

Supplementary Materials

Table 1S. Expressed “hypothetical” genes of *H. influenzae* strain Rd KW20

<i>Gene name</i> ^a		<i>Gene expression</i>		<i>Protein expression</i> ^b		<i>Essentiality</i> ^c	
<i>H. infl.</i>	<i>E. coli</i>	<i>H. infl.</i>	<i>E. coli</i>	+O ₂	-O ₂	<i>H. infl.</i>	<i>E. coli</i>
HI0017	<i>yfiD</i>	+	+	+	+	N	Y
HI0056	<i>ygdQ</i>	+	+	+	-	Y	N
HI0105	<i>ybgI</i>	+	+	+	-	N	N
HI0118	<i>ygdL</i>	+	+	±	+	Y	N
HI0119	<i>yebL</i>	+	+	+	+	Y	N
HI0121	<i>mpl</i>	+	+	+	-	Y	N
HI0146	<i>viaO</i>	+	+	+	+	Y	N
HI0147	n/a	+	n/a	+	±	nd	n/a
HI0148	<i>yjhT</i>	+	+	+	+	Y	N
HI0227	<i>yhchH</i>	+	+	+	+	nd	N
HI0241	<i>yajC</i>	+	+	+	-	nd	nd
HI0246	n/a	+	n/a	+	+	nd	n/a
HI0271	<i>yabhH</i>	+	+	±	+	N	N
HI0315	<i>yebC</i>	+	+	+	±	N	Y
HI0367	<i>yfgA</i>	+	+	±	+	Y	N
HI0370	<i>yfgM</i>	+	+	+	±	N	N
HI0396	<i>ycfD</i>	+	+	+	±	N	Y

HI0409	<i>yebA</i>	+	+	+	±	N	N
HI0442	<i>ybaB</i>	+	+	+	+	N	N
HI0466	<i>ygfZ</i>	+	+	-	+	Y	nd
HI0467	<i>yicC</i>	+	+	+	-	Y	nd
HI0491	<i>ygaG</i>	+	+	-	+	N	nd
HI0520	<i>yjwW</i>	+	+	±	+	N	N
HI0521	<i>yjiI</i>	+	+	+	-	nd	N
HI0595	<i>yqeA</i>	+	+	+	+	N	N
HI0668	<i>yiiU</i>	+	+	+	+	nd	Y
HI0700	<i>yjgD</i>	+	+	+	-	N*	nd
HI0719	<i>yjgF</i>	+	+	+	+	N*	Y
HI0753	<i>ydhH</i>	+	+	+	±	Y	N
HI0828	<i>yciI</i>	+	+	+	+	nd	N
HI0847	<i>yifE</i>	+	+	+	-	N	N
HI0912	n/a	+	n/a	+	±	Y	n/a
HI0918	<i>yaeL</i>	+	+	+	±	Y	N
HI1034	<i>yajQ</i>	+	+	+	+	N	N
HI1053	n/a	+	n/a	±	+	Y	n/a
HI1056	n/a	+	n/a	+	±	N*	n/a
HI1085	<i>yrbD</i>	+	+	+	±	N	N
HI1159	<i>ybbN</i>	+	+	+	±	N	N
HI1168	<i>yeeX</i>	+	+	+	-	N	N
HI1236	<i>ydgA</i>	+	+	+	±	nd	E
HI1349	<i>dps</i>	+	+	+	+	Y	N
HI1426	<i>ydaA</i>	+	+	+	+	N	N
HI1427	n/a	+	n/a	+	±	Y	n/a
HI1492	n/a	+	n/a	+	-		

HI1514	n/a	+	n/a	±	+	nd	n/a
HI1586	n/a	+	n/a	±	+	N*	n/a
HI1603	n/a	+	n/a	+	-	Y	n/a
HI1624	n/a	+	n/a	+	+	N	n/a
HI1647	n/a	+	n/a	+	+	N	n/a
HI1670	<i>proQ</i>	+	+	+	+		
HI1677	<i>ytfE</i>	+	+	±	+	N*	N
HI1681	<i>yccT</i>	+	+	+	+	N	N
HI1709	<i>ygiW</i>	+	+	+	+	nd	N
HI1720	<i>orfA</i>	+	+	+	-	Y	n/a

^a n/a indicates the absence of an *E. coli* ortholog for the given *H. influenzae* gene.

^b ± indicates less than 90% confidence in protein identification; - indicates absence of confidently identified proteins.

^c The data are from ref. [Akerley, 2002 #1; Gerdes, 2003 #209]; Y and N indicate, respectively, essential and nonessential genes; N* indicates genes that, while nonessential, appear to be required for optimal growth; nd - no data.

Table 2S. Sequence-based functional characterization of “conserved hypothetical” genes

<i>Gene</i>	<i>Swiss-Prot accession</i> ^a	<i>Size aa</i>	<i>COG no.</i>	<i>PDB entry</i> ^b	<i>Updated annotation, fold prediction</i>	<i>Annotation category</i> ^c
HI0017	P44455	127	3445	<i>2pfl</i>	Acid-induced glycyl radical enzyme	II
HI0056 ^d	P43932	237	861	-	Tellurium resistance protein TerC	II
HI0105	Q57354	279	327	-	NIF3 homolog, the central CutA-like domain is distantly related to the nitrogen regulator GlnB	III
HI0118	Q57097	261	1179	<i>1jwa</i> <i>1ngv</i>	NAD/FAD-binding enzyme involved in molybdopterin and thiamine biosynthesis	II
HI0119 ^d	P44526 <i>znua_haein</i>	337	803	1nry* <i>1toa</i>	Periplasmic component of an ABC-type metal ion transport system	I
HI0121	P43948 <i>mpl_haein</i>	453	773	1gqy* <i>1jhu</i>	UDP-N-acetylmuramate-L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	I
HI0146	P44542	329	1638	-	Periplasmic component of a TRAP-type dicarboxylate transport system	II
HI0147	P44543	633	3090 1593	-	A fusion of large and small subunits of a TRAP-type dicarboxylate transport system	II
HI0148 ^d	P44544	379	3055	-	Cell surface adhesion protein with a Kelch-like 7-bladed beta-propeller fold	III
HI0241	P44592	97	1862	-	Preprotein translocase subunit YajC	II
HI0271	P44607 <i>djla_haein</i>	288	1076	<i>1hdj</i>	DnaJ domain-containing protein	I
HI0367	Q57065	303	1426	-	Predicted transcriptional regulator with an N-terminal <i>xre</i> -type HTH domain	III
HI0396	P44683	404	2850	**	Cupin superfamily metalloenzyme, possible dioxygenase	III

HI0409	P44693 y409_haein	475	739	**	Membrane-bound metalloendopeptidase, M23/M37 family	I
HI0466	P44000	280	354	**	Predicted aminomethyltransferase related to GcvT	II
HI0467 ^d	P44726	287	1561	-	Uncharacterized stress-induced protein	II
HI0491	P44007 luxs_haein	167	1854	1j6w 1joe	S-ribosylhomocysteinase, LuxS protein involved in autoinducer AI2 synthesis	I
HI0520	P44743	262	1180	-	Pyruvate formate-lyase activating enzyme	II
HI0521	P44744	514	n/a	-	Glycyl radical enzyme, distantly related to pyruvate-formate lyase	III
HI0595	P44769 arcc_haein	310	549	1e19	Carbamate kinase	I
HI0719	P44839	130	251	1qd9	Putative translation inhibitor	II
HI0753	P44861	382	2377	**	Predicted molecular chaperone distantly related to HSP70-fold metalloproteases	II
HI0912	P44074	254	500	1g6q	SAM-dependent methyltransferase	II
HI0918	P44936 yael_haein	443	750	1lcy	Membrane-bound Zn-dependent protease EcfE, contains a central PDZ domain	I
HI1053	Q57498	113	599	1lw1	Carboxymuconolactone decarboxylase	II
HI1056	P71366 t3mh_haein	629	2189	1g60	Adenine-specific DNA methylase Mod	I
HI1084	P45028	214	2854	-	Periplasmic component of an ABC-type transport system involved in resistance to organic solvents	II
HI1085	P45029	167	1463	-	Permease component of an ABC-type transport system involved in resistance to organic solvents	II
HI1159	P43786	217	3118		Thioredoxin domain-containing protein	II
HI1349 ^d	P45173	160	783	1ji5	DNA-binding ferritin-like protein (oxidative damage protectant)	II

HI1426	P44195	340	589	<i>Imjh</i>	UspA-related nucleotide-binding protein	II
HI1427	P44196	278	5266	**	Periplasmic component of an ABC-type Co ²⁺ transport system	II
HI1514	P44236	631	3941	-	Mu-like prophage protein	I
	vg42_haein					
HI1586	P44263	519	1757	-	Na ⁺ /H ⁺ antiporter NhaC	II
HI1603 ^d	P44271	226	1392	-	Phosphate transport regulator	II
HI1624	P44276	227	5266	**	Periplasmic component of an ABC-type Co ²⁺ transport system	II
HI1647	P45293	291	214	**	Pyridoxine biosynthesis enzyme PDX1	II
HI1670	P44286	157	3109	<i>ldvo</i>	Activator of osmoprotectant transporter ProP	I
	proq_haein					
HI1677	P45312	223	2846	**	Regulator of cell morphogenesis and NO signaling	II
HI1720	Q57066	188	2963	**	Transposase	II

^a For proteins that are functionally annotated in SWISS-PROT, accession numbers are followed by the entry names.

^b Structures of *H. influenzae* proteins are shown in bold typeface, those on hold in PDB are marked with asterisks; structures of homologous proteins are in normal font, those covering less than a half of the *H. influenzae* target length are in italics. Two asterisks indicate targets of the structural genomics project, see <http://s2f.carb.nist.gov/>.

^c Annotation groups are as follows: I - SWISS-PROT-based annotation; II – CDD/COG-based annotation; III - annotation based on the sequence analysis performed in this work.

^d These genes were briefly discussed in our earlier paper (9).

Table 3S. Genomic context data for uncharacterized genes

<i>Gene</i>	<i>Putative co-regulated genes</i> ^a		<i>Functional cluster</i>
	<i>Characterized</i>	<i>Uncharacterized</i>	
HI0017 (<i>yfiD</i>)	<i>ung, mutH, seqA, metJ</i>	<i>oapA (ytfB), yjaG, yjgD, yihA, yihD, yejL</i>	I
HI0056 (<i>ygdQ</i>)	<i>mutH, terZ</i>	<i>yegH, BS_yceH, yjgJ</i>	I
HI0105 (<i>ybgI</i>)	<i>dnaG, rpoD</i>	<i>BS_yqf, BS_yqfO, DR0291</i>	
HI0118 (<i>ygdL</i>)	<i>mltA, sufE</i>		
HI0271 (<i>yabH</i>)	<i>nifS, iscA, iscU, fdx, dnaK</i>	<i>yfgA, yfgM, yfhJ, yfhP, yfhQ</i>	II
HI0367 (<i>yfgA</i>)	<i>gcpE, pilF, pgsA</i>	<i>yfgB, yfgM, yfhJ, yfgL</i>	II
HI0370 (<i>yfgM</i>)	<i>gcpE, pilF, rdgC</i>	<i>yfgA, yfgB, yfhJ, yjeT</i>	II
HI0396 (<i>ycfD</i>)	<i>purB, xceA</i>	<i>BS_yxbD, ycfC, yjfF</i>	
HI0409 (<i>yebA</i>)			III
HI0466 (<i>ygfZ</i>)	<i>gmk</i>	<i>yadR, yoaB, yicC</i>	IV
HI0467 (<i>yicC</i>)	<i>gmk, rph, rpoZ</i>	<i>ygfZ</i>	IV
HI0520 (<i>yjjW</i>)	<i>pflB, pflD, nrdD</i>	<i>yjjI</i>	V
HI0521 (<i>yjjI</i>)	<i>pflA, pflC</i>	-	V
HI0753 (<i>ydhH</i>)	-	<i>yfeU, yebA, yacG</i>	III
HI0912	<i>ushA</i>		
HI1159 (<i>ybbN</i>)	-	<i>ybbO, ycaR, PA1841, PA4012</i>	
HI1681 (<i>yccT</i>)	<i>mgsA</i>	<i>yccF, yccS, yccX</i>	(I, III)

^a The *E. coli* gene names are listed, where available. Genes from other organisms that are expected to be co-expressed with the given *H. influenzae* genes but are absent from *E. coli*, come from *Bacillus subtilis* (BS), *Deinococcus radiodurans* (DR), or *Pseudomonas aeruginosa* (PA).

Fig. 1S (Supplementary Materials). Conserved gene neighborhoods of HI0146-HI0148 and HI0227 genes.

