A Second-generation Protein–Protein Interaction Network of Helicobacter pylori*§

Roman Häuser†, Arnaud Ceol§¶, Seesandra V. Rajagopala∥, Roberto Mosca§, Gabriella Siszler†, Nadja Wermke†, Patricia Sikorski∥∥, Frank Schwarz‡, Matthias Schick†, Stefan Wuchty‡‡, Patrick Aloy***, and Peter Uetz§§¶¶

Helicobacter pylori infections cause gastric ulcers and play a major role in the development of gastric cancer. In 2001, the first protein interactome was published for this species, revealing over 1500 binary protein interactions resulting from 261 yeast two-hybrid screens. Here we roughly double the number of previously published interactions using an ORFeome-based, proteome-wide yeast two-hybrid screening strategy. We identified a total of 1515 protein–protein interactions, of which 1461 are new. The integration of all the interactions reported in 2001 results in 3004 unique interactions that connect about 70% of its proteome. Excluding interactions of promiscuous proteins we derived from our new data a core network consisting of 908 interactions. We compared our data set to several other bacterial interactomes and experimentally benchmarked the conservation of interactions using 365 protein pairs (interologs) of E. coli of which one third turned out to be conserved in both species. Molecular & Cellular Proteomics 13: 10.1074/mcp.O113.033571, 1318–1329, 2014.

Helicobacter pylori is a Gram-negative, microaerophilic bacterium that colonizes the stomach, an unusual highly acidic niche for microorganisms. In 1983, Warren and Marshall found it to be associated with gastric inflammation and duodenal ulcer disease (1, 2). A chronic infection with H. pylori can lead to development of stomach carcinoma and MALT lymphoma (reviewed in (3)). Hence, the World Health Organization has classified H. pylori as a class I carcinogen (4). It is estimated that half of the world’s population harbors H. pylori but with large variations in the geographical and socioeconomic distribution while causing annually 700,000 deaths worldwide (reviewed in (3)).

The pathogenesis of H. pylori has been extensively studied, including the effector CagA, cytotoxin VacA, its adhesins and urease (reviewed in (3, 5–7)). The latter allows the bacterium to neutralize the stomach acid through ammonia production. However, H. pylori is not a classical model organism and thus many gaps in our knowledge still exist.

The genome of H. pylori reference strain 26695 was completely sequenced in 1997 (8) and encodes 1587 proteins of which about 950 (61%) have been assigned functions (excluding “putatives”; Uniprot, CMR (9)). These numbers indicate that a large fraction of the proteins of H. pylori has not been functionally characterized.

Protein–protein interactions (PPIs)1 are required for nearly all biological processes. Unbiased interactomes are helpful to understand proteins or pathways and how they are linking poorly or uncharacterized proteins via their interactions. For instance, our study of the Treponema pallidum interactome (10) has led to the characterization of several previously “unknown” proteins such as YbeB, a ribosomal silencing factor (11), or TP0658, a regulator of flagellar translation and assembly (12, 13). However, only a few other comprehensive bacterial interactome studies have been published to date, including Campylobacter jejuni (14), Synechocystis sp. (15), Mycobacterium tuberculosis (16), Mesorhizobium loti (17), and recently Escherichia coli (18). In addition, partial interactomes are available for Bacillus subtilis (19) and H. pylori (20). Most of them used the yeast two-hybrid (Y2H) screening technology.

1 The abbreviations used are: PPI, protein–protein interaction; 3-AT, 3-amino-1,2,4-triazole; AD, activation domain; DBD, DNA-binding domain; ORF, open reading frame; PIM, protein interaction map; PRS, positive reference set; TAP-MS, Tandem affinity purification–mass spectrometry; Y2H, Yeast two-hybrid.

© 2014 by The American Society for Biochemistry and Molecular Biology, Inc. This paper is available online at http://www.mcponline.org

From the †German Cancer Research Center (Deutsches Krebsforschungszentrum), Technologiepark 3, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany; §Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), c/Baldiri Reixac 10-12, 08028 Barcelona, Spain; ¶Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia (IIT), Via Adamello 16, 20139 Milan, Italy; ‡J. Craig Venter Institute (JCVI), 9704 Medical Center Drive, Rockville, Maryland 20850, USA; **Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain; ‖Department of Computer Science, University of Miami, Coral Gables, Florida 33146, USA; §§Proteros Biostructures, Am Klopferspitz 19, D-82152 Martinsried, Germany; ¶¶Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23284, USA

Received August 16, 2013, and in revised form, March 11, 2014 Published, MCP Papers in Press, March 13, 2014 DOI 10.1074/mcp.O113.033571

Author contributions: R.H., P.A., and P.U. designed research; R.H., S.V.R., G.S., N.W., P.S., and M.S. performed research; M.S. contributed new reagents or analytic tools; R.H., A.C., S.V.R., R.M., F.S., S.W., P.A., and P.U. analyzed data; R.H., S.W., P.A., and P.U. wrote the paper.
(21) which allows the pairwise detection of PPIs. Furthermore, a few other studies (22–25) systematically identified protein complexes and their compositions in bacteria. In 2001, Rain and colleagues have established a partial interactome of *H. pylori*, the first published protein interaction network of a bacterium (20). In this study, 261 bait constructs were screened against a random prey pool library resulting in the detection of over 1500 PPIs. Although this network likely represents a small fraction of all PPIs that occur in *H. pylori*, many downstream studies were motivated by these results (see below).

Recent studies have disproved the notion that Y2H data sets are of poor quality (26, 27). Similarly, a high false-negative rate can be avoided by multiple Y2H expression vector systems (28–30) or protein fragments as opposed to full-length constructs (31). The aim of this study was to systematically screen the *H. pylori* proteome for binary protein interactions using a complementary approach to that of Rain et al. to produce an extended protein–protein interaction map of *H. pylori*. As a result, we have roughly doubled the number of known binary protein–protein interactions for *H. pylori* in this study.

**EXPERIMENTAL PROCEDURES**

**Cloning, Yeast Two-hybrid Screens, and Pairwise Tests—**

**Cloning**—Full-length ORFs of *H. pylori* were shuttled from pENTR221 entry clones (PFGRC, JCVI, Rockville, MD, USA) into Y2H plasmids pGBK7T7g, pGADT7g (32, 33), pGBK9-GW, pDEST32, and pDEST22 (Invitrogen, Carlsbad, CA) using Gateway® cloning (see supplemental Table S1 for details). For interolog tests *E. coli* ORFs (32) were shuttled into plasmids pGBK7T7g, pGADT7g, and pGBKClagt (18). Clones were checked for correctness by PCR.

Individual bait plasmids were transformed into haploid yeast strain CG-1945 and prey plasmids into Y187 (Clontech, Mountain View, CA) as described (34).

The prey library was created by growing all plasmid strains of the *H. pylori* entry clone library individually in selective LB medium, followed by pooling and plasmids isolation. The resulting entry clone pool was shuttled into the prey plasmids pGADT7 and pDEST22 using Gateway cloning (see supplemental Table S1 for details). For interolog tests *E. coli* ORFs (32) were shuttled into plasmids pGBK7T7g, pGADT7g, and pGBKClagt (18). Clones were checked for correctness by PCR.

**Y2H Screens**—Yeast baits and prey libraries were grown and mated as described in (36) with the following adjustments: for liquid mating corresponding volumes of OD600 = 2 of each individual bait strain and the prey pool were mixed. Selection of positive diploids was carried out on plates (15 cm diameter) containing agar medium with S.D. medium (MP Biomedicals, Solon, OH) without the amino acids Leu, Trp, and His supplemented with 0.1 mM 3-Amino-1,2,4-triazole (3-AT). The screening plates were incubated for 3–5 d at 30 °C. To check the mating efficiency, a 1:10,000 dilution was plated on -Leu-Trp S.D. agar in parallel to the screens and the number of diploid colonies was determined. A screen was repeated if the number of colonies was <200,000. For auto-activating baits, screens were repeated on 1 and 10 mM 3-AT.

Y2H positive preys were identified by colony PCR after Zymolyase (amsbio) treatment using BIOTAQ™ Red DNA Polymerase (BIOLINE) following enzymatic purification as described (37). PCR products were verified by agarose gel electrophoresis and analyzed by Sanger sequencing (GATC, Köln, Germany). The identities of the sequences were identified by blastn analysis. The sequences were blasted against a sequence database with ORF sequences of *H. pylori* strain 26695 and J99 and *E. coli* K12 (control).

**Retests/interlogs Pairwise Tests**—Bait and prey pairs were re-tested individually and tested against a self-activation negative control (empty prey plasmid) as described (38).

**Data Set Definitions and Analyses—**

**PRS**—A set of literature-curated interactions of *H. pylori* are used here as positive reference set (PRS) (supplemental Table S3). Small-scale studies and protein structures were considered only.

**Definition of the Proteomes**—The proteome of *H. pylori* is defined as the complete proteome provided by Uniprot (39) (strain ATCC 700392/26695) complemented with proteins from strain J99 (Fig. 1A).

**Protein Interaction Networks from Literature**—The literature protein networks and large scale studies in other species used for the network comparisons have been compiled from BIND (40), BioGRID (41), DIP (42), INTACT (43), MINT (44), and MPIDB (45) (releases available in May 2012) following the method described in (46).

**Network Topology**—The networks derived in this and previously published studies were analyzed using Cytoscape 2.8.3 and plugins (47, 48).

**Connectivity Between Protein Functions**—We used the cellular subrole definition from the Comprehensive Microbial Resource (CMR) web pages (49). We calculated the number of direct connections between all annotations in the *H. pylori* combined core PIM (including both homo- and heteromers) and in 10,000 networks with randomized nodes. For clarity, we filtered out annotations with only one interacting protein and pairs linked by <3 PPIs. In addition, we removed annotations such as “Hypothetical proteins/Conserved” and “Unknown function/General.” Finally, we calculated the z-score for each pair of shared annotation by comparing the number of links in the experimental network to the average number of links in the random networks (10).

**Semantic Similarity of GO Terms**—We measured semantic similarity of GO terms using the R package GOSemSim (50). Applying the mgsim function we calculated the maximum similarity over all pairs of GO terms of interacting proteins as described (51). Homomers were excluded.

**Interactome Conservation**—We compared the *H. pylori* interactome to published binary interactomes as well as protein complexes (supplemental Table S4). For complexes we used the list of binary interactions inferred by the authors. We predicted orthologous interactions (“interologs”) by using the Inparanoid algorithm (52) but ignored all interologs with an Inparanoid score < 0.6 (53).

**Interologs Prediction and Benchmarking**—In order to collect potential interologs in *E. coli* we combined the predictions of four resources: Blast (54) best reciprocal hits, Inparanoid (52) (version 4.1 using BLOSUM45 as suggested for prokaryotes by the authors), OMA (55) (downloaded March 2012) and Roundup v2.0 (56). We identified the best reciprocal Blast hits with a sequence coverage of at least 50% and an e-value < 10⁻⁴. We defined three groups of orthologs based on the in-paranoid score: high (score > 1), medium (score > 0.6), and low (score < 0.6). For calculation of the tools’ sensitivity and specificity, see supplemental Table S7.

**Sequence Conservation Properties of Interologs**—We compared the identity between the interacting proteins in two species, their similarity, the evolutionary distance, and the e-value returned by a blast search. The similarity and identity of interologs were calculated using EMBoss (57). The distance was calculated with the EMBoss
The Helicobacter pylori Interactome

Fig. 1. Y2H screening strategy and coverage. A, Y2H screening strategy. For details, see main text. B, Screen coverage. The whole area represents all possible bait-prey test combinations in the H. pylori 26695 proteome. Green: experimentally tested combinations in the 2µ plasmid system; pink: untested combinations (28.4%).

RESULTS AND DISCUSSION

Generation of a Comprehensive Binary Interactome of H. pylori—Initially we screened a subset of baits against two well-defined prey libraries containing 1343 individual ORFs of H. pylori 26695 and another 280 ORFs of strain J99 using both 2µ- and CEN-based vectors (Fig. 1A). We have shown previously that different vectors detect different PPIs (28, 30) and thus increase the sensitivity. The two types of vector produce N-terminal DBD- or AD-fusions (i.e. DNA-binding domain and activation domain, respectively, of the GAL4 transcription factor individually fused to the N termini of the recombinant proteins). Screening of 221 2µ-based baits and 212 CEN-based baits (with 210 overlaps) resulted in 244 and 59 unique PPIs, respectively, with 22 interactions found in both. A total of 105 (47.5%) 2µ-baits turned out to be productive (i.e. at least one PPI was detected per bait) in contrast to CEN-based baits with 57 (26.9%) productive baits. The 2µ-based system detected 84.8% of PPIs (28, 30) hence we used the 2 µ-based approach for all remaining screens.

We successfully screened a total of 1329 bait strains corresponding to 99.0% of ORF clones available or 83.7% of all ORFs in the H. pylori 26695 genome (Fig. 1B). Because the H. pylori 26695 ORFeome collection lacked 16.3% of ORFs our screens covered 70.9% of all 2.52 × 10^6 possible bait-prey combinations corresponding to 1.26 × 10^6 possible interactions (excluding bait-prey redundancy) of which we have tested 0.903 × 10^6 (71.62%).

Overall we have collected ~14,000 positive yeast colonies (Fig. 1A) of which 6011 were sequenced. 5484 (91.2%) prey sequences of good quality could be assigned to a bait protein. This produced 2154 bait-prey pairs including 1151 pairs that were found multiple times and classified as reproducible bait-prey pairs (category A). By contrast, for 1003 pairs we obtained only one positive prey sequence (category B single hit). To ensure reproducibility, we retested these cases in 969 pairwise Y2H retest assays and confirmed two thirds of them. Finally, we combined the bait-prey pairs of category A and the positively verified PPIs of category B as positives (supplemental Table S1). Removing redundancies, bait-prey reciprocals, and overlaps from the different vector systems resulted in 1669 unique PPIs (supplemental Table S2). After removing putative pseudogenes (no Uniprot ID) 1515 PPIs remained.
This interaction data set connects 739 proteins (47.6% of the proteome) and we named it *H. pylori* protein interaction map 2 (PIM.2_all) following the terminology used by Rain and colleagues (“PIM.1”) (20).

**Data Filtering**—To derive a high-quality core network we excluded all PPIs involving promiscuous preys with a prey count (i.e., the number of all baits that detected a certain prey) >15 (Figs. 1A and 3B). This resulted in a core network (PIM.2_core) of 908 PPIs connecting 596 proteins (38.4% of the proteome, Fig. 2A). Such a specificity filtering strategy is commonly used in interactomic surveys (10, 15, 17, 20). The threshold was chosen because (1) nearly all interactions from a positive reference set mapped exclusively to PPIs with low prey counts and (2) a conspicuous increase of the prey count was observed >15 (Fig. 3B).

**Benchmarking**—To benchmark the screening sensitivity of our interactome data we identified published binary interactions of *H. pylori* excluding the data from Rain et al. We collected a total of 137 PPIs from small-scale studies and 3D protein structures (supplemental Table S3) and used this data as a positive reference set (PRS). We tested in our screens 108 of the 137 PRS interaction combinations and confirmed 26 PPIs (24.07%) as part of PIM.2_all and another 25 PPIs (23.15%) as part of PIM.2_core (Fig. 3A). The detection sen-
Fig. 3. Data benchmarks and functional groups. A, Screening sensitivity: detected interactions in PIM.2 (this study) and PIM.1 (20) when compared to interactions known from the *H. pylori* literature (PRS, supplemental Table S3); “direct” PPIs with proteins that are known to bind directly to each other, “3D” PPIs from three-dimensional protein structures, “Combined” either direct PPI or from structure. B, All PPIs of PIM.2 sorted by prey count. Overlaps with the PRS are highlighted by blue crosses. Except one interaction all of them map to low prey counts, that is, these PPIs are specific. The prey count threshold (≤15) to propose a core data set is indicated by dashed arrows. C, Enrichment of PPIs among functional protein groups in the combined core PIM (CMR cellular subrole (9)). The z-score compares the number of PPIs to the average number of links in randomized networks. Functional groups are given as nodes and the interactions as edges (color indicates the z-score). Numbers reflect the absolute numbers of interactions within a functional group or connecting different functional groups, respectively. D, Semantic similarity between GO terms of interacting proteins of *H. pylori*, *E. coli*, and yeast, focusing on biological processes, molecular function and cellular components. The interactomes of all three species are similar.
sitivity is similar to other studies (26, 28). Notably, we lost only one interaction known from the literature (UreE-UreE) by applying the prey count filtering strategy (Fig. 3B).

Interestingly, 50.77% (33 out of 65 tested) of the PRS interactions had been already found by Rain et al. (PIM.1_all). In fact, 19 out of 28 (67.9%) PRS PPIs appear to have been studied because of their detection in the PIM.1 (we looked for citations of Rain et al. in these studies). We conclude that the original PPI data provided by Rain and colleagues motivated many scientists to carry out downstream experiments and this also explains the remarkable high true-positive fraction of PPIs present in PIM.1.

There is no absolute measure to determine false positive interactions even though several attempts have been made to estimate such rates, e.g. using so-called “random reference sets” (26, 28) or other methods (61, 62). The false positive fraction was determined in these studies to be less than 5% (26) and 6.5% (28), respectively, depending on stringency (i.e. low 3-Amino-1,2,4-Triazole = 3-AT concentration). Given that we use very similar protocols, we estimate that our false positive rate is in the same ballpark although it might be higher for promiscuous proteins.

For further analyses we have combined both data sets and named the combined networks PIM_all (whole data of PIM.1 and PIM.2) and PIM_core (interactions in PIM.1 with PBS category A and B and these from PIM.2 with a prey count ≤ 15).

**Interactome Properties and Comparison—** When we compared our interactions with those of Rain et al. only 54 interactions overlapped (about 3.5% of either PIM.2_all or PIM.1_all, respectively). The limited overlap is likely because of the variation in the Y2H system, the different sets of baits, and the property of the prey clones used (31). However, this demonstrates that we have clearly enlarged the *H. pylori* PIM. Overlapping interactions in PIM.1_core and PIM.2_core were overrepresented by 7.9 and 3.8-fold, respectively, when compared with the remaining interactions. PIM.2_all and PIM.2_core connect 47.6% and 38.4% of the *H. pylori* proteome, compared with PIM.1_all and PIM.1_core which connect 47.7% and 32.8%, respectively (supplemental Table S4).

It is notable that Rain and colleagues obtained such a high connectivity of the proteome while screening only 261 baits. When both *H. pylori* PIMs are combined the proteome connectivity increases to 71.5% (all data) and 56.2% (core subsets combined). Only the raw interaction networks of *C. jejuni* (79.1%) as well as *M. tuberculosis* (72.8%) exhibit a slightly higher proteome connectivity. As other interactomes the *H. pylori* network follows a power law distribution (Fig. 2B).

**The Protein Function Network—** We wondered whether the *H. pylori* core network is enriched for PPIs among certain functional categories (using the CMR cellular subrole (9)). We found an enrichment for proteins belonging to chemotaxis/motility, central intermediary metabolism, or cell division process, but also for many others that are assigned to the same cellular role (Fig. 3C). This likely reflects the structure of certain enzyme complexes or macromolecular entities. The most highly interconnected group is “chemotaxis/motility” with 26 interactions that encompass large parts of the motility apparatus of *H. pylori* confirming the specificity of these PPIs. In addition, the chemotaxis and motility proteins serve as “functional hub” in the PIM, similar to proteins involved in “regulation,” “cell envelope,” or “protein synthesis.” The high connectedness suggests cross-regulation of these processes.

Functional similarity of interacting proteins can also be used to assess the quality of interaction data. As benchmark data, we collected 9875 large-scale yeast Y2H PPIs from HINT (63) most of which were confirmed by more than one publication. Furthermore, we used a recently published set of 2231 *E. coli* Y2H PPIs (18). In Fig. 3D we calculated the semantic similarities (50, 51) of all protein interactions in *S. cerevisiae, E. coli*, and *H. pylori* using GO terms of biological process, molecular function and cellular component. Specifically, we obtained values between 0 and 1, where values close to 1 indicate highest similarity of GO terms and vice versa. Similar results suggest that the *H. pylori* interactomics data is of similar quality compared with yeast and *E. coli*. Interestingly, most interactions show highest semantic similarity using cellular component terms, supporting the plausibility of our *H. pylori* PPIs although the two combined *H. pylori* data sets did not exhibit any notable differences (Fig. 3D, supplemental Fig. S1).

**Conserved Interactions of the *H. pylori* PIM—**

**Comparison of Interactomes—** To systematically explore the conservation of interactions in different bacteria we first predicted interologs from bacterial interactomes and compared the resulting numbers (Fig. 4). Although many interologs can be predicted (Fig. 4A) experimentally confirmed overlaps are small and rarely exceed a few dozen interactions or 13% (*H. pylori versus B. subtilis* and *Synechocystis* (Fig. 4B, 4C). As expected, the absolute number of conserved PPIs is highest between the close relatives *C. jejuni* and *H. pylori* (77 interactions). On average only 2.61% of PPIs seem to be conserved between available interactomes with overlaps ranging from 0 to ~15%. Reasons for that might be methodic variations or the fact that few interactions are conserved.

Thus, we have selected an unbiased set of interologs for experimental validation (see below). A more detailed comparison of *H. pylori* PIM data can be found in Fig. 4D.

**Experimental Benchmarking of Interologs and Prediction Tools—** In 2001 Matthews and colleagues tested which PPIs are conserved between yeast and *C. elegans* proteins (64). Because these authors used identical protocols to test homologous interactions, experimental variation was minimized. They found that six of the 19 (31%) tested interactions were conserved. In another study it was shown that 49 out of 173 motility-related PPIs (28.3%) are reproducible in *C. jejuni* when Y2H interactions from *T. pallidum* screens were tested (65).
To further validate our *H. pylori* interaction data and to get a more global number of conserved interactions we predicted from our new *H. pylori* interaction map (PIM.2_all) interologous pairs in *E. coli*. We only used ORFeome constructs and thus did not consider the data from Rain and colleagues who used prey random fragments. A total of 246 interacting *H. pylori* protein pairs predicted 375 *E. coli* interologs (see below). We constructed these Y2H clones using the *E. coli* ORFeome clone library (32) and tested all the pair-wise combinations of the predicted interactions by Y2H assays. To enhance assay sensitivity the Y2H screening was conducted using both N-terminal and C-terminal fusion bait proteins with an estimated assay sensitivity of ~35% (28). By this method, we identified 65 positive (out of 313 tested) *E. coli* interaction pairs (20.8%) that correspond to 27.3% of the source PPIs in *H. pylori* (Fig. 5A, 5B, for details see supplemental Table S5 and S6). In addition, we retrieved 29 confirmed *E. coli* interologs from interaction databases. That is, we found 25.9% and 32.4% of PPIs to be conserved in *E. coli* and *H. pylori*, respectively.

For the interolog prediction we applied different tools. This was done to check whether a certain tool predicts conserved interactions more reliably compared with the outcome of the experiments (discussed above). We couldn’t find any striking difference in sensitivity or specificity of the used prediction tools—the percentage of experimental verified interactions is very similar—but they predict different numbers and the subsets of truly conserved interactions differ (Fig. 5C, supplemental Table S7). We conclude that several tools predict interologs with similar sensitivity and thus may be used on their own although their combination allows prediction and experimental detection of the highest number of possible interologs.

Properties of Conserved PPIs—One would expect that highly conserved protein pairs are more likely to interact than less conserved pairs. To test this hypothesis we compared the joint values (the square root of the product of the similarity for each pair of orthologs) as described in (58) for sequence identity, similarity, evolutionary distance, and the blast e-value (Fig. 5D, supplemental Fig. S2) but did not observe a statistically significant difference between confirmed and nonconfirmed interologs. Although the number of confirmed interologs increases with the joint sequence identity (Fig. 5C, supplemental Table S7), the increase is not as high as previously shown by Yu et al. who suggested that interologs can be transferred reliably.
when a pair of proteins has a joint sequence identity >80%. In our data set no pair of proteins has such a high joint sequence identity. A significant difference was observed when we took into account the minimum (for sequence identity and similarity) and maximum (for e-value and distance) values for each pair (supplemental Fig. S2A–S2D, p values < 0.025) but there is not a clear threshold to separate conserved and nonconserved interactions. There seems to be a slight tendency of higher conserved protein pairs to bind. Because we could not define clear thresholds we suggest that not the overall se-

![Fig. 5](image-url) Conservation test of interologs and conserved subnetworks. A, Summary of conservation tests carried out in this study by pairwise Y2H assays using protein pairs of E. coli and H. pylori PIM.2_all as prediction source. In addition, interologs from PPI databases and the literature are compared. Detailed results are given in supplemental Table S5 and S6. B, Number of interactions tested as positive with E. coli pairs by different Y2H vectors. Note that both N- and C-terminal fusions were used as indicated, while H. pylori screens only used N-terminal fusions. C, Numbers of E. coli interologs predicted (first number) and verified (second number) by various prediction methods. The confirmation rate is given in percent. For details see supplemental Table S7. D, Protein pairs of the experimentally confirmed and nonconfirmed interactions from our pairwise E. coli Y2H tests and literature overlaps are plotted against their sequence identities (%). Each dot represents an interolog in E. coli and the axis represents the identity of each protein with its ortholog in H. pylori. Solid, red circles show positively verified interactions, blue crosses are not confirmed ones. See also supplemental Fig. S2. E, Comparison of the joint sequence identities versus predicted and confirmed interologs. See text for details. F, Example of a conserved subnetwork enriched for “signal transduction” and chemotaxis proteins (TlpA, CheW, CheA, CheY). “Red PPIs” were identified in H. pylori, E. coli, and C. jejuni and represent a part of the well understood chemotaxis pathway. “Blue PPIs” were found in H. pylori and E. coli, only, and involve the unexpected enzymes Mda66 (MdaB, quinone reductase), PncC (NMN amidohydrolase), and YhdJ (DNA adenine methyltransferase).
sequence divergence but rather specific binding epitopes are important for the conservation of PPIs (64). Based on our validated Helicobacter - E. coli interologs, we conclude that at least one third of all protein interactions are conserved between these two clades. Interestingly, this number is similar to the results obtained in the aforementioned studies (64, 65) but more data would be required to predict interologs more reliably. Our data suggest that PPIs evolve fast or that only local protein features need to be maintained such as domains or epitopes, a conclusion that was also drawn by Matthews et al. (64). The gain or loss of an interaction is certainly a fundamental driving force of evolution and it remains to be seen which of the two thirds of seemingly nonconserved interactions in our data set reflects a gain or loss of function.

The Urease Interaction Map—Interactome data should extend our understanding of many aspects of H. pylori biology and pathogenesis. For instance, urease produces ammonia from urea that neutralizes the stomach acid - it ensures H. pylori’s survival in the host (reviewed in (3)). The system is composed of the core components UreAB and the accessory maturation factors UreEFGH proteins that mediate the assembly and incorporation of nickel ions into the core complex. In addition, the acid activated UreI channel arranges urea import. Factors involved in nickel uptake (NixA) and storage (histidine rich proteins like HspA, Hpn), gene regulation (Fur, NikR), and nitrogen metabolism (amidases such as AmiE and AmiF, arginase RocF, and glutamate synthase) help to regulate levels of nickel and urea. Moreover, HypA and HypB are involved in the nickel insertion step and are required for full urease activity (66).

Stingl and colleagues have isolated eight protein complexes associated with urease components (67). Based on the known PPIs we investigated which binary interactions explain the complexes topology. Although there is evidence for direct physical interactions between UreA-UreB, UreF-UreH, and UreG-UreE (Fig. 6, supplemental Table S3) we did not find any accessory proteins in the interaction data sets except...
UreA-UreH. Nonetheless, we could link UreA to IspDF (a bifunctional enzyme involved in isoprene biosynthesis and part of the Stingl complexes II and III). Similarly, UreF binds to cysteine desulfurase IscS (complex I). NrdA (Ribonucleotide reductase subunit α) interacts with UreB in complex I and VI, and UreE interacts with HP1111 (ferrodoxin oxidoreductase β subunit) and NapA (Na+/H+ antiporter) in complex I. UreE turned out to be a highly promiscuous prey in our study and thus the latter two interactions were not included in the core data set of PIM.2. Nevertheless, a combination of complex purification and Y2H does support an interaction of these proteins.

Furthermore we checked if the complex components as well as other interacting proteins are co-regulated on the gene expression level. The urease core genes are known to comprise elevated expression levels under acidic growth conditions (then when urease is needed for acid neutralization) (68–75). In fact, a vast majority of the linked proteins is co-up-regulated suggesting similarly increased protein levels (Fig. 6). This would also ensure increased ammonia production in a concerted fashion.

The urease linkage map reveals several unexpected PPIs among the urease core components. Stingl et al. (67) already observed some unexpected proteins in their complexes, for example, cell division factors (MinD, Tig, FtsA, Fig. 6). Another interesting example is the PPI between UreH and CheY (chemotaxis response regulator) which suggests that chemotaxis and the urease process may be connected, for example, because motility depends on the proton motive force that is generated by a proton gradient. In fact, urease positive strains show increased chemotactic response to urea in contrast to urease negative strains that have normal flagella but exhibit reduced swarming (76). Although the urease has been studied already intensively the comparison of its unbiased interactions suggests many other proteins to be associated with the system.

CONCLUSIONS

We have identified 1451 novel PPIs in the H. pylori proteome and extended the protein–protein interaction map significantly using comprehensive Y2H screens. Our data will help to reveal novel important aspects of H. pylori biology. Recent studies have shown that many factors determine the outcome of a Y2H screen (such as the location of fused domains, plasmid copy number etc.) (28, 30, 31), so there is little doubt that the H. pylori interactome is still incomplete. Although an estimated third of all PPIs can be extrapolated to more distantly related species, the specifics of such interolog prediction still need to be worked out. Our approach and data set should provide a useful contribution to the understanding of protein–protein interactions and their evolution.

Acknowledgments—We thank the Pathogen Functional Genomics Resource Center (PFGRC, managed and funded by the Division of Microbiology and Infectious Diseases, NIAID, NIH, DHHS, used to be operated by the JCVI) at the J. Craig Venter Institute (JCVI) for providing the Gateway®-based pENTR221 entry clone library of Helicobacter pylori (now available from BEI Resources, Manassas, VA).

The protein interactions from this publication have been submitted to the IMEx (http://www.imexconsortium.org) consortium through IntAct (77) and assigned the identifier IM-22256.

* The project was funded by the European Union FP-7 grant HEALTH-F3-2009-223101 “AntiPathoGN” and NIH grant RO1GM79710.

This article contains supplemental Figs. S1 and S2 and Tables S1 to S7.

To whom correspondence should be addressed: Center for the Study of Biological Complexity, Virginia Commonwealth University, 1015 Floyd Ave, Richmond, VA 23284. Tel.: 1-804-827-4573; E-mail: peter@uetz.us.

REFERENCES

1. Warren, J. R., and Marshall, B. (1983) Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet 1, 1273–1275

2. Marshall, B. J., and Warren, J. R. (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1, 1311–1315

3. Kusters, J. G., van Vliet, A. H., and Kuipers, E. J. (2006) Pathogenesis of Helicobacter pylori infection. Clin Microbiol. Rev. 19, 449–490

4. (1994) Schistosomoses, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7–14 June 1994. IARC monographs on the evaluation of carcinogenic risks to humans / World Health Organization, International Agency for Research on Cancer 61, 1–241

5. Bauer, B., and Meyer, T. F. (2011) The human gastric pathogen Helicobacter pylori and its association with gastric cancer and ulcer disease. Ulcers 2011, doi:10.1155/2011/340157

6. Romero-Adrian, T. B., Leal-Montiel, J., Monsalve-Castillo, F., Menguial-Moreno, E., McGregor, E. G., Perini, L., and Antunez, A. (2010) Helicobacter pylori: Bacterial factors and the role of cytokines in the immune response. Curr. Microbiol. 60, 143–155

7. Lertsetthakarn, P., Ottmann, K. M., and Hendrixson, D. R. (2011) Motility and chemotaxis in Campylobacter and Helicobacter. Annu. Rev. Microbiol. 65, 389–410

8. Tomb, J. F., White, O., Kerlavagne, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalal, H. G., Glodek, A., McKenney, K., Fitzgerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cowton, M. D., Weidman, J. M., Fuji, C., Bowman, C., Wathney, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature 388, 539–547

9. Davidsen, T., Beck, E., Ganapathy, A., Montgomery, R., Zafar, N., Yang, Q., Madupu, R., Goetz, P., Galsinsky, K., White, O., and Sutton, G. (2010) The comprehensive microbial resource. Nucleic Acids Res. 38, D340–D345

10. Titz, B., Rajagopala, S. V., Goll, J., Hauser, R., McKevitt, M. T., Pakliti, K., and Uetz, P. (2008) The binary protein interactome of Treponema pallidum—the syphilis spirochete. PLoS One 3, e2292

11. Hauser, R., Pech, M., Kijek, J., Yamamoto, H., Titz, B., Naevé, F., Towch, T., Carmele, J., de, McGovern, E. G., Perini, L., and Antunez, A. (2010) Helicobacter pylori: Bacterial factors and the role of cytokines in the immune response. Curr. Microbiol. 60, 143–155

12. Tomb, J. F., White, O., Kerlavagne, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalal, H. G., Glodek, A., McKenney, K., Fitzgerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cowton, M. D., Weidman, J. M., Fuji, C., Bowman, C., Wathney, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature 388, 539–547

13. Mukherjee, S., Yakhnin, H., Kysela, D., Sokoloski, J., Babitzke, P., and Kearns, D. B. (2011) CsrA-FliW interaction governs flagellin homeostasis and a checkpoint on flagellar morphogenesis in Bacillus subtilis. Mol. Microbiol. 82, 447–461

14. Parrish, J. R., Yu, J., Liu, G., Hines, J. A., Chan, J. E., Mangiola, B. A., Zhang, H., Pacifico, S., Fotouhi, F., DiRita, V. J., Ideker, T., Andrews, P., and Finley, R. L., Jr. (2007) A proteome-wide protein interaction map for Campylobacter jejuni. Genome Biol. 8, R130

The Helicobacter pylori Interactome

Molecular & Cellular Proteomics 13.5
15. Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007) A large-scale protein-protein interaction analysis in Synecocystis sp. PCC6803. DNA Res. 14, 207–216

16. Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., Zhang, L., Gao, C., He, Y., Li, Y., Huang, F., Zeng, J., Huang, C., Yang, Q., Tian, Y., Zhao, C., Chen, H., Zhang, H., and He, Z. G. (2010) Global protein-protein interaction network in the human pathogen Mycobacterium tuberculosis H37Rv. J. Proteome Res. 9, 6665–6677

17. Shimoda, Y., Shinpo, S., Kohara, M., Nakamura, Y., Tabata, S., and Sato, S. (2008) A large scale analysis of protein-protein interactions in the nitrogen-fixing bacterium Mesorhizobium loti. DNA Res. 15, 13–23

18. Rajagopala, S. V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S. B., Phanse, S., Ceol, A., Hauser, R., Sizsler, G., Wuchty, S., Emili, A., Babu, M., Aloy, P., Piiper, R., and Uetz, P. (2009) Improving the yeast two-hybrid system with permutated fusions: proteins the Varicella Zoster Virus interactome. Proteome Sci. 8, 8

19. Rajagopala, S. V., Hughes, K. T., and Uetz, P. (2009) Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. Proteomics 9, 5296–5302

20. Boxem, M., Maigia, Z., Kiltgord, N., Li, N., Lemmens, I., Mana, M., de Lichtenveld, L., Mul, J. D., van de Peut, D., Devos, M., Simonis, N., Yildirim, M. A., Kokol, M., Kao, H. L., de Smet, A. S., Wang, H., Schlatz, A. L., Hao, T., Millestein, S., Fan, C., Tipsword, M., Drew, K., Galli, M., Rhrsorsrakkai, K., Drchsel, D., Koller, D., Roth, F. P., Iakoucheva, L. M., Dunker, A. K., Bonneau, R., Gunsalus, K. C., Hill, D. E., Piana, F., Tavernier, J., van den Heuvel, S., Hyman, A. A., and Vidal, M. (2008) A protein domain-based interactome network for C. elegans early embryo-genesis. Cell 134, 534–545

21. Rajagopala, S. V., Yamamoto, N., Zweifel, A. E., Nakamichi, T., Huang, H. K., Mendez-Rios, J. D., Franca-Koh, J., Boogurla, M. P., Fujita, K., Suzuki, K., Hu, J. C., Wanner, B. L., Mori, H., and Uetz, P. (2010) The Escherichia coli K-12 ORFeome: a resource for comparative molecular microbiology. BMC Genomics 11, 470

22. Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Stroebel, A., Richards, B., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature 433, 531–537

23. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Arata, T., Nakagahasi, K., Huang, H. C., Hirai, A., Tszuki, K., Nakamura, S., Altatt-Al-Minn, A., Oshima, T., Baba, T., Yamamoto, N., Kawasaki, T., Ioka-Nakamichi, T., Togawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) Large-scale identification of protein-protein interactions of Escherichia coli K-12. Genome Res. 16, 686–691

24. Hu, P., Janga, S. C., Babu, M., Diaz-Mejia, J. J., Butland, G., Yang, W., Pogouste, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulous, C., Nazarians-Armavil, A., Nasser, N. K., Musso, G., Ali, M., Mazevet, C., Eroukova, N., Ghoshali, A., Paccanaro, A., Greenblatt, J. F., Moreno-Hagelsieb, G., and Emili, A. (2008) Global functional atlas of Escherichia coli: encompassing previously uncharacterized proteins. PLoS Biol. 6, e97

25. Kuhn, N., van Noort, V., Betts, M. J., Leo-Macias, A., Batisse, C., Rode, M., Rajagopala, S., Janga, S. C., Babu, M., Dey, S., King, S., Woods, R. A., and Gietz, R. D. (2001) High-efficiency transformation of plasmid DNA into yeast. Methods Mol. Biol. 177, 85–97

26. Schnell, J. D., Dohmen, R. J., Strasser, A. W., Honer, C., and Hermjakob, H. (2011) The IntAct molecular interaction database in 2011. Nucleic Acids Res. 40, D841–D846

27. Tavernier, J., van den Heuvel, S., Hyman, A. A., and Vidal, M. (2008) A protein domain-based interactome network for C. elegans early embryo-genesis. Cell 134, 534–545

28. Cagney, G., and Uetz, P. (2001) High-throughput screening for protein-protein interactions using yeast two-hybrid arrays. Curr. Protoc. Protein Sci. Chapter 19, Unit 19 16

29. Rajagopala, S. V., Yamamoto, N., Zweifel, A. E., Nakamichi, T., Huang, H. K., Mendez-Rios, J. D., Franca-Koh, J., Boogurla, M. P., Fujita, K., Suzuki, K., Hu, J. C., Wanner, B. L., Mori, H., and Uetz, P. (2010) The Escherichia coli K-12 ORFeome: a resource for comparative molecular microbiology. BMC Genomics 11, 470
