This paper is available on line at http://www.mcponline.org

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 *H. pylori* 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 (5)).

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

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 Lluís 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.

[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.

(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 pGBGT7g, pGADT7g (32, 33), pGBT9-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 pGBGT7g, pGADT7g, and pGBKC-atg (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 plasmid pool was shuttled into the prey plasmids pGADT7g and pDEST22 using a Gateway LR reaction (Invitrogen). The reaction was then transformed into electrocompetent *E. coli* DH10B (ElectroMAX™, Invitrogen), grown in selective LB medium and plasmids isolated. Plasmid pools were then individually transformed into Y187 as described (35) and spread onto 24 × 24 cm dishes containing S.D. agar. Finally all colonies were scratched from the plates, resuspended in 25% glycerol, and stored as 50 μl aliquots at −80 °C.

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/interologs Pairwise Tests—Bait and prey pairs were retested 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. 1*A*).

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 *mgosim* 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 < $10e^{-4}$. 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

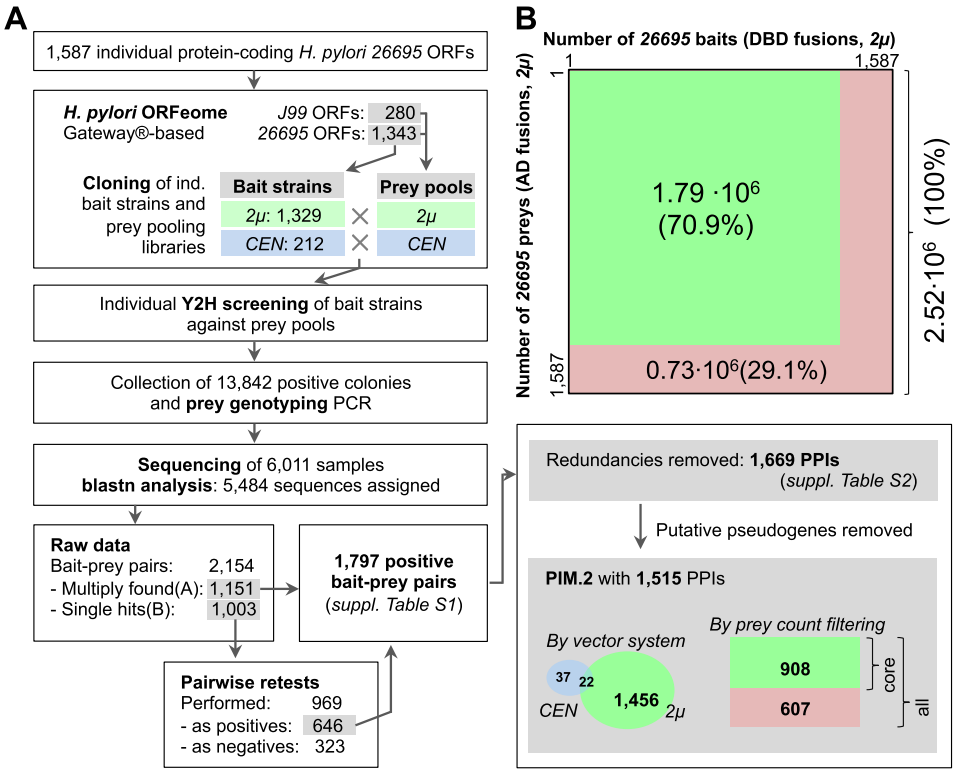

FIG. 1. **Y2H screening strategy and coverage.** *A,* Y2H screening strategy. For details, see main text. *B,* Screen coverage. The whole area represents all possible bait-prey test combinations in the *H. pylori* 26695 proteome. Green: experimentally tested combinations in the 2μ plasmid system; pink: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

distmat tool. We compared the confirmed and nonconfirmed interologs with the Wilcoxon rank test on the joint identity, joint similarity, joint *e*-value and joint distance as described (58) (joint values where calculated as $\sqrt{Pi * Pj}$, where *Pi* is the value of the property for the first protein and *Pj* is the value of the property for the second protein).

Conserved Subnetworks—To identify conserved subnetworks in the *H. pylori* PIM compared with other bacterial interactomes we used Netaligner (59, 60) and extracted alignments with $p < 0.01$. We then looked for enriched GO terms to identify the role of complexes.

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. 1*A*). 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. 1*B*). 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. 1*A*) 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.

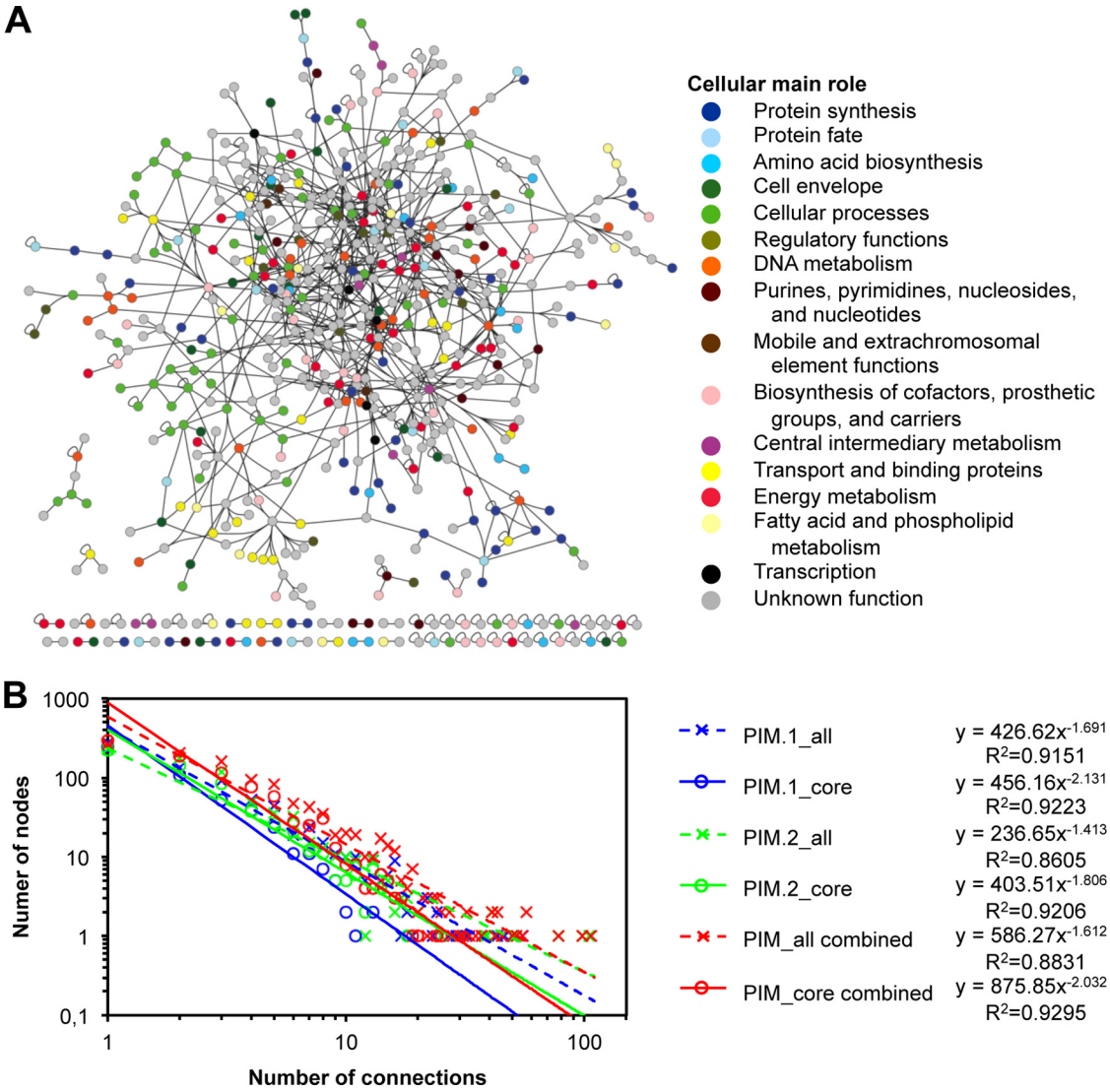

FIG. 2. The *H. pylori* protein interaction map. *A,* The core interaction map (PIM.2_core) involves 596 proteins connected by 908 interactions shown as nodes and edges, respectively. Proteins are colored according to their cellular main role (9). Graph drawn with Cytoscape 3.0.0-beta1(47). *B*, Degree distributions of different *H. pylori* PIM subsets (log. scale). The regression function and R^2 coefficients shown represent the power law fit yields. "PIM.1_all" all PPIs from (20); "PIM.1_core": PPIs in PBS categories A+B (20). "PIM.2_all" all PPIs from this study; "PIM.2_core": core PPIs from this study. The "combined" data sets contain either all or only the core data sets of both studies.

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. 1*A* and 3*B*). This resulted in a core network (PIM.2_core) of 908 PPIs connecting 596 proteins (38.4% of the proteome, Fig. 2*A*). 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. 3*B*).

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. 3*A*). The detection sen-

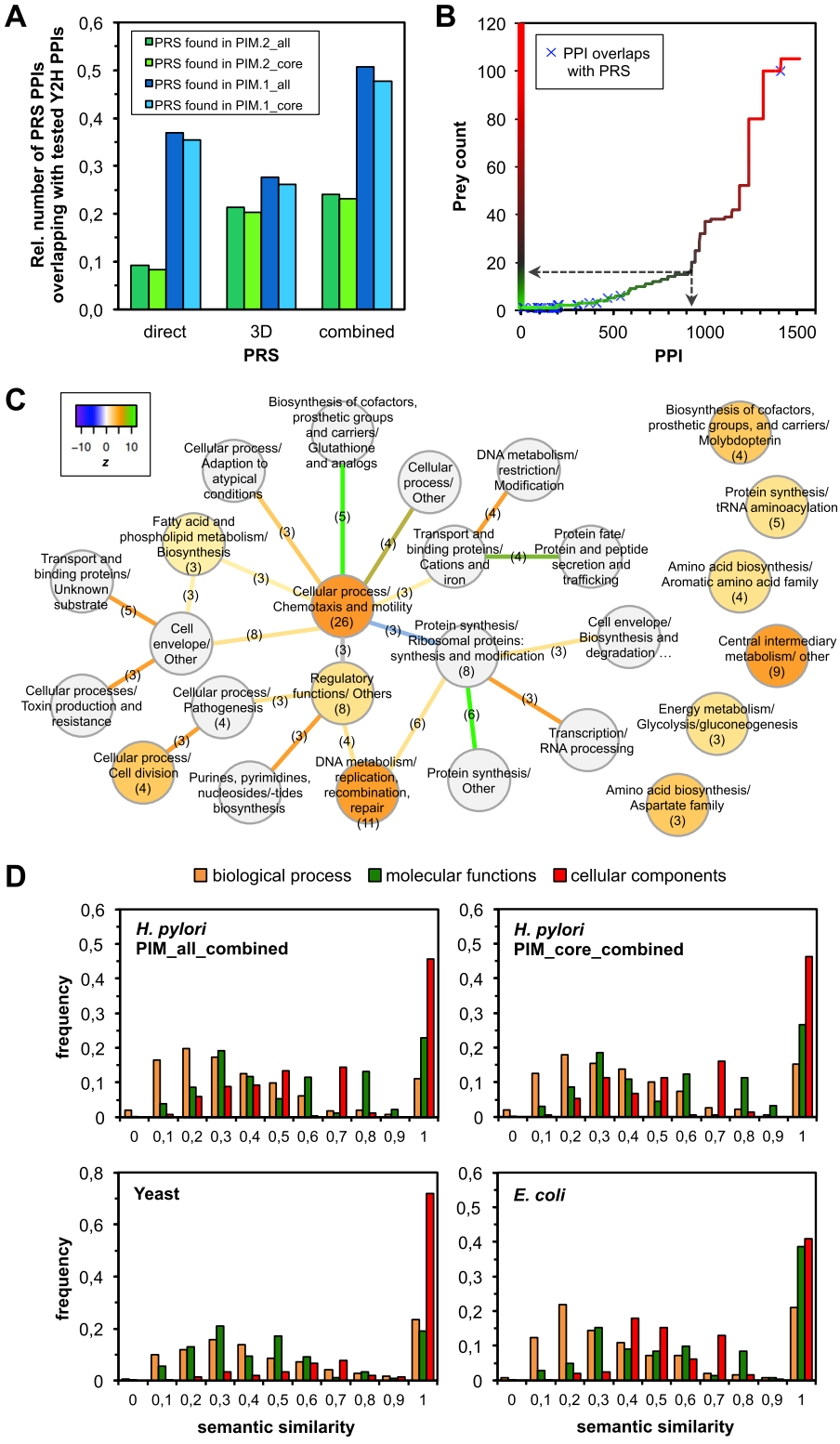

FIG. 3. **Data benchmarks and functional groups.** *A*, Screening sensitivity: detected interactions in PIM.2 (this study) and PIM.1 (20) when compared to interactions known from the *H. pylori* literature (PRS, supplemental Table S3); "direct" PPIs with proteins that are known to bind directly to each 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. 3*B*).

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 $\leq$ 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. 2*B*).

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. 3*C*). 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. 3*D* 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. 3*D*, 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. 4*A*) experimentally confirmed overlaps are small and rarely exceed a few dozen interactions or 13% (*H. pylori versus B. subtilis* and *Synechocystis* (Fig. 4*B*, 4*C*). 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. 4*D*.

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).

A

	H. pylori	*Synechocystis sp.*	*C. jejuni*	*B. subtilis*	*T. pallidum*	*M. pneumoniae*	*E. coli*	*M. tuberculosis*
H. pylori	3004	75	3365	37	136	278	1073	102
Synechocystis sp.	301	2689	2090	52	201	343	1535	330
C. jejuni	957	102	11765	42	262	365	1436	137
B. subtilis	361	218	3000	795	487	554	1944	466
T. pallidum	154	42	584	50	3640	323	748	54
M. pneumoniae	60	17	269	34	61	1059	400	38
E. coli	424	205	3447	132	466	471	5913	532
M. tuberculosis	233	245	1833	69	235	373	1562	7947

B

	H. pylori	*Synechocystis sp.*	*C. jejuni*	*B. subtilis*	*T. pallidum*	*M. pneumoniae*	*E. coli*	*M. tuberculosis*
H. pylori		10	76	5	4	1	12	1
Synechocystis sp.	10		11	2	0	1	7	0
C. jejuni	77	11		5	4	1	29	6
B. subtilis	5	2	5		2	5	6	0
T. pallidum	4	0	4	2		0	4	1
M. pneumoniae	1	1	1	5	0		24	1
E. coli	13	7	29	6	4	24		1
M. tuberculosis	1	0	6	0	1	1	1	

C

	H. pylori	*Synechocystis sp.*	*C. jejuni*	*B. subtilis*	*T. pallidum*	*M. pneumoniae*	*E. coli*	*M. tuberculosis*
H. pylori		13.3	2.3	13.5	2.9	0.4	1.1	1.0
Synechocystis sp.	3.3		0.5	3.8	0.0	0.3	0.5	0.0
C. jejuni	8.0	10.8		11.9	1.5	0.3	2.0	4.4
B. subtilis	1.4	0.9	0.2		0.4	0.9	0.3	0.0
T. pallidum	2.6	0.0	0.7	4.0		0.0	0.5	1.9
M. pneumoniae	1.7	5.9	0.4	14.7	0.0		6.0	2.6
E. coli	3.1	3.4	0.8	4.5	0.9	5.1		0.2
M. tuberculosis	0.4	0.0	0.3	0.0	0.4	0.3	0.1	

D

			H. pylori PIM							
			Number overlaps				% overlaps			
Dataset	Predicted interologs		.1_all	.2_all	_all	_core	.1_all	.2_all	_all	_core
Synechocystis sp.	all	75	3	8	10	8	4.0	10.7	13.3	10.7
	core	32	1	4	5	5	3.1	12.5	15.6	15.6
C. jejuni	all	3365	30	54	76	69	0.9	1.6	2.3	2.1
	core	751	24	45	62	58	3.2	6.0	8.3	7.7
B. subtilis		37	2	5	5	5	5.4	13.5	13.5	13.5
T. pallidum	all	136	1	3	4	4	0.7	2.2	2.9	2.9
	core	83	1	2	3	3	1.2	2.4	3.6	3.6
M. pneumoniae		278	0	1	1	1	0.0	0.4	0.4	0.4
E. coli		1073	5	8	12	10	0.5	0.7	1.1	0.9
M. tuberculosis		102	0	1	1	1	0.0	1.0	1.0	1.0

FIG. 4. **Comparison of interactome conservation.** *A*, Number of interologs predicted from published bacterial interactomes. White fields: PPIs per interactome (diagonal); the number of predicted interologs are plotted above or below (read by column), respectively. *B*, Number of interologs confirmed experimentally and *C*, Percentage of overlaps. Data sources used, see supplemental Table S4. The field color symbolizes an increasing number/fraction of predicted/conserved PPIs from black to green. *D*, Breakdown of *H. pylori* data sets. The figure compares the number of predicted interologs and conserved PPIs in the *H. pylori* PIMs from large scale screens of different species (used as prediction sources). "All" and "core" data (if available) were used as provided by the primary publications. Color symbolizes an increasing number of conserved PPIs from black to green. All studies used Y2H methods except of *M. tuberculosis* (bacterial 2-hybrid) and *M. pneumoniae* and *E. coli* (TAP-MS)). The highest overlaps tend to be found between the *H. pylori* PIMs with other Y2H studies. Note that our new data (PIM.2) provides increased numbers of conserved PPIs compared with PIM.1.

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. 5*A, 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. 5*C*, 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. 5*D*, 5*E*), the increase is not as high as previously shown by Yu *et al.* who suggested that interologs can be transferred reliably

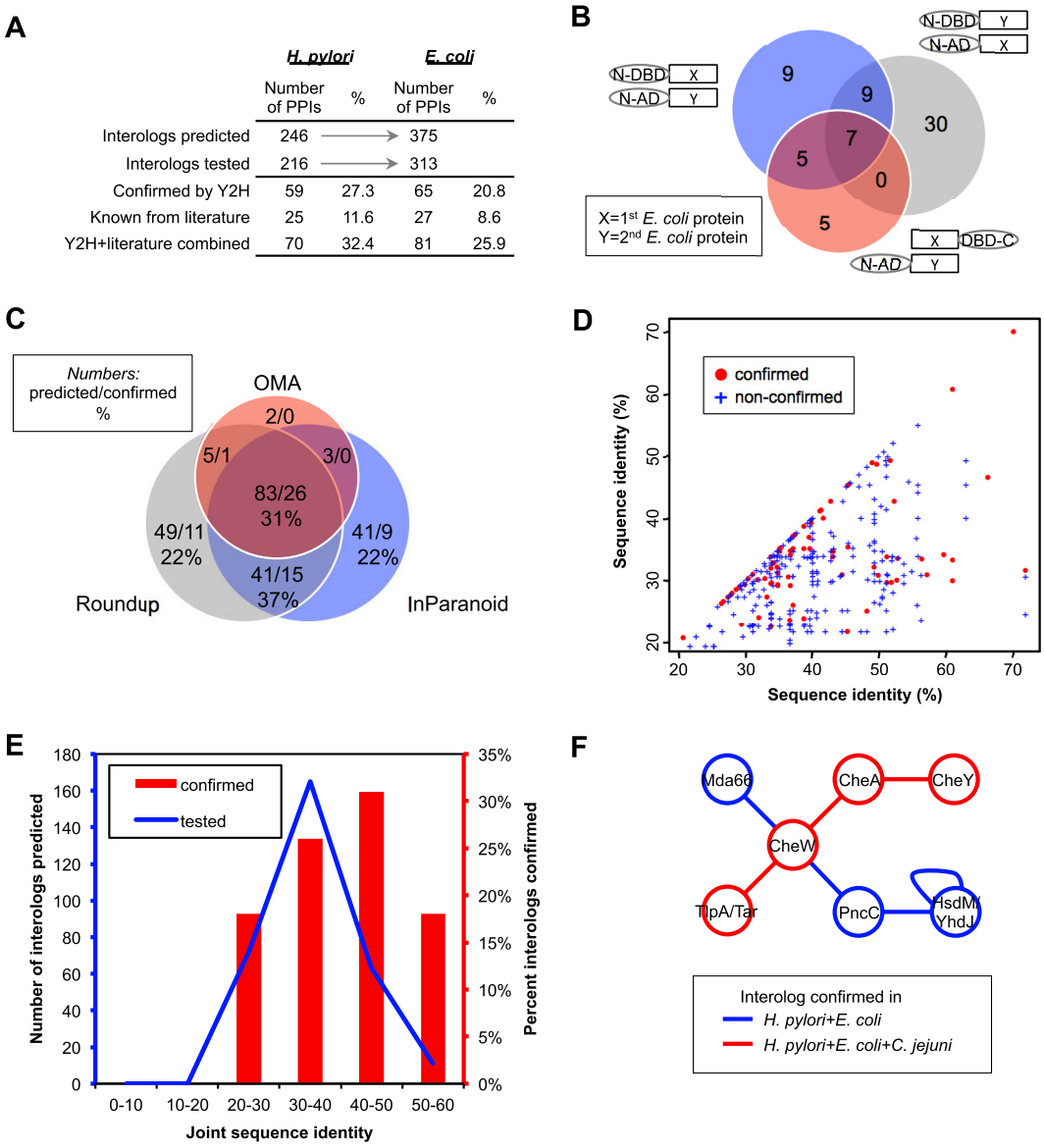

FIG. 5. **Conservation test of interologs and conserved subnetworks.** *A*, Summary of conservation tests carried out in this study by pairwise Y2H assays using protein pairs of *E. coli* and *H. pylori* PIM.2_all as prediction source. In addition, interologs from PPI databases and the literature are 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).

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. S2*A*–S2*D*, 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-

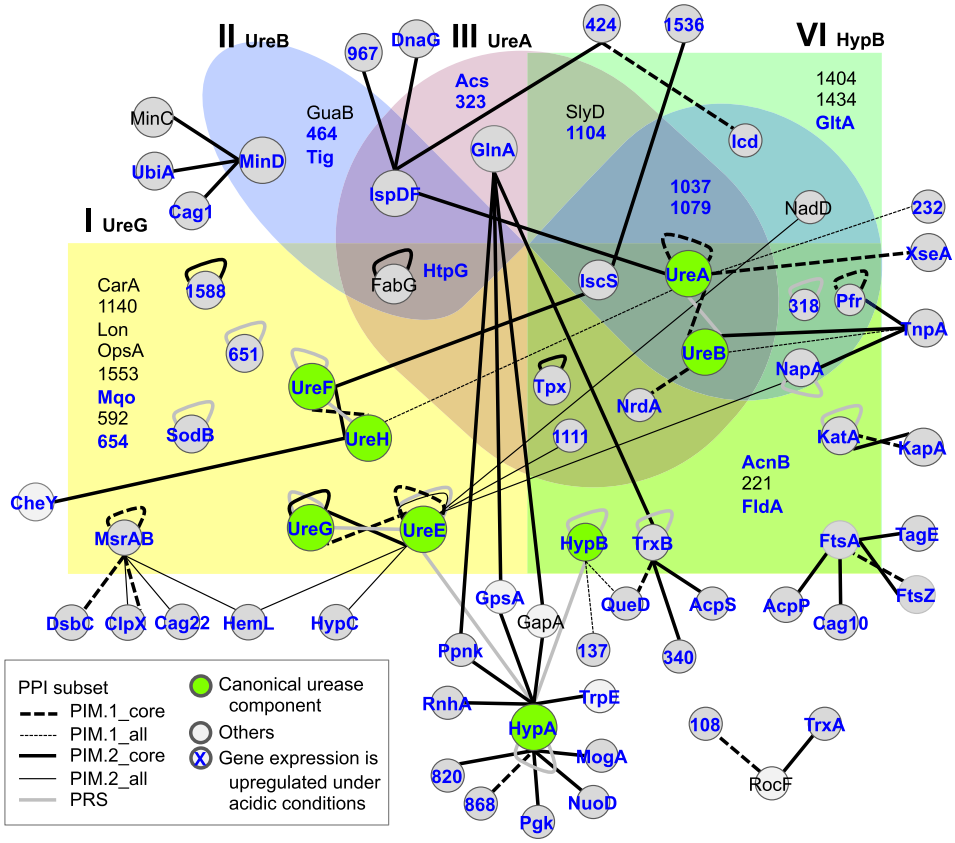

FIG. 6. **Urease linkage map: -omics data from protein complexes, binary interactions as well as gene expression is considered.** The map highlights the four most prominent protein complexes found by Stingl et al. (67) (*i.e.* the ones with the highest numbers of proteins and overlaps). Individual complexes are shown as fields of the Venn diagram. The bait protein used in Stingl *et al.* and the complex numbers are indicated at the top. To map interactions onto the complexes, we inferred binary interactions from the complexes (using a matrix approach) and then looked for overlaps with PIM.1, PIM.2, and the PRS. Proteins with known binary PPIs are drawn as nodes. Symbols added without nodes represent proteins for which no binary interaction is known within the complexes. For proteins for which no common name is available the primary locus tag number is indicated. Additionally, a selected subset of binary PPIs is shown that do not appear as a complex component but that could link the complexes to other processes. Protein symbols highlighted in blue font show an increased gene expression level under acidic growth conditions in a minimum of one experiment (68–75) (filtered data used as provided by the primary publications).

quence 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 drawn also 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 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 StingI 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. StingI *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.

S 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) Schistosomes, 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., Mengual-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. Lertsethtakarn, P., Ottemann, 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., Kerlavage, 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., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, 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., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, 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., Galinsky, 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., Palzkill, T., 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., Naeve, F., Tovchigrechko, A., Yamamoto, K., Szaflarski, W., Takeuchi, N., Stellberger, T., Diefenbacher, M. E., Nierhaus, K. H., and Uetz, P. (2012) RsfA (YbeB) proteins are conserved ribosomal silencing factors. *PLoS Genet.* **8,** e1002815
12. Titz, B., Rajagopala, S. V., Ester, C., Hauser, R., and Uetz, P. (2006) Novel conserved assembly factor of the bacterial flagellum. *J. Bacteriol.* **188,** 7700–7706
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

15. Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007) A large-scale protein protein interaction analysis in *Synechocystis* 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., Siszler, G., Wuchty, S., Emili, A., Babu, M., Aloy, P., Pieper, R., and Uetz, P. (2014) The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32,** 285–290
19. Marchadier, E., Carballido-Lopez, R., Brinster, S., Fabret, C., Mervelet, P., Bessieres, P., Noirot-Gros, M. F., Fromion, V., and Noirot, P. (2011) An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach. *Proteomics* **11,** 2981–2991
20. Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A., and Legrain, P. (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409,** 211–215
21. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245–246
22. Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., 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., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) Large-scale identification of protein-protein interaction 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., Pogoutse, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulos, C., Nazarians-Armavil, A., Nasseri, N. K., Musso, G., Ali, M., Nazemof, N., Eroukova, V., Golshani, A., Paccanaro, A., Greenblatt, J. F., Moreno-Hagelsieb, G., and Emili, A. (2009) Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* **7,** e96
25. Kuhner, S., van Noort, V., Betts, M. J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., Castano-Diez, D., Chen, W. H., Devos, D., Guell, M., Norambuena, T., Racke, I., Rybin, V., Schmidt, A., Yus, E., Aebersold, R., Herrmann, R., Bottcher, B., Frangakis, A. S., Russell, R. B., Serrano, L., Bork, P., and Gavin, A. C. (2009) Proteome organization in a genome-reduced bacterium. *Science* **326,** 1235–1240
26. Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J. M., Murray, R. R., Roncari, L., de Smet, A. S., Venkatesan, K., Rual, J. F., Vandenhaute, J., Cusick, M. E., Pawson, T., Hill, D. E., Tavernier, J., Wrana, J. L., Roth, F. P., and Vidal, M. (2009) An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6,** 91–97
27. Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I., Yildirim, M. A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J. M., Cevik, S., Simon, C., de Smet, A. S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R. R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M. E., Roth, F. P., Hill, D. E., Tavernier, J., Wanker, E. E., Barabasi, A. L., and Vidal, M. (2009) An empirical framework for binary interactome mapping. *Nat. Methods* **6,** 83–90
28. Chen, Y. C., Rajagopala, S. V., Stellberger, T., and Uetz, P. (2010) Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* **7,** 667–668
29. Stellberger, T., Hauser, R., Baiker, A., Pothineni, V. R., Haas, J., and Uetz, P. (2010) Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome. *Proteome Sci.* **8,** 8
30. 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
31. Boxem, M., Maliga, Z., Klitgord, N., Li, N., Lemmens, I., Mana, M., de Lichtervelde, L., Mul, J. D., van de Peut, D., Devos, M., Simonis, N., Yildirim, M. A., Cokol, M., Kao, H. L., de Smet, A. S., Wang, H., Schlaitz, A. L., Hao, T., Milstein, S., Fan, C., Tipsword, M., Drew, K., Galli, M., Rhrissorrakrai, K., Drechsel, D., Koller, D., Roth, F. P., Iakoucheva, L. M., Dunker, A. K., Bonneau, R., Gunsalus, K. C., Hill, D. E., Piano, F., Tavernier, J., van den Heuvel, S., Hyman, A. A., and Vidal, M. (2008) A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* **134,** 534–545
32. Rajagopala, S. V., Yamamoto, N., Zweifel, A. E., Nakamichi, T., Huang, H. K., Mendez-Rios, J. D., Franca-Koh, J., Boorgula, 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
33. Uetz, P., Dong, Y. A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S. V., Roupelieva, M., Rose, D., Fossum, E., and Haas, J. (2006) Herpesviral protein networks and their interaction with the human proteome. *Science* **311,** 239–242
34. Dohmen, R. J., Strasser, A. W., Honer, C. B., and Hollenberg, C. P. (1991) An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera. *Yeast* **7,** 691–692
35. Woods, R. A., and Gietz, R. D. (2001) High-efficiency transformation of plasmid DNA into yeast. *Methods Mol. Biol.* **177,** 85–97
36. Mohr, K., and Koegl, M. (2012) High-throughput yeast two-hybrid screening of complex cDNA libraries. *Methods Mol. Biol.* **812,** 89–102
37. Rajagopala, S. V., and Uetz, P. (2011) Analysis of protein-protein interactions using high-throughput yeast two-hybrid screens. *Methods Mol. Biol.* **781,** 1–29
38. 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
39. Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M. J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., Gardner, M., Laiho, K., Legge, D., Magrane, M., Pichler, K., Poggioli, D., Sehra, H., Auchincloss, A., Axelsen, K., Blatter, M.-C., Boutet, E., Braconi-Quintaje, S., Breuza, L., Bridge, A., Coudert, E., Estreicher, A., Famiglietti, L., Ferro-Rojas, S., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., James, J., Jimenez, S., Jungo, F., Keller, G., Lemercier, P., Lieberherr, D., Masson, P., Moinat, M., Pedruzzi, I., Poux, S., Rivoire, C., Roechert, B., Schneider, M., Stutz, A., Sundaram, S., Tognolli, M., Bougueleret, L., Argoud-Puy, G., Cusin, I., Duek-Roggli, P., Xenarios, I., and Apweiler, R. (2012) The UniProt-GO Annotation database in 2011. *Nucleic Acids Res.* **40,** D565–D570
40. Isserlin, R., El-Badrawi, R. A., and Bader, G. D. (2011) The Biomolecular Interaction Network Database in PSI-MI 2.5. *Database: the journal of biological databases and curation* 2011, baq037
41. Stark, C., Breitkreutz, B.-J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M. S., Nixon, J., Van Auken, K., Wang, X., Shi, X., Reguly, T., Rust, J. M., Winter, A., Dolinski, K., and Tyers, M. (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **39,** D698–D704
42. Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32,** D449–D451
43. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., and Hermjakob, H. (2011) The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. **40,** D841–D846
44. Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A. P., Santonico, E., Castagnoli, L., and Cesareni, G. (2011) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*. **40,** D857–D861
45. Goll, J., Rajagopala, S. V., Shiau, S. C., Wu, H., Lamb, B. T., and Uetz, P. (2008) MPIDB: the microbial protein interaction database. *Bioinformatics* **24,** 1743–1744

46. Mosca, R., Céol, A., and Aloy, P. (2013) Interactome3D: Adding structural details to protein networks. *Nat. Methods* **10,** 47–53
47. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* **27,** 431–432
48. Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., Lotia, S., Pico, A. R., Bader, G. D., and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods* **9,** 1069–1076
49. Peterson, J. D. (2001) The comprehensive microbial resource. *Nucleic Acids Res.* **29,** 123–125
50. Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26,** 976–978
51. Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., and Chen, C. F. (2007) A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23,** 1274–1281
52. Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D. N., Roopra, S., Frings, O., and Sonnhammer, E. L. L. (2010) InParanoid 7: New algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38,** D196–D203
53. Persico, M., Ceol, A., Gavrila, C., Hoffmann, R., Florio, A., and Cesareni, G. (2005) HomoMINT: An inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* **6,** Suppl 4, S21
54. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* **10,** 421
55. Schneider, A., Dessimoz, C., and Gonnet, G. H. (2007) OMA Browser–exploring orthologous relations across 352 complete genomes. *Bioinformatics (Oxford, England*) **23,** 2180–2182
56. DeLuca, T. F., Cui, J., Jung, J.-Y., St Gabriel, K. C., and Wall, D. P. (2012) Roundup 2.0: Enabling comparative genomics for over 1800 genomes. *Bioinformatics (Oxford, England*) **28,** 715–716
57. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Gen. TIG* **16,** 276–277
58. Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J. D., Bertin, N., Chung, S., Vidal, M., and Gerstein, M. (2004) Annotation transfer between genomes: Protein-protein interologs and protein-DNA regulogs. *Genome Res.* **14,** 1107–1118
59. Pache, R. A., Ceol, A., and Aloy, P. (2012) NetAligner–a network alignment server to compare complexes, pathways and whole interactomes. *Nucleic Acids Res.* **40,** W157–W161
60. Pache, R. A., and Aloy, P. (2012) A novel framework for the comparative analysis of biological networks. *PLoS One* **7,** e31220
61. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417,** 399–403
62. Deane, C. M., Salwinski, L., Xenarios, I., and Eisenberg, D. (2002) Protein interactions: Two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* **1,** 349–356
63. Das, J., and Yu, H. (2012) HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* **6,** 92
64. Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., Vincent, S., and Vidal, M. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs." *Genome Res.* **11,** 2120–2126
65. Rajagopala, S. V., Titz, B., Goll, J., Parrish, J. R., Wohlbold, K., McKevitt, M. T., Palzkill, T., Mori, H., Finley, R. L., Jr., and Uetz, P. (2007) The protein network of bacterial motility. *Mol. Syst. Biol.* **3,** 128
66. Olson, J. W., Mehta, N. S., and Maier, R. J. (2001) Requirement of nickel metabolism proteins HypA and HypB for full activity of both hydrogenase and urease in *Helicobacter pylori*. *Mol. Microbiol.* **39,** 176–182
67. Stingl, K., Schauer, K., Ecobichon, C., Labigne, A., Lenormand, P., Rousselle, J. C., Namane, A., and de Reuse, H. (2008) *In vivo* interactome of *Helicobacter pylori* urease revealed by tandem affinity purification. *Mol. Cell. Proteomics* **7,** 2429–2441
68. Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., Stadler, P. F., and Vogel, J. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464,** 250–255
69. Merrell, D. S., Goodrich, M. L., Otto, G., Tompkins, L. S., and Falkow, S. (2003) pH-regulated gene expression of the gastric pathogen Helicobacter pylori. *Infect. Immun.* **71,** 3529–3539
70. Pflock, M., Finsterer, N., Joseph, B., Mollenkopf, H., Meyer, T. F., and Beier, D. (2006) Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. *J. Bacteriol.* **188,** 3449–3462
71. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A., and Sachs, G. (2009) The pH-responsive regulon of HP0244 (FlgS), the cytoplasmic histidine kinase of *Helicobacter pylori. J Bacteriol* **191,** 449–460
72. Scott, D. R., Marcus, E. A., Wen, Y., Oh, J., and Sachs, G. (2007) Gene expression *in vivo* shows that *Helicobacter pylori* colonizes an acidic niche on the gastric surface. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 7235–7240
73. Bury-Mone, S., Thiberge, J. M., Contreras, M., Maitournam, A., Labigne, A., and De Reuse, H. (2004) Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori. Mol. Microbiol.* **53,** 623–638
74. Wen, Y., Marcus, E. A., Matrubutham, U., Gleeson, M. A., Scott, D. R., and Sachs, G. (2003) Acid-adaptive genes of *Helicobacter pylori*. *Infect. Immun.* **71,** 5921–5939
75. Ang, S., Lee, C. Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M. A., and Wang, J. T. (2001) Acid-induced gene expression in *Helicobacter pylori*: Study in genomic scale by microarray. *Infect. Immun.* **69,** 1679–1686
76. Nakamura, H., Yoshiyama, H., Takeuchi, H., Mizote, T., Okita, K., and Nakazawa, T. (1998) Urease plays an important role in the chemotactic motility of *Helicobacter pylori* in a viscous environment. *Infect. Immun,* **66,** 4832–4837
77. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R.C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., Hermjakob, H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40,** D841–D846 (Database issue)