Adv Exp Med Biol - Advances in Microbiology, Infectious Diseases and Public Health (2019) 1149: 1–16
https://doi.org/10.1007/5584_2019_356

Published online: 24 April 2019

The Story of *Helicobacter pylori*: Depicting Human Migrations from the Phylogeography

Langgeng Agung Waskito and Yoshio Yamaoka

Abstract

Helicobacter pylori is a spiral-shaped Gram-negative bacterium, which has infected more than half of the human population. Besides its colonisation capability, the genetic diversity of *H. pylori* is exceptionally well structured and belongs to several distinct genetic populations, depicting various prehistorical human migration events. The evolutionary relationship of *H. pylori* with its host had been started at least ~100,000 years ago. In addition, the discovery of the ancient *H. pylori* genome from a European Copper Age glacier mummy, "The Iceman", gave the idea that the second out of Africa migration resulted in the recombinant population of hpEurope at least about 5300 years ago. The advancement of next-generation genome sequencing discovered the prophage of *H. pylori* and could discriminate the big population of hpEurope into two different subpopulations. In addition, the implementation of the chromopainter/fineSTRUCTURE algorithm to the whole genome analysis of *H. pylori* provides a finer resolution population genetics of *H. pylori*; therefore it could also depict the recent migrations within the past 500 years after colonial expansion. This discovery shows that the genetic recombination of *H. pylori* strains is far more dynamic compared to its human host, but still maintains the similarity to its host, suggesting that *H. pylori* is a handy tool to reconstruct the human migration both in the past and the recent.

Keywords

Helicobacter pylori · Human migration · Phylogeography · Population genetics · Next-generation sequencing

L. A. Waskito
Department of Environmental and Preventive Medicine, Faculty of Medicine, Oita University, Yufu, Japan

Institute of Tropical Disease, Universitas Airlangga, Surabaya, Indonesia
e-mail: langgengaw@gmail.com

Y. Yamaoka (✉)
Department of Environmental and Preventive Medicine, Faculty of Medicine, Oita University, Yufu, Japan

Department of Medicine, Gastroenterology and Hepatology Section, Baylor College of Medicine, Houston, TX, USA
e-mail: yyamaoka@oita-u.ac.jp

1 Introduction

Helicobacter pylori is a spiral Gram-negative bacterium, which demonstrated its roles in gastritis, peptic ulcer and gastric cancer formation in humans (Uemura et al. 2001). Since its original revelation in the early 1980s, many ongoing studies had been conducted and showed that this bacterium infected more than half of the human population (Hooi et al. 2017). The route of infection is still controversial; however, the most

reasonable route is oral-oral, which indicates the possibility of transmission within families, mainly from parents to their children. In addition to the vertical transmission, *H. pylori* also could be transmitted between unrelated people living nearby (Kivi et al. 2003; Björkholm et al. 2001). These routes of transmission could become the fundamental evolution of this bacterium, which we are going to discuss later.

The evolutionary process resulting in genetic diversity of the organism involves mutation, recombination, migration, selection and genetic drift. These forces are typically sufficient to explain the genetic differences occurring in the evolutionary process. These forces drive the *H. pylori* to evolve as it had been colonising the highest selective pressure environment in the human stomach. The high selective pressure and long-term colonization, in addition to the nature of *H. pylori,* which has unusually high mutation rates (Björkholm et al. 2001) and homologous recombination rates (Falush et al. 2001; Suerbaum et al. 1998). This resulted in extremely high DNA sequence diversity that is much higher compared to that of other bacteria (Achtman et al. 1999) and even 50-fold higher than its human host (Li and Sadler 1991). These attributes are responsible for making *H. pylori* considerably as the most diverse pathogenic bacterium worldwide (Fischer et al. 2010). Interestingly, the high diversity of *H. pylori* genomes is very well structured, which could be divided into several distinct clusters. The population genetics of *H. pylori* is very consistent with the hosting people on different continents; therefore it reflects the human migration events in the past.

The multilocus sequence typing (MLST) approach has been used for many years to characterise the phylogeographic features of bacteria, including *H. pylori*. The combination of MLST sequence data and the STRUCTURE algorithm have discovered the population genetics of *H. pylori,* which correlates with the human migration events (Falush et al. 2003b; Linz et al. 2007). One of the most important virulence genes, *cagA* (Backert et al. 2010; Posselt et al. 2013), also has been proven to be a geographic predictor alongside with MLST and it could differentiate the isolates from East Asian and the Western isolates (Yamaoka 2010; Achtman et al. 1999). In recent years, the next-generation sequencing (NGS) methodology became less expensive, and thus available to generate more whole genome sequences. NGS analysis can obtain much more sequence data and in a shorter time, compared to the conventional Sanger sequencing approach. Therefore, NGS data can give us even better insights in terms of phylogeographic characterisation of *H. pylori* at the recent recombination point of view (Thorell et al. 2017) as well as the new phylogeographic tools, the bacteriophages (Vale et al. 2017). In addition, the discovery of a 5300-year-old *H. pylori* genome from a European Copper Age glacier mummy, "The Iceman", gave us an idea of the hybridisation between Asia and African populations of *H. pylori* (Maixner et al. 2016). Here, we summarise the current understanding of phylogeographic of *H. pylori* as genomic-based evidence to support the concept of human migration events, characterised by the MLST, specific virulence genes and entire genomes.

2 Characterization the Population Genetics of *H. pylori*

Population genetics characterisation is a useful technique to discover the genetic background of a given microorganism. Understanding the genetic background gives us an idea of the relationship of microorganisms and the nature of infection (discover the origin of the microbe and the way of transmission, for example in a particular outbreak); clinical outcomes (discovery of several virulence factors) and the geographic location of isolated microorganisms. In the case of *H. pylori*, which has an incredible highly and well-structured genetic diversity, the population genetics could reconstruct the evolutionary history.

We can characterise population genetics of bacteria in many ways. To characterise the population genetics of a given bacterium, the MLST was first proposed and applied to *Neisseria meningitidis* to determine the lineage from different invasive diseases and healthy carrier in 1998

(Maiden et al. 1998). For each cluster, the cluster analysis was applied to assign numbers arbitrarily and dendrograms by calculating the pairwise differences in multi-locus allelic profiles. This method resulting in ~470 bp fragments from altogether six genes identified meningococcal lineages associated with invasive diseases of humans (Maiden et al. 1998).

In 1999, Achtman and co-workers applied a similar approach to assign the population genetics to *H. pylori*. Their studies proposed the MLST using seven housekeeping genes (i.e., *atpA, efp, mutY, ppa, trpC, ureI*, and *yphC*) and two virulence-associated genes (*vacA* and *cagA*). This method resulted in a clonal descent, which could be observed, and this separation between one population compared to others reflected the geographical origin of the *H. pylori* isolates (Achtman et al. 1999). Currently, the method using seven housekeeping genes became a standard tool to describe the genetic populations of *H. pylori* using MLST analysis.

One of the most common approaches inferring population genetics based on the MLST data is the STRUCTURE algorithm (Pritchard et al. 2000). This algorithm was built using the Bayesian probability approach followed by Markov Chain Monte Carlo (MCMC), which inferred the posterior probability of the given MLST data into several population genetics. However, as *H. pylori* had a lot of interspecies recombination and mutation events, inferring the population genetics of this bacterium is challenging. Employing this condition, another model is implemented in the STRUCTURE software, known as the *linkage model* (Falush et al. 2003a). This model relies on the admixture linkage disequilibrium that is resulted when a gene flow (migration) occurs between genetically distinct populations. Therefore, it infers a finite number of "ancestral" or precursor number of populations (Falush et al. 2003a). In case of *H. pylori*, with the increasing number of investigated isolates from the original MLST study, this approach succeeded to build the population genetics of *H. pylori* and linear with the geographical origin of the isolates (Falush et al. 2003b).

The advancement of NGS put the genetic information of the microorganism to the next level. Increasing the high-throughput data, which contain high abundance of sequences, needed to consider the effectivity and efficiency of the analysis. One of the drawbacks of the STRUCTURE analysis is the time-consuming property. Therefore, a new approach was introduced, and it is called chromopainter and fineSTRUCTURE, which is less time consuming with the larger dataset. The chromopainter algorithm relies on the consideration that each sequence in a given sample is regarded in turn as a recipient, whose chromosomes are reconstructed using chunks of DNA donated by the other individuals. The result of this algorithm can be summarised as a matrix of co-ancestry, which reveals the critical information directly about the ancestral relationship among individuals (Lawson et al. 2012). Therefore, we can obtain a higher resolution of the genetic relationship between one strain and others at the donor-recipient wise. By applying these main methods, population genetics of *H. pylori* could reflect the major human migration events out of Africa, from the prehistorical migrations to the recent migrations.

3 The Phylogeographic of *H. pylori* from Out of Africa to the Pacific

There is accumulating evidence, which showed that modern humans migrated out of Africa to the Arabian Peninsula approximately 60,000–150,000 years ago (kya), then independently went to Europe and Asia (Cavalli-Sforza et al. 1994). By applying the Bayesian inferring algorithm implemented in the STRUCTURE programme, Linz and co-workers demonstrated the simulation, which indicated that *H. pylori* had spread from East Africa approximately ~58 kya, then the population was spreading through East Asia (Linz et al. 2007). The simulation resulted in several different populations, including hpEurope, hpAsia2, hpEastAsia and hpSahul with a characteristic of decreasing

genetic diversity by the increasing distance from Africa, reflecting the genetic drift of each population in the *H. pylori* (Linz et al. 2007).

The European population (hpEurope) is the biggest in terms of number and distribution in the *H. pylori* population genetics. The hpEurope cluster includes almost all *H. pylori* isolated from European countries as far east as Southeast Asian countries (Table 1). In addition, hpEurope also possessed very distinct properties from the other population genetic clusters in global observations. The European genetic properties of *H. pylori* were characterised by great genetic diversity, even higher than in Africa. Observations made with the MLST sequences using *linkage model* (Falush et al. 2003a) implemented in STRUCTURE algorithm showed that the hpEurope was formed by the hybridisation between two ancestral populations, known as Ancestral Europe 1 (AE1) and Ancestral Europe 2 (AE2) (Falush et al. 2003b). Originally, the AE1 was defined as the population found in Ladakh in north India, especially *H. pylori* from Ladakhi Muslim, which showed a distinct ancestral relation to either East Asian or European populations, respectively (Wirth et al. 2004), and currently known as hpAsia2; the AE2 is currently known as hpNEAfrica. The result showing that the current hpEurope population arose by the recombination between AE1 and AE2, suggests that the introduction of *H. pylori* into Europe occurred more than once in history (Moodley et al. 2012). Hence, it opens a question of how and when was the second introduction of *H. pylori* in the Eurasian land, especially in Europe, which needs to be solved in future studies.

In terms of where the second introduction of *H. pylori* in Eurasian land is coming from, implementation of the *linkage model* (Falush et al. 2003a) of STRUCTURE could capture the distribution of AE1 and AE2 of European *H. pylori*.

Table 1 The population genetics of *H. pylori* and geographic localization

Population	Sub-population	Locations	Main references
hpAfrica2	hspNorthSan	Namibia, Angola	Moodley et al. (2012) and Falush et al. 2003b)
	hspSouthSan	South Africa	
hpAfrica1	hspWAfrica	Senegal, the Gambia, Burkina Faso, Morocco, Algeria, Nigeria, Cameroon, South Africa	Nell et al. (2013), Linz et al. (2014), and Falush et al. (2003b)
	hspSAfrica	Namibia, Angola, South Africa, Madagascar	
	hspCAfrica	Cameroon, Namibia	
hpAfrica1/ hpEurope	hspAfrica1Nicaragua	Central America	Thorell et al. (2017)
	hspAfrica1NAmerica	North, Central and South America	
	hspMiscAmerica	Central and South America	
	hspEuropeColombia	South America	
hpNEAfrica	hspCentralNEAfrica	Sudan, Cameroon, Nigeria, Algeria	Linz et al. (2007), and Nell et al. (2013)
	hspEastNEAfrica	Sudan, Ethiopia, Somalia, Algeria	
hpEurope	hspNorthEurope	Europe as far east as Southeast Asia	Thorell et al. (2017) and Falush et al. (2003b)
	hspSouthEurope		
hpAsia2	hspIndia	India, Bangladesh, Malaysia, Thailand, the Philipines, Nepal	Linz et al. (2007), Wirth et al. (2004), and Tay et al. (2009)
	hspLadakh	India (Himalaya)	
hpSahul	hspAustralia	Australia	Moodley et al. (2009)
	hspNGuinea	New Guinea	
hpEastAsia	hspEasia	China, India, Malaysia, Singapore, Taiwan, Thailand, Cambodia, Vietnam, Japan, Korea	Moodley et al. (2009), Falush et al. (2003b), and Linz et al. (2007)
	hspMaori	Taiwan, the Philippines, Japan, Samoa, New Caledonia, Wallis and Futuna, Indonesia, New Zealand	
	hspAmerind	Canada, the USA, Venezuela, Colombia, Peru	

The implementation of the *linkage model* showed a different ratio between AE1/AE2 with the higher proportion of AE1 in northern Europe and a lower proportion of AE1 in southern Europe (Falush et al. 2003b). The utilisation of whole genome sequence data using chromopainter/fineSTRUCTURE algorithm also captured those kinds of information in a finer scale manner. The first study implementing chromopainter/fineSTRUCTURE algorithm to the *H. pylori* genome showed that the European population was divided into two subgroups, Europe_sg1 (mostly strains from Northern Europe) and Europe_sg2 (mostly strains from Southern Europe) (Yahara et al. 2013). Another study implementing chromopainter/fineSTRUCTURE to *H. pylori* also showed that the co-ancestry level of hpNEAfrica strains was higher in *H. pylori* isolates from Southern Europe than the Northern Europe ones (Maixner et al. 2016). In addition, the new phylogeographic characterization method generated from two prophage genes of *H. pylori* (integrase and holin) showed different European populations, present mainly in Northern Europe and Southern Europe (Vale et al. 2015). These data suggest that the introduction of the second *H. pylori* wave was coming from south of Eurasian continent as a "second out of Africa" event.

The time-wise of the second *H. pylori* introduction in the Eurasian land is somewhat puzzling until today. One assumption time is about 30–40 kya when modern humans settled Europe via two entering routes: from Turkey along the Danube corridor into Eastern Europe, and along the Mediterranean coast with early Neolithic farmers (Correa and Piazuelo 2012). Another assumption is, when after 52 kya based on the split time between hpNEAfrica (as the descendant of AE1) and the hpAfrica1, which was predicted 35–52 kya (Moodley et al. 2012). However, the discovery and characterization of *H. pylori* inside the stomach of "The Iceman", a European copper age mummified human, revealed intriguing evidence of the second introduction time of *H. pylori* in Europe. The MLST characterisation of "The Iceman" was almost pure of AE1 with the minimum admixture of AE2 (6.5%, probability interval 1.5–13.5%) (Maixner et al. 2016). The genome analysis showed a high co-ancestry relation to the Indian hpAsia2 *H. pylori* and with most of the European hpEurope *H. pylori* strains. "The Iceman" is believed to be killed in the Italian Ötztal Alps mountains ~5300 years ago, which is located in the European land. This discovery suggests, if *H. pylori* from the Iceman is representative of the time, the low level of AE2 admixture indicated there was no introduction of AE2 prior to *H. pylori* from the Iceman. Therefore, the AE2 ancestry observed in hpEurope today is a result of AE2 introgression into Europe after the Copper Age or at least later in Central Europe (Ötzal Alps) (Maixner et al. 2016). However, the limitation of this conclusion was due to the data from only one strain.

From all *H. pylori* populations, which have evolved outside Africa, hpAsia2 is probably the most intriguing. This population evolved among the people who either did not follow a southern coastal migration route or who settled in the early phase of migration and later began expanding to the Western and Central Asia (Fig. 1). The distribution of hpAsia2 was widespread through Asians and Europeans prior to the evolution of current hpEurope strains. In addition, hpAsia2 might have accompanied the humans who settled in Europe ~40 kya (Moodley 2016). However, in the modern days the hpAsia2 population was mostly replaced by the hpEurope, especially in western Eurasia, but it still could be found in the South and South East Asian countries, including Nepal, Bangladesh, Malaysia, Myanmar, Thailand and India (Wirth et al. 2004; Aftab et al. 2017; Miftahussurur et al. 2015a; Breurec et al. 2011a; Subsomwong et al. 2017; Tay et al. 2009). The hpAsia2 *H. pylori* can be differentiated into two subpopulations, hspLadakh from the isolated Himalayan region of northern India and hspIndia, which is found among Indians, Malays and Thais (Breurec et al. 2011a; Tay et al. 2009). The presence of hpAsia2 in Finland and Estonia showed the evidence of the remaining of AE1, which initially inhabited the Eurasian land before the introduction of AE2 from the south.

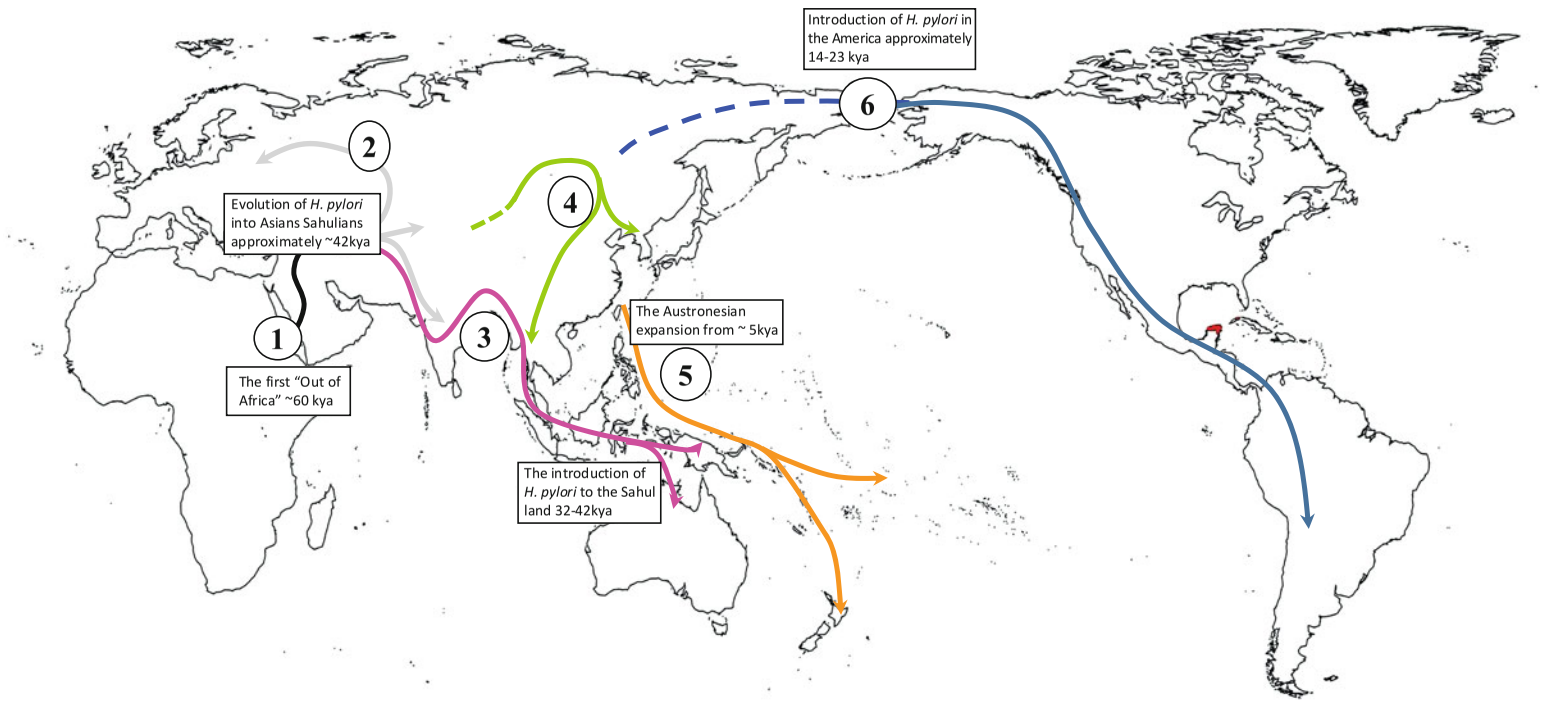

Fig. 1 The global human migration waves from africa to the pacific were constructed based on previous studies (Yamaoka et al. 2002; Falush et al. 2003b; Linz et al. 2007; Moodley et al. 2009). (*1*) The first major "Out of Africa" wave approximately happened ~60 kya; (*2*) The Asian *H. pylori*, especially hpAsia2 population, then occupying almost all western to a middle area of the Eurasian continent until the introduction of AE2, which would make the hybridisation with hpEurope; (*3*) A subsequent human population carries hpSahul and migrated towards southern coastal of Eurasian to the Sahul land; (*4*) The Chinese expansion during Zhou Dynasty in last 3000 years, which also expanding the hspEastAsia to East and South-East Asian countries; (*5*) The Austronesian expansion carrying hspMaori from Taiwan towards the island Southeast Asia and Polynesia; (*6*) The expansion sub-population of hpEastAsia, hspAmerind to America via Bering Strait

hpEastAsia is common among *H. pylori* isolates from East Asia and native American people. hpEastAsia is divided into three different subpopulations; hspMaori, hspAmerind, and hspEAsia and it was simulated to be split from hpAsia2 30–50 kya (Moodley et al. 2009). The hspEAsia is the biggest sub-population among the hpEastAsia population. This subpopulation was mainly observed in most East Asian countries, including China, South Korea, Japan, Thailand and Vietnam (Subsomwong et al. 2017; Breurec et al. 2011a; Linz et al. 2007; Falush et al. 2003a). The homogeneous distribution of the subpopulation of hspEAsia across those countries reflects the Chinese expansion language (family, Sino-Tibetan) during the last 3000 years, but mainly during the expansion of Zhou Dynasty (1100–211 BC) (Moodley 2016). Among three subpopulations of hpEastAsia, the hspAmerind was isolated among the diverse indigenous American population in North and South America, and it has its uniqueness characterised by very low genetic diversity compared to hpEastAsia, suggesting that the formation of Amerindian subpopulation was formed by a small group of people who migrated to America. This fact was different from the assumption of *H. pylori* in America was introduced by Chinese and Japanese people who migrated to America in more recent migration events. Instead, a subgroup was split from the big hpEastAsian population and went to America via crossing the Bering Strait approximately 12,000 years ago (Yamaoka et al. 2002). In addition, our previous data showed that four strains isolated from the Ainu ethnic group, living in Hokkaido, a northern island of Japan, belonged to the hspAmerind population (Gressmann et al. 2005), suggesting that the split between hpEastAsia and hspAmerind happened prior crossing the Bering Strait, possibly by human migrations in boats over the pacific ocean.

hspMaori was first observed in an isolate from Polynesians (Maoris, Tongans, and Samoans) of New Zealand, but was also observed in a small proportion of people in the Philippines and Japan (Linz et al. 2007; Falush et al. 2003b). In 2009, Moodley and co-workers discovered more hspMaori subpopulations among native Taiwanese, New Caledonia, and Torres Straits (an island located between Australia and New Guinea, which has been extensively visited by Polynesians), suggesting that the hspMaori is the marker for the entire Austronesian expansion rather than only for Polynesians (Moodley et al. 2009). Genetic analyses showed that the Taiwanese hspMaori have a significantly higher genetic diversity compared to the Pacific hspMaori ($\pi_{95} = 1.79–1.82\%$ vs 1.58–1.62%) and the indigenous Taiwanese isolates were isolated from the tribe that speak 5 of 10 subgroups of Austronesian family of languages, whereas the Pacific clades were isolated from individual speaking variants of Malayo-Polynesian, suggesting that the source of Austronesian expansion was in Taiwan. The split between indigenous Taiwanese hspMaori and Pacific hspMaori was predicted 4.9–5.0 kya (Moodley et al. 2009).

3.1 Observation of Asia2 and EastAsian Split from CagA Perspective

The split between hpAsia2 and hpEastAsia is also captured by the *cagA* gene, which encodes a highly immunogenic protein (CagA), located at the 3′ prime end of the *cag* pathogenicity island (PAI). The *cag*PAI encodes a type IV secretion system, through which CagA is delivered into host cells (Backert et al. 2015, Naumann et al. 2017). After delivery into gastric epithelial cells, CagA becomes tyrosine-phosphorylated at Glu-Pro-Ile-Tyr-Ala sequence motif (EPIYA) located in the 3′ region of the *cagA* gene (Backert and Selbach 2008; Zhang et al. 2015). Supporting the idea that *H. pylori* mirrors the human evolution, it was reported that the structure of the 3′ region of the *cagA* varies between strains from East Asian and Western countries (Yamaoka et al. 1998; Yamaoka et al. 1999; Yamaoka et al. 2000; Yamaoka et al. 2002). Isolates from the Western countries (Europe and Americas) mostly possess segments of the so-called EPIYA-ABC type,

which are well known as Western-type CagAs, whereas strains from East Asian countries possess segments of the EPIYA-ABD type, which were classified as East Asian-type CagAs. Recent studies in the borderline of Europe and Middle East countries (i.e. Turkey and Iran), reporting that the distribution Western-type CagA was predominant (Kocazeybek et al. 2015; Honarmand-Jahromy et al. 2015). The same pattern was also reported in Nepal (Miftahussurur et al. 2015a). Interestingly, in Bhutan, the country that shares a border with India, almost all strains possessing East Asian-type CagA with more than half having multiple repeats in the 3′ region, which is very rare in other countries (Matsunari et al. 2016). When moving to South-East Asian countries, the predominant CagA genotype was shifted from Western-type to East Asian-type. For example, in Vietnam and Malaysia, East Asian-type CagA is predominant by 96% and 56%, respectively (Uchida et al. 2009; Schmidt et al. 2009). However, Cambodia and Thailand showed a little different pattern with the neighbour countries, with the Western-type CagA being predominant (59% and 54%, respectively) (Breurec et al. 2011b; Chomvarin et al. 2012). The split between Western-type CagA and East Asian-type CagA was a result of sequence rearrangement within the *cagA* gene, including the CagA multimerisation sequence (CM) and the EPIYA-motifs. This rearrangement in the left half of the EPIYA-D segment, characteristic of East Asian CagA, was derived from Western-type EPIYA with Amerind-type EPIYA as intermediate, through recombination of specific sequences within the gene (Furuta et al. 2011; Correa and Piazuelo 2012).

3.2 Peopling of Australians and New Guineans: hpSahul

A subsequent split occurred from the hpAsia2 population, which migrated towards a South-East tip of the Sundaland (i.e., the Malay Peninsula, Sumatra, Java, Borneo, and Bali), and went to a continent called Sahul (i.e. Australia, New Guinea, and Tasmania). Observations on the human parental markers showed that the indigenous Australians are closely related to New Guineans (Hudjashov et al. 2007). However, the observations in *H. pylori* showed that a single population inhabits the indigenous people Australian and New Guinean as hpSahul. Furthermore, detailed analysis showed that hpSahul was divided into two subpopulations, hspAustralia and hspNGuinea (Moodley et al. 2009).

The split between the Asian population and the Sahulian population was somewhat mystifying. The evidence based on Pleistocene human archaeological sites showed that the human colonisation in the Pleistocene Sahul was approximately between 42–48 kya (Pope and Terrell 2008; Allen and O'Connell 2014; Gillespie 2002). This dating was a somewhat before the split dating observed by Moodley and co-workers, which was ~31–37 kya (Moodley et al. 2009). Interestingly, the observation of single phylogeny of hpSahul suggested that the introduction of *H. pylori* into Sahul occurred once, and was followed by the split between hspNGuenia and hspAustralia, because there was no sign of gene flow between those two split sub-populations (Moodley 2016).

Distinct genetic drift was also observed in the *H. pylori* strains isolated from individuals in the Highlander New Guineans, which was captured by the CagA characterisation. The CagA characterisation of highlander New Guinean strain PNG84A (Montano et al. 2015) showed the presence of the unique AB-type EPIYA CagA. Our data also showed the presence of the unique ABB-type CagA, predominantly in strains isolated from Papuan ethnic people (Miftahussurur et al. 2015c). In fact, this unique ABB-type CagA has a very similar B-segment compared to that observed in strain PNG84A. These data suggest a different genetic drift from the one demonstrated by Furuta and co-workers (Furuta et al. 2011)

4 The Migration Inside Africa and the Intimation with the Host

4.1 The Origin of *H. pylori* and Split of hpAfrica2

H. pylori is believed to accompany the humans who migrated out of Africa ~58 kya. This intriguing fact leads us to the assumption that, if the *H. pylori* and modern human migrated together out of Africa, then the initial colonisation of *H. pylori* in the human stomach has occurred far prior to these migration events. Moodley and co-workers demonstrated the possible initial colonisation of *H. pylori* in the human stomach using a rooted, fully resolved and calibrated global clonal phylogeny (Moodley et al. 2012). This approach resulted from the coalescent to single common ancestor, which has occurred approximately 100 kya (range: 88–116 kya, Fig. 2). However, lineage sorting and population bottleneck would wipe out the possible extended older lineage in this time frame. Therefore, it is possible that the association between human and *H. pylori* is older than this estimation (Moodley 2016).

Another intriguing fact about the origin of *H. pylori* is the very distinct relationship between hpAfrica2 and other genetic *H. pylori* populations. The hpAfrica2 population exhibits a great diversity compared to other populations and is exclusively possessed by individuals in the southern part of Africa, including South Africa, Namibia, and southern Angola. The distinct relationship between hpAfrica2 and other *H. pylori* was attributed to the different migration waves once the *H. pylori* successfully colonised modern humans. In addition, the exclusive possession on the specified location in Africa should have an

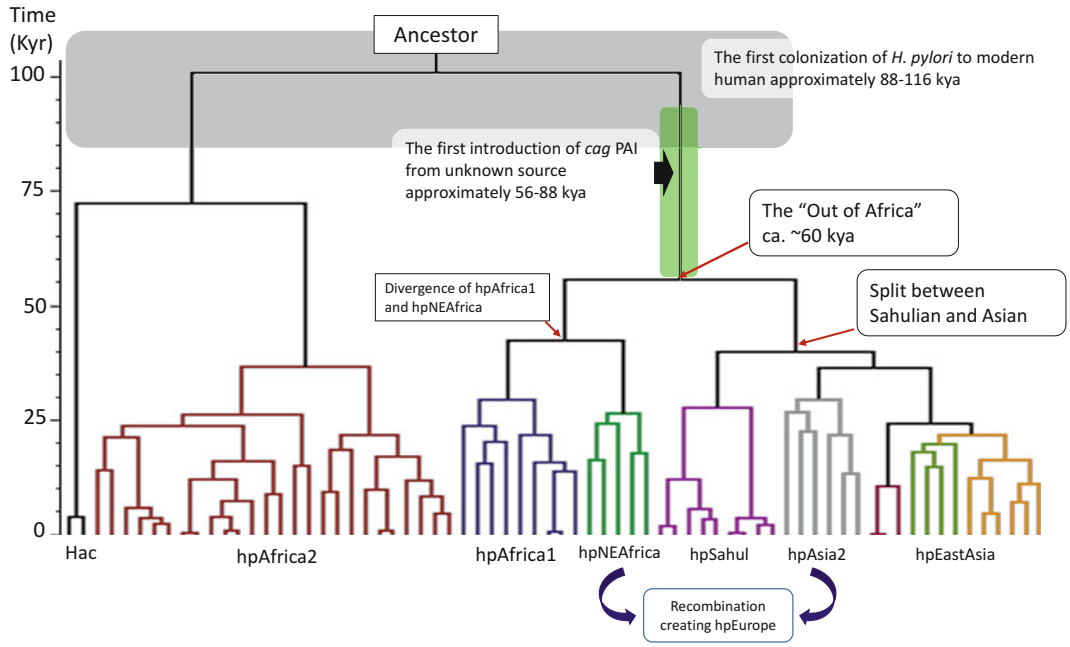

Fig. 2 Global phylogeny of *H. pylori* simulating the time split between populations and the estimated time in which the *cag*PAI was transferred to the *H. pylori* genome constructed based on previous papers (Moodley et al. 2009; Olbermann et al. 2010; Moodley et al. 2012). The housekeeping gene sequence diversity is structured into two super-lineages which coalescent approximately ~88–116 kya. The absence of the *cag*PAI in the hpAfrica2 population was attributed by the introduction of the *cag*PAI after a split into other *H. pylori,* but prior the first "out of Africa" wave approximately ~58–88 kya

association with the San Hunter-Gatherers. Indeed, hpAfrica2 was found in individuals from non-San people in Southern Africa (Falush et al. 2003b); however, there was a distinct relation inside the hpAfrica2 between Northern San and Southern San-Bantu speakers (Fig. 3), suggesting that the original host of hpAfrica2 was the Northern San Hunter-Gatherers and was then split into the Southern San Hunter-Gatherers approximately 32–47 kya (Moodley et al. 2012).

4.2 The hpNEAfrica and the Acquisition of *cag*PAI

Another independent migration wave from the human ancestor to the north, aside from the wave out of Africa, resulted in another genetic *H. pylori* population, called hpNEAfrica. This population was spread along the central Sahel and North Africa and shared its distribution with the population hpAfrica1 (Linz et al. 2007). Its frequency was

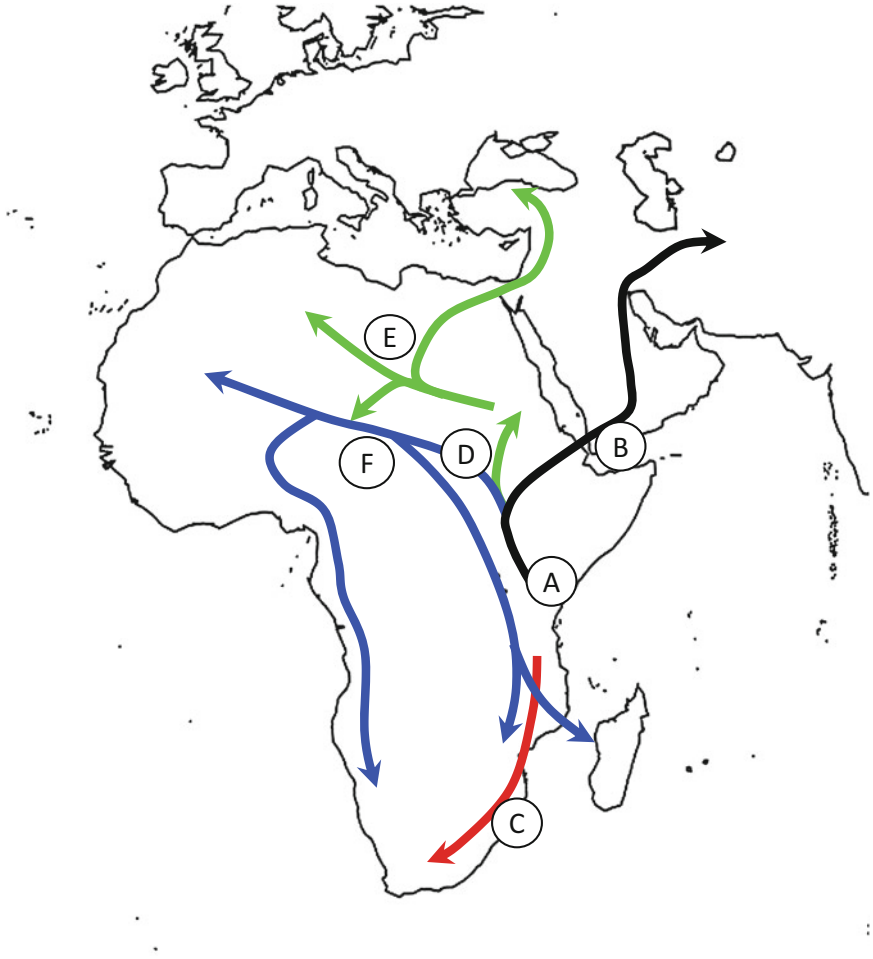

Fig. 3 The reconstruction of migration events and initial coalescent of *H. pylori* in Africa based on previous findings (Nell et al. 2013; Moodley et al. 2012; Linz et al. 2014; Maixner et al. 2016). (*A*) The first coalescent of *H. pylori* to human from the unknown ancestor; (*B*) The first out of Africa; (*C*) The other wave of the ancestor to the south, which evolved to the hpAfrica2 population; (*D*) *H. pylori* divergence of hpAfrica1 and hpNEAfrica.; (*E*) The expansion of Nilo-Saharan speakers to the western part of Africa (~6–9 kya). The subsequent population went to Eurasian via the Levant introducing the AE2 resulting hpEurope; (*F*) The expansion of Bantu speakers, including in the western and eastern coast of Africa, which introduced hpAfrica1 to Madagascar via Mozambique channel

increasing eastwards to the Nile and Horn of Africa, which is the home of the Nilo-Saharan speaking pastoralist society. Thus, the hpNEAfrica conveniently became the marker of Nilo-Saharan language in the shape of hpNEAfrica migration wave (Moodley 2016). The hpNEAfrica population was split into two subpopulations, hspCentralNEA and hspEastNEA and the split was demonstrated roughly along the Nile valley and Sudan which straddles the Nile containing both subpopulations at high frequency. The presence of hpsCentralNEA throughout Cameroon, Angola, and Nigeria suggested the migration wave of Nilo-Saharan speakers in the Holocene humid period (~6–9 kya) and carried their own language westwards from its home in northeast Africa into the waterlogged Sahara and beyond (Nell et al. 2013). As the introduction of second *H. pylori* into Europe occurred about the same time frame compared to the hpNEAfrica expansion during Holocene periods (~6–9 kya), there is a possibility that a subsequent population of hpNEAfrica migrated to Eurasian land via the Levant and introduced the AE2, resulting from the recombination population hpEurope (Fig. 3).

Another interesting observation of this superlineage distinct between hpAfrica2 and other *H. pylori* populations is the acquisition of the *cag*PAI. The *cag*PAI was hypothesised to be acquired from the unknown source prior to the out of Africa period and alongside with the host migrated up to Pacific and crossing Bering Strait; therefore it exhibits a specific genetic diversity pattern as determined by MLST (Olbermann et al. 2010). However, the *cag*PAI has not been observed in the hpAfrica2 population. The lack of the *cag*PAI in hpAfrica2 might be due to a different wave of human migration northwards, which acquired the *cag*PAI approximately ~88–58 kya just before the out of Africa migration events (Fig. 2) (Linz et al. 2007; Moodley et al. 2012).

4.3 The hpAfrica1: The Marker of Bantu Speakers

After settling in the Nile and at the horn of Africa, subsequent human migration westwards generated another *H. pylori* population, called hpAfrica1. The distribution of hpAfrica1 was spread alongside Morocco and Algeria in the north up to South Africa in the South (Fig. 3). Based on the geographical distribution of the hpAfrica1, cluster analysis divided the hpAfrica1 into three closely related subpopulations in the west and north of Africa, called hspWAfrica (Falush et al. 2003b); central Africa, hspCAfrica (Nell et al. 2013) and southern Africa (hspSAfrica) (Falush et al. 2003b) (Table 1). This distribution alongside those countries could be associated with the expansion of Bantu speakers people within last five kya from their original homeland in Nigeria/Cameroon. The observation of the subpopulation hspCAfrica in Cameroon and Angola, but not in South Africa and Namibia, supports the probability of a migration route alongside the west coast of Africa (Fig. 2). The hspSAfrica is presumed to be evolved during Bantu speakers expansion along the east coast that brought the Nguni speakers to southern Africa. Therefore, the observation of hpSAfrica in Madagascar, which is very closely related to hpSAfrica from South Africa, suggests the migration of Bantu speakers across the Mozambique Channel during or after the migration alongside the east coast of Africa (Linz et al. 2014).

5 The Recombination During Post Colonial Expansion

Population genetics of *H. pylori* not only capture the prehistoric migrations, but also capturing recent migration events. This idea was introduced by the observation of hpAfrica1, which could be found in America. The hspWAfrica was observed in several *H. pylori* isolates from several places in America, especially in African Americans in Louisiana and Tennessee (Falush et al. 2003b). This observation suggests the recent introduction of hpAfrica1 *H. pylori* in Americas by the transatlantic slave trades in the 16th–19th centuries. In addition, the hpEurope not only colonized the Eurasian area, but was also found in Australia, Africa, and America. The fact that several European Kingdoms (e.g. Portuguese, Spanish and Great Britain) explored the world since the

fifteenth century could be responsible for the export of hpEurope into several other places of the world (Falush et al. 2003b).

The high genetic recombination between more than two populations was also observed in *H. pylori* isolates from Portuguese speaking countries, including Portugal, Angola, Brazil and Cape Verde, as captured by MLST analysis (Oleastro et al. 2017). The hpEurope in Europe and Portuguese speaking countries revealed a distinct relation to each other. This distinct relation was attributed by the ancestral Africa 1 components. The observation of Ancestral Africa 1 in *H. pylori* strains from Portuguese speaking countries suggested a long history between those countries, which resulted in new recombination of *H. pylori* consisting of more than three ancestral populations. Also, the isolates from Brazil were mostly hpEurope. However, several isolates from Brazil, which were assigned as hpEurope, had a small components of EastAsian ancestral, from which hspAmerind is a subpopulation (Oleastro et al. 2017). This data suggests the replacement of hspAmerind from the native Brazilian people to the other *H. pylori* populations due to low genetic diversity compared to either hpEurope or hpAfrica1 (Dominguez-Bello et al. 2008).

Population separation was observed in *H. pylori* strains from Latin America countries, including Mexico, Nicaragua, and Colombia using MLST, Virtual Genome Fingerprint (VGF) and the *cag*PAI phylogenetic tree studies. These analyses resulted in the discovery of a new subpopulation, which is very different from the native *H. pylori* American population hspAmerind. The separation from the indigenous population could be attributed to *H. pylori*, which were isolated from the modern population of those countries instead of indigenous population. This would reflect the recent genetic admixture of pre-Columbian American groups with the European colonisers and African slaves last ~500 years due to the European expansion (Munoz-Ramirez et al. 2017). In addition, implementation of chromopainter/fineSTRUCTURE to the *H. pylori* genome isolated from Latin America discovered several new subpopulations, introduced as hspAfrica1NAmerica, hspEuropeColombia, hspAfrica1Nicaragua and hspMiscAmericas (Thorell et al. 2017). The formation of new subpopulations was attributed in the combination of recombination and drift events. The genetic drift was taken place as a result of a recent demographic bottleneck, which assumed to have a low sequence divergence. However, the pairwise sequence divergence of that new sub-population was as high as the recombinant population hpEurope. This high sequence divergence was maintained by the admixture between hpEurope and hpAfrica1 population. This high sequence divergence (which mostly attributed by the high frequency of African component) suggested that the bacteria of African origin have been particularly effective in colonizing the new continent (Thorell et al. 2017). These data suggest that the expansion of modern humans in the new environment lead to the rapid evolution of *H. pylori,* which was faster and more dynamic than the host.

6 Conclusions

In the recent years, there are remarkable findings of the human migration based on the fossil and human genetic studies, which could divide the migration waves into two major migrations, pre-60 kya and post-60 kya (reviewed by (Bae et al. 2017)). In case of *H. pylori*, the story mostly yielded to the post-60 kya "Out of Africa" era, which is regarded as the major migration event. Together, *H. pylori* has been regarded as a handy tool to trace human migrations, from the pre-historic migrations to the recent migrations. Moreover, the *H. pylori* genome with the chromopainting analysis gives us more detail in the high-resolution data of the recent migrations. The rapid and dynamic evolution compared to its host still become major features for *H. pylori,* which inform us comprehensively about the recent recombination events.

With these current data, however, it becomes obvious that several more places are needed to be discovered, such as Siberia, Mongolia, Northern part of Japan and Indonesia since those places are located at the "bridge" of the split of several populations. Interestingly, the ability of

H. pylori to colonize more than 50% of the human population became a "common sense", however, it seems not to fit to the tip of Sundaland countries such as Indonesia and Malaysia, which shows a very low prevalence of *H. pylori* (Syam et al. 2015; Miftahussurur et al. 2015b; Rahim et al. 2010). This fact leads us to the speculation that the people at the tip of Sundaland might not carry any *H. pylori* and perhaps the introduction of *H. pylori* to the Indonesian and Malay people was performed by the Chinese expansion (hspEAsia) in more recent migrations. In addition, since the "African expansion" was successfully demonstrated in America, it is also interesting to find out the "African expansion" in the different continent such as Europe and Asia.

Acknowledgements This work was supported in part by grants from the National Institutes of Health (DK62813) and the Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (221S0002, 16H06279, 15H02657 and 16H05191) (YY). It was also supported by the Japan Society for the Promotion of Science (JSPS) Institutional Program for Core-to-Core Program; B. Africa-Asia Science Platform (YY). LAW is a doctoral students supported by the Japanese Government (The Ministry of Education, Culture, Sports, Science and Technology [MEXT]) Scholarship Program for 2015.

References

Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ (1999) Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol Microbiol 32(3):459–470

Aftab H, Miftahussurur M, Subsomwong P, Ahmed F, Khan AKA, Matsumoto T, Suzuki R, Yamaoka Y (2017) Two populations of less-virulent *Helicobacter pylori* genotypes in Bangladesh. PLoS One 12(8): e0182947. https://doi.org/10.1371/journal.pone.0182947

Allen J, O'Connell J (2014) Both half right: updating the evidence for dating first human arrivals in Sahul. Aust Archaeol 79(1):86–108. https://doi.org/10.1080/03122417.2014.11682025

Backert S, Selbach M (2008) Role of type IV secretion in *Helicobacter pylori* pathogenesis. Cell Microbiol 10 (8):1573–1581. https://doi.org/10.1111/j.1462-5822.2008.01156.x

Backert S, Tegtmeyer N, Selbach M (2010) The versatility of *Helicobacter pylori* CagA effector protein functions: the master key hypothesis. Helicobacter 15 (3):163–176. https://doi.org/10.1111/j.1523-5378.2010.00759.x

Backert S, Tegtmeyer N, Fischer W (2015) Composition, structure and function of the *Helicobacter pylori* cag pathogenicity island encoded type IV secretion system. Future Microbiol 10(6):955–965. https://doi.org/10.2217/fmb.15.32

Bae CJ, Douka K, Petraglia MD (2017) On the origin of modern humans: Asian perspectives. Science 358 (6368):eaai9067. https://doi.org/10.1126/science.aai9067

Björkholm B, Sjölund M, Falk PG, Berg OG, Engstrand L, Andersson DI (2001) Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. Proc Natl Acad Sci 98(25):14607–14612

Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Oung C, Raymond J, Tan TS, Thiberge JM, Vong S, Monchy D, Linz B (2011a) Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. PLoS One 6(7): e22058. https://doi.org/10.1371/journal.pone.0022058

Breurec S, Guillard B, Hem S, Papadakos KS, Brisse S, Huerre M, Monchy D, Oung C, Sgouras DN, Tan TS, Thiberge JM, Vong S, Raymond J, Linz B (2011b) Expansion of European vacA and cagA alleles to East-Asian *Helicobacter pylori* strains in Cambodia. Infect Genet Evol 11(8):1899–1905. https://doi.org/10.1016/j.meegid.2011.08.007

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Chomvarin C, Phusri K, Sawadpanich K, Mairiang P, Namwat W, Wongkham C, Hahnvajanawong C (2012) Prevalence of cagA EPIYA motifs in *Helicobacter pylori* among dyspeptic patients in northeast Thailand. Southeast Asian J Trop Med Public Health 43(1):105–115

Correa P, Piazuelo MB (2012) Evolutionary history of the *Helicobacter pylori* genome: implications for gastric carcinogenesis. Gut Liver 6(1):21–28. https://doi.org/10.5009/gnl.2012.6.1.21

Dominguez-Bello MG, Perez ME, Bortolini MC, Salzano FM, Pericchi LR, Zambrano-Guzman O, Linz B (2008) Amerindian *Helicobacter pylori* strains go extinct, as european strains expand their host range. PLoS One 3(10):e3307. https://doi.org/10.1371/journal.pone.0003307

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A 98(26):15056–15061. https://doi.org/10.1073/pnas.251396098

Falush D, Stephens M, Pritchard JK (2003a) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567–1587

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Megraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S (2003b) Traces of human migrations in *Helicobacter pylori* populations. Science 299 (5612):1582–1585. https://doi.org/10.1126/science.1080857

Fischer W, Windhager L, Rohrer S, Zeiller M, Karnholz A, Hoffmann R, Zimmer R, Haas R (2010) Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. Nucleic Acids Res 38 (18):6089–6101. https://doi.org/10.1093/nar/gkq378

Furuta Y, Yahara K, Hatakeyama M, Kobayashi I (2011) Evolution of cagA oncogene of *Helicobacter pylori* through recombination. PLoS One 6(8):e23499. https://doi.org/10.1371/journal.pone.0023499

Gillespie R (2002) Dating the first Australians. Radiocarbon 44(2):455–472

Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M (2005) Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. PLoS Genet 1(4):e43. https://doi.org/10.1371/journal.pgen.0010043

Honarmand-Jahromy S, Siavoshi F, Malekzadeh R, Sattari TN, Latifi-Navid S (2015) Multiple repeats of *Helicobacter pylori* CagA EPIYA-C phosphorylation sites predict risk of gastric ulcer in Iran. Microb Pathog 89:87–92. https://doi.org/10.1016/j.micpath.2015.09.005

Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, Ng SC (2017) Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterology 153(2):420–429. https://doi.org/10.1053/j.gastro.2017.04.022

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc Natl Acad Sci U S A 104 (21):8726–8730. https://doi.org/10.1073/pnas.0702928104

Kivi M, Tindberg Y, Sorberg M, Casswall TH, Befrits R, Hellstrom PM, Bengtsson C, Engstrand L, Granstrom M (2003) Concordance of *Helicobacter pylori* strains within families. J Clin Microbiol 41(12):5604–5608

Kocazeybek BS, Caliskan R, Erdamar Cetin S, Ergin S, Kuskucu M, Kepil N, Oyku Dinc H, Ziya Erzin Y, Saribas S, Bahar Tokman H, Kalayci F, Akgul O, Yuksel P, Karakullukcu A, Ziver T, Sirekbasan S, Caglar E, Bal K (2015) Patterns of EPIYA motifs among cagA-positive *Helicobacter pylori* strains: a case-control study in a Turkish population with Eurasian geographical features. J Med Microbiol 64 (10):1117–1123. https://doi.org/10.1099/jmm.0.000141

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet 8(1):e1002453. https://doi.org/10.1371/journal.pgen.1002453

Li WH, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129(2):513–523

Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445 (7130):915–918. https://doi.org/10.1038/nature05562

Linz B, Vololonantenainab CR, Seck A, Carod JF, Dia D, Garin B, Ramanampamonjy RM, Thiberge JM, Raymond J, Breurec S (2014) Population genetic structure and isolation by distance of *Helicobacter pylori* in Senegal and Madagascar. PLoS One 9(1):e87355. https://doi.org/10.1371/journal.pone.0087355

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95 (6):3140–3145

Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, Kusebauch U, Vigl EE, Malfertheiner P, Megraud F, O'Sullivan N, Cipollini G, Coia V, Samadelli M, Engstrand L, Linz B, Moritz RL, Grimm R, Krause J, Nebel A, Moodley Y, Rattei T, Zink A (2016) The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science 351(6269):162–165. https://doi.org/10.1126/science.aad2545

Matsunari O, Miftahussurur M, Shiota S, Suzuki R, Vilaichone RK, Uchida T, Ratanachu-ek T, Tshering L, Mahachai V, Yamaoka Y (2016) Rare *Helicobacter pylori* virulence genotypes in Bhutan. Sci Rep 6:22584. https://doi.org/10.1038/srep22584

Miftahussurur M, Sharma RP, Shrestha PK, Suzuki R, Uchida T, Yamaoka Y (2015a) Molecular epidemiology of *Helicobacter pylori* infection in Nepal: specific ancestor root. PLoS One 10(7):e0134216. https://doi.org/10.1371/journal.pone.0134216

Miftahussurur M, Shiota S, Suzuki R, Matsuda M, Uchida T, Kido Y, Kawamoto F, Maimunah U, Adi P, Rezkitha Y, Nasronudin NI, Yamaoka Y (2015b) Identification of *Helicobacter pylori* infection in symptomatic patients in Surabaya, Indonesia, using five diagnostic tests. Epidemiol Infect 143(5):986–996. https://doi.org/10.1017/S095026881400154X

Miftahussurur M, Syam AF, Makmun D, Nusi IA, Zein LH, Zulkhairi AF, Uswan WB, Simanjuntak D, Uchida T, Adi P, Utari AP, Rezkitha YA, Subsomwong P, Nasronudin YY (2015c) *Helicobacter pylori* virulence genes in the five largest islands of Indonesia. Gut Pathogens 7:26. https://doi.org/10.1186/s13099-015-0072-2

Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, Moodley Y, Jensen JD (2015) Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. Genetics 200 (3):947–963. https://doi.org/10.1534/genetics.115.176404

Moodley Y (2016) *Helicobacter pylori*: genetics, recombination, population structure, and human migrations. In: Backert S, Yamaoka Y (eds) *Helicobacter pylori* research: from bench to bedside. Springer, Tokyo, pp 3–27. https://doi.org/10.1007/978-4-431-55936-8_1

Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhoft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M (2009) The peopling of the Pacific from a bacterial perspective. Science 323(5913):527–530. https://doi.org/10.1126/science.1166083

Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, Bernhoft S, Hale J, Suerbaum S, Mugisha L, van der Merwe SW, Achtman M (2012) Age of the association between *Helicobacter pylori* and man. PLoS Pathog 8(5):e1002693. https://doi.org/10.1371/journal.ppat.1002693

Munoz-Ramirez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Thorell K, Torres R, Aviles-Jimenez F, Camorlinga M, Canzian F, Torres J (2017) Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from Latin America have followed a unique evolution pathway. Front Cell Infect Microbiol 7:50. https://doi.org/10.3389/fcimb.2017.00050

Naumann M, Sokolova O, Tegtmeyer N, Backert S (2017) *Helicobacter pylori*: a paradigm pathogen for subverting host cell signal transmission. Trends Microbiol 25(4):316–328. https://doi.org/10.1016/j.tim.2016.12.004

Nell S, Eibach D, Montano V, Maady A, Nkwescheu A, Siri J, Elamin WF, Falush D, Linz B, Achtman M, Moodley Y, Suerbaum S (2013) Recent acquisition of *Helicobacter pylori* by Baka pygmies. PLoS Genet 9 (9):e1003775. https://doi.org/10.1371/journal.pgen.1003775

Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, Suerbaum S, Achtman M, Linz B (2010) A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. PLoS Genet 6(8):e1001069. https://doi.org/10.1371/journal.pgen.1001069

Oleastro M, Rocha R, Vale FF (2017) Population genetic structure of *Helicobacter pylori* strains from Portuguese-speaking countries. Helicobacter 22(4). https://doi.org/10.1111/hel.12382

Pope KO, Terrell JE (2008) Environmental setting of human migrations in the circum-Pacific region. J Biogeogr 35(1):1–21

Posselt G, Backert S, Wessler S (2013) The functional interplay of *Helicobacter pylori* factors with gastric epithelial cells induces a multi-step process in pathogenesis. Cell Commun Signal 11:77. https://doi.org/10.1186/1478-811X-11-77

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Rahim AA, Lee YY, Majid NA, Choo KE, Raj SM, Derakhshan MH, Graham DY (2010) *Helicobacter pylori* infection among Aborigines (the Orang Asli) in the northeastern region of Peninsular Malaysia. Am J Trop Med Hyg 83(5):1119–1122. https://doi.org/10.4269/ajtmh.2010.10-0226

Schmidt HM, Goh KL, Fock KM, Hilmi I, Dhamodaran S, Forman D, Mitchell H (2009) Distinct cagA EPIYA motifs are associated with ethnic diversity in Malaysia and Singapore. Helicobacter 14(4):256–263. https://doi.org/10.1111/j.1523-5378.2009.00684.x

Subsomwong P, Miftahussurur M, Vilaichone RK, Ratanachu-Ek T, Suzuki R, Akada J, Uchida T, Mahachai V, Yamaoka Y (2017) *Helicobacter pylori* virulence genes of minor ethnic groups in North Thailand. Gut Pathogens 9:56. https://doi.org/10.1186/s13099-017-0205-x

Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M (1998) Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci 95(21):12619–12624

Syam AF, Miftahussurur M, Makmun D, Nusi IA, Zain LH, Zulkhairi AF, Uswan WB, Simanjuntak D, Uchida T, Adi P, Utari AP, Rezkitha YA, Subsomwong P, Nasronudin SR, Yamaoka Y (2015) Risk factors and prevalence of *Helicobacter pylori* in five largest Islands of Indonesia: a preliminary study. PLoS One 10(11):e0140186. https://doi.org/10.1371/journal.pone.0140186

Tay CY, Mitchell H, Dong Q, Goh KL, Dawes IW, Lan R (2009) Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. BMC Microbiol 9:126. https://doi.org/10.1186/1471-2180-9-126

Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, Mendez A, Rizzato C, Bravo MM, Suzuki R, Yamaoka Y, Torres J, Sheppard SK, Falush D (2017) Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. PLoS Genet 13(2): e1006546. https://doi.org/10.1371/journal.pgen.1006546

Uchida T, Nguyen LT, Takayama A, Okimoto T, Kodama M, Murakami K, Matsuhisa T, Trinh TD, Ta L, Ho DQ, Hoang HH, Kishida T, Fujioka T, Moriyama M, Yamaoka Y (2009) Analysis of virulence factors of *Helicobacter pylori* isolated from a Vietnamese population. BMC Microbiol 9:175. https://doi.org/10.1186/1471-2180-9-175

Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, Schlemper RJ (2001) *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med 345 (11):784–789. https://doi.org/10.1056/NEJMoa001999

Vale FF, Vadivelu J, Oleastro M, Breurec S, Engstrand L, Perets TT, Megraud F, Lehours P (2015) Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. Sci Rep 5:14333. https://doi.org/10.1038/srep14333

Vale FF, Nunes A, Oleastro M, Gomes JP, Sampaio DA, Rocha R, Vitor JM, Engstrand L, Pascoe B, Berthenet E, Sheppard SK, Hitchings MD, Megraud F, Vadivelu J, Lehours P (2017) Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. Sci Rep 7:42471. https://doi.org/10.1038/srep42471

Wirth T, Wang X, Linz B, Novick RP, Lum JK, Blaser M, Morelli G, Falush D, Achtman M (2004) Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. Proc Natl Acad Sci U S A 101 (14):4746–4751. https://doi.org/10.1073/pnas.0306629101

Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I (2013) Chromosome painting in silico in a bacterial species reveals fine population structure. Mol Biol Evol 30 (6):1454–1464. https://doi.org/10.1093/molbev/mst055

Yamaoka Y (2010) Mechanisms of disease: *Helicobacter pylori* virulence factors. Nat Rev Gastroenterol Hepatol 7(11):629–641. https://doi.org/10.1038/nrgastro.2010.154

Yamaoka Y, Kodama T, Kashima K, Graham DY, Sepulveda AR (1998) Variants of the 3′ region of the cagA gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. J Clin Microbiol 36(8):2258–2263

Yamaoka Y, El-Zimaity HM, Gutierrez O, Figura N, Kim JG, Kodama T, Kashima K, Graham DY (1999) Relationship between the cagA 3′ repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. Gastroenterology 117(2):342–349

Yamaoka Y, Osato MS, Sepulveda AR, Gutierrez O, Figura N, Kim JG, Kodama T, Kashima K, Graham DY (2000) Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries. Epidemiol Infect 124(1):91–96

Yamaoka Y, Orito E, Mizokami M, Gutierrez O, Saitou N, Kodama T, Osato MS, Kim JG, Ramirez FC, Mahachai V, Graham DY (2002) *Helicobacter pylori* in North and South America before Columbus. FEBS Lett 517(1–3):180–184

Zhang XS, Tegtmeyer N, Traube L, Jindal S, Perez-Perez-G, Sticht H, Backert S, Blaser MJ (2015) A specific A/T polymorphism in Western tyrosine phosphorylation B-motifs regulates *Helicobacter pylori* CagA epithelial cell interaction. PLoS Pathog 11:e1004621. https://doi.org/10.1371/journal.ppat.1004621