Supplementary materials

Anna Zawilak-Pawlik *et al.*: "DiaA/HobA and DnaA: A Pair of Proteins Co-evolved to Cooperate During Bacterial Orisome Assembly"

1. Supplementary Results

Origin, phylogeny and taxonomic distribution of DiaA and HobA proteins

DiaA and HobA exhibit low sequence homology and therefore they were considered as non-related proteins. Independent experimental approaches were necessary to reveal that HobA and DiaA structures and function in *H. pylori* and *E. coli* orisome assembly are similar. This raised intriguing questions concerning the phylogeny of the two proteins and distribution of HobA/DiaA-like proteins among different prokaryotic species.

According to the CDD database, the domain of DiaA proteins (PRK10886) is a member of the SIS (Sugar ISomerase) superfamily (cl00389) which also includes 15 curated domain families. SIS domains are found in many phosphosugar isomerases, phosphosugar binding proteins and proteins regulating the expression of genes involved in synthesis of phosphosugars. Our BLAST searches of the non-redundant GenBank database using as a query *E. coli* DiaA protein followed by CDD analyses identified its closest homologues among phosphoheptose isomerases (GmhA). These analyses enabled us to identify as many as 916 non-redundant sequences representing DiaA/GmhA proteins. The sequences show a wide

range of taxonomic distribution and are represented in all main bacterial groups (Table SI). Some significant homologues were also found in Euryarchaeota. Out of 1121 complete prokaryotic genome sequences, 502 contained 714 GmhA/DiaA proteins (Table SI).

Clustering analysis performed on 916 DiaA/GmhA homologues (Fig. S4) distinguished several compact clusters created exclusively or in a majority by representatives of Actinobacteria, Firmicutes, Fusobacteria, Chlorobi, ε-Proteobacteria, β-Proteobacteria and γ-Proteobacteria. The last two groups form two or at least three clear clusters, respectively, which may reflect protein subdivision into two protein groups: GmhA, involved in phosphosugar metabolism, and DiaA, involved in initiation of chromosome replication. Indeed, one of the γ-Proteobacteria clusters, marked in Fig. S3, includes sequences that likely represent DiaA proteins. Their 87 sequences found as the top hit the PRK10886 domain annotated as DnaA initiator-associating protein DiaA in the CDD database. Among these sequences are also experimentally characterized DiaA proteins from *E. coli* and *Photobacterium profundum*. Since the DiaA domain record was described as provisional and was constructed solely on the basis of representatives of *Enterobacteriaceae*, the statistical domain characteristic (expressed by a position-specific scoring matrix) is likely too specific and is not able to recognize more distant DiaA proteins from other taxa. We marked in the graph sequences that found the DiaA domain as the second or third hit in CDD searches (30 and 33 cases, respectively). When we did this we found that the range of DiaA proteins became wider, including also some distant sequences of γ-Proteobacteria, members of β-Proteobacteria and Rhodobacterales bacterium HTCC2255 from α-Proteobacteria. Therefore, it cannot be excluded that sequences located in the wide neighborhood of the typical DiaA proteins also fulfill functions associated with initiation of replication. Results obtained by Keyamura *et al.*[1] indicated that Leu190 and Phe191 of *E. coli* DiaA play a crucial role in its association with DnaA. For this reason, it could be used to distinguish DiaA from GmhA.

Among 916 DiaA/GmhA proteins we identified 286 sequences containing conserved Leu-Phe residues and 86 sequences with Leu-Leu corresponding to *E. coli* DiaA residues (Table SIII). The sequences containing these motifs were widespread in many different groups including Euryarchaeota. However, these motifs were observed in typical GmhA proteins, so the presence of these residues is not likely to be a sufficiently distinctive feature of DiaA proteins.

In contrast to DiaA/GmhA proteins, HobA homologues are restricted exclusively to ε-Proteobacteria. We identified 48 complete HobA sequences in GenBank representing 29 species and 11 genera. This very narrow taxonomic distribution and origin of this protein family is very enigmatic. However, the first 22 structural homologues identified by DALI[2] represent DiaA/GmhA proteins. Interestingly, the highest Z-score (=12.9) was gained by three structures among which one belongs to DiaA (2yva-A). Although simple online BLAST searches of GenBank did not reveal any distant homologues to HobA, more sensitive searches made by ssearch35 versus DiaA/GmhA database identified homologues for 15 HobA sequences at E-value equal to or lower than the threshold 0.05. Five of them found DiaA/GmhA homologues from γ-Proteobacteria with E-value < 0.005, identity up to 25% and similarity up to 62% on the whole alignment. These structural and sequence searches together with experimental analyses showing functional similarity between HobA and DiaA in orisome formation suggest that HobA and DiaA proteins may be somehow related. To test this hypothesis we performed phylogenetic analyses on representatives of HobA and GmhA proteins including *E. coli* DiaA. As an outgroup we chose members of etherases, the family most closely related to GmhA and belonging to the same superfamily. All four approaches, three Bayesian (Fig. S4) and one maximum likelihood, gave very similar general topology. All of them placed the strongly supported diverged lineage of HobA among GmhA members belonging to γ-Proteobacteria including *E. coli* and *P. profundum* DiaA as well as sequences that found the DiaA domain as the second or third hit in CDD searches. The position of the

HobA clade is strongly supported at two nodes with posterior probabilities higher than 0.9, especially for the method with the covarion model (Fig. S4). This model assumes variation of the substitution rate at a site across time; therefore it is the most appropriate for phylogenetic analyses of very divergent sequences such as HobA versus GmhA.

2. Supplementary Materials and Methods

Construction of H. pylori hobA deletion mutants complemented by plasmid encoded diaA

diaA was amplified by PCR with the primer set AZP05-AZP02 (Table SII). The *diaA* PCR product was digested with SpeI/BamHI and inserted into pILL2150 and pILL2157 ([3] and Table I), giving plasmids pAZP03 and pAZP04. The *H. pylori* N6 was transformed with pAZP03 and pAZP04. The transformants were selected on chloramphenicol BA plates, leading to the selection of the *H. pylori* mutants N6 pAZP03 and N6 pAZP04. These strains were further used to construct the mutants with chromosomal deletions of *hobA* using pILL2283[4] or *hobA* exchanged to *cmychobA* using pAZP07-4. Transformed cells were plated on BA plates supplemented with chloramphenicol, kanamycin and 1 mM IPTG. The *H. pylori* mutants resistant to both antibiotics were analyzed by PCR with primers spanning the chromosomal region subjected to homologous recombination.

Bacterial two-hybrid analysis (BTH) vector construction

All genes were cloned into plasmids pAZP12, pAZP13 and pAZP14 (Table I and SIII). They are derivatives of pUT18[5], pUT18C[5] and pKT25[5], respectively, modified by insertion of hybridized AZP16 and AZP17 oligonucleotides between XbaI and KpnI sites (Table SII). The inserted linker modified the reading frame of the genes cloned into the BamHI restriction site and, additionally, introduced the *Xho*I restriction site into MCS of the original vectors. *H. pylori dnaA* was excised from pET28a(+)dnaA[5] using BamHI/XhoI and cloned into pAZP14, giving pAZP18. *hobA* was excised from pILL2286[4] by BamHI/XhoI and cloned into pAZP13 and pAZP14, giving pAZP16 and pAZP19, respectively. The *H. pylori* 5' *dnaA* region (1-336 bp), encoding domains I-II, was amplified by AZP18 and AZP19 primers and cloned into the BamHI site of pAZP12, giving pAZP23. *E. coli dnaA* was amplified with the primer pair AZP23-AZP24, digested with BamHI/SalI and cloned into BamHI/XhoI sites of pAZP13 and pAZP14, giving pAZP24 and pAZP26. $Ec^{I}Hp^{II-IV}$DnaA, excised from pAZP28

(see main text and Table I) with EcoRI, was cloned into pAZP24, giving pAZP30. *diaA* was excised from pAZP10 with BamHI/SalI and cloned into BamHI/XhoI sites of pAZP13 and pAZP14, giving pAZP25 and pAZP27.

Purification of proteins

EcDnaA was overexpressed from pdnaA 116 in WM2305 and purified as described previously[7]. HpDnaA was purified similarly as described in Zawilak-Pawlik *et al*.[8]; however, the overproduction of the protein was achieved in *E. coli* DH5alpha from pAZP29 plasmid (Table SIII). pAZP29, a pILL2150 derivative, contains *H. pylori dnaA*, amplified by AZP40 and AZP39, cloned into NdeI/BamHI restriction sites. $Ec^{I}Hp^{II-IV}$DnaA was purified from IPTG-induced DH5alpha containing pAZP28 (Table SIII) using the purification protocol exactly as for HpDnaA isolation.

diaA was expressed from pAZP10 (Table SIII). pAZP10 is a derivative of pET151/D-TOPO (Invitrogen), obtained by insertion of *diaA* amplified with the primer pair AZP01 and AZP02 (Table SII). 6HisDiaA was expressed and purified in a similar way as 6HisHobA was[4]. Fractions containing pure 6HisDiaA were pooled and subjected to TEV protease cleavage for 16 hrs at $4^{O}C$ in the presence of 1 mM DTT and 0.5 mM EDTA followed by buffer exchange to 50 mM Tris-HCl pH 8.0, 300 mM NaCl and 5% glycerol by gel filtration. Afterwards samples were passed through a Co^{2+}-NTA (Talon) column to separate DiaA from His-tagged proteins (i.e. uncleaved proteins, cleaved hexahistidine peptide, and TEV protease). For long-term storage, DiaA was supplemented with 50% glycerol and kept at $-20^{o}C$.

hobA was digested from pILL2280[4] with BamHI/XhoI and ligated to pAZP10 restricted with BamHI/SalI to give pAZP11 (Table SIII). 6HisHobA purification and TEV cleavage were carried out similarly to 6HisDiaA.

Collection of sequences and alignment construction

Analyzed sets of sequences were obtained from BLAST searches across the non-redundant GenBank database of protein sequences and the database of 1121 completely sequenced prokaryotic genomes. To verify the BLAST results and determine domain content in found homologues we additionally searched the Conserved Domain Database (CDD)[9]. We also performed highly sensitive searches between the set of HobA homologues versus DiaA/GmhA databases using ssearch35 from FASTA package[10] to evaluate the similarity between these proteins at the sequence level. Statistical parameters of the searches were estimated from 10,000 shuffled copies of each library sequence, which provides greater statistical accuracy. Final alignments were obtained in 3D-Coffee[11]) using the slow pair method and the TM-align program[12] to align 3D protein structures. This procedure is especially recommended for highly diverged sequences. The resulting alignments were edited manually and corrected in Jalview[13].

CLANS clustering and phylogenetic analyses

Clustering analysis was performed using CLANS (Cluster Analysis of Sequences) software[14], which uses a version of the Fruchterman–Reingold graph layout algorithm to visualize pairwise sequence similarities in either two-dimensional or three-dimensional space. The program performs all-against-all BLAST searches and calculates pairwise attraction values based on P-values of HSPs (high scoring segment pairs). Analyzed sequences are represented in the graph by vertices which are connected by edges reflecting attractive forces proportional to the negative logarithm of the HSP's P-value. To increase sensitivity, we used word-size = 1 in the pairwise BLASTP searches.

Phylogenetic trees were inferred by the maximum likelihood method in Phyml[15] and by the Bayesian approach in PhyloBayes[16] software. In Phyml, we used the LG+Γ(5)+F amino acid substitution model as proposed by the program ProtTest 2.4[17]. We assumed the best

heuristic search algorithms in Phyml, i.e. NNI and SPR. Edge support was assessed by bootstrap analysis with 1000 replicates and by the approximate likelihood ratio test (aLRT) based on χ^2 and the Shimodaira-Hasegawa-like procedure. The minimum of these two support values was shown at nodes in the presented tree. In PhyloBayes, we carried out three types of analyses. In the first one, we assumed the CAT-Poisson+Γ(5) model with number of components, weights and profiles inferred from the data. In the second approach we applied the LG+Γ(5) model, and in the third one, the CAT-Poisson model including Tuffley and Steel's covarion model and rates across sites modeled using a Dirichlet process. The covarion model assumed variation of the substitution rate at a site across time (heterotachy) and is especially appropriate for very divergent sequences. Two independent Markov chains were run for 1,000,000 cycles in the first approach, and 200,000 cycles in the second and third approaches. After getting a convergence, the last 500,000 (first approach) and 25,000 trees (the second and third approaches) from each chain were collected to compute the posterior consensus.

3. **Supplementary References**

1. Keyamura K, Fujikawa N, Ishida T, Ozaki S, Su'etsugu M, Fujimitsu K, Kagawa W, Yokoyama S, Kurumizaka H, Katayama T. 2007. The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev*. **21**:2083-99.
2. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. 2008. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**:2780-2781.
3. Boneca IG, Ecobichon C, Chaput C, Mathieu A, Guadagnini S, Prévost MC, Colland F, Labigne A, de Reuse H. (2008) Development of inducible systems to engineer conditional mutants of essential genes of *Helicobacter pylori*. *Appl Environ Microbiol.* 74,2095-102.

4. Zawilak-Pawlik A., Kois A, Stingl K, Boneca I.G, Skrobuk P, Jakimowicz P, Lurz R, Zakrzewska-Czerwińska J, Labigne A. (2007) HobA - a novel protein involved in initiation of chromosomal replication in *Helicobacter pylori. Mol. Microbiol.* 65, 979-994.

5. Zawilak-Pawlik AM, Kois A, Zakrzewska-Czerwinska J. (2006) A simplified method for purification of recombinant soluble DnaA proteins. Protein Expr Purif. 48:126-33

6. Karimova, G., J. Pidoux, A. Ullmann and D. Ladant. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. U.S.A*. **95**:5752-5756.

7. Krause M, Rückert B, Lurz R, Messer W. 1997. Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J Mol Biol.* **274**:365-80.

8. Zawilak A, Jakimowicz P, Durrant M, Backert S, Zakrzewska-Czerwinska J. (2003) DNA binding specificity of the replication initiator protein, DnaA from *Helicobacter pylori. J. Mol. Biol*. **334**: 933-947

9. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH. 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res*. (Database issue):D205-10.

10. Pearson W.R. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods in Molecular Biology*. **132**: 185–219.

11. O'Sullivan O, Suhre K, Abergel C, Higgins DG, Notredame C. 2004. 3DCoffee combining protein sequences and structures within multiple sequence alignments. *J Mol Biol.* **340**:385-95.

12. Zhang Y., Skolnick J. 2005. TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Research* **33**: 2302-2309

13. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M., Barton, G.J. 2009. Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench" *Bioinformatics* **25**: 1189-1191

14. Frickey T., Lupas A.N. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**:3702-3704

15. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. **52**:696-704

16. Lartillot N, Philippe H. 2004. A Bayesian mixture model for across site heterogeneities in the aminoacid replacement process. *Mol Biol Evol.* **21**:1095-1109

17. Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics.* **21**:2104-2105.

4. **Supplementary Tables**

Bio – biotin ; Pho – phosphate

Table SI. Taxonomic distribution of DiaA/GmhA homologues according to searches in GenBank and complete prokaryotic genome database.

Taxon	**Number of homologues found in GenBank**	**Number of homologues with:**[1]		**Number of genomes in which a given number of homologues was found:**		
		Leu-Phe	**Leu-Leu**	**1**	**2**	**3**
Acidobacteria	7			2		1
Actinobacteria	63	9	2	26	5	3
Aquificales	10			7		
Bacteroidetes/Chlorobi	38	3	1	13	2	1
CFB-group	23	3	1	4		1
Chlorobi	15			9	2	
Chlamydiae/Verrucomicrobia	15	3		3	2	
Chlamydiae	2	2		2		
Lentisphaerae	2	1				
Verrucomicrobia	11		1	1	2	
Chloroflexi	14	3		4	6	
Cyanobacteria	18	8		11	1	
Deferribacteres	2			2		
Deinococcus-Thermus	1		1	1		
Dictyoglomi	2	2		2		
Elusimicrobia	1			1		
Euryarchaeota	11	3		8		
Firmicutes	49	2		10		
Fusobacteria	19	7		2	1	
Gemmatimonadetes	1	1		1		

Nitrospirae	4	4		1		
Planctomycetes	8	1		2		
Proteobacteria	626	231	81	200	166	7
α-Proteobacteria	50	3		21	5	1
β-Proteobacteria	125	10	79	55	17	
γ-Proteobacteria	346	174	1	84	133	5
δ-Proteobacteria	40	32		20	4	1
ε-Proteobacteria	62	12		20	6	
ζ-Proteobacteria	1					
Unclassified	2		1		1	
Spirochetes	11	3		4	4	
Synergistetes	6	4		2		
unclassified Bacteria	9	2			1	
Eudicot *Ricinus communis*[2]	1	NA	NA	NA	NA	NA
In total	**906**			**302**	**188**	**12**

[1]Number of homologues that contain conserved Leu-Phe or Leu-Leu residues corresponding to Leu190 and Phe191 of *E. coli* DiaA.

[2]The *R. communis* protein (XP_002536928) consists of two domains: N-terminal phosphoheptose isomerase (SIS_GmhA) and C-terminal RfaE bifunctional protein domain I. This protein has the closest homologues to two proteins coded by two adjacent genes in the *Methylotenera mobilis* JLW8 (β-Proteobacteria) genome. The domains of *R. communis* protein show 77% and 63% identity to bacterial homologues, respectively. If this sequence is not a contamination artifact, it was likely gained from β-Proteobacteria by horizontal gene transfer, which was accompanied by a gene fusion.

Table SII Oligonucleotides used in this study

name	Sequence 5'-3'
AZP01	CACCGGATCCATGCAAGAAAGAATTAAAGCTTG
AZP02	GTCGACTTAATCATCCTGGTGAGGGAAAAG
AZP03	CGGGATCCGGAGGGAATAATGCAAGAAAGAATTAAAGCTTG
AZP04	CGGGATCCTTAATCATCCTGGTGAGGGAAAAG
AZP05	ACTAGTATGCAAGAAAGAATTAAAGCTTG
AZP06	CGGGATCCTCATTGATCTAATTGATACACCGC
AZP07	CCCAAGCTTGATCGGCAGAGACGGCAAAAC
AZP08	CGGGATCCTGACTAACTAGGAGGAATAAATG
AZP09	TTCTGCAGCATTATTCCCTCCAGGTACTA
AZP10	AACTGCAGATGGAACAAAAATTAATTAGCGAAGAAGATTTAGGCGGCGGTACCCC
AZP11	GGGGTACCGCCGCCTAAATCTTCTTCGCTAATTAATTTTTGTTCCATCTGCAGTT
AZP12	GGGGTACCATGAAAAATTTCTACGATTGGATC
AZP13	CGGAATTCCTAGAGAGAGATTCGCATCGT
AZP16	Pho-CTAGAGGGATCCCTCGAGCCCCGGGTAC
AZP17	Pho-CCGGGGCTCGAGGGATCCCT
AZP18	CGGGATCCATGGATACCAACAACAATATTG
AZP19	CGGGATCCGTCTTTGACGCTCGTTTTTATGGC
AZP23	GGATCCCATATGTCACTTTCGCTTTGGCAGC
AZP24	GGATCCGTCGACTTACGATGACAATGTTCTGATTAAA
AZP25	GGTGTTTTCACGAGCACTTCAC
AZP26	CGGATAACAATTAATTGTGAGCGG
AZP27	GTGAGCATGGAGGGCGTGGAAG

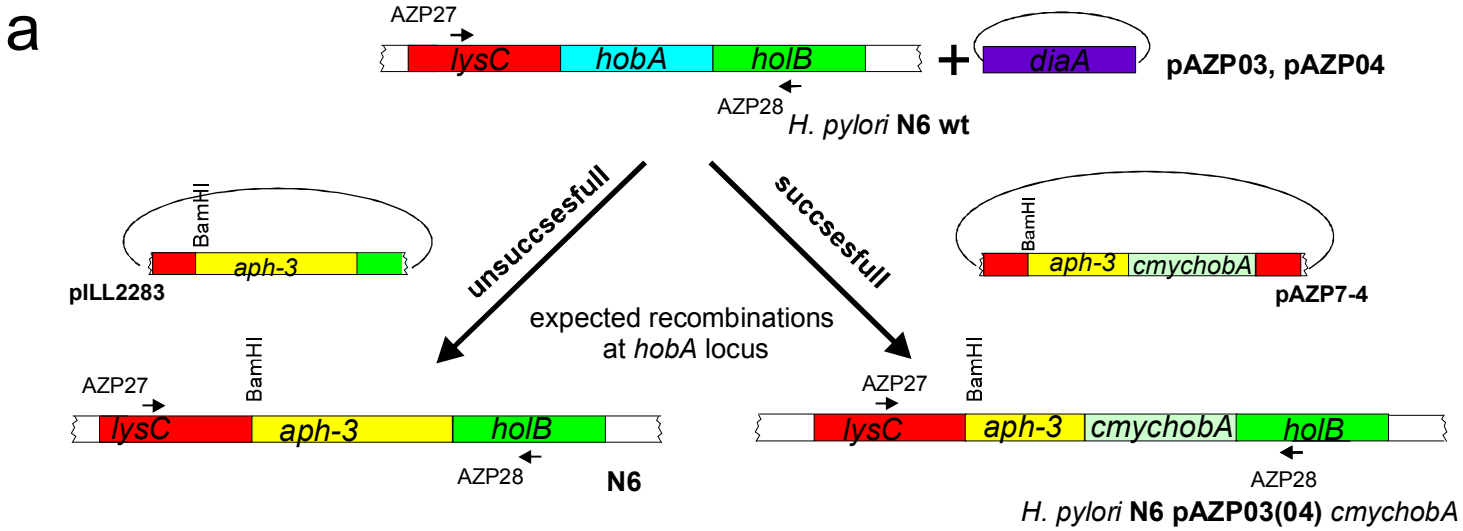
AZP28	CTCTGGAATCAAAATTTTCTTTGTC
AZP29	Bio- TGTGATCTCTTATTAGGATC
AZP30	ACTCAAATAAGTATACAGATC
AZP31	Bio- CCAGCGCAAAGCAGCATGAAAATC
AZP32	CAATATTGTTGTTGGTATCCATGG
AZP36	TTCTGCAGGGAGGAACATATGTCACTTTCGCTTTGGCAGC
AZP37	CGGGATCCACATGCATGCTTTGGTGCCGACTTCAAAACGC
AZP38	ACATGCATGCATCCAAATTAACGCCCAATCTAA
AZP39	CGGGATCCTCATTCACTTGAATTGAAAGCGGT
AZP40	GAAAAGCATATGGATACCAACAACAATATTGAAAAAG

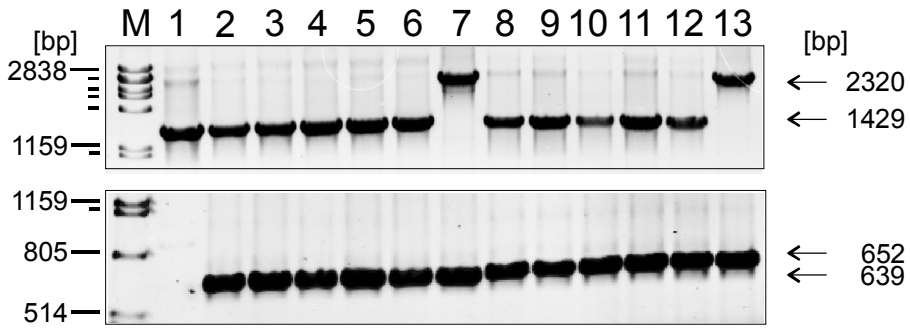
Bio – biotin; Pho – phosphate

Table SIII Strains and plasmids used/described in supplementary work

Strain/ plasmid	**Relevant genotype/feature**	**Reference/ source**
N6 pAZP03	N6 pAZP03 *(diaA$^+$)*	This work
N6 pAZP04	N6 pAZP03 *(diaA$^+$)*	This work
N6 pAZP03 *c-mychobA*	N6 *hobA*Ω*c-mychobA,* pAZP03 *(diaA$^+$)*	This work
N6 pAZP04 *c-mychobA*	N6 *hobA*Ω*c-mychobA,* pAZP04 *(diaA$^+$)*	This work
pILL2283	pUC18 derivative containing *hobA* 5' and 3' flanking regions, used for allelic exchange of *hobA* to *aphA-3*, Amp^R, Kan^R	4
pILL2286	pET28a(+), expressing *hobA* with a hexahistidine tag at N-terminus	4

pUT18C	BACTH expression vector, Amp^R	6
pET28a(+)	IPTG inducible *E. coli* expression vector, Kan^R	Novagen
pET151/D-TOPO	expression vector, IPTG inducible, Amp^R	Invitrogen
pdnaA116	*E. coli* wt DnaA overexpression vector, IPTG inducible, Amp^R	7
pAZP03	pILL2150 derivative, expressing *diaA*	This work
pAZP10	pET151/D-TOPO derivative, expressing *diaA* with a hexahistidine tag at N-terminus	This work
pAZP11	pET151/D-TOPO derivative, expressing *hobA* with a hexahistidine tag at N-terminus	This work
pAZP12	pUT18 modified by insertion of AZP16 and AZP17 annealed ds oligonucleotide between XbaI and KpnI sites, XhoI site introduced into MCS	This work
pAZP13	pUT18C modified by insertion of AZP16 and AZP17 annealed ds oligonucleotide between XbaI and KpnI sites, XhoI site introduced into MCS	This work
pAZP14	pKT25 modified by insertion of AZP16 and AZP17 annealed ds oligonucleotide between XbaI and KpnI sites, XhoI site introduced into MCS	This work
pAZP24	pUT18C derivative, expressing *E. coli* DnaA with an N-terminal T18	This work
pAZP29	pILL2150 derivative, expressing *H. pylori dnaA*	This work





H. pylori N6	kan/chl resistance	PCR analysis *diaA* plasmid locus	PCR analysis *hobA* chromosomal locus (primers AZP27-AZP28)	*hobA* locus
wt	-	-	1429 bp	wt
pAZP03, *aph3*	+	639 bp*	expected 1737 bp observed 1429 bp	wt
pAZP04, *aph3*	+	639 bp*	expected 1737 bp observed 1429 bp	wt
pAZP03 ,*cmychobA*	+	652 bp#	2320 bp	recombined
pAZP04, *cmychobA*	+	652 bp#	2320 bp	recombined

primers: *AZP25-AZP2, #AZP26-AZP2

Fig. S1. Analysis of *diaA* ability to complement *H. pylori hobA* deletion mutant.

a) Scheme representing the strategy used to delete the chromosomal *hobA* after introduction of a plasmid-borne *diaA* copy. The pAZP03 or pAZP04, encoding IPTG-inducible *diaA*, were successfully introduced into *H. pylori* N6. Then the suicide vectors: pILL2283 or pAZP07-4 were used to delete *hobA* from its chromosomal locus *or introduce cmychobA,* respectively, by recombination between homologous fragments. Successful recombination should lead to deletion of *hobA* or introduction of *cmychobA,* while accompanying insertion of *aph-3* cassette allowed for mutants selection. Transformation with pILL2283 led to introduction of *aph-3* cassette into unknown chromosomal locus leaving *hobA* locus as in the wild type strain. Transformation with pAZP07-4 gave colonies which, after PCR analysis, were shown to be the designed mutants.

b) Analysis of the *H. pylori* N6 mutant strains. Top: agarose gels showing the results of PCR fragments obtained with AZP27-AZP28 primers (upper) and the primer pairs: AZP25-AZP02 and AZP26-AZP02 (lower). Bottom: table summarizing analyzed *H. pylori* mutants.

M, DNA marker ; 1, N6 wt; 2-6, N6 pAZP03 *aph-3*; 7, N6 pAZP03 *cmychobA*; 8-12, N6 pAZP04 *aph-3*, 13, pAZP04 *cmychobA*.

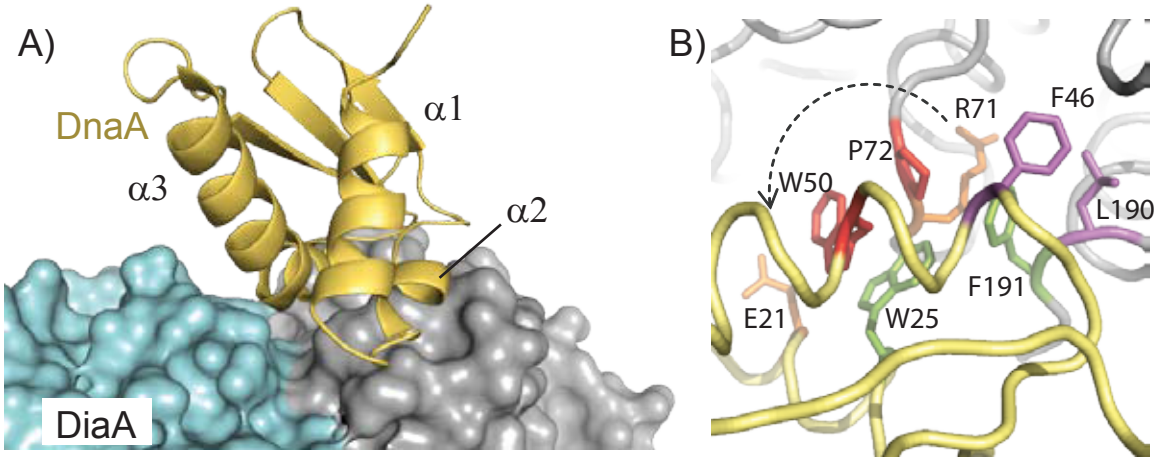


Fig. S2 A putative model of the DiaA/DnaA^{I} complex

A) Picture depicting DiaA crystal structure (surface representation) with subunit A colored in blue and B in grey bound to one EcDnaA^{I} (displayed as a yellow ribbon).

B) Detailed view of the binding interface where the side chains of residues of DiaA and DnaA interacting together are represented in ball and stick and colored identically. This picture illustrates that the model we generated based on HobA/HpDnaA crystal structure is in agreement with the interacting study described in Keyamura *et al* (2009). DiaA residues Arg71, Pro72, Leu190, Phe191 interact with DnaA Glu21, Trp50, Phe46 and Trp25, respectively. Noteworthy is the position of Arg71 side chain from the DiaA ERP loop that, with a 180 degree rotation (shown by an arrow), could form a salt bridge with Glu21.

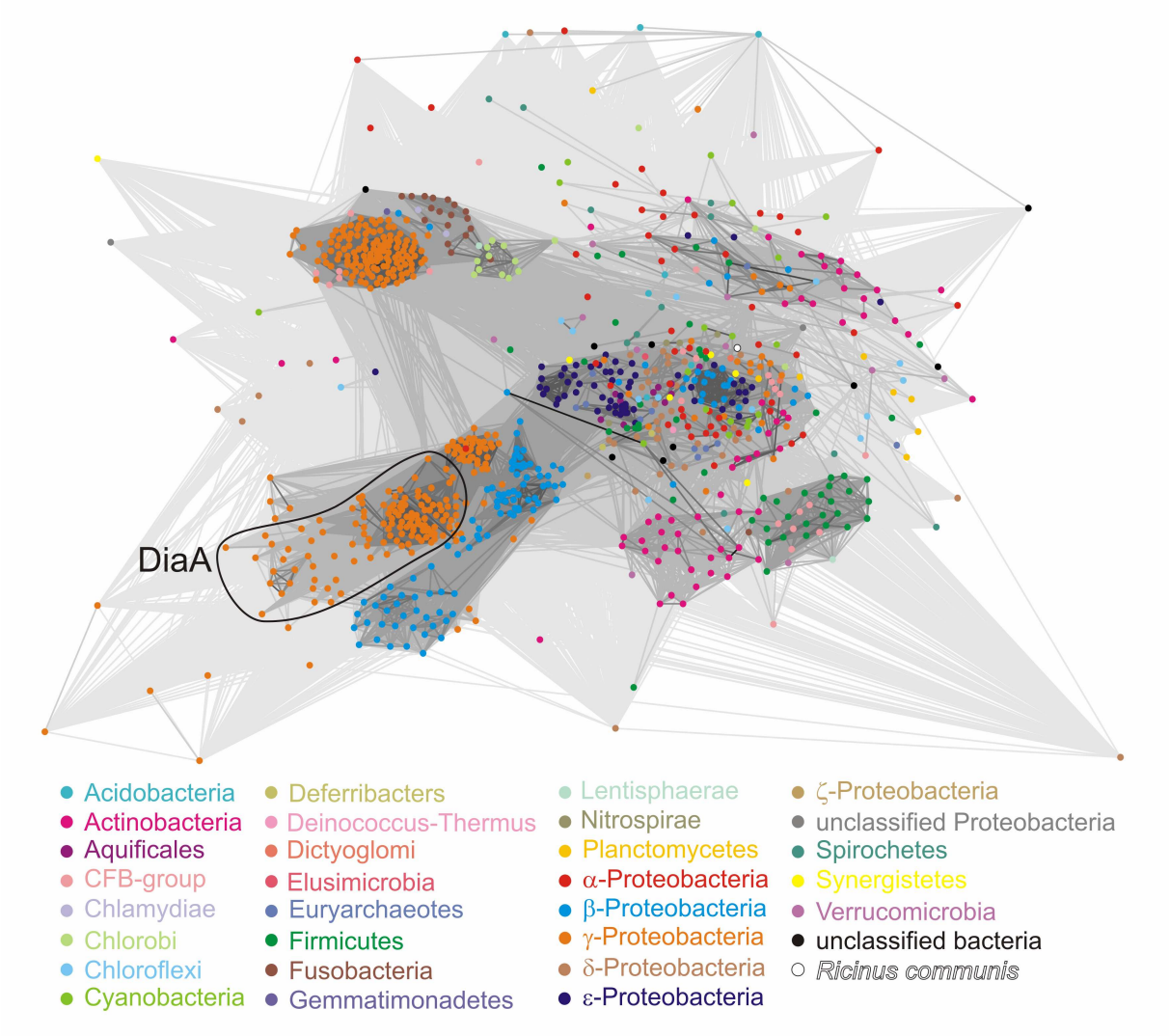


Fig. S3. Cluster analysis of 916 DiaA/GmhA proteins made in CLANS software. Analyzed sequences are represented by vertices connected by edges reflecting attractive forces proportional to the negative logarithm of the HSP's P-value. The grayness intensity of the connections is proportional to these forces. The outlined clusters contains sequences that found as the best top hit the domain PRK10886 annotated as DnaA initiator-associating protein DiaA in CDD database.

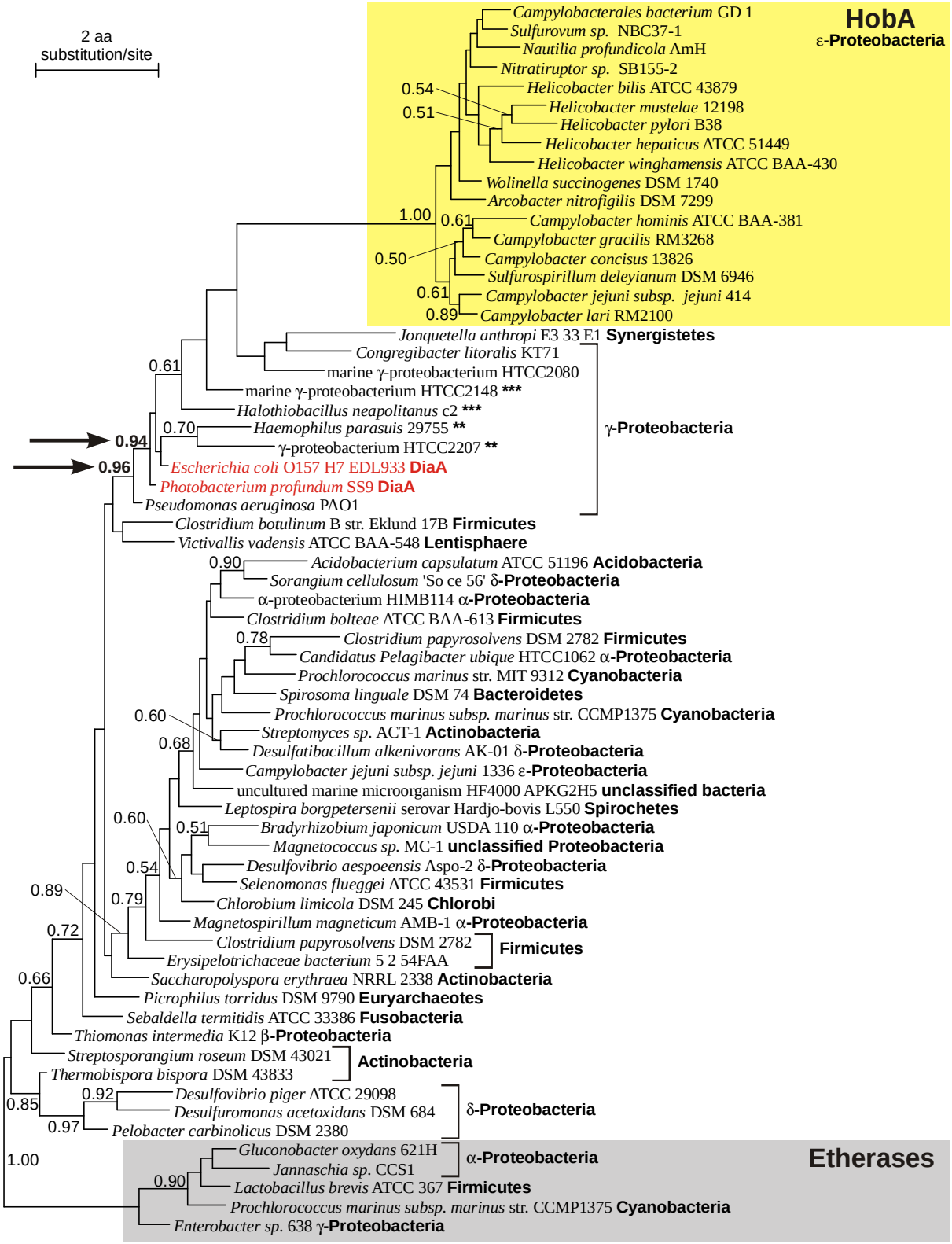


Fig. S4. The Bayesian tree obtained in PhyloBayes under the covarion model for representatives of GmhA and HobA proteins (placed in the yellow rectangle). Etherases (in the grey rectangle) were chosen as an outgroup. *E. coli* and *P. profundum* DiaA are indicated in red while sequences that found DiaA domain as the second and third top CDD hit

are marked by ** two and *** three asterisk, respectively. Numbers at nodes correspond to posterior probabilities (PP). The PP values lower or equal to 0.50 were omitted or indicated by a dash "-". Arrows indicate high PP values supported position of HobA clade among γ-Proteobacteria GmhA proteins.