Infection, Genetics and Evolution 60 (2018) 109–116

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Review

Genetic populations and virulence factors of *Helicobacter pylori*

Evariste Tshibangu Kabamba[a,c,1], Vo Phuoc Tuan[a,d,1], Yoshio Yamaoka[a,b,*]

[a] Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine, 1-1 Idaigaoka, Hasama-machi, Yufu-City, Oita 879-5593, Japan
[b] Department of Medicine-Gastroenterology, Baylor College of Medicine and Michael E. Debakey Veterans Affairs Medical Center, 2002 Holcombe Blvd., Houston, TX 77030, USA
[c] Department of Internal Medicine, University of Mbujimayi Faculty of Medicine, Mbujimayi, The Democratic Republic of Congo
[d] Department of Endoscopy, Cho Ray Hospital, Ho Chi Minh, Viet Nam

ARTICLE INFO

Keywords:
Genetic population
Helicobacter pylori
Virulence factor

ABSTRACT

Helicobacter pylori is a bacterium that has infected more than half of the human population worldwide. This bacterium is closely associated with serious human diseases, such as gastric cancer, and identifying and understanding factors that predict bacterial virulence is a priority. In addition, this pathogen shows high genetic diversity and co-evolution with human hosts. *H. pylori* population genetics, therefore, has emerged as a tool to track human demographic history. As the number of genome sequences available is increasing, studies on the evolution and virulence of *H. pylori* are gaining momentum. This review article summarizes the most recent findings on *H. pylori* virulence factors and population genetics.

1. Introduction

Helicobacter pylori (*H. pylori*) is a Gram-negative spiral-shaped bacterium found in the gastric epithelium of humans (Suerbaum and Michetti, 2002). Since its first description by Marshall and Warren (1984), this species has gained considerable research attention because of its clinical and evolutionary importance (Marshall and Warren, 1984; Suerbaum and Achtman, 1999; Yamaoka, 2010).

The bacterium is primarily transmitted within families and acquired during childhood. In the absence of adequate treatment, life-long gastric colonization can result in several diseases such as chronic gastritis, peptic ulcer, and gastric cancer (Yamaoka, 2010). The association between gastric cancer—one of the most common malignancies in the world—and *H. pylori* infection has attracted great interest worldwide, with the International Agency for Research on Cancer (IARC), a subordinate organization of the World Health Organization (WHO), classifying *H. pylori* as a "group 1 (definite carcinogen)" in 1994. Therefore, many studies have explored bacterial factors to explain the link between gastric cancer and this bacterium. Several bacterial virulence factors have been identified that predict severe clinical outcomes and explain the global geographic distribution of gastric cancer (Yamaoka, 2010). Thus far, the biological function and structure of these virulence factors, as well as the discovery of new candidate virulence factors, has continued to occupy most *H. pylori*-related research.

H. pylori and humans have co-evolved for at least 100,000 years, long before human ancestors left Africa (Moodley et al., 2012). During this long history in its hostile gastric niche in humans, *H. pylori* has developed a wide spectrum of strategies to persist in and adapt to changing conditions in and around its host (Suerbaum and Achtman, 1999). Thus, *H. pylori* is one of the most diverse bacterial species and arguably the most successful human pathogen known to date (Falush et al., 2003b; Suerbaum and Achtman, 1999). However, despite of their high genetic diversity, *H. pylori* strains appear genetically structured, exhibiting phylogeographic patterns that consistently correlate with that of their human hosts. Therefore, the population genetics and phylogenetic relationships among isolates have enabled accurate mapping of the local and global demographic histories of human evolution (Falush et al., 2003b; Linz et al., 2007; Moodley et al., 2012; Moodley et al., 2009). *H. pylori* genetics is promising to shed light on yet unknown dynamics of human evolution. Therefore, an increasing amount of resources are being devoted to detailed analyses of *H. pylori* populations with the aim of describing human history.

In addition, the increasing availability of *H. pylori* whole-genome sequences is enabling more genomic analyses than ever before. Such analyses are empower efforts to detect new virulence factors as well as detailed studies of population genetics. The present literature review addresses the most recent and important findings on bacterial virulence factors and genetic populations of *H. pylori*. Scientific data, mostly that

* Corresponding author at: Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine, 1-1 Idaigaoka, Hasama-machi, Yufu-City, Oita 879-5593, Japan.
E-mail address: yyamaoka@oita-u.ac.jp (Y. Yamaoka).
[1] Evariste Tshibangu Kabamba and Vo Phuoc Tuan contributed equally in this article.

https://doi.org/10.1016/j.meegid.2018.02.022
Received 6 November 2017; Received in revised form 15 February 2018; Accepted 16 February 2018
Available online 19 February 2018

reported in the last three years, are summarized, with the aim of highlighting expected future developments in *H. pylori*-related molecular epidemiological research.

2. New insights on *H. pylori* virulence factors

The pathogenesis of *H. pylori* is driven by several virulence factors that facilitate colonization, induce inflammation, and damage host cells. These virulence factors have been linked to the risk of developing severe gastric diseases and include the *cag* pathogenicity island (PAI), vacuolating cytotoxin (VacA), outer membrane proteins (OMPs), serine protease HtrA, and many others.

2.1. VacA

VacA is an exotoxin that was named for its capacity to induce host cell vacuolation (Cover and Blaser, 1992). At the time of its discovery, no bacterial toxin with similar activity had yet been described, and since this discovery, many studies have been conducted to clarify its function and structure.

VacA have been described as a multi-receptor protein that has pleiotropic effects, including membrane depolarization, mitochondrial dysfunction, autophagy, activation of mitogen-activated protein kinases, inhibition of T cell function, and induction of apoptosis (Foegeding et al., 2016). These functions contribute to the persistent colonization of *H. pylori* and its pathogenesis in several upper digestive tract diseases. Recently, additional VacA-related pathways and functions have been reported. Amilon et al. (2015) described a putative stem-loop structure in the 5′ untranslated region that affects transcription of *vacA* and leads to higher expression and toxicity of VacA (Amilon et al., 2015). An extra-digestive location of functional VacA in lungs has led to the suggestion that VacA plays a role in the pathogenesis of respiratory diseases by inducing IL-8 and IL-6 (Nakashima et al., 2015). In addition, new host factors that interact with or regulate the VacA-induced apoptosis have been reported. Yahiro et al. (2015) described a new signaling pathway for VacA-induced apoptosis that is mediated by cytoplasmic accumulation of connexin 43 (Cx43), a ubiquitous connexin family member that plays a role in gap junction and cell-cell channel formation (Yahiro et al., 2015). In addition, Chang et al. (2016) described the role of cortactin, an actin-binding protein, in the regulation of apoptosis induced by VacA (Chang et al., 2016).

VacA includes a 33-kDa N-terminal domain associated with cytotoxicity and a 55-kDa C-terminal domain involved in binding of the bacterium to cell surface receptors (Yahiro et al., 2015). Almost all *H. pylori* strains harbor the *vacA* gene, and allelic polymorphisms of the protein show clinical significance and toxic activity that are associated with specific combination of its three regions: the signal peptide (s1 and s2 variants), the intermediate region (i1, i2, and i3 variants), and the middle region (m1 and m2 variants). Molecular epidemiological studies have revealed two novel polymorphic sites, the deletion (d1 and d2 variants) and c-region (c1 and c2 variants) located in the 3′- end region of VacA (Fig. 2) (Thi Huyen Trang et al., 2016). Similar to sites described previously, some variants of these two novel regions have been associated with high risk of gastric cancer (Bakhti et al., 2016; Ogiwara et al., 2009). However, the contributions of these regions to different VacA functions such as vacuole formation have not yet been identified.

2.2. cag PAI

cag PAI is a chromosomic region of approximately 37 kb that encodes the *cag* type IV secretion system (*cag*-T4SS), including cytotoxin-associated gene A (*cagA*) (Fig. 1A). CagA is a 120–140-kDa cellular effector that is translocated into host cells through the *cag*-T4SS and interacts with a large repertoire of cellular signaling pathways, including those leading to carcinogenesis (Tegtmeyer et al., 2017a). CagA was discovered prior to the *cag* PAI and was named for its presumed link with the vacuolating cytotoxin activity of the VacA protein that had been discovered two years before (Tummuru et al., 1993). Since then, CagA and *cag* PAI have shown no effect on vacuolating cytotoxin production, suggesting the possible functional independence of *vacA* and *cagA*, two genes located approximately 300 kb apart on the *H. pylori* chromosome (Tummuru et al., 1994).

The *cag* PAI is currently the most extensively studied *H. pylori* virulence factor. Its epidemiological role has been discussed previously (Suzuki et al., 2012; Yamaoka, 2010), and numerous studies have shown that *cag* PAI-positive *H. pylori* strains are associated with an increased risk of gastric cancer compared to strains that lack *cag* PAI (Yamaoka, 2010). The risk of gastric cancer is determined by several *cag* PAI-related features, including *cag* PAI-positivity, sequence variation within CagA (such as the number and type of EPIYA motifs), and the presence or absence of a functional *cag* type IV secretion system (which translocates CagA into host cells) (Suzuki et al., 2012; Tegtmeyer et al., 2017a; Yamaoka, 2010). Host c-Src and c-Abl kinases control hierarchic phosphorylation and CagA function in Western and East Asian *H. pylori* strains (Mueller et al., 2012). Recently, the function and structure of the CagA and *cag*-T4SS have been elucidated further.

First, the crystal structure of the N-terminal segment of the CagA molecule, which harbors a unique structure with no sequence homology to any known proteins, was recently obtained (Hayashi et al., 2012). The structured N-terminal part of CagA consists of several domains and harbors the putative integrin-binding region (Hayashi et al., 2012; Kaplan-Turkoz et al., 2012). The unstructured C-terminal region displays recognized repeated sections, EPIYA (Glu-Pro-Ile-Tyr-Ala), and CM (CagA multimerization) or CRPIA (conserved repeat responsible for phosphorylation-independent activity) motifs, as well as a region that binds to the secretion chaperone CagF, and a C-terminal secretion signal (Fig. 1B) (Schindele et al., 2016).

Second, further insights in the molecular mechanisms regulating *cagA* function through the *cag* PAI have been released. In fact, the *cagA* promoter region, which had been described previously (Loh et al., 2011), was further characterized by Ferreira et al. (2016). This study identified specific sequence motifs located in the promoter region (+59 AATAAGATA and −10 TATAATGA sequence motifs) that are linked to CagA expression levels and interleukin-8 (IL-8) secretion in an infected gastric cell line, as well as to severe clinical outcomes (Ferreira et al., 2016). Because these sequence variations can be used to discriminate between two different levels of gastric cancer risk associated with Colombian strains and those with European and African origins, the discussion should be extended in future studies to strains from other geographical origins. Another important *cagA*-related feature identified recently is the number of copies of this gene found in different strains. Jang et al. (2017) showed that *H. pylori* isolates can carry multiple tandem copies of *cagA* that affect CagA expression and activity and may impact the development of gastric disease (Jang et al., 2017). Consistent with the findings of Jang et al. (2017), Draper et al. (2017) showed, using close strains named PMSS1 and SS1, that the number of *cagA* changes dynamically and modulates CagA activity (Draper et al., 2017). Thus, future epidemiological studies should address not only the sequence variation within CagA (EPIYA and CM/CRPIA motifs) but also the functionality of the whole *cag* PAI/T4SS in determining the biological effects of CagA, as well *cagA* promoter variants and the number of *cagA* copies as useful markers for disease risk. A β-lactamase-dependent reporter system that allows a precise and quantitative determination of CagA translocation into host cells has just been developed (Schindele et al., 2016). This phosphorylation-independent assay has opened the door to further insights into the in vivo function and epidemiological role played by *H. pylori cag*-T4SS.

Finally, an integrative model of the activity of translocated CagA was recently developed (Tegtmeyer et al., 2017a). This model summarizes the data available on highly complex signaling pathways altered by translocated CagA through multiple receptor kinases (c-Met and EGFR) and non-receptor kinases (Src, Abl, Csk, aPKC, Par1, PI3K,