The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: Evolution during disease progression

Jung D. Oh*†‡, Helene Kling-Bäckhed*†‡§¶, Marios Giannakis*†, Jian Xu*‖, Robert S. Fulton‖, Lucinda A. Fulton‖, Holland S. Cordum‖, Chunyan Wang‖, Glendoria Elliott‖, Jennifer Edwards‖, Elaine R. Mardis‖**, Lars G. Engstrand§¶, and Jeffrey I. Gordon*†,††**

*Center for Genome Sciences, ‖Genome Sequencing Center, and Departments of †Molecular Biology and Pharmacology and **Genetics, Washington University School of Medicine, St. Louis, MO 63108; §Department of Microbiology, Tumor, and Cell Biology, Karolinska Institute, 171 77 Stockholm, Sweden; and ¶Swedish Institute for Infectious Disease Control, 171 82 Solna, Sweden

Contributed by Jeffrey I. Gordon, May 8, 2006

***Helicobacter pylori* produces acute superficial gastritis in nearly all of its human hosts. However, a subset of individuals develops chronic atrophic gastritis (ChAG), a condition characterized in part by diminished numbers of acid-producing parietal cells and increased risk for development of gastric adenocarcinoma. Previously, we used a gnotobiotic transgenic mouse model with an engineered ablation of parietal cells to show that loss of parietal cells provides an opportunity for a *H. pylori* isolate from a patient with ChAG (HPAG1) to bind to, enter, and persist within gastric stem cells. This finding raises the question of how ChAG influences *H. pylori* genome evolution, physiology, and tumorigenesis. Here we describe the 1,596,366-bp HPAG1 genome. Custom HPAG1 Affymetrix GeneChips, representing 99.6% of its predicted ORFs, were used for whole-genome genotyping of additional *H. pylori* ChAG isolates obtained from Swedish patients enrolled in a case-control study of gastric cancer, as well as ChAG- and cancer-associated isolates from an individual who progressed from ChAG to gastric adenocarcinoma. The results reveal a shared gene signature among ChAG strains, as well as genes that may have been lost or gained during progression to adenocarcinoma. Whole-genome transcriptional profiling of HPAG1's response to acid during *in vitro* growth indicates that genes encoding components of metal uptake and utilization pathways, outer membrane proteins, and virulence factors are among those associated with *H. pylori*'s adaptation to ChAG.**

acid regulation | comparative microbial genomics | ecogenomics | functional genomics | gastric cancer

In the United States and Canada, as well as in Northern and Western Europe, 5–15% of children and 10–60% of adults harbor *Helicobacter pylori*. The prevalence is much higher elsewhere. For example, in Bangladesh, >50% of 2- to 9-year-old children from rural families are infected (1–3). Once acquired in childhood, this bacterium is able to establish a life-long relationship with its host (4).

H. pylori infection presents a therapeutic conundrum: The vast majority of hosts are asymptomatic and do not develop severe pathology. Moreover, *H. pylori* may benefit us by protecting against gastroesophageal reflux disease (5) and esophageal cancer (6). However, the risk of gastric cancer, which caused 10% of all cancer deaths worldwide in the year 2000 (7), is twice as high for infected individuals (8). Thus, one challenge is to identify *H. pylori*-bearing hosts who are at greatest risk for developing severe pathology and to target treatment to this population.

Virtually all individuals who become infected with *H. pylori* develop acute superficial gastritis; a subset progress to chronic atrophic gastritis (ChAG). ChAG is associated with loss of two of the three principal epithelial lineages in the stomach: acid-producing parietal cells and pepsinogen-expressing zymogenic (chief) cells. Both *H. pylori* infection and ChAG are associated with increased risk of gastric cancer (9, 10). Reports have appeared describing regression of histopathology after *H. pylori* eradication, leading to the suggestion that screening and treatment of these at-risk patients may be a cost-effective strategy for reducing gastric cancer (11, 12).

Although the frequency and persistence of *H. pylori* infection in humans make it an attractive model for examining the coevolution and coadaptation of a gut bacterium and its host over a significant fraction of an individual's life span, genetic and environmental variations among humans and colonizing strains have made it difficult to develop hypotheses about the contributions of microbial and host factors to the evolution of *H. pylori*-associated pathology.

Just as gastric pathology can evolve during what is typically a lifelong infection, the organism is also strongly suspected of being able to adapt to the changes that it induces in its gastric environment. This view is supported by reports of rapid loss of clonality during infection, the high rates of mutation and recombination observed in *H. pylori*, and the bacterium's natural competence (13–15). In a pioneering study, Israel *et al.* (13) examined the evolution of *H. pylori* in the setting of acid-peptic disease. At the time, two *H. pylori* genome sequences were available: 26695, obtained from a patient with superficial gastritis, and J99, obtained from a patient with duodenal ulcer disease (16, 17). The patient from whom J99 was obtained refused antimicrobial treatment and underwent repeat esophagogastroduodenoscopy 6 years later. The patient's gastric pathology exhibited no significant change and no evidence of ChAG was reported. DNA microarrays containing PCR-amplified ORFs from J99 and 26695 were used to characterize 13 strains isolated at the time of the second esophagogastroduodenoscopy. The results confirmed that all isolates were related to the original J99 isolate, yet all were distinct. Each strain had lost some genes that were present in the J99 isolate but had also acquired genes that are similar to those found in the genome of strain 26695. One tantalizing explanation is that a strain similar to 26695 was transiently present in the patient colonized with J99 (13). Based on these findings, we can envision a scenario in which (*i*) J99 was well adapted to the gastric habitat of this patient with stable duodenal ulcer disease, (*ii*) variation in the *H. pylori* population was dominated by gain or loss of "variable" genes, whereas mutation in

MICROBIOLOGY

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviations: ChAG, chronic atrophic gastritis; COG, Cluster of Orthologous Groups; PAI, pathogenicity island; R-M, restriction-modification.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database [accession nos. CP000241 (HPAG1 chromosome) and CP000242 (HPAG1 plasmid)].

‡J.D.O. and H.K.-B. contributed equally to this work.

††To whom correspondence should be addressed at: Washington University School of Medicine, 4444 Forest Park, Campus Box 8510, St. Louis, MO 63108. E-mail: jgordon@wustl.edu.

Table 1. ChAG-associated *H. pylori* genes

Function	Gene no.	Gene description/annotation
COG category		
Amino acid transport and metabolism (E)	HPAG1_0680	Hydantoin utilization protein
	HPAG1_0681	*N*-methylhydantoinase
Carbohydrate transport and metabolism (G)	HPAG1_0917	Proline and betaine transporter
Cell motility (N)	HPAG1_0103	Methyl-accepting chemotaxis protein
	HPAG1_0291	Putative vacuolating cytotoxin (VacA) paralog
	HPAG1_0579	Hemolysin secretion protein
	HPAG1_0903	Vacuolating cytotoxin (VacA) paralog
Cell wall/membrane/ envelope biogenesis (M)	HPAG1_0157	Lipopolysaccharide 1,2-glycosyltransferase
	HPAG1_1064	Peptidoglycan-associated lipoprotein
	HPAG1_1288	Siderophore-mediated iron transport protein
Coenzyme transport and metabolism (H)	HPAG1_0168	Molybdopterin biosynthesis protein
	HPAG1_0753	Molybdenum cofactor biosynthesis protein
	HPAG1_0783*	**Molybdenum cofactor biosynthesis protein**
	HPAG1_0784	**Molybdopterin biosynthesis protein**
	HPAG1_0785	**Molybdopterin converting factor, subunit 2**
Defense mechanisms (V)	HPAG1_0591	ABC transporter, permease
	HPAG1_0592	ABC transporter, permease
	HPAG1_0593	ABC transporter, ATP-binding protein
	HPAG1_0594	ABC transporter, ATP-binding protein
	HPAG1_0831	Type I restriction enzyme R protein
	HPAG1_1394	Type III R-M system restriction enzyme
	HPAG1_1395	Type IIS restriction-modification protein
Energy production and conversion (C)	HPAG1_0885	Phosphotransacetylase
	HPAG1_0626	NAD(P)H-flavin oxidoreductase
	HPAG1_0627	NAD(P)H-flavin oxidoreductase
	HPAG1_0985	Biotin sulfoxide reductase
Inorganic ion transport and metabolism (P)	**HPAG1_0420**	**Ferric uptake regulation protein**
	HPAG1_0451	Molybdenum ABC transporter
	HPAG1_0452	Molybdenum ABC transporter
	HPAG1_0669	Iron(III) dicitrate transport protein
	HPAG1_1469	**Iron(III) dicitrate transport protein**
Intracellular trafficking, secretion, and vesicular transport (U)	HPAG1_1143	Preprotein
	HPAG1_1286	Biopolymer transport protein
	HPAG1_1287	Biopolymer transport protein
Lipid transport and metabolism (I)	HPAG1_0402	Acetyl-CoA synthetase
	HPAG1_0538	Acyl carrier protein
Nucleotide transport and metabolism (F)	HPAG1_0838	Guanosine 5′-monophosphate oxidoreductase
Posttranslational modification, protein turnover, chaperones (O)	HPAG1_1457	Thioredoxin
Replication, recombination and repair (L)	HPAG1_0046	Adenine-specific DNA methyltransferase
	HPAG1_0047	**Cytosine-specific DNA methyltransferase**
	HPAG1_0262	Type III adenine methyltransferase
	HPAG1_0455	Type II adenine-specific methyltransferase
	HPAG1_0460	**Type II DNA modification enzyme**
	HPAG1_0671	Hypothetical protein
	HPAG1_0674	Hypothetical protein
	HPAG1_1300	Adenine-specific DNA methyltransferase
	HPAG1_1313	Type III restriction enzyme M protein
	HPAG1_1393	Type III R-M system modification enzyme
Secondary metabolites biosynthesis, transport and catabolism (Q)	HPAG1_0682	Acetone carboxylase, γ-subunit
	HPAG1_1451	ABC transport system substrate binding protein
Signal transduction mechanisms (T)	**HPAG1_1312**	**Response regulator**
Transcription (K)	HPAG1_1226	Putative transcriptional regulator
Translation, ribosomal structure and biogenesis (J)	HPAG1_0467	Ribosomal protein
	HPAG1_1251	Ribosomal protein
General function prediction only (R)	HPAG1_0212	Cysteine-rich protein
	HPAG1_0296	Aliphatic amidase
	HPAG1_0493	Hypothetical protein
	HPAG1_0745	Hypothetical protein
	HPAG1_0887	Phosphotransacetylase
	HPAG1_1035	Short-chain alcohol dehydrogenase
	HPAG1_1055	Cysteine-rich protein
	HPAG1_1180	Formamidase
	HPAG1_1458	Hypothetical protein
	HPAG1_1468	Hypothetical protein

conserved genes was selected against, and (*iii*) other incoming strains were unable to establish a foothold in the habitat. This picture is one of neutral genetic drift within a highly adapted population in a stable habitat. When the habitat changes, we would expect to see directional selection acting to accumulate mutations and genes that increase fitness (18, 19); this could, potentially, result in loss of diversity as selective sweeps go through the population (20, 21). Thus, searching for adaptive selection and loss of diversity in strains that have survived a transition in gastric pathology would be an excellent method for identifying candidate biomarkers (single nucleotide mutations or entire genes) whose presence predicts pathology. Such biomarkers would be valuable for clinical diagnostics and for understanding the molecular interplay between *H. pylori* and the host that results in pathology. In an extreme case, it may be that progression of gastric pathology allows a transient or cocolonizing strain that is preadapted to the new gastric habitat to totally displace the initial infecting strain. Testing these hypotheses relies critically on a characterization, over time, of the "pan-genome" (22) of the *H. pylori* population that resides within individuals with different (evolving) gastric pathologies.

A finished genome sequence has not been described for a *H. pylori* strain from a patient with ChAG or gastric adenocarcinoma. We have recently used a gnotobiotic transgenic mouse model of ChAG to show that (*i*) loss of acid-producing parietal cells stimu-

Table 1. (continued)

Function	Gene no.	Gene description/annotation
Function unknown (S)	HPAG1_0030	Hypothetical protein
	HPAG1_0401	**Hypothetical protein**
	HPAG1_0419	Hypothetical protein
	HPAG1_0694	Hypothetical protein
	HPAG1_0835	**Hypothetical protein**
	HPAG1_1084	**Hypothetical protein**
	HPAG1_1227	Conserved hypothetical secreted protein
	HPAG1_1410	Hypothetical protein
	HPAG1_1536	Hypothetical protein
Not assignable to COGs		
Outer membrane/envelope structure	HPAG1_0449	Outer membrane protein
	HPAG1_0079	Outer membrane protein
	HPAG1_0009	Outer membrane protein
	HPAG1_0454	Outer membrane protein
	HPAG1_0256	Outer membrane protein
	HPAG1_0255	Outer membrane protein
	HPAG1_0023	Outer membrane protein
	HPAG1_1379	Outer membrane protein
	HPAG1_1448	Outer membrane protein
	HPAG1_0468	Neuraminyllactose-binding hemagglutinin
	HPAG1_1459	Membrane-associated lipoprotein
	HPAG1_0782	Flagellar sheath adhesin
Restriction-modification	HPAG1_0264	Type II DNA modification enzyme
	HPAG1_1315	Type III restriction enzyme R protein
	HPAG1_1299	Type II restriction endonuclease
	HPAG1_1485	Putative type II methylase protein
Metal-binding	HPAG1_1352	Histidine-rich metal-binding polypeptide
	HPAG1_1357	Histidine and glutamine-rich metal-binding protein

These genes were identified by comparing 1,025 ORFs present in all ChAG strains with 1,150 ORFs that are represented in all 56 global *H. pylori* isolates characterized by Gressmann *et al.* (33). ORFs that are present in the former data set and absent from the latter are listed above, and the hypothetical proteins without assignable COGs are listed in Table 7. Because Gressmann *et al*. (33) used DNA microarrays consisting of 26695 and J99 ORFs, we converted the 1,150 ''universally present'' ORFs to HPAG1 gene identifiers to perform the comparison. Among the 1,150 ORFs, 1,146 are represented in the HPAG1 genome.

*Genes in boldface indicate ChAG-associated genes that are also acid-regulated in HPAG1.

lates proliferation of gastric epithelial stem cells, expanding their census in the stomach, and (*ii*) a *H. pylori* isolate (HPAG1) from a patient with ChAG (23) can establish residency in these progenitors (24). Here we present the complete genome sequence of strain HPAG1. In addition to identifying HPAG1 strain-specific genes, we use custom HPAG1 GeneChips containing probesets representing 99.6% of its protein-coding genes to identify acid-regulated genes during growth at pH 5.0 and 7.0 and to genotype five other ChAG strains from the same Swedish case-control study of gastric cancer that yielded HPAG1. We have also genotyped isolates recovered from a patient with ChAG and from the same patient 4 years later, when pathology had progressed to gastric adenocarcinoma. The results provide a genome-wide view of *H. pylori*'s adaptations to ChAG and its sequelae.

Results and Discussion

Features of the HPAG1 Genome. HPAG1 was isolated from an 80-year-old female enrolled in a Swedish case-control study of gastric adenocarcinoma (23) and was subjected to minimal passage before we prepared DNA for whole-genome shotgun sequencing. HPAG1 is a type 1 strain containing two well described virulence factors, *vacA* (genotype s1b/m1) and *cagA*, and is able to induce interleukin-8 production by a human gastric adenocarcinoma cell line (AGS) (25). Studies using germ-free transgenic mice that express a human α-1,3,4 fucosyltransferase in their mucus-producing pit cell lineage and using germ-free transgenic mice with an attenuated diphtheria toxin A fragment (*tox*176)-mediated ablation of parietal cells that results in amplification of gastric stem cells disclosed that HPAG1 binds to two classes of epithelial glycan receptors: Lewis^b, which is produced by gastric surface mucous cells in the majority of humans, and sialyl-Lewis^x, which is synthesized by gastric epithelial progenitors (25).

HPAG1 has a 1,596,366-bp circular chromosome and a single 9,369-bp plasmid, pHPAG1 (Fig. 3 and Table 2, which are published as supporting information on the PNAS web site). The chromosome contains 1,536 predicted protein-coding genes. Forty-three of these genes are either not detectable at all or incompletely represented in the 26695 and J99 genomes (see Table 3, which is published as supporting information on the PNAS web site). Three of the 43 genes (HPAG1_0313, HPAG1_0314, and HPAG1_1485) have BLASTX best hits to members of restriction-modification (R-M) systems from other bacterial species. Components of R-M systems are frequently exchanged between microbes and undergo rapid evolution. Every clinical *H. pylori* isolate is believed to have a unique complement of type II R-M systems, each of which consists of two enzymes, a restriction endonuclease for degrading foreign DNA, and a methyltransferase for protecting endogenous DNA (26).

The predicted products of 15 of the 43 genes have high homology (BLASTX best hit *e* value $< 1e{-}05$) to other proteins reported in various *H. pylori* strains. Two are encoded by ORFs in the HPAG1 cytotoxin-associated gene (*cag*) pathogenicity island (PAI): HPAG1_0496, which is present in many Swedish *H. pylori* isolates (27), and HPAG1_0523, a protein of unknown function first identified in strain NCTC11638 (28). Four others are related to R-M system components. Another, HPAG1_1382, has high homology to CagY but is located outside HPAG1's *cag* PAI. CagY forms part of the core pilus-like structure of the type IV secretion system required for CagA translocation into gastric epithelial cells; introduction of CagA leads to activation of a kinase cascade that produces morphological changes in host cells and induces interleukin-8 (29, 30).

HPAG1 lacks 29 of the 1,408 genes present in both 26695 and J99; 5 of these missing genes are members of R-M systems, 2 are members of the *cag* PAI, 1 is an outer membrane protein, 1 is a putative integrase/recombinase, and 20 have unknown functions. The 26695- or J99-specific genes missing in HPAG1 include transposases (e.g., members of IS605 and IS606) and R-M system-related components, but most do not have assigned functions and are located in their "plasticity zones" (Table 4, which is published as supporting information on the PNAS web site).

The results of our analysis of synteny, evidence for plasmid-mediated gene transfers, plus comparisons of Cluster of Orthologous Groups (COG)- and Kyoto Encyclopedia of Genes and Genomes (KEGG)-based functional annotations of the HPAG1, 26695, and J99 proteomes are presented in *Supporting Text*, Figs. 3 and 4, and Table 5, which are published as supporting information on the PNAS web site.

HPAG1 GeneChip. We designed a custom Affymetrix HPAG1 GeneChip containing probe sets to 1,530 of the strain's predicted 1,536 chromosomal protein-coding genes and 7 of the 8 predicted plasmid-associated genes (average number of perfect-match/single base-mismatch oligonucleotide probe pairs per predicted ORF = 11; see Table 6, which is published as supporting information on the PNAS web site). Our first objective was to perform whole-genome genotyping of additional isolates obtained from Swedish patients

MICROBIOLOGY

with ChAG who were enrolled in the same case-control study that yielded HPAG1 (23) as well as ChAG- and cancer-associated single-colony isolates recovered from a single Swedish patient. This patient had been enrolled in a study whose original purpose was to design a valid esophagogastroduodenoscopy survey for a general adult population (the Kalixanda study; refs. 31 and 32), and had progressed from ChAG to gastric adenocarcinoma during a 4-year interval between endoscopies. We hoped that we could obtain a ChAG gene signature by comparing the genotypes of all these ChAG strains to 56 isolates that had been recovered from individuals living throughout the world, without regard to their gastric pathology (33), and to characterize *H. pylori* genome evolution during the transition from ChAG to gastric adenocarcinoma. Our second objective was to use the GeneChips for whole-genome transcriptional profiling of HPAG1 to verify our gene predictions and to identify genes regulated by exposure to acidic conditions *in vitro*, including those that are associated with ChAG.

Whole-Genome Genotyping of ChAG Strains. We analyzed HPAG1, five other ChAG isolates from the case-control study that had provided us with HPAG1, and two ChAG isolates from the patient in the Kalixanda study obtained before progression to gastric adenocarcinoma. Whole-genome genotyping with our HPAG1 GeneChip revealed 1,025 genes that were present in all of the ChAG strains and 12 genes (encoding hypothetical proteins) that were unique to HPAG1. Components of R-M systems, transposases, and *cag* PAI genes were included among the HPAG1 genes that were variably present in the other seven ChAG strains (see Figs. 5 and 6, which are published as supporting information on the PNAS web site).

The 1,025-member ChAG gene "signature" was further distilled by comparing it to the *H. pylori* "pan-genome" defined by Gressmann *et al.* (33) from their analysis of 56 "global" *H. pylori* strains. Their analysis was based on DNA microarrays containing PCR products from 98% of the ORFs present in the 26695 and J99 genomes.

One hundred and twenty-one of the 1,025 genes represented in all eight of our ChAG strains were not on the list of 1,150 genes identified as being present in all 56 global isolates. These 121 genes, which we defined as "ChAG-associated," are listed in Table 1 and Table 7, which is published as supporting information on the PNAS web site (Table 7 lists genes encoding hypothetical proteins that are not assignable to COGs). The ChAG-associated genes include Hop family members predicted to function as porins and adhesins (such as HopZ; ref. 34), genes involved in forming the molybdenum cofactor for metalloenzymes that participate in metabolism of nitrogen, sulfur, and carbon-containing substrates, as well as genes encoding components of R-M systems.

Another group of ChAG-associated genes is related to utilization of iron and other metals: the ferric uptake regulator (*fur*; HPAG1_0420; refs. 35 and 36), an iron (III) dicitrate transport protein (HPAG1_1469), and a nickel storage protein (histidine-rich metal-binding polypeptide; *hpn*; HPAG1_1352; ref. 37). Acid keeps ferric iron in solution until it reaches sites of absorption in the small intestine (38). Presumably, in the hypo- or achlorhydric ChAG environment, *H. pylori* must cope with a change in ferric iron availability and must compete for metals with members of the intestinal microbiota that are now able to reside in the stomach because of loss of the acid barrier to colonization.

pH-Regulated HPAG1 Genes. To characterize the impact of pH on expression of the "ChAG-associated" and other HPAG1 genes, we conducted whole-genome transcriptional profiling of HPAG1 during *in vitro* growth at pH 5.0 and 7.0. To do so, we grew HPAG1 in liquid culture at pH 7.0 to mid-log phase. The culture was then divided; one half was exposed to pH 5.0, whereas the other half was maintained at pH 7.0. Samples were collected 1 h and 3 h later, and transcriptional profiles from the two pH conditions (at each time point) were compared with each other. The experiment was performed on three separate occasions, providing triplicate data sets for each time point and pH. Combining "present" calls from all experimental conditions, we found that 99.9% of the predicted ORFs in HPAG1 were transcribed, providing validation for our HPAG1 gene annotation.

We used DNA-CHIP ANALYZER (DCHIP) software (39) to identify differentially expressed genes at pH 5.0 compared with pH 7.0 at each of the two time points. A total of 12 genes were defined as up-regulated, and 177 were identified as down-regulated at 1 h and/or 3 h after the shift from pH 7.0 to pH 5.0 (selection criteria: 2-fold change in expression; absolute difference in expression ≥ 100, p value for paired t test ≤ 0.05; 100% "present" call in the up-regulated condition for a given probe set) (see Fig. 1 for a subset of the genes and Table 8, which is published as supporting information on the PNAS web site, for a complete list of these acid-regulated genes).

Transcripts up-regulated at pH 5.0 included the ChAG-associated gene *fur*, two quinone-reactive Ni/Fe hydrogenases, four heat shock-responsive genes/chaperones, and a plasmid-associated gene of unknown function (HPAG1_p006). Genes down-regulated at pH 5.0 include those involved in coenzyme transport and metabolism, especially components engaged in synthesis of molybdenum cofactor (molybdenum covalently bound to molybdopterin), as well as genes that participate in cell wall/membrane biogenesis (e.g., lipid A disaccharide synthetase, type 1 capsular polysaccharide biosynthesis protein J, and a predicted LPS 1,2-glycosyltransferase) (Fig. 1). Two members of the *cag* PAI, VacA and cytotoxin-associated gene PAI protein 4, were represented in this group of gene products down-regulated under more acidic conditions; the latter protein is essential for CagA translocation and interleukin-8 induction in epithelial cells (29).

Given that hypochlorhydria is a major feature of ChAG, it seems reasonable to consider acid-regulated HPAG1 genes that are also components of the ChAG-associated gene signature as important for *H. pylori*'s adaptation/transition to this environment. Genes present in both datasets (shown in boldface in Table 1) include *fur* (HPAG1_0420), iron (III) dicitrate transport protein (HPAG1_1469), and three molybdenum cofactor biosynthesis genes (HPAG1_0783, HPAG1_0784, HPAG1_0785). Molybdenum cofactor-containing bacterial enzymes are involved in a variety of global metabolic reactions important for anaerobic growth. These enzymes include nitrate reductase, formate dehydrogenase, and trimethylamine *N*-oxide reductase (40); their increased expression under pH conditions resembling those encountered in a host with ChAG may help ChAG-associated strains to maintain their representation in a gastric microbiota that now contains intestinal microbes.

Stability of the HPAG1 Genome. We designed an experiment to assess HPAG1's potential for genetic diversification in a gastric ecosystem devoid of acid (and other microbial species) versus one that is fully capable of acidification. To do so, we surveyed HPAG1 strain-specific genes in 40 isolates of HPAG1 retrieved after a 56-week colonization of the stomachs of 12 germ-free *tox*176 transgenic mice with an engineered ablation of their parietal cells (32 isolates), and four normal littermates (8 isolates). *tox*176 animals and their nontransgenic littermates were colonized at 8 weeks of age with a single gavage of a common culture of HPAG1 started from a single colony. This experiment provided a highly controlled test of the effects of acid. (*i*) Animals were housed in a single gnotobiotic isolator but were grouped into cages based on their genotype (all *tox*176 or all normal littermates) to prevent exchange of strains from acid-containing to acid-free stomachs by way of coprophagy. (*ii*) All animals consumed the same autoclaved diet. (*iii*) Surveillance cultures verified that animals from each group were colonized only with *H. pylori* throughout the year-long experiment.

Sequencing of both strands of amplicons, generated by PCR of

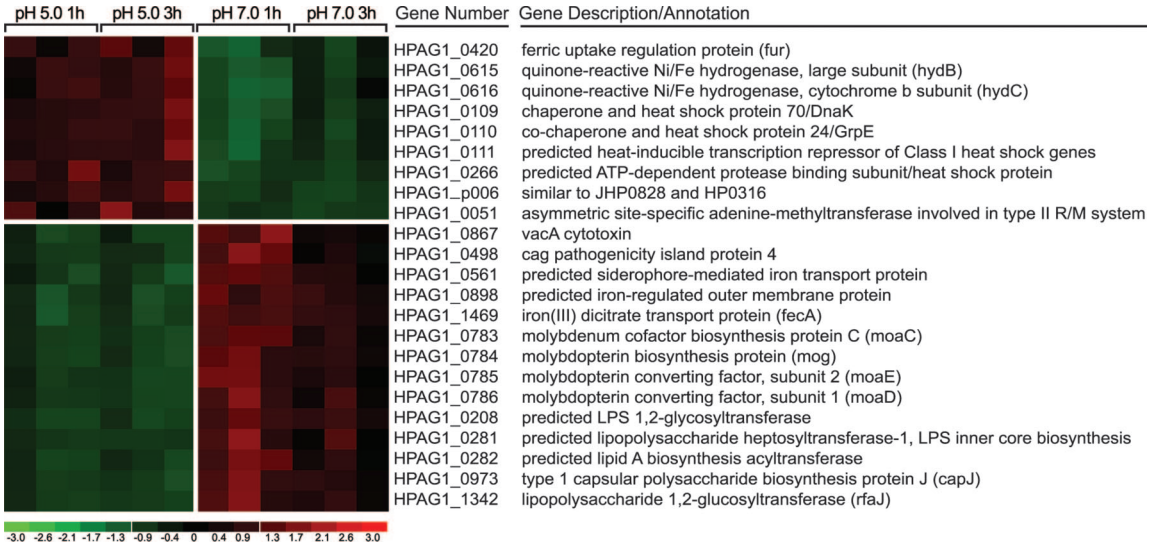

Fig. 1. Acid-regulated genes in HPAG1. Selected transcripts, differentially expressed in mid-log phase cells after incubation at pH 5.0 versus pH 7.0 for 1 and/or 3 h, are shown. Three independent experiments were performed. Numbers at the bottom indicate standard deviations above (red) or below (green) the mean level of expression (black) of a gene.

26 HPAG1 strain-specific genes in the 40 isolates recovered from the two groups of mice, indicated that none of the genes were lost from any isolates and none had nucleotide sequence alterations (note that two of the genes surveyed are also members of the 121-member "ChAG-associated" gene list; see Table 9, which is published as supporting information on the PNAS web site). We randomly selected four of the isolates (from two normal and two *tox*176 mice) for follow-up whole-genome genotyping with HPAG1 GeneChips. Consistent with the PCR data, there was no loss of any HPAG1 chromosomal genes in any of the isolates under these experimental conditions (data not shown).

Genes Associated with Progression from ChAG to Gastric Adenocarcinoma. We next used our HPAG1 GeneChips to compare the genotypes of the two ChAG strains obtained from the Kalixanda study patient before progression to gastric adenocarcinoma with the genotypes of two isolates recovered 4 years later, at the time of diagnosis of gastric cancer. The *cag* PAI genes present in the patient's ChAG isolates were also present in the cancer isolates (see Figs. 5 and 6). The nine genes "gained" in cancer-associated strains (i.e., not detected in the ChAG isolates) included D-alanine: D-alanine ligase A, a metalloprotease, a methionyl-tRNA formyltransferase, and a putative ribonuclease N. Genes involved in DNA repair (uracil-DNA glycosylase, a transcription-repair coupling factor, endonuclease III), and an outer membrane protein (P1) were among the six genes that were "lost" in the cancer-associated isolates (i.e., present in the patient's ChAG isolates) (Fig. 2).

Loss of DNA repair-related genes in these cancer strains suggests an increased propensity for genomic instability. Studies have shown how *H. pylori*'s predisposition for diversity can be used to subjugate and circumvent the host immune response to achieve persistence (41). It is intriguing that a trait that may evolve through adaptive selection or "sweeps" during the transition from ChAG to gastric adenocarcinoma is *H. pylori*'s ability to diversify. Presumably, in the ChAG stomach *H. pylori* can diversify through a number of mechanisms, including loss of DNA repair enzymes and gain of new genes from a gastric "microbiome" (42) that now may be expanded because of the presence of intestinal microbes. This diversification could provide a fitness advantage that allows the organism to adapt more readily to a gastric ecosystem that is changing as a result of host- and microbe-mediated disease progression. Isolates of *H. pylori* that possess relatively stable genomes may represent "dead-ends" in terms of their capacity for promoting further adverse gastric pathology. Looking at an individual's *H. pylori* pan-genome and its potential for diversification (i.e., loss of certain R-M systems, DNA repair enzymes, and recombinases) may be useful as a predictor of future adverse outcomes.

Prospectus. We have also sequenced HPAG1 by using the recently introduced, highly parallel Genome Sequencer 20 System (GS-20) from 454 Life Sciences (Branford, CT) (J.X., E.R.M., and J.I.G, unpublished observations). In a single run of this instrument, 447,626 short-reads were collected (average quality value 20 read length, 106 bp). The "short-read" genome assembly, generated by using the Newbler assembler (43), contained 58 sequence contigs >600 bp, which together total 1,567,482 bp (N50 contig size = 51 kb; N50 contig number = 9). We were able to align 1,561,158 contig bases (99.6% of all of the contig bases) to 1,562,663 finished bases on the chromosome, i.e., the contigs generated by the GS-20 instrument covered 97.9% of the chromosome. The contiguity of

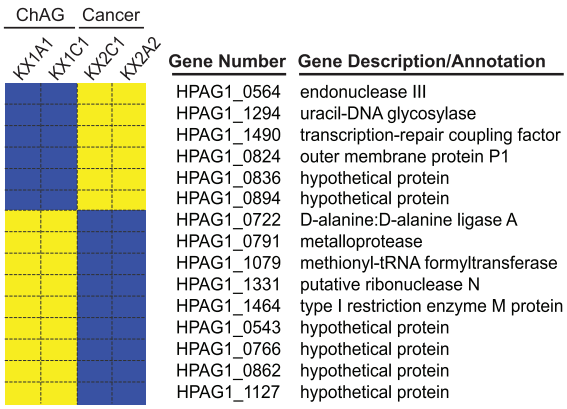

Fig. 2. HPAG1 GeneChip-based genotyping of *H. pylori* strains isolated from a single patient who progressed from ChAG to gastric adenocarcinoma. *H. pylori* isolates were obtained from the antrum (HPAG-KX1A1) and corpus (HPAG-KX1C1) of a patient with ChAG. HPCa-KX2A2 (antrum) and HPCa-KX2C1 (corpus) are isolates obtained from the same patient 4 years later, when ChAG had progressed to gastric adenocarcinoma. Blue and yellow indicate the presence and absence of a gene, respectively.

this assembly was comparable with 8× whole-genome shotgun assembly of reads from an Applied Biosystems 3730Xl capillary sequencer. The overall accuracy in aligned regions was 99.99%.

These results indicate that this instrument provides an opportunity to rapidly sequence multiple individual *H. pylori* isolates (or pooled groups of isolates) from single individuals and to define the organism's pan-genome as a function of the host and his/her evolving pathology without having to rely on DNA microarrays based on a limited number of previously sequenced isolates. The results should provide information about qualitative and quantitative changes in genetic makeup (new mutations, allele frequencies) associated with progression to gastric adenocarcinoma, as well as general insights about *H. pylori* diversity and population structure/fitness. An obvious starting point for such a study would be to obtain deep draft sequences of *H. pylori* isolates recovered at each of the two time points described above for the patient who progressed from ChAG to adenocarcinoma, as well as of isolates from patients who did or did not progress from normal gastric histology to ChAG or who maintained their ChAG status over an extended period.

Materials and Methods

Bacterial Strains. *H. pylori* strains HPAG1 (formerly CAG7:8; 24), HPAG-20:5, HPAG-27:1, HPAG-34:1, HPAG-61:4, and HPAG-72:3 were part of a panel of isolates obtained from an already completed Swedish case-control study of gastric cancer (23). Strains HPAG-KX1A1, HPAG-KX1C1, HPCa-KX2A2, and HPCa-KX2C1 were from the Kalixanda study (31, 32). HPAG-KX1 strains were from an de-identified patient with ChAG (HPAG-KX1A1 from the antrum and HPAG-KX1C1 from the corpus). HPCa-KX2 strains were recovered from the same patient 4 years later after progression to gastric adenocarcinoma (HPCa-KX2A2 from the antrum and HPCa-KX2C1 from the corpus). The Kalixanda study was approved by the Ethics Committee of Umeå University in May, 1998. The histologic criteria used to score the patient's gastric biopsy are described by Storskrubb *et al.* (32). According to then prevailing Swedish medical practices and the Institutional Review Board-approved study protocol, a histologic diagnosis of ChAG at the time of initial esophagogastroduodenoscopy was not considered an indication of *H. pylori* eradication.

All strains were grown under microaerophilic conditions (5% O_2, 10% CO_2, and 85% N_2) for 48–72 h at 37°C on brain–heart infusion agar, supplemented with 10% calf blood, vancomycin (6 μg/ml), trimethoprim (5 μg/ml), and amphotericin B (8 μg/ml). For liquid culture, bacteria were grown under microaerophilic conditions in brain–heart infusion broth supplemented with 5% FCS (Sigma) and 1% IsoVitaleX (Becton Dickinson) (adjusted to pH 7.0).

Sequencing of the HPAG1 Genome. Two whole-genome shotgun libraries were constructed from HPAG1 DNA: (*i*) a plasmid library with an average insert size of 4 kb and (*ii*) a fosmid library with an average insert size of 40 kb. A total of 9.5× Phred quality value 20 (Q20) sequence coverage was obtained with an Applied Biosystems 3730XL capillary machine (7.4× coverage from the plasmid library and 2.1× coverage from the fosmid library). Traditional methods for finishing the genome sequence were used (for details, see *Materials and Methods* in *Supporting Text*).

Details about using HPAG1 GeneChips for whole-genome genotyping and transcriptional profiling can be found in *Supporting Text*.

We thank Maria Karlsson and David O'Donnell for maintaining gnotobiotic mice; Janaki Guruge for assembling *H. pylori* strain panels from colonized gnotobiotic mice; Magnus Bjursell, Peter Turnbaugh, and Douglas Leip for software support; and Laura Kyro for graphics assistance. This work was supported in part by National Institutes of Health Grants DK58529 and DK63483 and the Swedish Cancer Society Grant 4518-B05-06XAC.

1. Brown, L. M. (2000) *Epidemiol. Rev.* **22,** 283–297.
2. Torres, J., Perez-Perez, G., Goodman, K. J., Atherton, J. C., Gold, B. D., Harris, P. R., la Garza, A. M., Guarner, J. & Munoz, O. (2000) *Arch. Med. Res.* **31,** 431–469.
3. Frenck, R. W., Jr., & Clemens, J. (2003) *Microbes Infect.* **5,** 705–713.
4. Mitchell, H. M., Hu, P., Chi, Y., Chen, M. H., Li, Y. Y. & Hazell, S. L. (1998) *Gastroenterology* **114,** 256–261.
5. Blaser, M. J. (1999) *J. Infect. Dis.* **179,** 1523–1530.
6. Ye, W., Held, M., Lagergren, J., Engstrand, L., Blot, W. J., McLaughlin, J. K. & Nyren, O. (2004) *J. Natl. Cancer Inst.* **96,** 388–396.
7. Parkin, D. M. (2001) *Lancet Oncol.* **2,** 533–543.
8. Huang, J. Q., Zheng, G. F., Sumanac, K., Irvine, E. J. & Hunt, R. H. (2003) *Gastroenterology* **125,** 1636–1644.
9. Ohata, H., Kitauchi, S., Yoshimura, N., Mugitani, K., Iwane, M., Nakamura, H., Yoshikawa, A., Yanaoka, K., Arii, K., Tamai, H., *et al.* (2004) *Int. J. Cancer* **109,** 138–143.
10. Lahner, E., Bordi, C., Cattaruzza, M. S., Iannoni, C., Milione, M., Delle Fave, G. & Annibale, B. (2005) *Aliment. Pharmacol. Ther.* **22,** 471–481.
11. Malfertheiner, P., Megraud, F., O'Morain, C., Bell, D., Bianchi Porro, G., Deltenre, M., Forman, D., Gasbarrini, G., Jaup, B., Misiewicz, J. J., *et al.* (1997) *Eur. J. Gastroenterol. Hepatol.* **9,** 1–2.
12. Ley, C., Mohar, A., Guarner, J., Herrera-Goepfert, R., Figueroa, L. S., Halperin, D., Johnstone, I. & Parsonnet, J. (2004) *Cancer Epidemiol. Biomarkers Prev.* **13,** 4–10.
13. Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S. & Peek, R. M., Jr. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 14625–14630.
14. Aras, R. A., Kang, J., Tschumi, A. I., Harasaki, Y. & Blaser, M. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 13579–13584.
15. Blaser, M. J. & Berg, D. E. (2001) *J. Clin. Invest.* **107,** 767–773.
16. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., *et al.* (1997) *Nature* **388,** 539–547.
17. Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., *et al.* (1999) *Nature* **397,** 176–180.
18. Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9721–9726.
19. Cooper, T. F., Rozen, D. E. & Lenski, R. E. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 1072–1077.
20. Novick, A. & Szilard, L. (1950) *Proc. Natl. Acad. Sci. USA* **36,** 708–719.
21. Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989) *Genetics* **123,** 887–899.
22. Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., *et al.* (2005) *Proc. Natl. Acad. Sci. USA* **102,** 13950–13955.
23. Enroth, H., Kraaz, W., Engstrand, L., Nyren, O. & Rohan, T. (2000) *Cancer Epidemiol. Biomarkers Prev.* **9,** 981–985.
24. Oh, J. D., Karam, S. M. & Gordon, J. I. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 5186–5191.
25. Bjorkholm, B. M., Guruge, J. L., Oh, J. D., Syder, A. J., Salama, N., Guillemin, K., Falkow, S., Nilsson, C., Falk, P. G., Engstrand, L. & Gordon, J. I. (2002) *J. Biol. Chem.* **277,** 34191–34197.
26. Xu, Q., Morgan, R. D., Roberts, R. J. & Blaser, M. J. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9671–9676.
27. Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. & Lundeberg, J. (2004) *Gene* **328,** 85–93.
28. Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14648–14653.
29. Fischer, W., Puls, J., Buhrdorf, R., Gebert, B., Odenbreit, S. & Haas, R. (2001) *Mol. Microbiol.* **42,** 1337–1348.
30. Rohde, M., Puls, J., Buhrdorf, R., Fischer, W. & Haas, R. (2003) *Mol. Microbiol.* **49,** 219–234.
31. Aro, P., Ronkainen, J., Storskrubb, T., Bolling-Sternevald, E., Carlsson, R., Johansson, S. E., Vieth, M., Stolte, M., Engstrand, L., Talley, N. J. & Agreus, L. (2004) *Scand. J. Gastroenterol.* **39,** 1280–1288.
32. Storskrubb, T., Aro, P., Ronkainen, J., Vieth, M., Stolte, M., Wreiber, K., Engstrand, L., Nyhlin, H., Bolling-Sternevald, E., Talley, N. J. & Agreus, L. (2005) *Scand. J. Gastroenterol.* **40,** 302–311.
33. Gressmann, H., Linz, B., Ghai, R., Pleissner, K. P., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T. F. & Achtman, M. (2005) *PLoS Genet.* **1,** e43.
34. Peck, B., Ortkamp, M., Diehl, K. D., Hundt, E. & Knapp, B. (1999) *Nucleic Acids Res.* **27,** 3325–3333.
35. Bereswill, S., Greiner, S., van Vliet, A. H., Waidner, B., Fassbinder, F., Schiltz, E., Kusters, J. G. & Kist, M. (2000) *J. Bacteriol.* **182,** 5948–5953.
36. Bijlsma, J. J., Waidner, B., Vliet, A. H., Hughes, N. J., Hag, S., Bereswill, S., Kelly, D. J., Vandenbroucke-Grauls, C. M., Kist, M. & Kusters, J. G. (2002) *Infect. Immun.* **70,** 606–611.
37. Gilbert, J. V., Ramakrishna, J., Sunderman, F. W., Jr., Wright, A. & Plaut, A. G. (1995) *Infect. Immun.* **63,** 2682–2688.
38. Conrad, M. E., Umbreit, J. N. & Moore, E. G. (1999) *Am. J. Med. Sci.* **318,** 213–229.
39. Li, C. & Wong, W. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 31–36.
40. Schwarz, G. (2005) *Cell Mol. Life Sci.* **62,** 2792–2810.
41. Cooke, C. L., Huff, J. L. & Solnick, J. V. (2005) *FEMS Immunol. Med. Microbiol.* **45,** 11–23.
42. Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J. & Relman, D. A. (2006) *Proc. Natl. Acad. Sci. USA* **103,** 732–737.
43. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., *et al.* (2005) *Nature* **437,** 376–380.