



Comparative Analysis of Two *Helicobacter pylori* Strains using Genomics and Mass Spectrometry-Based Proteomics

Roger Karlsson^{1,2*}, Kaisa Thorell^{3,4†}, Shaghayegh Hosseini⁴, Diarmuid Kenny⁵, Carina Sihlbom⁵, Åsa Sjöling^{3†}, Anders Karlsson¹ and Intawat Nookaew^{4,6*}

OPEN ACCESS

Edited by:

Jessy L. Labbé,
Oak Ridge National Laboratory, USA

Reviewed by:

Robin Anderson,
Agricultural Research Service (USDA),
USA

Nico Jehmlich,
Helmholtz-Centre for Environmental
Research (UFZ), Germany

*Correspondence:

Roger Karlsson
roger.karlsson@nanoxisconsulting.com
Intawat Nookaew
inookaew@uams.edu

†Present address:

Kaisa Thorell and Åsa Sjöling,
Department of Microbiology, Tumor
and Cell Biology, Karolinska Institutet,
Stockholm, Sweden

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 05 August 2016

Accepted: 19 October 2016

Published: 11 November 2016

Citation:

Karlsson R, Thorell K, Hosseini S, Kenny D, Sihlbom C, Sjöling Å, Karlsson A and Nookaew I (2016)
Comparative Analysis of Two *Helicobacter pylori* Strains using Genomics and Mass Spectrometry-Based Proteomics.
Front. Microbiol. 7:1757.
doi: 10.3389/fmicb.2016.01757

¹ Nanoxis Consulting AB, Gothenburg, Sweden, ² Department of Infectious Diseases, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, ³ Department of Microbiology and Immunology, University of Gothenburg, Gothenburg, Sweden, ⁴ Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, ⁵ Proteomics Core Facility, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, ⁶ Department of Biomedical Informatics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Helicobacter pylori, a gastroenteric pathogen believed to have co-evolved with humans over 100,000 years, shows significant genetic variability. This motivates the study of different *H. pylori* strains and the diseases they cause in order to identify determinants for disease evolution. In this study, we used proteomics tools to compare two *H. pylori* strains. Nic25_A was isolated in Nicaragua from a patient with intestinal metaplasia, and P12 was isolated in Europe from a patient with duodenal ulcers. Differences in the abundance of surface proteins between the two strains were determined with two mass spectrometry-based methods, label-free quantification (MaxQuant) or the use of tandem mass tags (TMT). Each approach used a lipid-based protein immobilization (LPITM) technique to enrich peptides of surface proteins. Using the MaxQuant software, we found 52 proteins that differed significantly in abundance between the two strains (up- or downregulated by a factor of 1.5); with TMT, we found 18 proteins that differed in abundance between the strains. Strain P12 had a higher abundance of proteins encoded by the cag pathogenicity island, while levels of the acid response regulator ArsR and its regulatory targets (KatA, AmiE, and proteins involved in urease production) were higher in strain Nic25_A. Our results show that differences in protein abundance between *H. pylori* strains can be detected with proteomic approaches; this could have important implications for the study of disease progression.

Keywords: Genomics, *Helicobacter pylori*, MaxQuant, protein abundance, proteomics, TMT

INTRODUCTION

Helicobacter pylori is a Gram-negative, spiral shaped ε-proteobacterium that colonizes the stomachs of roughly half the world's population (Brown, 2000). Infection with *H. pylori* typically causes a mild, mixed gastritis; however, chronic infections cause severe clinical outcomes, such as duodenal and gastric ulcers, in approximately 15% of infected individuals. Gastric adenocarcinomas and mucosa-associated lymphoid tissue lymphomas occur in an additional 1–2%. Notably, the coexistence of gastric cancer and duodenal ulcer in the same individual is

rare. Chronic *H. pylori* infection also contributes to gastric issues that are associated with the development of gastric cancer, such as the loss of acid secretion and the loss of acidic mammalian chitinase expression (Nookae et al., 2013). The factors leading to this divergence in clinical outcomes are not entirely known, but host genetics that regulate the potency of the immune response toward the infection, bacterial genetics, and environmental factors, such as diet and smoking, contribute (Ahn and Lee, 2015).

The different strains of *H. pylori* are very diverse; there is a much higher frequency of recombination than point mutations. At the same time, the phylogeny of this bacterium is clearly traceable and reflects the ancestry of the carrier and the migration of ancient human groups (Falush et al., 2003; Suerbaum and Achtman, 2004; Thorell et al., 2016). During evolution, some genes remain present in all strains and make up the core genome of a species, but other genes occur variably among different isolates. It is beneficial for a pathogen to be able to adapt to new environmental niches. The genes that show the greatest sequence variations are often those that encode surface-exposed proteins, including outer membrane proteins (OMPs), that are the first to interact with the host (Suerbaum and Josenhans, 2007; Jabbour et al., 2010). The surface-exposed proteins have different functional roles; they function as adhesion factors (e.g., BabA and SabA), nutrient transporters, secreted toxins, iron-chelating proteins, and proteases, for example (Solis and Cordwell, 2011). A study comparing the frequency of recombination across the *H. pylori* genome showed that genes for OMPs are hotspots of recombination compared to, for example, metabolic genes (Yahara et al., 2012). This motivates studies that focus on how surface-exposed proteins vary in abundance across different *H. pylori* strains and how this variation affects clinical outcome.

The recently developed technology called Lipid-based Protein Immobilization (LPITM) immobilizes bacteria within a flow cell, and exposed proteins are digested with an enzyme, such as trypsin (Chooneea et al., 2010). Combining LPI with mass spectrometry-based proteomics targets surface-exposed proteins for analysis. One study that performed proteomic typing of *H. pylori* strains with the LPITM technology found that 60% of the strain-specific peptides that were found to be unique biomarkers of the *H. pylori* J99 strain are membrane-associated proteins, several of which belong to the outer membrane (Karlsson et al., 2012). Different methodologies can be used to quantitatively compare samples. These methods include labeling peptides with stable isotopes in culture (stable isotope labeling with amino acids in cell culture, SILAC), labeling prior to digesting samples (isotope-coded affinity tags, ICAT), or labeling after digestion (isobaric tags for relative and absolute quantitation, iTRAQ; or tandem mass tags, TMT). TMT labeling enables the relative quantification, with good sensitivity, of uniquely labeled peptides in samples that are pooled for MS analysis (Thompson et al., 2003). MS/MS-based tag detection is often used to generate quantitative proteomic profiles between mammalian cell lines (Paulo et al., 2013). However, there are also methods and data analysis platforms, such as MaxQuant, that rely on label-free quantitation (Cox and Mann, 2008) where the amount of detected protein is derived from the sum of the peak intensities

of the observed peptides. This analysis was also included in the present study.

Here, our aim was to evaluate and compare genetic content and protein abundance between two pathogenic strains of *H. pylori*; strain P12 was isolated in Europe from a patient with duodenal ulcers, and strain Nic25_A was isolated in Nicaragua from the antrum of a patient with atrophy and intestinal metaplasia. The selected whole-genome sequence of Nic25_A (Thorell et al., 2016) was compared with the complete genome sequence of *H. pylori* strain P12 (Fischer et al., 2010) to identify similarities and differences in the genomic potential. The two strains were cultured under controlled conditions and analyzed with mass spectrometry-based proteomics to compare their cell-surface proteomes.

Due to the genetic heterogeneity of *H. pylori* strains, we used the peptides found in both strains (common peptides) to enable for a better comparison of protein abundance as one part of the study. The common peptides were assayed with label-free MaxQuant analysis or with TMT labels for relative quantification. Another part of the study focused on unique proteins detected in either strain. Using these strategies, we were able to observe differences in the abundance of virulence factors between the two *H. pylori* strains that will be useful for evaluating the pathogenicity of these bacteria.

MATERIALS AND METHODS

Data Acquisition and Analysis of Genome Sequences

The genomic information for *H. pylori* strain P12 (Fischer et al., 2010) was retrieved from the RefSeq database at NCBI. The genomic information for Nic25_A was retrieved from our previous work (Thorell et al., 2016) deposited in the Sequence Read Archive¹) under accession number SRP04449; the sequence of Nic25_A, from a patient with intestinal metaplasia and atrophy, (SRS679647) was used in this project.

Comparative Genomics

To ensure that the annotation between both genomes was compatible for comparative analyses, we used the Prokka pipeline v1.9 (Seemann, 2014) to annotate both genomes. This pipeline predicts open reading frames (ORFs) using Prodigal (Hyatt et al., 2010), specifically developed for Gram-negative bacteria; it also allows for rRNA, tRNA, and signal-peptide prediction. Prokka was applied to the assembled contigs of the Nic25_A strain and the P12 chromosome and plasmid. As the primary annotation source in Prokka, we used the strain 26695 genome with the most recent reannotation (Resende et al., 2013) and with manually curated OMP annotation (Alm et al., 2000). For predicted ORFs with no closely related (blastp *e*-value < 10⁻⁹) match in the 26695 genome, annotation was based on the global *H. pylori* reference databases (see **Supporting File 6** for complete annotation of the two strains). To compare the genome sequences of the two strains, we used MAUVE (Darling

¹<http://www.ncbi.nlm.nih.gov/sra>

et al., 2010) to perform entire-genome pairwise alignment, and the alignment was visualized with BRIG (Alikhan et al., 2011). The protein-coding sequences of the two genomes were further clustered to identify groups of orthologous proteins with the UClust software (identity threshold of 0.8) (Edgar, 2010). Then, the representative protein sequences of each cluster were further predicted for their membrane-associated properties with LipoP v1.0 (Juncker et al., 2003) and PSORTb v3.02 (Yu et al., 2010).

***Helicobacter pylori* Strain Isolation and Cultivation**

Helicobacter pylori strains P12 and the Nicaraguan clinical isolate Nic25_A were cultured on Columbia blood agar plates under microaerophilic conditions (80% N₂, 10% CO₂, 10% H₂), which is standard for *H. pylori* (Janzon et al., 2009). The strains were replated and cultured for 24 h before the colony biomass was scraped from the plates and suspended in 500 μL of phosphate buffered saline.

Proteomics – Sample Processing and Generation of Peptides with LPI™ HexaLane

An overview of the workflow for sample processing and peptide generation is presented in **Figure 1**. The bacterial biomass for each strain was washed with PBS, centrifuged for 8 min at 4000g, and resuspended in PBS. This procedure was repeated three times before the bacteria were resuspended in PBS to an optical density (600 nm) of approximately 0.5 (NanoDrop-1000 Spectrophotometer, Thermo Fisher Scientific). The bacterial suspension was immediately loaded into the LPI™ HexaLane FlowCell (Nanoxis Consulting AB) (**Figure 1**, step 1). An excess of bacteria was applied to the flow cell by adding a bacterial suspension (100 μL) to fill the LPI™ Flow Cell channel, which has a volume of 50 μL. Excess bacterial suspension was removed from the inlet and outlet ports to ensure that there was an equal amount of bacterial material in each flow cell channel. The flow cells containing bacteria were incubated for 2 h at room temperature to allow cells to attach. The channels were subsequently washed with 1.0 mL of triethylammonium bicarbonate (TEAB) buffer (100 mM) to remove excess, unbound bacteria. Bacterial membrane proteins were digested by injecting 100 μL of trypsin (20 μg/mL) into the LPI™ HexaLane FlowCell channels and incubating for 30 min at room temperature.

The cleaved peptides were eluted by injecting TEAB buffer (200 μL, 100 mM) into the channels and collecting the flow-through at the outlet ports with a pipette. The eluate was incubated overnight at 37°C to allow for complete digestion. Afterward, the samples were frozen at -20°C until analysis by MS. The peptide samples were split into two equal parts and analyzed by two different approaches. One aliquot was analyzed by LC-MS/MS, as described previously (Karlsson et al., 2012), in order to generate protein lists for each sample; each sample was analyzed in triplicate injections (**Figure 1**, step 3). Relative quantification of peptides was performed on the second

aliquot with Tandem Mass Tag (TMT), MS-based quantification (**Figure 1**, step 4).

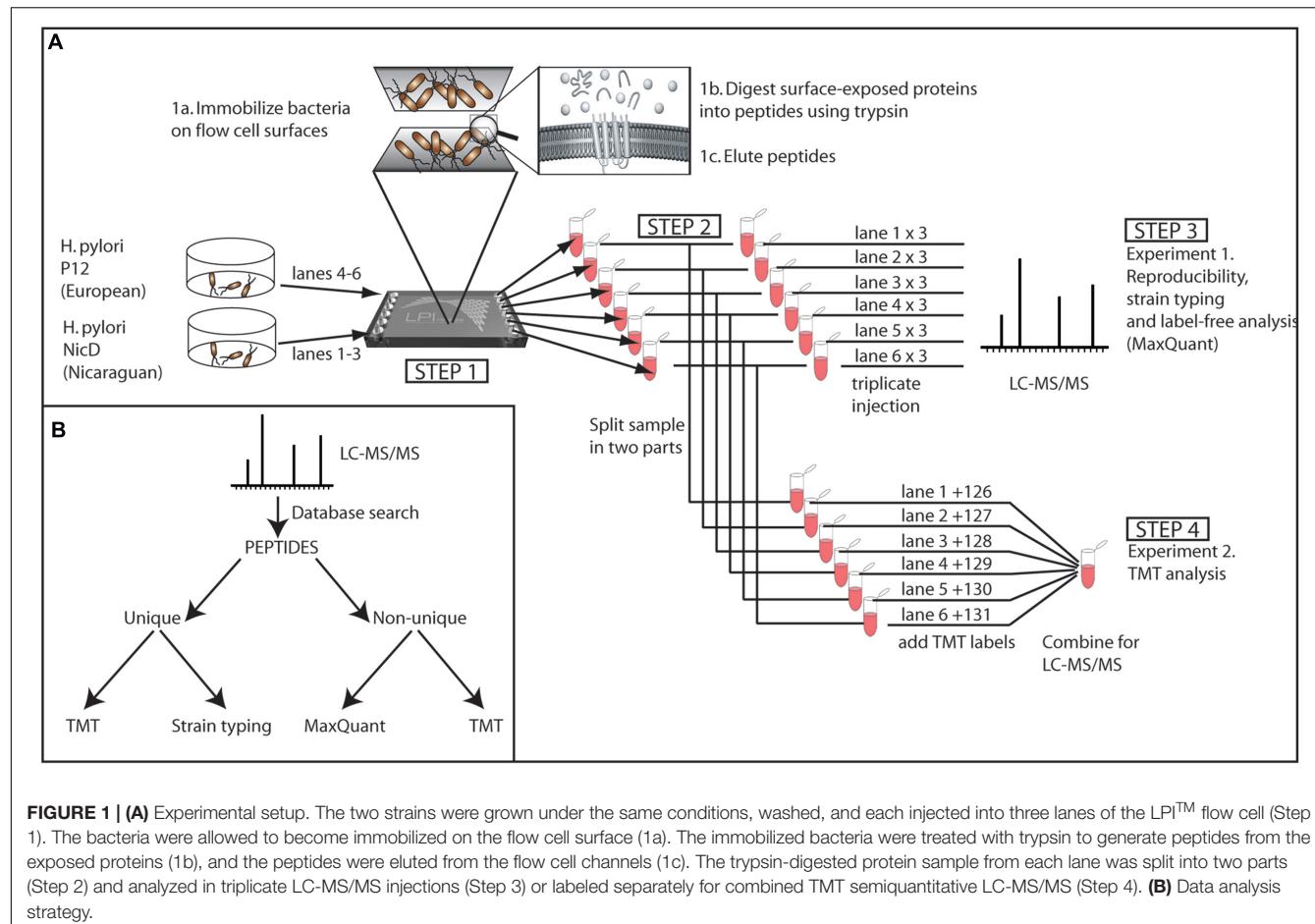
Mass Spectrometry-Based Analysis of Peptides

The peptide samples were analyzed with two different LC-MS/MS approaches using the LTQ-Orbitrap XL at the Proteomics Core Facility, University of Gothenburg², as described previously (Karlsson et al., 2012) (with some modifications). The length of the MS analysis was 120 min, and the LC gradient was adjusted to increase from 5 to 37% acetonitrile during 100 min. The samples were first analyzed individually in triplicate injections, (**Figure 1**, step 3). In addition to this, we quantified the relative abundance of peptides with Tandem Mass Tag (TMT), MS-based quantification using NHS chemistry to target primary amines (Proteome Science, Surrey, UK) (**Figure 1**, step 4). For quantifying the relative abundance of proteins, the LTQ-Orbitrap XL was switched between CID (collision-induced dissociation) and HCD (high-energy collision dissociation) in the data-dependent mode of the three most abundant doubly, triply, and quadruply protonated ions from each FT-MS scan. The CID scan was used for identification and the HCD scan was used to quantify the TMT reporter ions. The settings for the MS2 was as follows: 1 microscan for HCD-MS2 at 7500 resolution (at m/z 400), mass range m/z 100–2000 with a collision energy of 50%, 1 microscan for CID-MS2 with a collision energy of 30%. Dynamic exclusion of a precursor selected for MS2 was used for 60 s after one repeat. The TMT-labeled sample was analyzed two additional times with an exclusion list of all m/z already passing the peptide identification criteria in order to increase the number of identified proteins. All raw files generated in this study can be found in the PRIDE database under accession number PXD005074.

Protein Quantification with TMT

MS/MS data were analysed with Proteome Discoverer version 1.4 (Thermo Fisher Scientific) incorporating Mascot version 2.3 (Matrix Science, London, UK). Mascot was configured to search against the custom-made database of the Nic25_A and P12 proteomes (3,066 total entries). The MS peptide tolerance was set as 10 ppm, MS/MS tolerance as 0.5 Da, trypsin digestion allowing 1 missed cleavage with variable modifications, methionine oxidation, and, where relevant, fixed modifications N-terminal TMT6-plex label, lysine TMT6-plex label. The detected peptide threshold in the software was set to a detection confidence using a false discovery rate (FDR) of 1% (the FDR determined via target-decoy). For quantification, the ratios of TMT reporter ion intensities in MS/MS spectra (m/z 126.12, 127.13, 128.13, 129.14, 130.14) from raw data sets were used to calculate fold changes for common peptides (not unique) that are isoforms or proteins of the same family between the strains. Peptide measurements were not grouped, missing values were replaced, and normalization was used. The

²<http://www.proteomics.cf.gu.se>



ratios were then exported into Microsoft Excel for manual data interpretation.

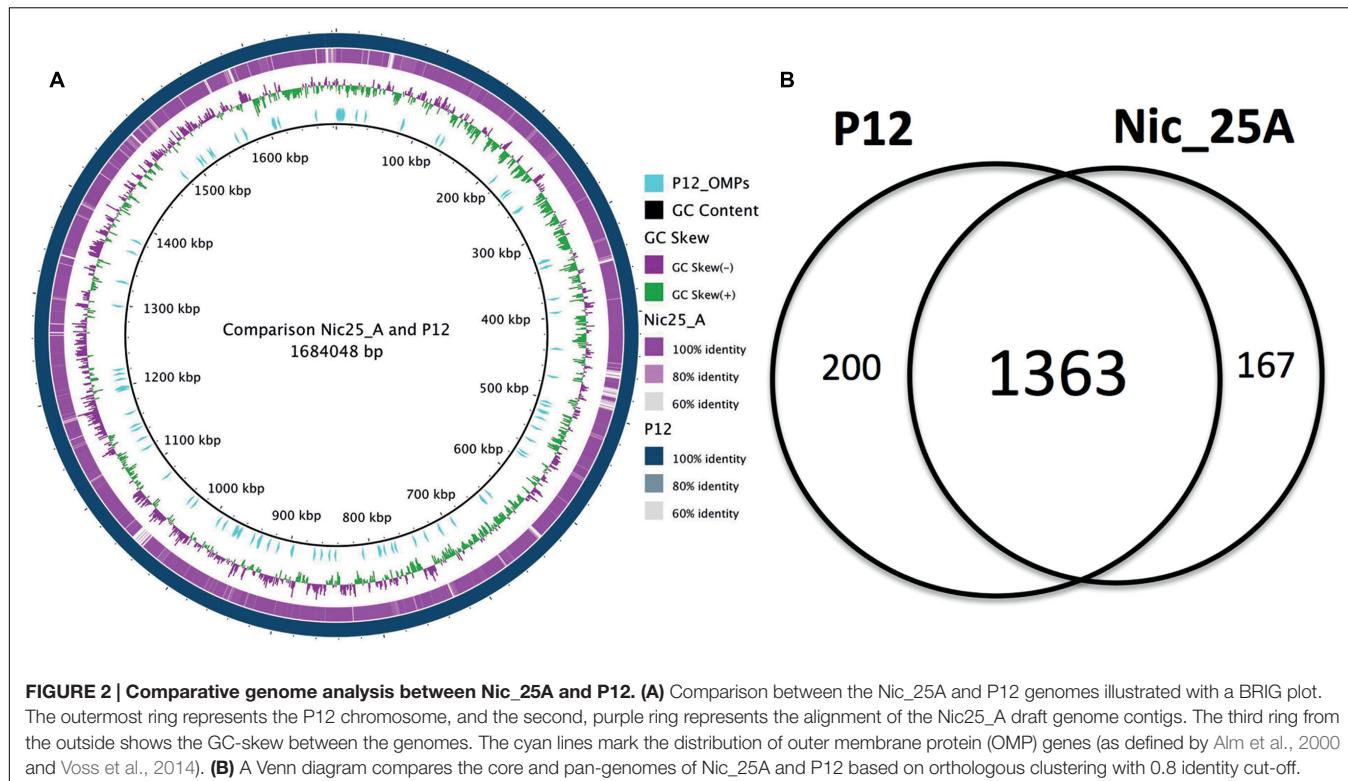
Label-Free Protein Quantification

Individual raw sample files were processed with MaxQuant (version 1.5.0.30). The data from the three samples of each strain were searched against a custom database compiled from the Prokka-annotated protein FASTA files for each respective strain (i.e., one database for P12 and one for Nic25_A). The searches were split into separate experiments, with each experiment consisting of the raw files corresponding to three injections per sample. The variable modifications for the searches were acetyl (N-terminal) and oxidation (methionine), with a maximum of five modifications per peptide. The digestion enzyme was trypsin, and a maximum of one missed cleavage was accepted. For the primary search, the precursor peptide tolerance was 10 ppm, and the MS/MS tolerance was 0.5 Da. The match between runs was selected for each experiment with a match time window of 1 min. A minimum of razor peptide was required for identification, and a minimum score of 40 was accepted for modified peptides. The FDR was set to 0.01 for peptide spectral matches, proteins and site and calculated using reverse decoy. The maximal peptide PEP was 0.01. Minimal peptide length was seven amino acids.

Protein quantification was performed using intensity-based absolute quantification (iBAQ), which sums the peak intensities of all matching peptides and divides this by the number of theoretical observable peptides to provide a proxy quantification of the protein levels (Schwanhausser et al., 2011). The minimum reporter PIF (precursor intensity fraction) was 0.75. The percentage of each protein in the pool was determined by dividing the individual iBAQ value for each protein by the summed iBAQ values of all the quantified proteins identified in the sample. The three individual P12 and Nic25_A samples were grouped, and the coefficient of variation and median percentage were determined.

H. pylori Strain Identification with Peptide Mapping

One hundred and sixty genome sequences of *H. pylori* strains (complete and draft) were retrieved from RefSeq and the amino acid sequences were collected and combined with the amino acid sequences of *H. pylori* Nic25_A to become a database for peptide mapping. The high quality peptides identified with mass spectrometry were mapped using blast+ v2.2.27 (blastp, word size = 2, score matrix = PAM30). The results from blast+ were further used for strain-typing analysis with the scoring method proposed by Karlsson et al. (2012).



Protein Abundance Analysis

Protein intensities from MaxQuant for P12 and Nic25_A were combined into the same file for differential-abundance analysis with the 26695 gene IDs. If the protein was annotated with the same gene ID in the P12 and Nic25_A, respectively, the intensities were tested for differential abundance against each other with the Qprot algorithm (Choi et al., 2008). This was also used for identifying proteins uniquely abundant in each strain. The criterion set for this analysis was: protein found in at least 2 of 3 runs (triplicate analysis) for one strain but absent in all three runs in the other strain.

RESULTS AND DISCUSSION

H. pylori Nic25_A Genome Properties and Its Comparison with *H. pylori* P12

Based on the genomic information, the Nic25_A genome was found to contain 1627 predicted ORFs, similar to other *H. pylori* strains. The *H. pylori* Nic25_A genome was further compared to the *H. pylori* P12 genome through entire genome alignment with MAUVE (Darling et al., 2010) to produce a BRIG plot (Figure 2A). Several single nucleotide variants (SNVs) between the two *H. pylori* strains were distributed along the chromosome and accounted for about 1.1% of the total nucleotides in the *H. pylori* P12 chromosome (Supporting Files 1–3).

The amino acid sequences of the two strains were subjected to orthologous clustering with UClust (Edgar, 2010) to identify

protein families (Figure 2B). Using an identity cut-off of 0.8, we identified 1730 pan-protein families with 200 (P12) or 167 (Nic25_A) unique protein sequences (Supporting Files 1–3). Of these, 1363 proteins were found in common between the two strains. In total, 53 proteins were identical between the two strains; among these were OMPs, HorE, HopK, HopN, SabB, and LPP20.

Unique Peptides Predicted from Bioinformatics Analyses Are Identified with MS

More than half of the membrane-associated protein families identified by bioinformatics analysis were also identified with high-quality peptides using an identity cut-off of 0.99. The high-quality peptides derived from individual LPITM lanes and individual MS injections were used to detect unique peptides from the two strains. Proteomic strain-typing based on unique peptides was performed according to Karlsson et al. (2012) (i.e., by matching the identified peptides to the genomes of different *H. pylori* strains and ranking the results according to the highest number of matched peptides per strain). Thus, the analysis of the proteomic results verifies that the top-ranked strain is the strain being analyzed, as illustrated for the strain ranking analysis for Nic25_A and P12 (Figure 3). We achieved high reproducibility when comparing different individual injections or different LPITM lanes with a minimum Spearman rank correlation of 0.83 (Supplementary Figure S1).

One factor contributing to the demonstrated reproducibility is that the same amount of material is immobilized in each

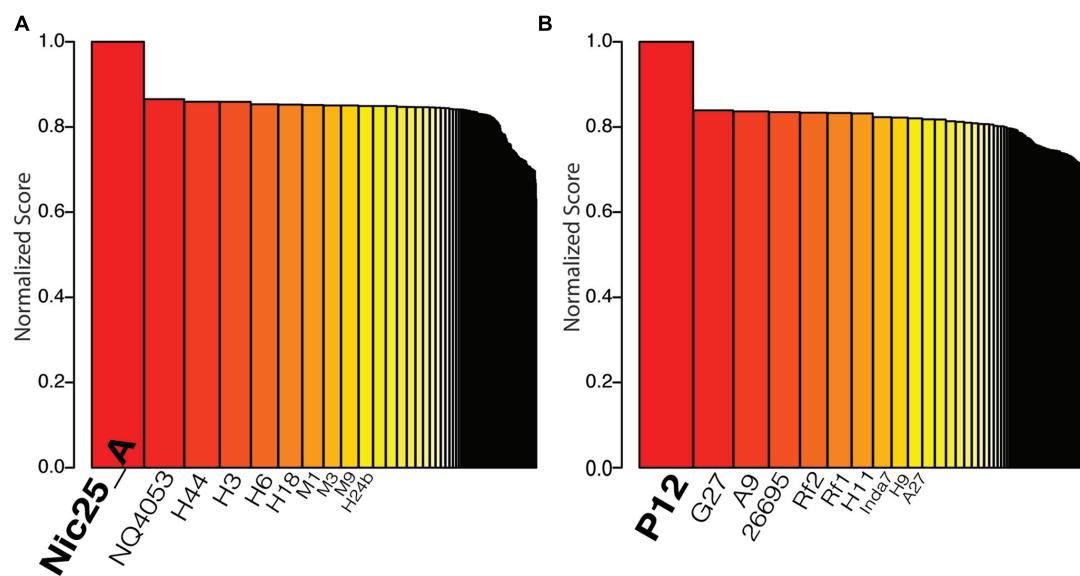


FIGURE 3 | Strain typing results using the method proposed by Karlsson et al., 2012. Bar plot of sorted normalized score based on the collection of 160 genomes from different *Helicobacter pylori* strains. The strains with the top ten highest scores are labeled on the x-axis. **(A)** The result when peptide data from *H. pylori* Nic25_A was used for calculating the score. **(B)** The result when peptide data from *H. pylori* P12 was used for calculating the score.

lane, due to the same capturing surface-area of each lane. This demonstrates one of the benefits of using the LPI approach for comparative proteomics.

Analysis of Differential Protein Abundance

Using the protein intensities calculated from the MaxQuant and TMT analyses as input into the QProt software, we identified 52 proteins that differed in abundance (FDR 0.05) between Nic25_A and P12 using the MaxQuant data and 18 proteins using the TMT data (Figure 4). Processed data can be found in Supporting Files 4 and 5.

Analysis of Unique Proteins

In order to detect proteins that were uniquely abundant in either of the two strains, we used MaxQuant to evaluate the triplicate runs with the following criteria: unique for Nic25_A—found in 2 out of the 3 runs for Nic25_A but in 0 of the P12 runs; unique for P12—found in 2 out of the 3 runs for P12 but in 0 of the Nic25_A runs. This produced 59 unique proteins for both Nic25_A and P12 (Supporting Files 7 and 8).

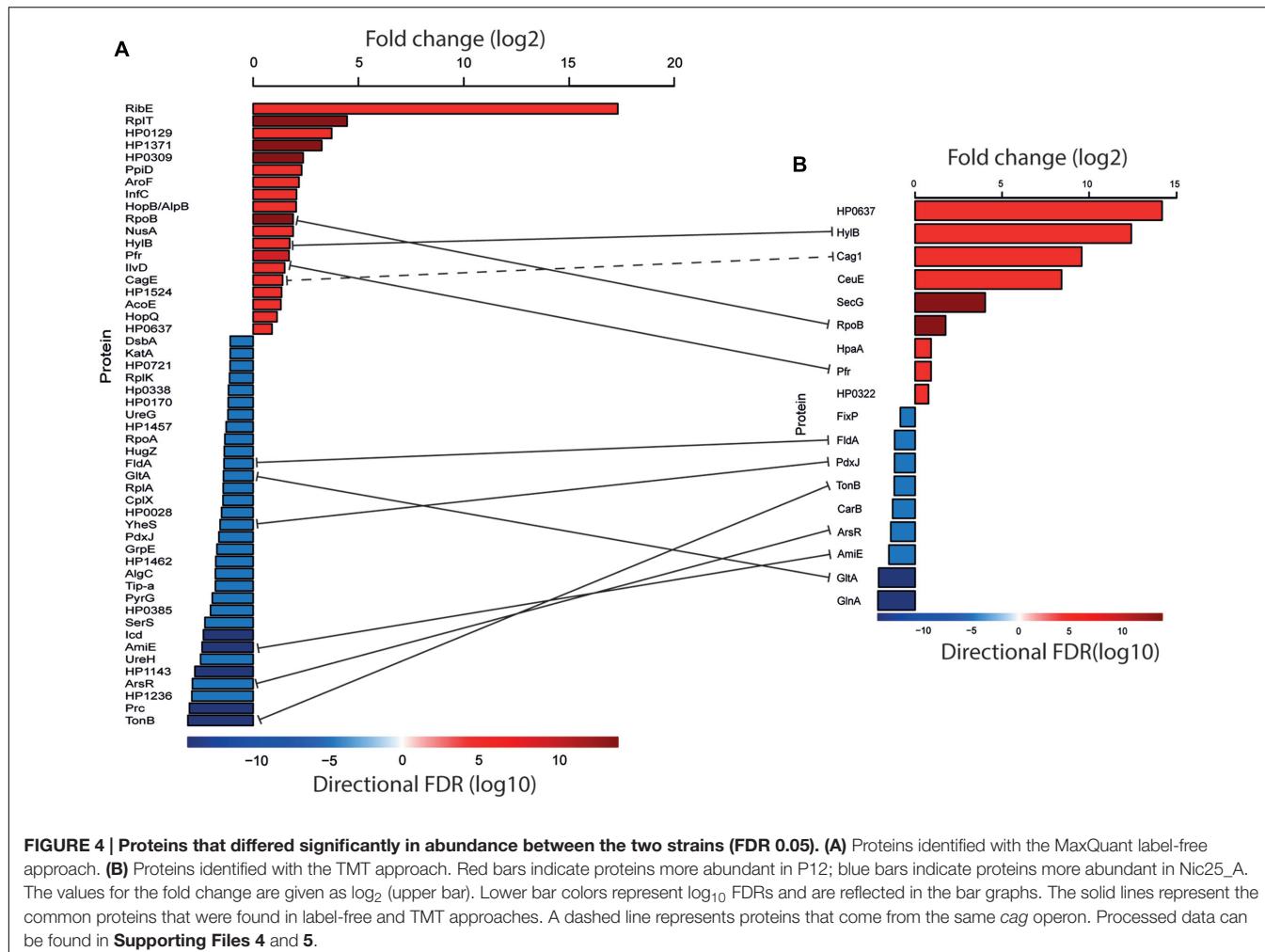
Workflow of Comparative Proteomics of the Two *H. pylori* Strains

The goal of this study was to evaluate the feasibility of using MS-based proteomics (and genomic analyses) to quantify differences in protein abundance in the human pathogen *H. pylori*, which displays high intraspecies sequence variability. The driving hypothesis is that it is possible to identify differences in protein abundance between *H. pylori* strains isolated from different clinical and geographical origins, and this information can be

used to evaluate biologically important features. To demonstrate proof-of-principle, we chose two *H. pylori* strains isolated from patients with different clinical manifestations and from different geographical locations. The P12 strain has been analyzed in several studies and is sequenced. We isolated and sequenced strain Nic25_A previously (Thorell et al., 2016). We used these genome sequences to determine unique peptide/protein markers for each specific strain, and we quantified the relative abundance of proteins in both strains.

Using the LPITM HexaLane FlowCell and the TMT 6-plex for relative quantification proved to work well, as did the label-free approach. The LPI HexaLane FlowCell has six channels for immobilizing bacteria, and all six samples can be processed and analyzed in parallel. All of the channels have the same surface area for capturing bacteria; therefore, the channels bind the same amount of sample in each channel, enabling us to reliably compare protein levels. In this study, three channels were used for *H. pylori* Nic25_A and three for *H. pylori* P12. We found that the data generated from the three replicates were highly reproducible, hence it was possible to produce statistically significant data on protein abundance using the flow cells.

A factor clearly complicating the analysis of data from shotgun proteomics approaches is the problem of annotation. Alm et al. (2000) did the first thorough annotation of *H. pylori* OMPs in 2000; they compared strains J99 and 26695, the only strains for which complete sequences were available (Alm et al., 2000). Despite this effort, the current databases are filled with divergent annotations, and it requires significant effort to identify homologies and corresponding locus names between strains. The fact that comparative proteomics and genomics rely on well-annotated databases is a major hurdle. To overcome this, we used the recently reannotated genome of strain 26695 (Resende



et al., 2013) as a reference for our annotation of both the P12 and Nic25_A proteomes with the Prokka annotation pipeline (Seemann, 2014). Because we are focusing on membrane-bound and surface exposed proteins, we also curated the reference manually with regards to the OMPs described by Alm et al. (2000). Using the two newly annotated proteomes, we searched the peptides from Nic25_A injections against the Nic25_A proteome and the P12 peptides against the P12 proteome. This allowed us to compare the intensities of the detected peptides with the harmonized annotation and not introduce bias into the analysis (due to ambiguities in matching to the database).

The Biological Significance

We found that the abundance of several known virulence factors and OMPs differed between strains P12 and Nic25_A. Based on the shared protein hits and relative quantification, *cag* pathogenicity island (CagPAI) proteins were more abundant in P12 than Nic25_A as shown in Figure 4. CagA is a well-described *H. pylori* virulence factor that is injected into mammalian cells through a type IV secretion system encoded by several other ORFs in the CagPAI. CagE is an essential component of this secretion system and, based on homology, functions as an

ATPase. Loss of CagE leads to the incomplete assembly of the secretion system (Zhang et al., 2002). As seen in the Figure 4, the adhesions AlpB, HpaA, and HopQ were also found to be more abundant in the P12 strain. Based on the unique proteins identified in the triplicate runs (Supporting File 8), the adhesins BabB and SabB were more abundant in the P12 strain, as were OMPs of the Hop and Hom families (Voss et al., 2014) and other virulence factors including vacuolating cytotoxic protein A (VacA), elongation factor (EF-Tu), and flagellar basal body protein FliL.

With respect to Nic25_A, the transcription factor ArsR was more abundant in this strain than in P12 as shown in Figure 4; ArsR is part of a two-component acid-response system (Wen et al., 2003). The Nic25_A isolate was also more abundant in several other major virulence factors; these include UreG and UreH, two urease accessory proteins, and KatA, involved in the protection against reactive oxidative species that is also regulated by ArsR (Wen et al., 2003). From the list of unique proteins found in Nic25_A, there were also some known virulence factors present (Supporting File 7), including UreB and heat shock protein GroEL. It is interesting to note that the Nic25_A strain, isolated from an individual with atrophy and

intestinal metaplasia and commonly thought to be associated with hypochlorhydria (pH 4–7), overexpressed several genes involved in acid resistance and oxidative stress compared to P12, which was isolated from a patient with duodenal ulcer. Our results are, however, corroborated by other proteome studies where gastric cancer is associated with a higher abundance of urease genes and antioxidant proteins such as KatA and TrxA, as well as GroEL (Park et al., 2006; Momynaliev et al., 2010; Repetto et al., 2014). Thus, the increase in ArsR levels might be associated with a greater capacity to respond to oxidative stress rather than pH changes.

CONCLUSION

In this study, we analyzed two strains as a proof-of-concept that, with proper controls for annotation and the identification of peptide hits, an LC-MS-based proteomics approach is useful for detecting differences in *H. pylori* proteomes. Importantly, we want to stress that the high degree of strain variability among *H. pylori* protein sequences requires each strain to be compared to its own sequences with correct annotations for proper detection and accurate quantification of surface proteins.

AUTHOR CONTRIBUTIONS

All authors (RK, KT, SH, DK, CS, ÅS, AK, and IN) contributed to the writing of this manuscript, and all authors have given approval to the final version of this manuscript.

ACKNOWLEDGMENTS

We thank Professor Rainer Haas, University of Munich, for sharing the P12 strain with us. The Proteomics Core

REFERENCES

- Ahn, H. J., and Lee, D. S. (2015). *Helicobacter pylori* in gastric carcinogenesis. *World J. Gastrointest. Oncol.* 7, 455–465. doi: 10.4251/wjgo.v7.i12.455
- Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402
- Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E., and Trust, T. J. (2000). Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect. Immun.* 68, 4155–4168. doi: 10.1128/IAI.68.7.4155-4168.2000
- Brown, L. M. (2000). *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol. Rev.* 22, 283–297. doi: 10.1093/oxfordjournals.epirev.a018040
- Choi, H., Fermin, D., and Nesvizhskii, A. I. (2008). Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* 7, 2373–2385. doi: 10.1074/mcp.M800203-MCP200
- Chooneea, D., Karlsson, R., Encheva, V., Arnold, C., Appleton, H., and Shah, H. (2010). Elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilising a lipid-based protein immobilization technique. *BMC Microbiol.* 10:44. doi: 10.1186/1471-2180-10-44
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., et al. (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585. doi: 10.1126/science.1080857
- Fischer, W., Windhager, L., Rohrer, S., Zeiller, M., Karnholz, A., Hoffmann, R., et al. (2010). Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res.* 38, 6089–6101. doi: 10.1093/nar/gkq378
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Jabbour, R. E., Wade, M. M., Deshpande, S. V., Stanford, M. F., Wick, C. H., Zulich, A. W., et al. (2010). Identification of *Yersinia pestis* and *Escherichia coli* strains by whole cell and outer membrane protein extracts with mass spectrometry-based proteomics. *J. Proteome Res.* 9, 3647–3655. doi: 10.1021/pr100402y
- Janzon, A., Bhuiyan, T., Lundgren, A., Qadri, F., Svenssonholm, A. M., and Sjoling, A. (2009). Presence of high numbers of transcriptionally active *Helicobacter pylori* in vomitus from Bangladeshi patients suffering

Facility at Sahlgrenska Academy, University of Gothenburg (<http://proteomics.cf.gu.se/Proteomics>), is acknowledged for providing expertise and MS proteomics analyses. The purchase of the LTQ-OrbitrapXL was made possible through grant KAW2007.0118 from the Knut and Alice Wallenberg Foundation. The Swedish Research Council (VR-2013-4504) is acknowledged for financial support of IN. This manuscript was edited by the Office of Grants and Scientific Publications at the University of Arkansas for Medical Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01757/full#supplementary-material>

FIGURE S1 | High reproducibility was observed when comparing different individual injections and LPI lanes with a minimum Spearman rank correlation between 0.83 and 0.94. (Top left) Three different LPI sample lanes (Lanes 1, 2, and 3) for *H. pylori* D (corresponding to Nic25_A) and the corresponding three injections (D1.1, D1.2, D1.3; D2.1, D2.2, D2.3; D3.1, D3.2, and D3.3) for each LPI sample lane. (Top right) Corresponding LPI sample lanes (Lanes 1, 2, and 3) for *H. pylori* P12 and the corresponding three injections (P1.1, P1.2, P1.3; P2.1, P2.2, P2.3; P3.1, P3.2, and P3.3) for each LPI sample lane.

Supporting File 1 | Core genome of Nic25_A and P12.

Supporting File 2 | Unique genes of Nic25_A.

Supporting File 3 | Unique genes of P12.

Supporting File 4 | Processed data MaxQuant.

Supporting File 5 | Processed data TMT.

Supporting File 6 | Prokka translation.

Supporting File 7 | Unique proteins found in Nic25_A.

Supporting File 8 | Unique proteins found in P12.

- from acute gastroenteritis. *Helicobacter* 14, 237–247. doi: 10.1111/j.1523-5378.2009.00692.x
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703
- Karlsson, R., Davidson, M., Svensson-Stadler, L., Karlsson, A., Olesen, K., Carlsohn, E., et al. (2012). Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. *J. Proteome Res.* 11, 2710–2720. doi: 10.1021/pr2010633
- Momynaliev, K. T., Kashin, S. V., Chelysheva, V. V., Selezneva, O. V., Demina, I. A., Serebryakova, M. V., et al. (2010). Functional divergence of *Helicobacter pylori* related to early gastric cancer. *J. Proteome Res.* 9, 254–267. doi: 10.1021/pr900586w
- Nookae, I., Thorell, K., Worah, K., Wang, S., Hibberd, M. L., Sjovall, H., et al. (2013). Transcriptome signatures in *Helicobacter pylori*-infected mucosa identifies acidic mammalian chitinase loss as a corpus atrophy marker. *BMC Med. Genomics* 6:41. doi: 10.1186/1755-8794-6-41
- Park, J. W., Song, J. Y., Lee, S. G., Jun, J. S., Park, J. U., Chung, M. J., et al. (2006). Quantitative analysis of representative proteome components and clustering of *Helicobacter pylori* clinical strains. *Helicobacter* 11, 533–543. doi: 10.1111/j.1523-5378.2006.00456.x
- Paulo, J. A., Kadiyala, V., Banks, P. A., Conwell, D. L., and Steen, H. (2013). Mass spectrometry-based quantitative proteomic profiling of human pancreatic and hepatic stellate cell lines. *Genomics Proteomics Bioinformatics* 11, 105–113. doi: 10.1016/j.gpb.2013.01.009
- Repetto, O., Zanussi, S., Casarotto, M., Canzonieri, V., De Paoli, P., Cannizzaro, R., et al. (2014). Differential proteomics of *Helicobacter pylori* associated with autoimmune atrophic gastritis. *Mol. Med.* 20, 57–71. doi: 10.2119/molmed.2013.00076
- Resende, T., Correia, D. M., Rocha, M., and Rocha, I. (2013). Re-annotation of the genome sequence of *Helicobacter pylori* 26695. *J. Integr. Bioinform.* 10:233. doi: 10.2390/biecoll-jib-2013-233
- Schwanhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Solis, N., and Cordwell, S. J. (2011). Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* 11, 3169–3189. doi: 10.1002/pmic.201000808
- Suerbaum, S., and Achtman, M. (2004). *Helicobacter pylori*: recombination, population structure and human migrations. *Int. J. Med. Microbiol.* 294, 133–139. doi: 10.1016/j.ijmm.2004.06.014
- Suerbaum, S., and Josenhans, C. (2007). *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* 5, 441–452. doi: 10.1038/nrmicro1658
- Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi: 10.1021/ac0262560
- Thorell, K., Hosseini, S., Palacios Gonzales, R. V., Chaotham, C., Graham, D. Y., Paszat, L., et al. (2016). Identification of a Latin American-specific BabA adhesin variant through whole genome sequencing of *Helicobacter pylori* patient isolates from Nicaragua. *BMC Evol. Biol.* 16:53. doi: 10.1186/s12862-016-0619-y
- Voss, B. J., Gaddy, J. A., McDonald, W. H., and Cover, T. L. (2014). Analysis of surface-exposed outer membrane proteins in *Helicobacter pylori*. *J. Bacteriol.* 196, 2455–2471. doi: 10.1128/JB.01768-14
- Wen, Y., Marcus, E. A., Matruthum, U., Gleeson, M. A., Scott, D. R., and Sachs, G. (2003). Acid-adaptive genes of *Helicobacter pylori*. *Infect. Immun.* 71, 5921–5939. doi: 10.1128/IAI.71.10.5921-5939.2003
- Yahara, K., Kawai, M., Furuta, Y., Takahashi, N., Handa, N., Tsuru, T., et al. (2012). Genome-wide survey of mutual homologous recombination in a highly sexual bacterial species. *Genome Biol. Evol.* 4, 628–640. doi: 10.1093/gbe/evs043
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249
- Zhang, Z. W., Dorrell, N., Wren, B. W., and Farthing, M. J. (2002). *Helicobacter pylori* adherence to gastric epithelial cells: a role for non-adhesive virulence genes. *J. Med. Microbiol.* 51, 495–502. doi: 10.1099/0022-1317-51-6-495

Conflict of Interest Statement: Roger Karlsson and Anders Karlsson are part-time employees at Nanoxis Consulting AB that holds the patent rights for the LPI technology described in this manuscript. All the other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Karlsson, Thorell, Hosseini, Kenny, Sihlbom, Sjöling, Karlsson and Nookae. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.