Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer

Wolfgang Fischer1,*, Lukas Windhager2, Stefanie Rohrer1, Matthias Zeiller1, Arno Karnholz1, Reinhard Hoffmann3, Ralf Zimmer2 and Rainer Haas1

1Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Ludwig-Maximilians-Universität, Pettenkoferstr. 9a, D-80336 München, 2Institut für Informatik, LFE Bioinformatik, Ludwig-Maximilians-Universität, Amalienstr. 17, D-80333 München and 3Institut für Medizinische Mikrobiologie, Immunologie und Hygiene, Technische Universität, Trogerstr. 30, 81675 München, Germany

Received March 25, 2010; Revised April 23, 2010; Accepted April 26, 2010

ABSTRACT

The availability of multiple bacterial genome sequences has revealed a surprising extent of variability among strains of the same species. The human gastric pathogen *Helicobacter pylori* is known as one of the most genetically diverse species. We have compared the genome sequence of the duodenal ulcer strain P12 and six other *H. pylori* genomes to elucidate the genetic repertoire and genome evolution mechanisms of this species. In agreement with previous findings, we estimate that the core genome comprises about 1200 genes and that *H. pylori* possesses an open pan-genome. Strain-specific genes are preferentially located at potential genome rearrangement sites or in distinct plasticity zones, suggesting two different mechanisms of genome evolution. The P12 genome contains three plasticity zones, two of which encode type IV secretion systems and have typical features of genomic islands. We demonstrate for the first time that one of these islands is capable of self-excision and horizontal transfer by a conjugal process. We also show that excision is mediated by a protein of the XerD family of tyrosine recombinases. Thus, in addition to its natural transformation competence, conjugal transfer of genomic islands has to be considered as an important source of genetic diversity in *H. pylori*.

INTRODUCTION

The evolutionary interplay between pathogenic or commensal microorganisms and their hosts is a constant process of adaptation. Microbial genomes reflect these adaptive processes by continuously developing new variants of host adaptation factors, but also by acquisition and loss of genes via horizontal gene transfer. It has become clear in recent years that the entire gene pool available in a given species (the pan-genome) may be much larger than the genome of a single organism or the core genome common to all isolates of the respective species (1). Thus, genome comparison of many isolates of the same species is required to obtain a picture of the genetic repertoire of this species. The human gastric pathogen *Helicobacter pylori*, a Gram-negative bacterium that colonizes the stomach of half of the world’s population and causes diseases ranging from mild gastritis to gastric cancer (2), is considered as highly adapted to its host, and it has a comparatively small genome (3). *Helicobacter pylori* is usually acquired in childhood by intrafamilial transmission and establishes a lifelong infection in the absence of treatment, although horizontal transmission and co-infection with multiple strains are also possible (4). Due to high mutation rates and frequent recombination events, *H. pylori* is one of the most genetically variable pathogenic bacteria (5). Thus, basically each individual host harbors a different *H. pylori* strain, a fact that has been exploited to define geographically distinct populations and to reconstruct prehistoric and historic human migrations by comparing their *H. pylori* sequence types in housekeeping genes (6,7). This genetic diversity is postulated to be required for persistent colonization of the stomach, where different microenvironments and changing conditions are likely to be encountered (8–10). A comparison between the first two *H. pylori* genome sequences (3,11) suggested that about 7% of all genes are strain-specific, and a strain isolated from a gastric atrophy patient was shown to lack many further genes (12). A microarray-based study with a larger strain collection has suggested that up to 25% of all genes in an individual isolate may be accessory genes (13).
to the limited number of complete H. pylori genome sequences, an estimation of this flexible H. pylori gene pool has not been possible so far.

While it is clear that H. pylori is unusual in its genetic diversity, the mechanisms of horizontal gene transfer leading to this diversity are not well-understood. It is assumed that natural transformation competence plays a major role in genetic exchange, but conjugative processes may also have an important function (14). Many H. pylori strains contain cryptic plasmids, and the plasmid sequences suggest that extensive gene shuffling among plasmids or between plasmids and chromosome occur during their evolution and spread (15). In contrast to other naturally competent bacteria, H. pylori uses a specialized type IV secretion system, the ComB system, for DNA uptake during transformation (16). One additional type IV secretion system that may have a function in horizontal DNA transfer has also been described (17), indicating a preference of Helicobacter for this type of macromolecule transport systems.

We have recently determined the complete genome sequence of H. pylori strain P12, a strain isolated from a duodenal ulcer patient in Germany. Using this sequence and other complete H. pylori genome sequences, we performed a comparative analysis with respect to the flexible gene pool of this species. Most strain-specific genes are either clustered in genome plasticity regions, or located at integration hot spots that are often coincident with genome rearrangement sites. In contrast to all other strains, the P12 genome features three different plasticity zones, two of which are integrated as genomic islands into restriction–modification system pseudogenes. One plasticity zone harbors a complete set of genes encoding a novel type IV secretion system which seems to be involved in horizontal transfer of DNA between H. pylori strains. We show that this plasticity zone represents a genomic island capable of self-transfer by a conjugative mechanism, and that this transfer involves an XerCD-family tyrosine recombinase encoded on the island.

MATERIALS AND METHODS

Bacterial strains and culture conditions

Helicobacter pylori strains were grown on GC agar plates (Difco) supplemented with vitamin mix (1%), horse serum (8%), vancomycin (10 mg/l), trimethoprim (5 mg/l), and nystatin (1 mg/l) (serum plates), and incubated for 16–60 h in a microaerobic atmosphere (85% N₂, 10% CO₂, 5% O₂) at 37°C. Escherichia coli strains were grown on Luria–Bertani (LB) agar plates supplemented with ampicillin (100 mg/l), chloramphenicol (30 mg/l), or kanamycin (40 mg/l), as appropriate.

Whole-genome sequencing, annotation and bioinformatic procedures

The genome of H. pylori strain P12 was sequenced by a whole-genome shotgun method. Bacteria were subjected to minimal passage before DNA was prepared by the cesium chloride method. Two shotgun libraries of 2 kb and of 5 kb were constructed in pSMART-hcKan (Lucigen). High-throughput sequencing of the libraries was performed using ABI 3730 technology to a 5-fold theoretical coverage. Raw sequence data were assembled with the SeqMan program of the LaserGene software package (DNASTAR). Metacontigs obtained by SeqMan were arranged based on BLAST (18) alignments to the genomic sequences of strains 26695, HPAG1 and J99. Gaps were filled by primer walking on shotgun library plasmids or by sequencing PCR fragments. Coding sequences were predicted based on homology to 26695, HPAG1 and J99 proteins as well as EasyGene1.0 (19), GeneMarkS (20) and Reganor (21) predictions. Open reading frames were assumed to be coding if they were predicted by at least two of the three prediction-tools or if their translated sequences showed significant similarities to 26695, HPAG1 or J99 proteins (using tblastn; thresholds: bit-score >100, e-value <10⁻¹⁰, identity >30%, proportion of aligned ORF-sequence >66%) and were at least 40 amino acids long. Start codons (ATG, GTG or TTG) were chosen using majority vote (tool predictions and homology approach); if non-unique, the most upstream start codon was chosen. Gene predictions were partly manually curated. Assignment of protein functions was based on the annotations of H. pylori strain HPAG1, H. acinonychis strain Sheeba as well as PyloriGene annotations of H. pylori strains 26695 and J99 (22), BLASTP searches versus nr- and cdd-databases (thresholds: E-value <10⁻²⁰, identity >35%) as well as manual curation. Groups of orthologous genes were derived based on homology and synteny, allowing identification of orthologs despite possibly low sequence homology, insertion/deletion or recombination events, frameshifts, or premature stops. An orthologous group contains all genes or gene fragments presumably originating from a single ancestor gene, although this ancestor gene may be split in all organisms considered. Dendrograms of VirB4, VirB9 and VirB10 homologs of H. pylori 26695, HPAG1, J99, Shi470, G27 and P12, H. acinonychis Sheeba, and C. jejuni 81-176 (pTet and pVir) were created using ClustalW2 (23). Pairwise distances of homologs were obtained from these dendrograms, and distances between whole type IV secretion systems were calculated as averages of gene-to-gene distances. Type IV systems were clustered using Neighbor Joining. A phylogenetic tree of H. pylori and H. acinonychis strains based on nucleotide similarity of the core genome was calculated as a consensus tree of 1117 independent phylogenetic trees for each orthologous group of genes within the core genome of the strains examined.

Plasmid constructions

All cloning procedures were performed according to standard protocols (24). Deletion and replacement of genes or genomic regions was achieved using the streptomycin susceptibility contraselection strategy (25). An rpsL-erm cassette cloned into the BamHI restriction site of pBluescript II SK+, which was a kind gift of Douglas Berg, was used for all cloning procedures. For the deletion of the tfs3 system, the corresponding upstream and downstream regions were amplified by PCR using primers.
SR17/SR18, and SR30/SR35 (Supplementary Table S7), respectively. Likewise, tfs4 upstream and downstream regions were amplified using primers SR13/SR14 and SR32/SR34, respectively. Upstream and downstream regions were restricted with XhoI/ClaI and NotI/SacI, respectively, and cloned together with or without the rpsL-erm cassette into the corresponding sites of pBluescript II SK+. The plasmids obtained were used in sequential transformations to generate marker-free deletions of the tfs3 and/or tfs4 systems. For reconstitution of frameshifted virB4rfs4 and topA1rfs4 alleles of P12, a fragment spanning both mutations was deleted using a plasmid obtained by PCR amplification of the flanking regions with primer pairs SR48/SR49 and SR50/SR51, respectively, and cloning together with rpsL-erm, as above. In a second step, the rpsL-erm cassette was replaced by a fragment obtained with primers SR48 and SR51 from chromosomal DNA of strain ATCC43526. The correct replacement was confirmed by sequencing. To monitor DNA transfer, donor strains were constructed by transformation of the virB4topA1rfs4-reconstituted P12 strain with plasmids containing either an insertion of a catGC chloramphenicol resistance cassette in the intergenic region between hpp12_453 and hpp12_454, or replacement of the xerD (hpp12_437) or virD4 (hpp12_454) genes with the catGC cassette. These plasmids were generated by inverse PCR using primer pairs WS395/WS396, WS433/WS434, or WS348/WS352 from appropriate plasmids of the 2 kb or 5 kb shotgun plasmids were generated by inverse PCR using primer pairs WS395/WS396, WS433/WS434, or WS348/WS352 from appropriate plasmids of the 2 kb or 5 kb shotgun libraries, and insertion of catGC via BsmHI and SacI.

Bacterial transformation and co-cultivation experiments

Shuttle plasmids and suicide plasmids were introduced into H. pylori strains by conjugation or transformation, as described (26,27). Helicobacter pylori transformants were selected on serum agar plates containing 6 mg/l chloramphenicol, 10 mg/l erythromycin, 500 mg/l streptomycin, or 8 mg/l kanamycin. For co-cultivation experiments, suspensions of donor and recipient strains in Brucella broth were adjusted to a density of 6 x 10^8 cells/ml and preincubated with 0.5 mg/ml DNase I and 0.5 mM MgCl2 for 30 min at 37° C in the presence of 10% CO2. Equal amounts of donor and recipient strains (1.5 x 10^7 cells in 25 µl each) were mixed in 20 mM Tris–HCl, pH 7.5, 1 mM MgCl2 in the presence of DNase I (1 mg/ml), spotted onto serum agar plates and incubated at 37° C, 10% CO2 for 12–16 h. Co-cultivation mixtures were resuspended in Brucella broth and plated on double selective media. To determine live cell counts, serial dilutions were also plated on chloramphenicol and kanamycin-containing plates, respectively. All plates were incubated under standard conditions for 4–6 days.

Microarray hybridization

Genomic DNA of H. pylori strains was isolated using the cesium chloride method. DNA fragmentation, labeling and hybridization to microarray DNA chips was performed according to standard protocols (Affymetrix). Briefly, 5 µg of purified DNA were fragmented by limited DNase I digestion to 50–200 bp products, and fragmentation was confirmed by acrylamide gel electrophoresis. Biotinylation of DNA fragments was performed using the GeneChip DNA labeling reagent and terminal deoxynucleotidyl transferase, and labeling efficiency was monitored by a gel-shift assay after incubation with NeutrAvidin. Logarithmic intensities derived from the microarray hybridizations were used to calculate a matrix indicating the presence or absence of individual genes in each strain. A gene was considered as absent in a strain if the intensity of the corresponding probe set was at least 2 SD values lower than the intensity of the same probe set for P12.

RESULTS

Strain-specific genes and the H. pylori pan-genome

We have determined the complete genome sequence of H. pylori strain P12 (formerly designated 888-0), a type I strain containing a vacA s1/m1 genotype (28) and a functional cag pathogenicity island (29) that was isolated from a duodenal ulcer patient in Germany in 1992 (26). The P12 genome consists of a circular 1673813 bp chromosome (GenBank accession CP001217) and a 10225 bp plasmid termed pHel12 (GenBank accession CP001218). Using an integrated approach of different gene prediction algorithms (see ‘Materials and Methods’ section), 1567 protein-coding genes were predicted on the chromosome, and 11 on the plasmid (Table 1 and Figure 1A).

A comparison of the gene content of P12 with that of other published H. pylori genome sequences (strains 26695, J99, HPAG1, Shi470, G27 and B38) revealed the presence of 1223 genes (orthologous groups; see ‘Materials and Methods’ section for details) common to all H. pylori strains (the core genome; Figure 1B). This value corresponds well to the core genome size estimated recently for five H. pylori genomes (30), but it may drop to about 1100–1150 genes if many more genomes are considered (13). The pan-genome of these seven complete H. pylori genome sequences (the total number of distinct genes/orthologous groups present in these genomes) comprises 2070 genes. Thus, 847 genes are absent from at least one genome and might thus be considered as strain-specific. Two individual strains may differ in as much as 13% of their gene content (Figure 1B), and each strain contains a fraction of 5–12% of all coding sequences that is absent from at least four other H. pylori genomes (Table 1), and a fraction of 2–9% that is unique for this strain. However, since many unique genes have no orthologs in the databases, it is not clear at this point how many of them are in fact pseudogenes. For strain P12, strain-specific genes include genes encoding active components of restriction–modification systems, as described previously (31), as well as plasmid-related genes and genes encoding membrane proteins or ATPases (Supplementary Table S1).

To obtain an estimate of the H. pylori pan-genome, the number of newly identified genes was calculated for the seven complete genome sequences analyzed, and plotted versus the number of genomes, as done previously for other species (1). This resulted in a very slow decrease with increasing number of genomes (Figure 1C),
suggesting that a considerable number of further strain-specific genes are expected to be found upon sequencing of further *H. pylori* genomes, and that *H. pylori* is accordingly a species with an open pan-genome.

**Strain-specific genes localize preferentially to plasticity zones or genome rearrangement sites**

Analysis of the localization of strain-specific genes on the chromosome of strain P12 showed a preference for the replicophile halves surrounding the origin, and a clustering in three large (>10 kb) plasticity zones (PZs) with significantly lower GC content that we refer to as PZ1, PZ2 and PZ3 (Figure 1A and Table 1). Comparison of the overall genome revealed that many of them contain fragments of IS elements or other repetitive sequences, which strongly suggests that both genome rearrangements and insertion of specific genes into these sites depend on recombination events, indicating that rearrangements and gene insertion events might be linked. Therefore we asked whether other hot spots for insertion of strain-specific genes exist and where they are located.

To estimate this, we analyzed 101 genes of strain P12 that are absent from at least four out of the seven complete *H. pylori* genome sequences analyzed. About half of these genes are located in one of the PZs, and the remainder at one of 37 other integration sites (Table 2). Most integration sites outside the PZs contain different genes in different strains and were thus considered as hot spot sites, whereas only 10 sites are unique integration sites where always the same gene is inserted. To examine whether integration sites coincide with potential genome rearrangement sites, we considered all rearrangement sites between the *H. pylori* genomes (Supplementary Table S2) and also between *H. pylori* and *H. acinonychis* (Supplementary Table S3), reasoning that the latter synteny breakpoints might be potential rearrangement sites for *H. pylori* as well. Although insertion sites were as frequent within synteny regions as at synteny breakpoints, the majority of insertion sites at synteny breakpoints (18 out of 21) was classified as insertion hot spots (Table 2). Most integration sites outside the PZs contain different genes in different strains and were thus considered as hot spot sites, whereas only 10 sites are unique integration sites where always the same gene is inserted. To examine whether integration sites coincide with potential genome rearrangement sites, we considered all rearrangement sites between the *H. pylori* genomes (Supplementary Table S2) and also between *H. pylori* and *H. acinonychis* (Supplementary Table S3), reasoning that the latter synteny breakpoints might be potential rearrangement sites for *H. pylori* as well. Although insertion sites were as frequent within synteny regions as at synteny breakpoints, the majority of insertion sites at synteny breakpoints (18 out of 21) was classified as insertion hot spots (Table 2), suggesting that potential genome rearrangement sites represent preferential insertion sites for novel genes. A closer examination of rearrangement sites revealed that many of them contain fragments of IS elements or other repetitive sequences, which strongly suggests that both genome rearrangements and insertion of specific genes into these sites depend on recombination

### Table 1. Comparison of the P12 genome with other complete *H. pylori* genomes

<table>
<thead>
<tr>
<th>Strain</th>
<th>P12</th>
<th>26695</th>
<th>J99</th>
<th>HPAG1</th>
<th>Shi470</th>
<th>G27</th>
<th>B38</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chromosome size (bp)</strong></td>
<td>1673813</td>
<td>1667867</td>
<td>1643831</td>
<td>1596366</td>
<td>1608548</td>
<td>1652982</td>
<td>1576758</td>
</tr>
<tr>
<td><strong>Plasmid size (bp)</strong></td>
<td>10225</td>
<td>–</td>
<td>–</td>
<td>9370</td>
<td>–</td>
<td>10031</td>
<td>–</td>
</tr>
<tr>
<td><strong>GC content (%)</strong></td>
<td>38.8%</td>
<td>38.9%</td>
<td>39.2%</td>
<td>39.1%</td>
<td>38.9%</td>
<td>38.9%</td>
<td>39.2%</td>
</tr>
<tr>
<td><strong>Number of CDS</strong></td>
<td>1567+11</td>
<td>1566c</td>
<td>–</td>
<td>1536+8</td>
<td>1569</td>
<td>1493+11</td>
<td>1528</td>
</tr>
<tr>
<td><strong>Average CDS length</strong></td>
<td>958 bp</td>
<td>954 bp</td>
<td>997 bp</td>
<td>954 bp</td>
<td>913 bp</td>
<td>959 bp</td>
<td>946 bp</td>
</tr>
<tr>
<td><strong>Plasticity zones (PZ)</strong></td>
<td>54+56 (7.0%)</td>
<td>83+48 (8.4%)</td>
<td>33+40 (4.9%)</td>
<td>68+34 (6.6%)</td>
<td>158+36 (12.4%)</td>
<td>45+41 (5.8%)</td>
<td>96+20 (7.6%)</td>
</tr>
<tr>
<td><strong>PZ1</strong> (40.7 kb)</td>
<td>PZ (31.0 kb)</td>
<td>PZ (45.4 kb)</td>
<td>PZ (4.1 kb)</td>
<td>PZ (7.9 kb)</td>
<td>PZ (12.0 kb)</td>
<td>PZ (18.1 kb)</td>
<td></td>
</tr>
<tr>
<td><strong>PZ2</strong> (14.1 kb)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>PZ3</strong> (30.0 kb)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Number of CDS</strong></td>
<td>90</td>
<td>56a</td>
<td>38</td>
<td>6</td>
<td>46</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td><strong>PZ GC content (%)</strong></td>
<td>33.4%b</td>
<td>33.2%</td>
<td>34.1%</td>
<td>n.d.</td>
<td>32.5%c</td>
<td>32.7%d</td>
<td>n.d.</td>
</tr>
<tr>
<td><strong>Cag-PAI (TFS1)</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td><strong>ComB (TFS2)</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>TFS3/TFS4</strong></td>
<td>+/+</td>
<td>+/+</td>
<td>–</td>
<td>–</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
</tr>
</tbody>
</table>

*GenBank accession numbers of chromosome sequences: P12, CP001217; 26695, AE000511; J99, AE001439; HPAG1, CP000241; Shi470, CP001072; G27, CP001173; B38, FM991728.*

*Coding sequences on chromosome + plasmid.*

*Revised annotation according to (22).*

*CDs that are absent from all other six genomes + additional CDS absent from at least four of the other genomes. Numbers in brackets indicate the percentage of these genes in relation to all CDS.*

Parts of a single PZ2 that was split by a genome rearrangement and contains PZ1- and PZ3-like elements.

*PZ2 that contains PZ1- and PZ3-like elements.*

*Excluding IS elements.*

*PZ1, 33.8%; PZ2, 33.9%; PZ3, 32.8%.*

*PZ1-like regions only.*

*Presence or absence of type IV secretion systems (TFS) is indicated by (+) or (–), respectively.*

*Only fragments present.*

---

**Strain-specific genes localize preferentially to plasticity zones or genome rearrangement sites**

Analysis of the localization of strain-specific genes on the chromosome of strain P12 showed a preference for the replicophile halves surrounding the origin, and a clustering in three large (>10 kb) plasticity zones (PZs) with significantly lower GC content that we refer to as PZ1, PZ2 and PZ3 (Figure 1A and Table 1). Comparison of the overall genome revealed that many of them contain fragments of IS elements or other repetitive sequences, which strongly suggests that both genome rearrangements and insertion of specific genes into these sites depend on recombination
Figure 1. Strain-specific genes in the *H. pylori* P12 genome. (A) Circular representation of the genome. Genes predicted on the plus and minus strands are shown as bars on the outer circles. Third and fourth circles: Positions of strain-specific genes (being absent from at least three out of six complete genome sequences). Fifth and sixth circles: GC content and GC skew calculated with a window size of 10,000 and steps of 100 bp. Positions of the putative origin and terminus of replication, the *cag* pathogenicity island (cagPAI), the *comB* genes and the plasticity zones (PZ1-3) with their corresponding type IV secretion systems (tfs) are indicated. (B) Venn diagrams showing numbers of common and strain-specific genes for the complete genome sequences indicated. (C) Power law regression fit for the identification of new genes in *H. pylori* genomes [according to (1)]. The *x*-axis indicates the number *N* of complete genome sequences examined, and the *y*-axis the number *n* of newly identified genes in each genome. The straight green line, representing a regression fit of the mean numbers of *n*, indicates a power law progression (*n* ∼ *N*^−α^), and a power law coefficient (*α* = 0.339) < 1 (dashed line) indicates that the rate of newly identified genes is decreasing very slowly and that *H. pylori* has thus an open pan-genome.
Plasticity zones contain distinct type IV secretion systems

All *H. pylori* strains have PZs of different sizes in a defined location on the chromosome (11,32) (Supplementary Figure S4; Table 1), but P12 is the only strain known so far that contains two additional PZs. Both PZ1 and PZ3 are inserted into non-functional restriction–modification system genes (Figure 2A and Supplementary Figure S5), and both harbor complete type IV secretion systems and genes encoding coupling protein homologs, relaxases, and several other DNA-processing enzymes, such as topoisomerases, helicases and XerCD-family tyrosine recombinases (Figure 2A and Supplementary Tables S4 and S5). As noted previously (32), different PZ1-like elements contain both conserved and variable regions that suggest a composition of different modules (Figure 2A). Since the PZ3 type IV secretion system is very similar to the TFS3 type IV secretion system described previously (17,32), whereas the PZ1 type IV secretion system shows much less similarity, we refer to these systems as TFS3 and TFS4, respectively. To determine relationships between the different type IV secretion systems, we constructed multiple sequence alignments and neighbor joining trees of the corresponding coupling protein homologs, the relaxase homologs, and all secretion apparatus components. In all cases, the TFS3 and the TFS4 representatives clustered in distinct groups, and the TFS3 components were more closely related to the corresponding ComB system components than to the TFS4 components (data not shown). To confirm this finding, we constructed a phylogenetic tree based on the VirB4, VirB9 and VirB10 homologs of type IV secretion systems in *Helicobacter* and *Campylobacter* including the different TFS4 variants (Figure 2B). Again, we found that the TFS3 system is more closely related to the ComB system than to the TFS4 system, and all three systems (ComB, TFS3 and TFS4) are more closely related to the *Campylobacter* pVir type IV secretion system than to the Cag type IV secretion system, indicating that the Cag system has a completely different evolutionary origin. Taken together, these data suggest that the TFS3 and TFS4 systems are distinct and independent secretion systems.

Since fragments of PZ1- or PZ3-like elements are also present in the PZs of strains 26695 and J99 (Figure 2A), we tested the frequency and integrity of PZs in a set of 37 *H. pylori* strains including clinical isolates and reference strains (Supplementary Table S6). Southern blot hybridizations were positive for a TFS3 probe in 24% and for a TFS4 probe in 30% of the strains. In contrast, probes from other PZ regions hybridized with higher percentages (Supplementary Table S6), suggesting that genes encoding type IV secretion components are more often subject to loss after integration of a PZ. To find out more precisely whether type IV secretion system genes are collectively lost and which genes are retained in the PZs, we designed custom DNA microarray chips containing all genes present in the three PZs of strain P12. Using these microarray chips, we performed whole-genome genotyping with a set of representative *H. pylori* strains, including two reference strains, mouse- and Mongolian gerbil-adapted strains and several clinical isolates. Consistent with previous studies, many genes in the PZs are variably present (Figure 3). Although complete TFS3 and/or TFS4 type IV secretion systems were found in some strains (ATCC43526, B128), partial absence of type IV secretion system genes seems to be more common. This may indicate that PZ1 and PZ3 genes other than secretion apparatus genes provide stronger selective advantages. However, we cannot exclude that strains lacking parts of TFS4 may be complemented by the corresponding parts from either TFS3 (or a divergent TFS4 variant) to obtain functional type IV secretion systems. Furthermore, frameshift mutations in the *virB4*<sub>TFS4</sub> and *topA*<sub>TFS4</sub> genes of different strains (Figure 2A) suggest that even complete systems are frequently non-functional.

**Plasticity zone 1 is a transferable genomic island**

The integration of PZ1 and PZ3 into restriction–modification system genes or other genomic regions suggested that these regions are in fact genomic islands or conjugative transposons that may be transferred as independent entities (32). Sequence analysis of PZ1 borders in P12 showed that eight nucleotides of the hp464 gene ortholog (AAAGAATG) are duplicated at the left and right borders of the integration site (Supplementary Figure S6), suggesting that these elements might be integrated by site-specific recombination. Therefore, we asked whether PZ1 or parts of it can be transferred between *H. pylori* strains. To examine this, we constructed suitable donor and recipient strains and performed co-cultivation experiments. Since the *virB4*<sub>TFS4</sub> and *topA*<sub>TFS4</sub> genes in strain P12 contain frameshift mutations that are likely to render both proteins nonfunctional, we first reconstituted both genes in P12. Sequence analysis of the *virB4*<sub>TFS4</sub> and *topA*<sub>TFS4</sub> genes of strains B128 and ATCC43526 revealed no corresponding frameshifts in these strains (GenBank accession ABSY0100034.1, position 4607–6517 and data not shown). Therefore, we constructed a *virB4*<sub>TFS4</sub>-reconstituted P12 strain in which the region containing the frameshifts was replaced with a corresponding DNA fragment of strain ATCC43526, using a marker-free gene

---

**Table 2. Analysis of integration sites of 101 strain-specific genes in strain P12**

<table>
<thead>
<tr>
<th></th>
<th>No. of genes</th>
<th>No. of sites</th>
<th>Unique sites&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Hot spots&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntenic regions&lt;sup&gt;c&lt;/sup&gt;</td>
<td>20</td>
<td>16</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Synteny breakpoints&lt;sup&gt;d&lt;/sup&gt;</td>
<td>32</td>
<td>21</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Plasticity zones</td>
<td>49</td>
<td>3</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>40</td>
<td>13</td>
<td>27</td>
</tr>
</tbody>
</table>

<sup>a</sup>Insertion sites containing either a certain specific gene or no insertion at all.

<sup>b</sup>Insertion sites containing different strain-specific genes in different strains.

<sup>c</sup>Insertion sites within regions of gene synteny.

<sup>d</sup>Insertion sites at breakpoints of gene synteny between *H. pylori* strains or between *H. pylori* and *H. acinonychis.*
Figure 2. Type IV secretion systems in plasticity zones. (A) Gene arrangement in PZ1 and comparison with corresponding regions in other genome sequences. PZ1 of strain P12 is inserted into a restriction–modification system pseudogene (similar to gene hp464 in strain 26695). Genes encoding type IV secretion system components are indicated as black arrows, and frameshift mutations are indicated by asterisks. Genes encoding proteins with 90–95% sequence similarity to the P12 proteins are shown in full colour, genes encoding proteins with 50–75% sequence similarity are hatched. Note that tfs4 has been termed tfs3a (for strains Shi470 and G27) or tfs3b (for strain P12) previously (32). PeCan18B plasticity zone, GenBank accession AF487344.3. (B) Neighbor-joining tree showing relationships between PZ type IV secretion systems and other Helicobacter and Campylobacter type IV secretion systems, based on average distances of the corresponding VirB4, VirB9 and VirB10 homologs. The TFS4 systems of strains P12 (tfs4-HPP12), G27 (tfs4-G27) and Shi470 (tfs4-HPSH) are depicted individually to show their mutual relationships. pTet, C. jejuni 81–176 pTet plasmid; pVir, C. jejuni 81–176 pVir plasmid.
deletion/replacement procedure which leaves the genetic environment intact (25). Using the same technique, we constructed isogenic \textit{tfs3} and \textit{tfs4} deletion mutants lacking the complete respective type IV secretion gene clusters, as well as double mutants. Respective experiments with these reconstituted and mutated strains showed that neither the \textit{tfs3} nor the \textit{tfs4} system is involved in DNA uptake by natural transformation or in type IV translocation of the CagA protein to AGS cells (data not shown).

Next we constructed a \textit{virB4/topA}_{TFS4}-reconstituted P12 strain containing a chloramphenicol resistance gene cassette in a PZ1 intergenic region and additionally a \textit{recA} deletion to prevent backward transformation by recipient strain DNA (Figure 4A), and used this strain as a donor strain in co-cultivation experiments. As recipients, we used either wild-type or \textit{tfs4}-deleted P12 strains with kanamycin resistance markers. Co-cultivation experiments were performed in the presence of DNase to select for conjugative transfer and against transformation events. Transconjugant clones selected by double resistance to chloramphenicol and kanamycin were obtained at a frequency of \(4.2 \times 10^{-7}\) from wild-type recipients (Figure 4B). When a \(\Delta tfs4\) strain was used as a recipient, transconjugants were obtained at a similar average rate (\(5.5 \times 10^{-7}\)). These data are indicative of a conjugative transfer process, but they do not discriminate between transfer of small DNA fragments surrounding the resistance marker, or of the whole PZ. To obtain evidence for the transfer of the whole island or at least of the \textit{tfs4} genes, transconjugant clones were further analyzed by PCR and sequencing to determine their \textit{virB4/topA} alleles and thus whether a co-transfer of the reconstituted \textit{virB4} allele had occurred. Wild-type recipients had acquired the reconstituted \textit{virB4} allele together with the \textit{cat} resistance marker in 64\% of the cases examined, and \(\Delta tfs4\) recipients in all cases (Figure 4C). This shows that transconjugants resulted from transfer of large genome fragments (>13 kb), even in cases where integration of the \textit{cat} resistance marker would have been possible via homologous recombination of small fragments. Thus, a fragmented or deleted PZ1 type IV secretion system can easily be reconstituted by horizontal gene transfer from \textit{tfs4}-positive donor strains.
PZ1 uses a XerCD-like tyrosine recombinase as genomic island excisionase

Having established that PZ1 genes can be transferred between \textit{H. pylori} strains, we next examined whether the PZ1-encoded type IV secretion system is involved in this process. We constructed isogenic \textit{virD4} mutants in the \textit{virB4/topA}$_{\text{TFS4}}$-reconstituted P12 strain, provided them with a \textit{recA} mutation and used them as donors for co-cultivation experiments, as before. In comparison to the experiments with wild-type donors, the rate of
double-resistant clones decreased to $1.7 \times 10^{-7}$ when wild-type recipients were used, and to $0.9 \times 10^{-8}$ when $\Delta tfs4$ recipients were used (Figure 4B). Despite these lower transfer rates, some wild-type recipients and all $\Delta tfs4$ recipients had acquired the donor $virB4$ allele (Figure 4C), suggesting that PZ1 transfer was still possible in the absence of the $virD4$ gene $hpp12\_437$.

Since PZ1 contains a gene encoding an XerC/XerD-like tyrosine recombinase, which is unlikely to be involved in resolution of chromosome dimers (32,33), we further asked whether site-specific excision activity of this protein may play a role in PZ transfer. To examine this, we constructed a donor strain with an isogenic deletion of the PZ1 $xerD$ homolog $hpp12\_437$. When this strain was used in co-cultivation experiments, we still obtained double-resistant clones, but sequencing showed that the reconstituted $virB4$ allele had not been transferred in any case (Figure 4C). Thus, transfer of the $\Delta xerD::cat$ mutation did not involve horizontal transfer of the entire PZ, which suggests that PZ transfer is dependent on the activity of the PZ1-encoded XerD-like recombinase.

To further corroborate the conclusion that XerD mediates excision from the host chromosome and thus enables PZ self-transfer, we sought to obtain direct evidence for the excision event. Therefore, we prepared chromosomal DNA from donor strains and from co-cultivation mixtures of donor and recipient strains, and used different primer sets targeting PZ1 left and right junctions to amplify circular PZ intermediates or empty-site fragments by PCR. Circular intermediates were detected from both wild-type and $\Delta tfs4$ recipients and all $\Delta tfs4$ recipients had acquired the donor $virB4$ allele (Figure 4C), suggesting that PZ1 transfer was still possible in the absence of the $virD4$ gene $hpp12\_437$.

SEQUENCE TYPES OF HOUSEKEEPING GENES HAVE BEEN USED TO GROUP $H.\ pylori$ STRAINS INTO GEOGRAPHICALLY DISTINCT POPULATIONS AND TO TRACE HUMAN MIGRATIONS (6,7). THE UNDERLYING GENETIC DRIFT IS SUPERIMPOSED BY ADDITIONAL DIVERSITY CAUSED BY HORIZONTAL GENE TRANSFER EVENTS (5). OUR DATA INDICATE THAT GEOGRAPHICAL DISTANCE IS NEITHER REFLECTED IN THE EQUIPMENT WITH STRAIN-SPECIFIC GENES NOR IN OVERALL GENE SYNTENY. THIS CONFIRMS THE PREVIOUS NOTATION THAT COMMON POOLS OF FLEXIBLE GENES HAVE BEEN MAINTAINED IN DIFFERENT POPULATIONS INDEPENDENTLY (13). THE OBSERVATION THAT SOME GENOME REARRANGEMENTS ARE INCONGRUENT WITH STRAIN HISTORIES SUGGESTS THAT REARRANGEMENTS PARTLY TAKE PLACE IN A NON-RANDOM MANNER, WHICH HAS ALSO BEEN FOUND FOR OTHER BACTERIAL GENOMES (39). INTERESTINGLY, POTENTIAL SYNTENY BREAKPOINTS SEEM TO ACT FREQUENTLY AS HOT SPOTS FOR INTEGRATION OF STRAIN-SPECIFIC GENES, INDICATING THAT GENOME REARRANGEMENTS AND INSERTION OF SPECIFIC GENES RELY ON SIMILAR MECHANISMS. LARGE GENOME REARRANGEMENTS OFTEN OCCUR IN A SYMMETRIC FASHION WITH RESPECT TO THE REPLICATION ORIGIN AND TERMINUS, SINCE THEY ARE UNDER SELECTION FOR REPlicHORE BALANCE (40). IT HAS BEEN SUGGESTED THAT THESE REARRANGEMENTS OCCUR BY REPLICATION-DEPENDENT RECOMBINATION (41). GENERALLY, SPECIES SUCH AS $H.\ pylori$ THAT HAVE A HIGH DENSITY OF REPEATS ARE SUBJECT TO MORE FREQUENT REARRANGEMENTS (42). SINCE VESTIGES OF PREVIOUS IS ELEMENT INSERTIONS OR OTHER REPETITIVE DNA SEQUENCES ARE FREQUENTLY FOUND AT SYNTENY BREAKPOINTS AND INTEGRATION HOT SPOTS, IT IS VERY LIKELY THAT RECOMBINATION BETWEEN SUCH SEQUENCES PLAYS A MAJOR ROLE FOR REARRANGEMENTS AND GENE INTEGRATION. WHILE RECOMBINATION EVENTS AT SEQUENCE REPEATS HAVE PREVIOUSLY BEEN ASSOCIATED WITH INTRAGENOMIC DELETION EVENTS (43), WE PROPOSE THAT THEY PLAY A MAJOR ROLE FOR GENOME REARRANGEMENTS AND INSERTION EVENTS OF STRAIN-SPECIFIC GENES, AND FOR $H.\ pylori$ GENOME EVOLUTION.

DISCUSSION

The $H.\ pylori$ pan-genome and genome evolution

With a rapidly increasing number of bacterial genome sequences available, it has become evident that genetic variability and horizontal gene transfer are important evolutionary requirements (34). Bacterial genomes are subject to constant evolution involving gene gain, gene loss, or genome rearrangements, and different strains are often equipped with strain-specific genes that may provide them with an adaptive potential while maintaining large pan-genome pools (1). The ability to adapt to host populations is particularly important for pathogenic microorganisms such as $H.\ pylori$, where genes under positive selection have indeed been identified in certain geographic groups (35). Although $H.\ pylori$ is considered as a species well-adapted to its niche in the human host, our data indicate that it possesses an open pan-genome, as observed for other bacterial species (36,37). A recent estimation of the $H.\ pylori$ pan-genome size suggested the presence of 4000–5000 genes, which would exceed the size of the core genome by about 4-fold (38). This underscores the requirement for maintaining a diverse gene pool, possibly to colonize different human populations or to adapt to changing environments within a single host (9).

Imported DNA fragments carrying strain-specific genes might contain sequences similar to the IS or other repetitive elements that may be used for rearrangements. Alternatively, those strain-specific genes that are consistently integrated in defined locations might be acquired occasionally, as suggested for the minimal mobile elements in Neisseria species (44). In both cases, flanking sequences would not need to be very long, since they might be processed by the action of restriction enzymes, which have been suspected to be involved in determining crossover endpoints (45). Thus, active restriction–modification systems may modulate integration of foreign plasmid or chromosomal DNA (45–47), and non-functional systems may facilitate the acquisition of further strain-specific genes or gene clusters, such as the genomic islands identified here.
Plasticity zones as transferable genomic islands

While there is no evidence for correlations among individual sets of strain-specific genes, it is obvious that PZ genes and other strain-specific genes define two independent flexible gene pools. This might also reflect two separate pathways of horizontal gene transfer. Whereas PZs of extant genomes consistently contain complete or partial type IV secretion systems and are thus likely to have been spread by conjugative processes, single strain-specific genes or small gene groups are most probably transferred by transformation, which is considered as the main mechanism of horizontal gene transfer in H. pylori. Based on the comparison of sequential isolates from the same patients, it has been estimated that natural transformation and recombination are able to replace large portions of the H. pylori genome in very short time intervals, with a mean size of recombined fragments of roughly 400 bp (48). In contrast, transformation experiments in vitro suggest that DNA fragments integrated by homologous recombination have a somewhat larger average size (1300–4000 bp) and may be interrupted by short fragments of recipient DNA (45,49). It has been suggested that this relatively short average fragment length might be due to restriction enzyme activity. Thus, transfer of larger genomic fragments containing functional modules such as complete type IV secretion systems would require a different transformation-independent and restriction-resistant mechanism.

Although DNase-resistant transfer from H. pylori donor cells has been shown for RP4 origin-containing plasmids (50) and for antibiotic resistance-conferring point mutations (51,52), conjugation mechanisms in H. pylori are still only poorly understood (14). We show here for the first time that transfer and integration of large PZs is possible by a DNase-resistant process. Since plasmid DNA uptake during conjugation of H. pylori depends on the ComB transformation system (S. Rohrer, W. Fischer and R. Haas, unpublished data), we cannot completely rule out that PZ transfer occurs by a transformation-like process as well. In any case, it is likely that this process also occurs in vivo, as both loss and acquisition of PZ genes have been reported for sequential isolates from the same patients (9,53). Although the GC content of the PZs is considerably below the genome average, the broad distribution of PZ genes in the Helicobacter population (including H. acinonychis) suggests that these genes have been present in the Helicobacter gene pool for a considerable time, notably longer than the cag pathogenicity island (13). Therefore, their low GC content does not necessarily reflect recent acquisition from a distinct species, but might rather be indicative of a general horizontal gene transfer capability (54).

Although the PZs are genomic elements that are transferable as a whole, both our microarray data and those previously published (13) show that they often contain only subsets of the genes present on complete islands. Subsequent gene transfer and rearrangement events are obviously able to reshape these regions considerably. Although the islands might frequently lose their transfer capability due to frameshift mutations or gene deletions within the type IV secretion system clusters, our results also demonstrate that functional systems can easily be reconstituted after horizontal gene transfer from a PZ-positive donor strain. Thus, the PZs seem to form a distinct flexible gene pool that is maintained within a population and distributed by constant horizontal gene transfer, which also suggests that PZ genes provide a selective advantage to the recipient bacteria.

Site-specific integration of PZs may involve inverted repeats at their junction sites (32), and the duplicated target sequence motifs probably represent coupling sequences of the putative att sites (55). PZ1 has previously been considered as a transposable element and termed TnPZ (32), but mobilization within a strain has not been shown so far. Mobilization from donor to recipient strains, as shown here, is rather reminiscent of genomic island transfer, although there is no clear-cut distinction between genome islands and (conjugative) transposons. In most proteobacteria, integrases encoded on genomic islands cluster in a phylogenetic group (IntG) that is distinct from integrases of phage, conjugative transposon or ICE (integrative conjugative element) origin (56). Interestingly, only e-proteobacteria seem to lack such genomic island integrases. We show here that in H. pylori, PZ excision is performed by a XerCD-family tyrosine recombinase [previously termed XerT (32)] instead. Usually, XerC and XerD recombinases catalyze resolution of chromosome dimers at dif sites (57), a function which is probably mediated by the chromosomal XerH recombinase in H. pylori (33). Besides their role in dimer resolution, XerC and XerD proteins have also been described to control integration of foreign DNA. They are, for instance, involved in integration of the CTX phage into the Vibrio cholerae dif site (58), or of the gonococcal genetic island into the Neisseria gonorrhoeae genome (59). However, in these cases the mobile elements take advantage of the corresponding host Xer recombinases to accomplish integration into chromosomal dif sites. Although Xer family recombinases have been found on other genome islands (60), our data show for the first time that Xer family tyrosine recombinases are exploited for horizontal transfer of genomic islands.

Taken together, our findings underscore the importance of horizontal gene transfer for the maintenance of genome plasticity. Different H. pylori strains are not only highly variable in terms of sequence microdiversity, but they contain also individual sets of strain-specific genes that are subject to transfer either by natural transformation, or by a novel conjugative transfer process involving XerD-family proteins as site-specific excisionases. We speculate that these different mechanisms enable H. pylori to adapt to different host organisms (during transmission) or to spatially and temporally varying niches within a single host, and thus provide the basis for the persistence of infection.

ACCESSION NUMBERS
CP001217, CP001218.
SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Evelyn Weiss and Silvia Weidner for excellent technical assistance, to Uladzimir Antonenka for helpful discussions, to Luisa F. Jiménez Soto for critical reading of the manuscript, and to Douglas Berg for the gift of the rpsL-erm contraselection system.

FUNDING

Deutsche Forschungsgemeinschaft (HA2697/10-1) and from ERA-NET PathoGenoMics (0313930D to R.H.). Funding for open access charge: Deutsche Forschungsgemeinschaft (HA2697/10-1) and from ERA-NET PathoGenoMics (0313930D to R.H.).

Conflict of interest statement. None declared.

REFERENCES

13. Soto for critical reading of the manuscript, and to Douglas Berg for the gift of the rpsL-erm contraselection system.
