gastroenteritis. A similar finding was also observed in three other patients in Switzerland.[2] In a multi-centre prospective study using cefoperazone MacConkey agar as the selective medium,[3] we confirmed that *L hongkongensis* is associated with community-acquired gastroenteritis and traveller's diarrhoea.[4] Furthermore, freshwater fish were also confirmed to be a reservoir of *L hongkongensis*.[4,5] *L hongkongensis* is likely to be globally distributed, as travel histories from patients suggested that it was present in at least four continents, including Asia, Europe, Africa and Central America.[2,4] *L hongkongensis* has also been reported from another coastal province in mainland China. Although the causative role of *L hongkongensis* in gastroenteritis is yet to be established, these data provide strong evidence that the bacterium is a potential diarrhoeal pathogen that warrants further investigation.

Sequencing the complete genome of microorganisms has revolutionised the study of microbiology and infectious disease. However, due to the relative high cost of sequencing a complete genome, sample sequences of bacterial genomes have been used for characterisation of the microorganism. Recently, we published a sample sequence of *Penicillium marneffei*, which has facilitated further molecular research and provided the foundation for complete genome sequencing of this dimorphic fungus. In this study, random exploration of the *L hongkongensis* genome was performed, and 1957 random sequence tags (RSTs) were sequenced and analysed, which represents about 26% of the *L hongkongensis* genome. This study has laid down the foundation of complete genome sequencing of this bacterium.[6]

Methods

This study was conducted from July 2006 to December 2007. In order to have better understanding of the biology and putative virulence mechanisms of *L hongkongensis*, we sequenced and annotated 1957 RSTs, which represents about 26% of the *L hongkongensis* genome. The DNA sequences of the resultant contigs and singlets were analysed using the Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information server at the National Library of Medicine (http://www.ncbi.nlm.nih.gov). Special attention was paid to comparison with genomes and other known virulence genes responsible for adhesion, survival, diarrheogenesis in *Chromobacterium violaceum* (a bacterium phylogenetically closely related to *L hongkongensis* with its complete genome sequence available) and other diarrhoeal pathogens.

Single sequencing reads, with an average length of about 500 bp, were obtained for one end of 2000 clones, using the T3 vector primer of pBK-CMV.

Using the sequence information of the sample sequences, additional

polymerase chain reaction (PCR) primers (Life Technologies Corp, CA, USA) were designed to amplify the gaps of the urease cassette of *L hongkongensis*.

DNA sequencing was performed with an ABI 3700 automated sequencer according to the manufacturer's instructions (Life Technologies Corp, CA, USA). To remove low-quality traces, the raw sequence reads were processed by *ad hoc* scripts. After filtering, the entire sets of RSTs were assembled into contigs using the Phred/Phrap/Consed software. The DNA sequences of the resultant contigs and singlets were analysed by BLAST search. The searches were performed at both the protein and DNA levels. All results were inspected and interpreted manually.

Phylogenetic tree construction of the predicted protein products of the urease cassette was performed using ClustalX (version 1.81) and the neighbour-joining method with GrowTree (Genetics Computer Group, Inc, WI, USA).

Results

General features of the random sequence tags

A random analysis of 1957 RSTs representing about 26% of the genome of *L hongkongensis* was performed. Clusters of Orthologous Groups (COG) classification showed that 6.2% of the genes had putative functions related to translation, 5.1% to transcription, 4.9% to replication, recombination and repair, 1.2% to cell cycle control, mitosis and meiosis, 1.3% to defence mechanisms, 5% to signal transduction mechanisms, 6.2% to cell wall and cell membrane biogenesis, 2.2% to cell motility, 1.8% to intracellular trafficking and secretion, 4.1% to posttranslational modification, protein turnover and chaperones, 6.4% to energy production and conversion, 2.7% to carbohydrate transport and metabolism, 8.3% to amino acid transport and metabolism, 2.1% to nucleotide transport and metabolism, 4.1% to coenzyme transport and metabolism, 3.1% to lipid transport and metabolism, 5.1% to inorganic ion transport and metabolism, 1.5% to secondary metabolites biosynthesis, transport and catabolism, 10.5% to general function prediction only, 7.5% to unknown functions, and 10.8% were not in the COGs. Phylogenetically, a large proportion of the genes were homologous to those of other members of the *Neisseriaceae* family of beta-proteobacteria, most commonly, *C violaceium*, the most closely related bacterium whose complete genome is available. Similar to *C violaceum*, *L hongkongensis* possessed a higher proportion of genes with putative functions related to signal transduction mechanisms (COG:T).

Potential virulence factors

Different classes of potential virulence factors were observed. They included genes that putatively encode urease, haemolysins, RTX toxin, phospholipase and adhesin. They also included a diverse array of protein secretion, iron uptake, efflux pump and lipopolysaccharide production systems.

Using the information of the sample sequences, additional PCR primers were designed to amplify the gaps of the urease cassette of *L hongkongensis*. The complete urease cassette, occupying a 7556 bp region, includes eight ORFs which encode three urease structural proteins (UreA, UreB and UreC) and five accessory proteins (UreE, UreF, UreG, UreD and UreI). Similar to the urease of other bacteria, the urease of *L hongkongensis* is presumably a nickel-containing enzyme. The histidine residues at the carboxyl terminal of UreE are supposed to bind to the nickel ions that are transported into *L hongkongensis* through a nickel transporter, and donate the nickel ions to UreC during urease activation. Phylogenetically, all seven genes in the urease cassette are closely related to the corresponding homologues in *Brucella* species, *Yersinia* species and *Photorhabdus luminescens* subsp *laumondii*.

Discussion

The complete sequencing of bacterial genomes has revolutionised microbiology. However, the current high cost of completely sequencing genomes has limited application of such technology to important pathogens and commercially important bacteria. The majority of this cost is incurred owing to the labour-intensive methods, which must still be used to close gaps covering the last few percent of the genome and to reduce error rates to below 0.1%. In contrast, a partial sequence of a bacterial genome can be obtained at low cost, which makes it possible to consider sample sequencing of multiple genomes within a species, genus, or family. When a completely sequenced genome and a closely related sample-sequenced genome are compared, it is possible to identify sequences in the sampled genome that are already characterised in the completely sequenced genome. The vast GenBank database can be considered a huge collection of sample sequences for these purposes.

In addition to the cost advantage of partial genome sequencing, the limited partial sequences obtained in this study revealed ample evidence about the genetic composition and the potential virulence factors of *L hongkongensis*. With about 26% of the genome sequenced, the predicted coding ORFs can be classified into all COG functional categories, except for genes of RNA processing and modification. As for *C violaceum*, analyses indicated that a substantial proportion of the genes in the *L hongkongensis* genome had putative functions related to signal transduction (COG T), which are related to bacterial survival under diverse environmental conditions. This is in line with the lifestyles of the bacteria. *C violaceum* is a highly versatile, soil- and water-borne, free-living bacterium and therefore requires the highest proportion of genes devoted to this COG group. *Neisseria meningitidis* and *Neisseria gonorrhoeae*

are strictly aerobic bacteria with humans as the only host, and therefore they require the lowest proportion of genes. *L hongkongensis* can survive in human and freshwater fish and therefore require an intermediate proportion of genes with putative functions related to signal transduction.

Different classes of potential virulence factors were identified. Among these, a complete urease cassette, which is a major virulence factor in another gastrointestinal pathogen, *Helicobacter pylori*, was observed in the *L hongkongensis* genome. Urease is able to hydrolyse the limited amount of urea available in the stomach to generate carbon dioxide and ammonia, which increases the pH. Therefore, the complete urease cassette observed in the *L hongkongensis* genome probably relates to its survival in the gastrointestinal tract, as the bacterium has to pass through the highly acidic environment of the stomach before reaching the intestine. Bioinformatic analysis showed that all seven genes in the urease cassette are closely related to the corresponding homologues in *Brucella* species (α-proteobacteria), *Yersinia* species (γ-proteobacteria) and *P luminescens* subsp *laumondii* (γ-proteobacteria), instead of those in other members of β-proteobacteria. This indicates that *L hongkongensis* has probably acquired the genes through horizontal gene transfer after evolution into a distinct species.

Acknowledgement

This study was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#05050012).

References

1. Yuen KY, Woo PC, Teng JL, Leung KW, Wong MK, Lau SK. *Laribacter hongkongensis* gen. nov., sp. nov., a novel Gram-negative bacterium isolated from a cirrhotic patient with bacteremia and empyema. J Clin Microbiol 2001;39:4227-32.
2. Woo PC, Kuhnert P, Burnens AP, et al. *Laribacter hongkongensis:* a potential cause of infectious diarrhea. Diagn Microbiol Infect Dis 2003;47:551-6.
3. Lau SK, Woo PC, Hui WT, et al. Use of cefoperazone MacConkey agar for selective isolation of *Laribacter hongkongensis*. J Clin Microbiol 2003;41:4839-41.
4. Woo PC, Lau SK, Teng JL, et al. Association of *Laribacter hongkongensis* in community-acquired human gastroenteritis with travel and with eating fish: a multicentre case-control study. Lancet 2004;363:1941-7.
5. Teng JL, Woo PC, Ma SS, et al. Ecoepidemiology of *Laribacter hongkongensis*, a novel bacterium associated with gastroenteritis. J Clin Microbiol 2005;43:919-22.
6. Woo PC, Lau SK, Tse H, et al. The complete genome and proteome of *Laribacter hongkongensis* reveal potential mechanisms for adaptations to different temperatures and habitats. PLoS Genet 2009;5:e1000416.