Genomic Adaptation to Acidic Environment: Evidence from *Helicobacter pylori*

Xuhua Xia* **and Gareth Palidwor†**

Department of Biology, University of Ottawa, 150 Louis Pasteur, P.O. Box 450, Station A, Ottawa, Ontario K1N 6N5, Canada

Submitted February 2, 2005; Accepted July 18, 2005; Electronically published October 4, 2005

Online enhancement: table.

ABSTRACT: The origin of new functions is fundamental in understanding evolution, and three processes known as adaptation, preadaptation, and exaptation have been proposed as possible evolutionary pathways leading to the origin of new functions. Here we examine the origin of an acid resistance mechanism in the mammalian gastric pathogen *Helicobacter pylori*, with reference to these three evolutionary pathways. The mechanism involved is that *H. pylori*, when exposed to the acidic environment in mammalian stomach, restricts the acute proton entry across its membrane by its increased usage of positively charged amino acids in the inner and outer membrane proteins. The results of our comparative genomic analysis between *H. pylori*, the two closely related species *Helicobacter hepaticus* and *Campylobacter jejuni*, and other relevant proteobacterial species are incompatible with the hypotheses invoking preadaptation or exaptation. The acid resistance mechanism most likely arose by selection favoring an increased usage of positively charged lysine in membrane proteins.

Keywords: acid resistance, adaptation, preadaptation, exaptation, *Helicobacter*, membrane proteins.

It is of fundamental importance for evolutionary biologists to demonstrate not only how fitness-enhancing traits, also called aptations (Gould and Vrba 1982), have been maintained by natural selection, but also how these aptations have been acquired during the course of evolution. As asserted by Darwin himself, evolutionary theory would be weak unless we understand how the innumerable species have acquired those aptations (Darwin 1859, p. 3).

Aptations can originate through three processes: adaptation in response to a selection regime (Williams 1966, p. 6); preadaptation that involves a trait originally selected for one function but that subsequently gained a different function beneficial to the carrier of the trait (e.g., rudimentary feathers that presumably have been selected for thermoregulation in nonavian dinosaurs preadapted their carriers to subsequent evolution of flight; Prum and Brush 2002, 2003); and exaptation, which we consider to refer to an originally neutral trait that has subsequently acquired a beneficial function, for example, the brain-specific RNA gene BC200 resulting from the exaptation of a presumably neutral SINE repeat (Smit 1999). Although Gould (2002, p. 1223) had condemned the term preadaptation as "the most unfortunate name" in favor of lumping the last two processes together as exaptation (Gould and Vrba 1982), we have kept the trichotomy because the three processes do have differences. Thus, our usage of exaptation in this article is different from its original.

While the origin of new aptations is essential in understanding how organisms invade and colonize new ecological niches resulting in new components of biodiversity (Kolbe et al. 2004), few evolutionary studies consider all three processes as alternative hypotheses, partly because the three processes are often all lumped together as adaptation (e.g., Ridley 2004, pp. 6, 287). Here we present a detailed analysis of an aptation in *Helicobacter pylori* that helps the bacterium to successfully colonize the acidic environment of the mammalian stomach with reference to the three evolutionary processes.

Helicobacter pylori is one of the terminal lineages in the highly invasive *Helicobacter* complex. It thrives in the acidic environment of mammalian stomach, causing gastric and duodenal ulcers and gastric cancer in human (Correa 1997; Hamajima et al. 2004; Hunt 2004; Menaker et al. 2004; Siavoshi et al. 2004). Being an acid-resistant neutralophile (Sachs et al. 1996; Bauerfeind et al. 1997; Rektorschek et al. 2000; Scott et al. 2002), it is capable of surviving for at least 3 h at pH 1 with urea (Stingl et al. 2001) and maintaining a nearly neutral cytoplasmic pH

* E-mail: xxia@uottawa.ca.

† Present address: Ontario Genomics Innovation Centre, Ottawa Health Research Institute, 501 Smyth, Ottawa, Ontario K1H 8L6, Canada.

Am. Nat. 2005. Vol. 166, pp. 776–784.

between pH 3.0 and 7.0 (Matin et al. 1996; Scott et al. 2002; Stingl et al. 2002*b*). These properties allow it to survive and reproduce in the human stomach, where the gastric fluid has a pH averaging about 1.4 over a 24-h period (Sachs et al. 2003). The buffering action of the gastric epithelium and limited acid diffusion through the gastric mucus were previously thought to protect the bacterium against stomach acidity, but both empirical studies (Allen et al. 1993) and theoretical modeling (Engel et al. 1984) have suggested that the protection is rather limited (Matin et al. 1996; Sachs et al. 2003). Recently, it has also been shown that mucus does not hinder proton diffusion, and a transmucus pH gradient is abolished when the luminal pH drops to <2.5 (Baumgartner and Montrose 2004). It is therefore necessary for *H. pylori* to have acid resistance mechanisms to colonize the gastric mucosa successfully (Sachs et al. 2003).

One of the three mechanisms for acid resistance in *H. pylori* (Sachs et al. 2003) is the restriction of acute proton entry across its membranes by having a high frequency of positively charged amino acids in the inner and outer membrane proteins (Scott et al. 1998; Sachs et al. 2003; Valenzuela et al. 2003). This is supported by the recent discovery of a basic proteome (Tomb et al. 1997) as well as a set of basic membrane proteins (Baik et al. 2004) in *H. pylori*. The mechanism has been interpreted to result from adaptation (Tomb et al. 1997; Sachs et al. 2003). This hypothesis will be referred to as the acid adaptation hypothesis hereafter.

One might argue that if lysine is favored in *H. pylori* by natural selection for its positive charge, then we should expect other positively charge amino acids such as arginine and histidine to be favored too. The fact that arginine is not overused in *H. pylori* (Tomb et al. 1997) seems to weaken the acid adaptation hypothesis. However, there is a reason why arginine is not overused in *H. pylori*. The long-term intracellular pH homeostasis in *H. pylori* depends almost entirely on the buffering of ammonia produced from urea, catalyzed by the highly expressed urease (Mobley et al. 1991; Rektorschek et al. 2000; Stingl et al. 2002*a*; Sachs et al. 2003). To maintain the supply of urea, arginine is efficiently converted to urea by arginase, which is highly active in *H. pylori* (Mendz and Hazell 1996). Thus, arginine is efficiently channeled away from protein synthesis, and the acid adaptation hypothesis does not require the overuse of arginine.

Another argument against the acid adaptation hypothesis is as follows. The *H. pylori* genome is AT rich, and AT-rich genomes always contain many AAR (where R stands for either A or G) codons coding for the basic amino acid lysine. Thus, the basic proteome may simply be a consequence of a high genomic AT%, which may result from factors completely unrelated to the acid resistance.

Two factors unrelated to the acid resistance can result in a high genomic AT% and consequently high lysine codons. First, an AT-biased spontaneous mutation spectrum has been shown in mitochondrial genomes (Marcelino et al. 1998), prokaryotic genomes (Wang et al. 1996), and pseudogenes in mammalian nuclear genomes (Gojobori et al. 1982; Li et al. 1984). A possible biochemical basis of this AT-biased mutation spectrum is the spontaneous deamination (Sancar and Sancar 1988; Frederico et al. 1990; Frederico et al. 1993; Lindahl 1993). Biased mutations can change nucleotide frequencies not only at the third codon position but also at the first and the second codon positions and consequently change the amino acid usage (Sueoka 1961; Gu et al. 1998; Hickey and Singer 2004; Lobry 2004; Wang et al. 2004). If the high genome AT% and the associated basic proteome result from AT-biased mutation spectrum, then the basic proteome is a result of exaptation. We will designate this explanation for the high frequency of lysine in proteins as the AT-biased mutation hypothesis.

Second, the CTP concentration is generally the lowest among the four nucleotides (Colby and Edlin 1970), suggesting that CTP may be limited and that a reduced use of CTP may be evolutionarily advantageous. Consistent with this interpretation, the protozoan parasite *Trypanosoma brucei* maintains its de novo synthesis pathway for CTP, and inhibiting its CTP synthetase effectively eradicates the parasite population in the host (Hofer et al. 2001). In contrast, the parasite does not have de novo synthesis pathways for purines, suggesting that the parasite can obtain the purines by its salvage pathway. This suggests that little CTP can be salvaged from the host. The relevance of these observations is highlighted by the fact that *H. pylori* maintains an active biosynthesis pathway, and a much less active salvage pathway, for pyrimidine nucleotides (Mendz et al. 1994). Thus, it might be evolutionarily beneficial for a mammalian parasite or symbiont to minimize the use of CTP in its DNA (and consequently GTP because of the complementary nature of the DNA double helix) in building its genomes and in transcription (Xia 1996; Rocha and Danchin 2002). If the high genomic AT% in *H. pylori* has resulted from the selection to minimize the usage of the presumably rare C, then the basic proteome should be interpreted as resulting from preadaptation. We will designate this explanation for the basic proteome in some bacterial species as the C minimization hypothesis.

One additional piece of information against the acid adaptation hypothesis is that the high genomic AT% and high protein pI values are not unique in *H. pylori*; they are also found in other bacterial species in the proteobacterial cluster to which *H. pylori* belongs. For example, of the proteobacterial genomes available in GenBank as of

January 2005, four (*Wigglesworthia glossinidia* and three different *Buchnera aphidicola* strains) are even more AT rich than *H. pylori*, use more lysine, and have even higher average pI values for the genome-derived proteome. Note that both *W. glossinidia* and *B. aphidicola* are obligate endosymbionts and are not exposed to an acidic environment. If nature happened to dip *W. glossinidia* or *B. aphidicola* into the acidic mammalian gastric fluid, then their positively charged proteins may well come in handy and contribute to the limitation of the H^+ influx into their cells. In such cases, we would have interpreted the acid resistance mechanism as resulting from preadaptation or exaptation instead of adaptation. It seems reasonable to apply the same argument to *H. pylori.*

The acid adaptation hypothesis and the other two hypotheses have different predictions. According to the acid adaptation hypothesis, the first and second codon position should contribute more to the AT richness of the *H. pylori* genome than the third codon positions because increasing AT richness at the third codon position does not increase lysine usage, whereas increasing AT richness at the first and second codon positions does. However, according to the AT-biased mutation hypothesis, the third codon position, which is prone to spontaneous mutation and neutral fixation, should contribute more than the first and second codon position toward the genomic AT richness. Similarly, the C minimization hypothesis would predict a reduction of C and G (and a consequent increase of A and T) at the third codon position because such a reduction allows the minimization of C without involving any deleterious nonsynonymous mutations. Thus, the prediction from the AT-biased mutation spectrum and that from the C minimization hypothesis are the same, and both are different from that based on the acid adaptation hypothesis.

For bacterial species with an extremely high genomic AT% but not exposed to an acidic environment, such as *W. glossinidia* and *B. aphidicola*, the acid resistance hypothesis is obviously inapplicable. We expect the nucleotide and dinucleotide frequencies at different codon sites in the protein-coding genes of these species to follow the prediction of either the AT-biased mutation hypothesis or the C minimization hypothesis. In contrast, if the acid adaptation hypothesis is correct for *H. pylori*, that is, acid resistance demands increased usage of the positively charged lysine whose AAR codons drag up the genomic AT%, then we should expect nucleotide A and dinucleotide AA to be found mostly at the first and second codon positions. In this article, we perform a comparative genomic analysis to examine these different predictions.

Material and Methods

The complete genomes used in this article are listed in table A1 (available in the online edition of the *American Naturalist*), together with basic descriptions of genomic features. In addition to the two strains of *Helicobacter pylori*, two species related to *H. pylori*, that is, *Helicobacter hepaticus* and *Campylobacter jejuni* (with two genomes from two different strains), are also included for comparison. *Helicobacter hepaticus* is a mammalian liver parasite (Suerbaum et al. 2003) occasionally found in the intestine, and *C. jejuni* is a parasite of mammalian intestine (Parkhill et al. 2000). Four other proteobacterial genomes (*Wigglesworthia glossinidia* and three strains of *Buchnera aphidicola*) were included because of their extremely high genomic AT% and high average protein pI. These different genomes may allow an evaluation of the relative importance of the three hypotheses in different evolutionary lineages.

The coding sequences (CDSs) were extracted by using the DAMBE software (Xia 2001; Xia and Xie 2001). The genomic AT%, nucleotide frequencies at the three codon positions, and dinucleotide frequencies at the codon positions (1,2), (2,3), and (3,1) were also calculated by using DAMBE.

The CDSs were translated into amino acid sequences, and the amino acid usage was calculated from these genome-derived proteins. Protein pI's were calculated with the pK values for arginine, lysine, histidine, tyrosine, cysteine, glutamate, aspartate, NH_2, and COOH being 12.50, 10.79, 6.50, 10.95, 8.30, 4.25, 3.91, 8.56, and 3.56, respectively. These calculations are also done with DAMBE.

The *H. pylori* membrane proteins were based on a previous study (Baik et al. 2004). The CDSs coding for the membrane proteins were extracted from the genome of the strain 26695 of the *H. pylori* and translated into amino acid sequences, and amino acid usage and protein pI were quantified.

Results and Discussion

Comparative Study Based on Genome-Derived Proteomes

The distribution of nucleotide frequencies by codon positions (table 1) exhibits a pattern consistent with the acid adaptation hypothesis. For *Helicobacter pylori*, the frequency of nucleotide A is the highest at the first and second codon positions but relatively rare at the third codon position (table 1). This supports the interpretation that the necessity of increasing protein pI favors the increase of lysine codons (AAA and AAG) and consequently an increase of the frequency of nucleotide A at the first and second codon positions.

The pattern contradicts the AT-biased mutation hy-

Table 1: Distribution of nucleotide frequencies by codon positions (CP)

Species and CP	A	C	G	U
Helicobacter pylori J99:[a]				
1	.3383	.1464	.2996	.2157
2	.3598	.1766	.1479	.3157
3	.2594	.2210	.2055	.3140
Helicobacter hepaticus:				
1	.3301	.1788	.2979	.1932
2	.3459	.1843	.1522	.3176
3	.3120	.1266	.1499	.4115
Campylobacter jejuni NCTC 11168:[a]				
1	.3596	.1277	.2976	.2151
2	.3665	.1642	.1408	.3285
3	.3616	.0916	.1033	.4436
Buchnera aphidicola APS:[a]				
1	.3873	.1289	.2449	.2389
2	.3600	.1743	.1315	.3343
3	.4200	.0635	.0789	.4376
Wigglesworthia glossinidia:				
1	.4318	.0887	.2205	.2590
2	.3741	.1614	.1203	.3441
3	.4679	.0513	.0672	.4136

[a] Different strains of the same species differ little in their frequencies by codon positions, and we present results from only one strain.

pothesis, which predicts that the third codon position, not constrained by purifying selection, would be more affected by the biased mutations and should have more nucleotide A than the first and second codon positions. We may conclude that the high genomic AT% in *H. pylori* cannot be solely explained by the AT-biased mutation hypothesis. This is corroborated by a recent mutation study (Torii et al. 2003) showing that all characterized mutations in *H. pylori* were A → G transitions. This weakens the foundation of the AT-biased mutation hypothesis; that is, *H. pylori* does not seem to have an AT-biased mutation spectrum.

Nucleotide C is more frequent at the third codon position than at the first and second codon positions in *H. pylori* (table 1). This contradicts the C minimization hypothesis, which would predict the opposite; that is, the selection for minimizing the use of C should be more effective at the third codon position because it does not need to fight against the stronger purifying selection at the first and second codon positions.

The distribution of nucleotide A among the three codon positions is relatively even in the three related genomes, one from *Helicobacter hepaticus* and two others from the two strains of *Campylobacter jejuni* (table 1). For the other four genomes (one from *Wigglesworthia glossinidia* and three from the three strains of *Buchnera aphidicola*), nucleotides A and U are both more frequent at the third codon position than at the first and second codon positions. This is consistent with the prediction of the AT-biased mutation hypothesis. However, the pattern for these genomes is also generally consistent with the C minimization hypothesis. Except for *H. pylori*, the frequencies of nucleotides C and G are consistently lower at the third codon position than at the two other codon positions (table 1). In short, while the AT-biased mutation hypothesis and the C minimization hypothesis are unsatisfactory to account for the frequency distribution of nucleotides among the three codon positions in *H. pylori*, they both can explain the observed pattern of nucleotide distribution among the three codon positions for the rest of the species quite satisfactorily.

The distribution of nucleotide A among the three codon sites in *H. pylori* is more similar to that of *H. hepaticus* than to that of other species (table 1). However, even this smaller difference between *H. hepaticus* and *H. pylori* is still highly significant (χ^2 contingency table test with two species and three codon positions, $\chi^2 = 7{,}428.05$, df $= 2$, $P = .0000$).

The distribution of dinucleotide AA over the codon positions (1,2), (2,3), and (3,1) (i.e., the last codon position and the first codon position of the next codon) also supports the acid adaptation hypothesis for *H. pylori* strains (table 2). More AA dinucleotides are found at codon positions (1,2) and (2,3) than at codon position (3,1) in *H. pylori* CDSs. The distribution is roughly even in other species. Although the *H. hepaticus* CDSs exhibit a pattern similar to that of *H. pylori* (table 2), the difference is still highly significant ($\chi^2 = 984.18$, df $= 2$, $P = .0000$ for the test between *H. pylori* J99 and *H. hepaticus*; $\chi^2 = 911.12$, df $= 2$, $P = .0000$ for the test between *H. pylori* 26695 and *H. hepaticus*).

One may note that AA at codon position (2,3), instead of at codon position (1,2), is the most common in table 2 for the two *H. pylori* strains. The reason is as follows. Both Lys and Asn (coded by AAY) codons contribute to AA dinucleotides at codon position (1,2); that is, the number of AA dinucleotides at position (1,2) is the sum of AAA, AAG, AAC, and AAU codons. Similarly, the number

Table 2: Distribution of dinucleotide AA at three codon positions

Species	(1,2)	(2,3)	(3,1)
Helicobacter pylori J99	4.9001	5.0366	3.3028
H. pylori 26695	4.9218	5.0793	3.3567
Helicobacter hepaticus	4.2728	4.4777	3.6305
Campylobacter jejuni NCTC 11168	5.2465	5.6433	5.1828
C. jejuni RM1221	5.3576	5.7590	5.2666
Buchnera aphidicola APS	5.6772	5.6964	5.7063
B. aphidicola Bp	5.7286	5.2055	5.6889
B. aphidicola Sg	5.9636	5.9508	6.1857
Wigglesworthia glossinidia	6.7045	5.9885	7.2147

of AA dinucleotides at codon position (2,3) is the sum of AAA Lys codons and non-Lys CAA, GAA, and UAA codons. Therefore, the numbers do not depend entirely on the Lys codons. To clarify this point, we have detailed these respective contributions in table 3, contrasting between the two *H. pylori* strains and that of *H. hepaticus*. All three genomes have more AAR codons than AAY codons, but this is more conspicuous in the two *H. pylori* strains than in *H. hepaticus* (table 3). Similarly, there are more (CGU)AA codons than AAA codons in the three genomes, but AAA codons are relatively more abundant in the two *H. pylori* strains than in *H. hepaticus* (table 3). In particular, many more AA dinucleotides are at codon position (3,1) in *H. hepaticus* than in the two strains of *H. pylori* (table 3).

The lysine codons (AAA and AAG) account for 8.82% and 8.91% of the total codons in *H. pylori* J99 and *H. pylori* 26695, respectively, but only 7.5% in the sister species *H. hepaticus*. The difference can be tested by the contingency table test with two species and two codon categories (AAR and non-AAR codons) and is highly significant statistically ($\chi^2 = 581.19$, df = 1, $P = .0000$ for the test between *H. pylori* J99 and *H. hepaticus*; $\chi^2 = 667.73$, df = 1, $P = .0000$ for the test between *H. pylori* 26695 and *H. hepaticus*). This is particularly significant in light of the fact that the *H. hepaticus* genome is more AT rich than the two *H. pylori* species and is consequently expected to have more AAR codons if nucleotides assemble randomly to form codons.

For simplicity of presentation, we designate $N_{o.ijk}$ as the observed number of codon *ijk* (where *i*, *j*, and *k* are any of the four nucleotides), and $N_{e.ijk}$ and $R_{o/e.ijk}$ as follows:

$$N_{e.ijk} = NP_iP_jP_k,$$

$$R_{o/e.ijk} = \frac{N_{o.ijk}}{N_{e.ijk}}, \quad (1)$$

where P_i, P_j, and P_k are the frequencies of nucleotide *i*, *j*, and *k*, respectively, and *N* is the total number of codons in the genome. The plot of $N_{o.AAR}$ versus $N_{e.AAR}$ shows that the percentage of AAR codons in the two *H. pylori* strains is substantially higher than the regression line (fig. 1).

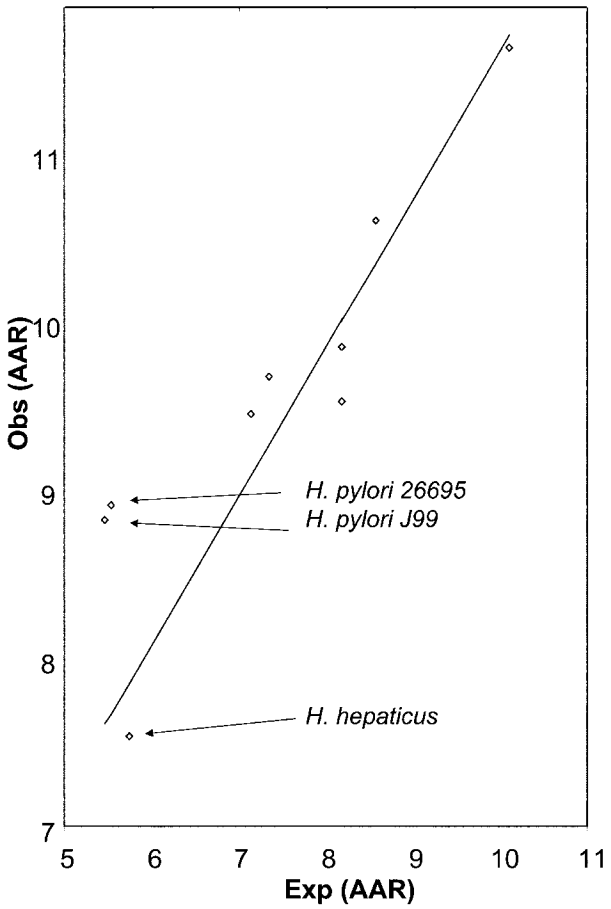

Figure 1: Observed versus expected number of AAR codons as a percentage over the total number of codons. The regression line is fitted after excluding the two points from the two *Helicobacter pylori* strains.

One might argue that if lysine is favored in *H. pylori* by natural selection for its positive charge, then we should expect other positively charged amino acids such as histidine to be favored too (we have already explained why arginine should not be overused in the introduction). Histidine is coded by CAY codons, and the $R_{o/e.CAY}$ value is 0.978 for *H. hepaticus*, but it increased to 1.252 and 1.267, respectively, for *H. pylori* strains J99 and 26695. This is consistent with the acid adaptation hypothesis.

Table 3: Contrasting the distribution of AA dinucleotides at codon positions (1,2), (2,3), and (3,1) together with the total number of codons (N_{codon}) in the genome

	$AA_{1,2}$		$AA_{2,3}$			
Species	AAR	AAY	AAA	(CGU)AA[a]	$AA_{3,1}$	N_{codon}
Helicobacter pylori J99	43,702	29,075	33,471	41,333	49,053	495,564
H. pylori 26695	44,813	29,379	34,548	42,018	50,599	502,993
Helicobacter hepaticus	42,176	29,510	32,279	42,846	60,910	559,874

[a] The sum of CAA, GAA, and UAA.

To summarize results obtained by the analysis of CDSs from different species, we may conclude that *H. pylori* has increased the AAR codons significantly from its sister species *H. hepaticus* and that the increase is not due to AT-biased mutation or C minimization but is consistent with the predictions from the acid adaptation hypothesis. In short, it is reasonable to interpret the basic genome-derived proteome in the *H. pylori* strains as resulting from adaptation in response to the acidic stomach fluid.

Comparative Study Based on Membrane Proteins

The membrane proteins have long been suspected to play an important role in acid resistance in *H. pylori* (Huynen et al. 1998; Alm et al. 2000; Yamaoka et al. 2002; Solnick et al. 2004). This is corroborated by a recent characterization of membrane proteins of *H. pylori* STR 26695 showing membrane proteins to be mainly basic (Baik et al. 2004). We calculated the protein pI for the 34 identified membrane proteins (35 in the original article, but there are only 34 unique ones). Four proteins (HP0243, HP0072, HP0512, and HP1563) have pI values ranging from 5.86 to 6.25, whereas the other 30 have pI > 7. The average pI is 8.9221 for these 34 membrane proteins, whereas the average calculated pI value for the other 1,542 proteins is 8.2147. The two average pI values are significantly different by a two-sample *t*-test ($t = 2.075$, $df = 1{,}572$, $P = .0382$, two-tailed test). Thus, membrane proteins are significantly more basic than the rest of the proteins in *H. pylori*.

A comparison of the *H. pylori* membrane proteins with those in the related *H. hepaticus* may shed light on whether the basic membrane proteins in *H. pylori* result from adaptation in response to the acidic environment. If *H. hepaticus*, which is not acid resistant, also features a set of equally basic or even more basic membrane proteins, then the set of basic membrane proteins in *H. pylori* is likely an ancestral trait evolved before *H. pylori* became a stomach parasite and consequently should not be interpreted as resulting from adaptation in response to the acidic stomach environment. In contrast, if the set of basic membrane proteins is unique in *H. pylori*, then we would have more confidence in the acid adaptation hypothesis.

We searched homologs of these 34 *H. pylori* membrane proteins by BLASTing (Altschul et al. 1990, 1997) their corresponding CDSs against the 1,875 CDSs in the related *H. hepaticus* genome, with a cutoff *e* value of 0.0001. The four *H. pylori* membrane proteins with pI < 7 all have homologs in the *H. hepaticus* genome. In contrast, among the other 30 membrane proteins with pI > 7, only one has a homolog in the *H. hepaticus* genome. It is important to note that, out of the 1,576 predicted protein-coding genes in the genomes of *H. pylori* strain 26695, 938 found matches in the genome of *H. hepaticus*. Similarly, 941 out of 1,492 predicted protein-coding genes in the genome of *H. pylori* strain J99 have matches in the genome of *H. hepaticus* (Suerbaum et al. 2003). With this reference in mind, the number of matches for *H. hepaticus* 26695 membrane proteins in the *H. hepaticus* genome is relatively small. This suggests that nearly all those positively charged membrane proteins in *H. pylori* are unique and most likely result from the evolution along the *H. pylori* lineage. These basic membrane proteins either may have evolved quickly along the *H. pylori* lineage, so as to be beyond recognition in the *H. hepaticus* genome, or represent differential gain of genes in the *H. pylori* lineage or differential loss of genes in the *H. hepaticus* lineage. In any case, this result lends more support to the interpretation that the basic proteome and its function to decrease the influx of H^+ into the *H. pylori* cell have resulted from adaptation.

The rapid evolution of the membrane proteins in *H. pylori* from *H. hepaticus* is not surprising because *H. pylori* is known to evolve rapidly, leading to many genetically variable strains differing not only in the genome (Owen et al. 1991; Hazell et al. 1997; Alm and Trust 1999; Axon 1999; Suerbaum 2000; Enroth and Engstrand 2001; Israel et al. 2001) but also in the protein production (Jungblut et al. 2000). For example, among the more than 20 putative DNA restriction modification (RM) systems, all those strain-specific RM genes are active and functional, whereas most shared genes are inactive and presumably nonfunctional (Lin et al. 2001). This implies that evolution is quickly erasing the shared ancestry of different *H. pylori* strains.

Helicobacter pylori is a highly adaptable proteobacterial species. For example, an *H. pylori* strain that initially had grown only weakly in piglets became quickly adapted to them by spontaneous mutation and selection after only 12 serial passages, resulting in an increase in yield from about 10^3 to $>10^7$ bacteria per gram of mucosa (Akopyants et al. 1995). A number of factors may contribute to the great adaptive potential of *H. pylori*. The population size is large, and the mutation rate is high (Wang et al. 1999; Bjorkholm et al. 2001). Different strains often cocolonize the same host (Akopyants et al. 1995; Cao and Cover 1997; Suerbaum and Achtman 2004). This, coupled with a high recombination rate (Suerbaum et al. 1998; Kersulyte et al. 1999; Bjorkholm et al. 2001), would create a great diversity of genotypes for selection to act on. In addition, most strains of *H. pylori* are naturally competent for uptake of chromosomal DNA (Wang et al. 1993), leading to the horizontal gene transfer (Censini et al. 1996; Covacci et al. 1997; Alm and Trust 1999; Axon 1999). The joint effect of mutation, recombination, and horizontal transfer will contribute to the generation of new genotypes and phenotypes in *H. pylori* for natural selection to act on.

Acknowledgments

This study is supported by grants from the University of Ottawa and from the Natural Sciences and Engineering Research Council's Discovery, Research Tools and Instruments, and Strategic grants to X.X. We thank G. Sachs for suggestions and references and J. Popescu and H. Xiong for discussion. Comments from two anonymous reviewers significantly improved the readability of the article.

Literature Cited

Akopyants, N. S., K. A. Eaton, and D. E. Berg. 1995. Adaptive mutation and cocolonization during *Helicobacter pylori* infection of gnotobiotic piglets. Infection and Immunity 63:116–121.

Allen, A., G. Flemstrom, A. Garner, and E. Kivilaakso. 1993. Gastroduodenal mucosal protection. Physiological Reviews 73:823–857.

Alm, R. A., and T. J. Trust. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. Journal of Molecular Medicine 77:834–846.

Alm, R. A., J. Bina, B. M. Andrews, P. Doig, R. E. Hancock, and T. J. Trust. 2000. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. Infection and Immunity 68:4155–4168.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389–3402.

Axon, A. T. 1999. Are all helicobacters equal? mechanisms of gastroduodenal pathology and their clinical implications. Gut 45:I1–I4.

Baik, S. C., K. M. Kim, S. M. Song, D. S. Kim, J. S. Jun, S. G. Lee, J. Y. Song, et al. 2004. Proteomic analysis of the sarcosine-insoluble outer membrane fraction of *Helicobacter pylori* strain 26695. Journal of Bacteriology 186:949–955.

Bauerfeind, P., R. Garner, B. E. Dunn, and H. L. Mobley. 1997. Synthesis and activity of *Helicobacter pylori* urease and catalase at low pH. Gut 40:25–30.

Baumgartner, H. K., and M. H. Montrose. 2004. Regulated alkali secretion acts in tandem with unstirred layers to regulate mouse gastric surface pH. Gastroenterology 126:774–783.

Bjorkholm, B., A. Lundin, A. Sillen, K. Guillemin, N. Salama, C. Rubio, J. I. Gordon, et al. 2001. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. Infection and Immunity 69:7832–7838.

Cao, P., and T. L. Cover. 1997. High-level genetic diversity in the vapD chromosomal region of *Helicobacter pylori*. Journal of Bacteriology 179:2852–2856.

Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, et al. 1996. cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proceedings of the National Academy of Sciences of the USA 93:14648–14653.

Colby, C., and G. Edlin. 1970. Nucleotide pool levels in growing, inhibited, and transformed chick fibroblast cells. Biochemistry 9: 917.

Correa, P. 1997. *Helicobacter pylori* as a pathogen and carcinogen. Journal of Physiology and Pharmacology 48:19–24.

Covacci, A., S. Falkow, D. E. Berg, and R. Rappuoli. 1997. Did the inheritance of a pathogenicity island modify the virulence of *Helicobacter pylori*? Trends in Microbiology 5:205–208.

Darwin, C. 1859. On the origin of species. J. Murray, London.

Engel, E., A. Peskoff, G. L. Kauffman Jr., and M. I. Grossman. 1984. Analysis of hydrogen ion concentration in the gastric gel mucus layer. American Journal of Physiology 247:G321–G338.

Enroth, H., and L. Engstrand. 2001. An update on *Helicobacter pylori* microbiology and infection for the new millennium. Scandinavian Journal of Infectious Diseases 33:163–174.

Frederico, L. A., T. A. Kunkel, and B. R. Shaw. 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29:2532–2537.

———. 1993. Cytosine deamination in mismatched base pairs. Biochemistry 32:6523–6530.

Gojobori, T., W.-H. Li, and D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. Journal of Molecular Evolution 18:360–369.

Gould, S. J. 2002. The structure of evolutionary theory. Belknap, Cambridge, MA.

Gould, S. J., and E. S. Vrba. 1982. Exaptation: a missing term in the science of form. Paleobiology 8:4–15.

Gu, X., D. Hewett-Emmett, and W. H. Li. 1998. Directional mutational pressure affects the amino acid composition and hydrophobicity of proteins in bacteria. Genetica 102/103:383–391.

Hamajima, N., Y. Goto, K. Nishio, D. Tanaka, S. Kawai, H. Sakakibara, and T. Kondo. 2004. *Helicobacter pylori* eradication as a preventive tool against gastric cancer. Asian Pacific Journal of Cancer Prevention 5:246–252.

Hazell, S. L., R. H. Andrews, H. M. Mitchell, and G. Daskalopoulous. 1997. Genetic relationship among isolates of *Helicobacter pylori*: evidence for the existence of a *Helicobacter pylori* species-complex. FEMS Microbiology Letters 150:27–32.

Hickey, D. A., and G. A. Singer. 2004. Genomic and proteomic adaptations to growth at high temperature. Genome Biology 5:117.

Hofer, A., D. Steverding, A. Chabes, R. Brun, and L. Thelander. 2001. *Trypanosoma brucei* CTP synthetase: a target for the treatment of African sleeping sickness. Proceedings of the National Academy of Sciences of the USA 98:6412–6416.

Hunt, R. H. 2004. Will eradication of *Helicobacter pylori* infection influence the risk of gastric cancer? American Journal of Medicine 117:86S–91S.

Huynen, M., T. Dandekar, and P. Bork. 1998. Differential genome analysis applied to the species-specific features of *Helicobacter pylori*. FEBS Letters 426:1–5.

Israel, D. A., N. Salama, U. Krishna, U. M. Rieger, J. C. Atherton, S. Falkow, and R. M. Peek Jr. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proceedings of the National Academy of Sciences of the USA 98: 14625–14630.

Jungblut, P. R., D. Bumann, G. Haas, U. Zimny-Arndt, P. Holland, S. Lamer, F. Siejak, et al. 2000. Comparative proteome analysis of *Helicobacter pylori*. Molecular Microbiology 36:710–725.

Kersulyte, D., H. Chalkauskas, and D. E. Berg. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. Molecular Microbiology 31:31–43.

Kolbe, J. J., R. E. Glor, L. Rodriguez Schettino, A. C. Lara, A. Larson,

and J. B. Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431:177–181.

Li, W. H., C.-I. Wu, and C. C. Luo. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. Journal of Molecular Evolution 21:58–71.

Lin, L. F., J. Posfai, R. J. Roberts, and H. Kong. 2001. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. Proceedings of the National Academy of Sciences of the USA 98:2740–2745.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. Nature 362:709–715.

Lobry, J. R. 2004. Life history traits and genome structure: aerobiosis and G+C content in bacteria. Lecture Notes in Computer Science 3039:679–686.

Marcelino, L. A., P. C. Andre, K. Khrapko, H. A. Coller, J. Griffith, and W. G. Thilly. 1998. Chemically induced mutations in mitochondrial DNA of human cells: mutational spectrum of N-methyl-N′-nitro-N-nitrosoguanidine. Cancer Research 58:2857–2862.

Matin, A., E. Zychlinsky, M. Keyhan, and G. Sachs. 1996. Capacity of *Helicobacter pylori* to generate ionic gradients at low pH is similar to that of bacteria which grow under strongly acidic conditions. Infection and Immunity 64:1434–1436.

Menaker, R. J., A. A. Sharaf, and N. L. Jones. 2004. *Helicobacter pylori* infection and gastric cancer: host, bug, environment, or all three? Current Gastroenterology Reports 6:429–435.

Mendz, G. L., and S. L. Hazell. 1996. The urea cycle of *Helicobacter pylori*. Microbiology 142:2959–2967.

Mendz, G. L., B. M. Jimenez, S. L. Hazell, A. M. Gero, and W. J. O'Sullivan. 1994. De novo synthesis of pyrimidine nucleotides by *Helicobacter pylori*. Journal of Applied Bacteriology 77:1–8.

Mobley, H. L., L. T. Hu, and P. A. Foxal. 1991. *Helicobacter pylori* urease: properties and role in pathogenesis. Scandinavian Journal of Gastroenterology 187(suppl.):39–46.

Owen, R. J., J. Bickley, M. Costas, and D. R. Morgan. 1991. Genomic variation in *Helicobacter pylori*: application to identification of strains. Scandinavian Journal of Gastroenterology 181(suppl.):43–50.

Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, et al. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 403:665–668.

Prum, R. O., and A. H. Brush. 2002. The evolutionary origin and diversification of feathers. Quarterly Review of Biology 77:261–295.

———. 2003. Which came first, the feather or the bird? Scientific American 288:84–93.

Rektorschek, M., A. Buhmann, D. Weeks, D. Schwan, K. W. Bensch, S. Eskandari, D. Scott, et al. 2000. Acid resistance of *Helicobacter pylori* depends on the UreI membrane protein and an inner membrane proton barrier. Molecular Microbiology 36:141–152.

Ridley, M. 2004. Evolution. Blackwell, Malden, MA.

Rocha, E. P., and A. Danchin. 2002. Base composition bias might result from competition for metabolic resources. Trends in Genetics 18:291–294.

Sachs, G., K. Meyer-Rosberg, D. R. Scott, and K. Melchers. 1996. Acid, protons and *Helicobacter pylori*. Yale Journal of Biology and Medicine 69:301–316.

Sachs, G., D. L. Weeks, K. Melchers, and D. R. Scott. 2003. The gastric biology of *Helicobacter pylori*. Annual Review of Physiology 65:349–369.

Sancar, A., and G. B. Sancar. 1988. DNA repair enzymes. Annual Review of Biochemistry 57:29–67.

Scott, D., D. Weeks, K. Melchers, and G. Sachs. 1998. The life and death of *Helicobacter pylori*. Gut 43:S56–S60.

Scott, D. R., E. A. Marcus, D. L. Weeks, and G. Sachs. 2002. Mechanisms of acid resistance due to the urease system of *Helicobacter pylori*. Gastroenterology 123:187–195.

Siavoshi, F., R. Malekzadeh, M. Daneshmand, D. T. Smoot, and H. Ashktorab. 2004. Association between *Helicobacter pylori* infection in gastric cancer, ulcers and gastritis in Iranian patients. Helicobacter 9:470.

Smit, A. F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Current Opinion in Genetics and Development 9:657–663.

Solnick, J. V., L. M. Hansen, N. R. Salama, J. K. Boonjakuakul, and M. Syvanen. 2004. Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. Proceedings of the National Academy of Sciences of the USA 101:2106–2111.

Stingl, K., E.-M. Uhlemann, G. Deckers-Hebestreit, R. Schmid, E. P. Bakker, and K. Altendorf. 2001. Prolonged survival and cytoplasmic pH homeostasis of *Helicobacter pylori* at pH 1. Infection and Immunity 69:1178–1180.

Stingl, K., K. Altendorf, and E. P. Bakker. 2002*a*. Acid survival of *Helicobacter pylori*: how does urease activity trigger cytoplasmic pH homeostasis? Trends in Microbiology 10:70–74.

Stingl, K., E.-M. Uhlemann, R. Schmid, K. Altendorf, and E. P. Bakker. 2002*b*. Energetics of *Helicobacter pylori* and its implications for the mechanism of urease-dependent acid tolerance at pH 1. Journal of Bacteriology 184:3053–3060.

Sueoka, N. 1961. Correlation between base composition of deoxyribonucleic acid and amino acid composition of proteins. Proceedings of the National Academy of Sciences of the USA 47:1141–1149.

Suerbaum, S. 2000. Genetic variability within *Helicobacter pylori*. International Journal of Medical Microbiology 290:175–181.

Suerbaum, S., and M. Achtman. 2004. *Helicobacter pylori*: recombination, population structure and human migrations. International Journal of Medical Microbiology 294:133–139.

Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, et al. 1998. Free recombination within *Helicobacter pylori*. Proceedings of the National Academy of Sciences of the USA 95:12619–12624.

Suerbaum, S., C. Josenhans, T. Sterzenbach, B. Drescher, P. Brandt, M. Bell, M. Droge, et al. 2003. The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. Proceedings of the National Academy of Sciences of the USA 100:7901–7906.

Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388:539–547.

Torii, N., T. Nozaki, M. Masutani, H. Nakagama, T. Sugiyama, D. Saito, M. Asaka, et al. 2003. Spontaneous mutations in the *Helicobacter pylori* rpsL gene. Mutation Research 535:141–145.

Valenzuela, M., O. Cerda, and H. Toledo. 2003. Overview on chemotaxis and acid resistance in *Helicobacter pylori*. Biological Research 36:429–436.

Wang, G., M. Z. Humayun, and D. E. Taylor. 1999. Mutation as an

origin of genetic variability in *Helicobacter pylori*. Trends in Microbiology 7:488–493.

Wang, H. C., G. A. Singer, and D. A. Hickey. 2004. Mutational bias affects protein evolution in flowering plants. Molecular Biology and Evolution 21:90–96.

Wang, R. F., W. Campbell, W. W. Cao, C. Summage, R. S. Steele, and C. E. Cerniglia. 1996. Detection of *Pasteurella pneumotropica* in laboratory mice and rats by polymerase chain reaction. Laboratory Animal Science 46:81–85.

Wang, Y., K. P. Roos, and D. E. Taylor. 1993. Transformation of *Helicobacter pylori* by chromosomal metronidazole resistance and by a plasmid with a selectable chloramphenicol resistance marker. Journal of General Microbiology 139:2485–2493.

Williams, G. C. 1966. Adaptation and natural selection. Princeton University Press, Princeton, NJ.

Xia, X. 1996. Maximizing transcription efficiency causes codon usage bias. Genetics 144:1309–1320.

———. 2001. Data analysis in molecular biology and evolution. Kluwer Academic, Boston.

Xia, X., and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. Journal of Heredity 92:371–373.

Yamaoka, Y., M. Kita, T. Kodama, S. Imamura, T. Ohno, N. Sawai, A. Ishimaru, et al. 2002. *Helicobacter pylori* infection in mice: role of outer membrane proteins in colonization and inflammation. Gastroenterology 123:1992–2004.

Associate Editor: Chung-I Wu
Editor: Jonathan B. Losos