

Table S1 Strains, two amino acid positions before the first EPIYA motif, NCBI accession numbers, and other details of 115 *H. pylori* strains with EPIYA-ABD motif pattern and one *cagA* copy number. Denote that Strain, Country and Source data were extracted from each GenBank file associated with its accession number.

NO	Strain extracted from GenBank file	Two amino acid positions before the first EPIYA motif	NCBI accession number	EPIYA motif pattern	Copy number of <i>cagA</i>	Country extracted from GenBank file	Source extracted from GenBank file	Genome size (bp)	Start position of <i>cagA</i> on Genome	End position of <i>cagA</i> on Genome
1	F30	--	AP011941	ABD	1	-	-	1570564	831324	834848
2	Hpfe028	--	CP094145	ABD	1	China	gastric biopsy	1559488	652819	656343
3	Hpfe029	--	CP094144	ABD	1	China	gastric biopsy	1569933	940378	943902
4	Hpfe040	--	CP094135	ABD	1	China	gastric biopsy	1566721	654415	657936
5	Hpfe068	--	CP094115	ABD	1	China	gastric biopsy	1566753	932717	936232
6	Hpfe081	--	CP094098	ABD	1	China	gastric biopsy	1573286	925685	929218
7	Hpfe082	--	CP094096	ABD	1	China	gastric biopsy	1613588	928012	931539
8	Hpfe084	--	CP094094	ABD	1	China	gastric biopsy	1629886	713019	716528
9	Hpfe088	--	CP094091	ABD	1	China	gastric biopsy	1558458	928866	932378
10	Hpfe106	--	CP094071	ABD	1	China	gastric biopsy	1590902	961637	965179
11	Hpfe123	--	CP094058	ABD	1	China	gastric biopsy	1615068	695420	698944
12	H-137	--	CP025474	ABD	1	South Korea	gastric biopsy	1637442	661619	665143
13	F13	--	AP017329	ABD	1	Japan	Human stomach	1591903	848461	851991
14	Hpfe074	NA	CP094107	ABD	1	China	gastric biopsy	1602783	970194	973739
15	JSHR 3	NE	AP023344	ABD	1	Japan	-	1624837	704923	708441
16	JSHR3	NE	NZ_AP023344	ABD	1	Japan	-	1624837	704923	708441
17	GD63	NE	CP031558	ABD	1	Viet Nam	Gastric antrum	1581994	653613	657158
18	Hpfe0002	NE	CP094172	ABD	1	China	gastric biopsy	1615483	697209	700748
19	Hpfe0008	NE	CP094179	ABD	1	China	gastric biopsy	1607782	795904	799446
20	Hpfe0012	NE	CP094164	ABD	1	China	gastric biopsy	1627796	937871	941380
21	Hpfe0020	NE	CP094159	ABD	1	China	gastric biopsy	1576980	654328	657861
22	Hpfe027	NE	CP094147	ABD	1	China	gastric biopsy	1581642	671971	675513
23	Hpfe036	NE	CP094139	ABD	1	China	gastric biopsy	1561930	928560	932105
24	Hpfe048	NE	CP094134	ABD	1	China	gastric biopsy	1613194	939199	942747
25	Hpfe051	NE	CP094132	ABD	1	China	gastric biopsy	1616176	657749	661300

NO	Strain extracted from GenBank file	Two amino acid positions before the first EPIYA motif	NCBI accession number	EPIYA motif pattern	Copy number of <i>cagA</i>	Country extracted from GenBank file	Source extracted from GenBank file	Genome size (bp)	Start position of <i>cagA</i> on Genome	End position of <i>cagA</i> on Genome
26	Hpfe052	NE	CP094131	ABD	1	China	gastric biopsy	1560058	660777	664325
27	Hpfe061	NE	CP094120	ABD	1	China	gastric biopsy	1613185	698837	702385
28	Hpfe062	NE	CP094177	ABD	1	China	gastric biopsy	1560534	923192	926716
29	Hpfe064	NE	CP094119	ABD	1	China	gastric biopsy	1615760	1022913	1026464
30	Hpfe066	NE	CP094116	ABD	1	China	gastric biopsy	1586299	653435	656983
31	Hpfe069	NE	CP094113	ABD	1	China	gastric biopsy	1580144	683295	686828
32	Hpfe070	NE	CP094111	ABD	1	China	gastric biopsy	1580056	683287	686820
33	Hpfe071	NE	CP094110	ABD	1	China	gastric biopsy	1618643	700889	704413
34	Hpfe080	NE	CP094099	ABD	1	China	gastric biopsy	1563024	655798	659343
35	Hpfe085	NE	CP094093	ABD	1	China	gastric biopsy	1574428	936926	940459
36	Hpfe097	NE	CP094082	ABD	1	China	gastric biopsy	1640211	733915	737460
37	Hpfe102	NE	CP094076	ABD	1	China	gastric biopsy	1607515	968913	972413
38	Hpfe103	NE	CP094074	ABD	1	China	gastric biopsy	1597927	653776	657258
39	Hpfe107	NE	CP094070	ABD	1	China	gastric biopsy	1620835	740796	744335
40	Hpfe115	NE	CP094063	ABD	1	China	gastric biopsy	1618823	930305	933835
41	Hpfe076	NE	CP094105	ABD	1	China	gastric biopsy	1573039	974968	978372
42	Hpfe104	NE	CP094073	ABD	1	China	gastric biopsy	1574424	287219	290557
43	Hpfe072	NN	NZ_CP094109	ABD	1	China	gastric biopsy	1610658	938674	942249
44	3401	NT	NZ_AP024599	ABD	1	Japan	-	1591853	669232	672762
45	#5	NT	NZ_AP024963	ABD	1	-	-	1608283	651157	654669
46	#6	NT	AP024964	ABD	1	-	-	1630060	660999	664529
47	F32	NT	AP011943	ABD	1	-	-	1578824	567055	570576
48	F57	NT	AP011945	ABD	1	-	-	1609006	594999	598511
49	UM066	NT	CP005493	ABD	1	Malaysia	biopsy	1658047	960491	964018
50	Hpfe0001	NT	CP094173	ABD	1	China	gastric biopsy	1649959	691314	694829
51	Hpfe0003	NT	CP094171	ABD	1	China	gastric biopsy	1564612	934138	937686
52	Hpfe0004	NT	CP094170	ABD	1	China	gastric biopsy	1560176	657078	660616
53	Hpfe022	NT	CP094150	ABD	1	China	gastric biopsy	1617425	658886	662419

NO	Strain extracted from GenBank file	Two amino acid positions before the first EPIYA motif	NCBI accession number	EPIYA motif pattern	Copy number of <i>cagA</i>	Country extracted from GenBank file	Source extracted from GenBank file	Genome size (bp)	Start position of <i>cagA</i> on Genome	End position of <i>cagA</i> on Genome
54	Hpfe023	NT	CP094149	ABD	1	China	gastric biopsy	1608056	693663	697193
55	Hpfe024	NT	CP094148	ABD	1	China	gastric biopsy	1595545	684482	688030
56	Hpfe032	NT	CP094143	ABD	1	China	gastric biopsy	1626209	705163	708693
57	Hpfe033	NT	CP094142	ABD	1	China	gastric biopsy	1559352	655496	659029
58	Hpfe039	NT	CP094136	ABD	1	China	gastric biopsy	1626226	942390	945920
59	Hpfe049	NT	CP094133	ABD	1	China	gastric biopsy	1613080	647147	650671
60	Hpfe054	NT	CP094129	ABD	1	China	gastric biopsy	1563078	929271	932819
61	Hpfe055	NT	CP094128	ABD	1	China	gastric biopsy	1572908	659606	663136
62	Hpfe056	NT	CP094126	ABD	1	China	gastric biopsy	1645557	1006102	1009632
63	Hpfe057	NT	CP094124	ABD	1	China	gastric biopsy	1577941	684825	688355
64	Hpfe058	NT	CP094122	ABD	1	China	gastric biopsy	1615882	712834	716364
65	Hpfe059	NT	CP094121	ABD	1	China	gastric biopsy	1588133	658531	662064
66	Hpfe060	NT	CP094178	ABD	1	China	gastric biopsy	1576166	935894	939421
67	Hpfe065	NT	CP094117	ABD	1	China	gastric biopsy	1568330	934590	938141
68	Hpfe073	NT	CP094108	ABD	1	China	gastric biopsy	1607927	652576	656112
69	Hpfe075	NT	CP094106	ABD	1	China	gastric biopsy	1652404	962771	966301
70	Hpfe078	NT	CP094102	ABD	1	China	gastric biopsy	1609466	534700	538236
71	Hpfe079	NT	CP094101	ABD	1	China	gastric biopsy	1622213	927291	930824
72	Hpfe083	NT	CP094095	ABD	1	China	gastric biopsy	1577172	659704	663234
73	Hpfe086	NT	CP094092	ABD	1	China	gastric biopsy	1588502	924688	928215
74	Hpfe092	NT	CP094088	ABD	1	China	gastric biopsy	1620255	982787	986317
75	Hpfe095	NT	CP094085	ABD	1	China	gastric biopsy	1598616	652807	656325
76	Hpfe101	NT	CP094078	ABD	1	China	gastric biopsy	1572519	654654	658181
77	Hpfe110	NT	CP094066	ABD	1	China	gastric biopsy	1664142	1010855	1014397
78	Hpfe118	NT	CP094062	ABD	1	China	gastric biopsy	1637345	982175	985705
79	Hpfe120	NT	CP094061	ABD	1	China	gastric biopsy	1609748	651406	654933
80	Hpfe122	NT	CP094059	ABD	1	China	gastric biopsy	1559659	647058	650585
81	Hpfe124	NT	CP094056	ABD	1	China	gastric biopsy	1548938	630253	633000

NO	Strain extracted from GenBank file	Two amino acid positions before the first EPIYA motif	NCBI accession number	EPIYA motif pattern	Copy number of <i>cagA</i>	Country extracted from GenBank file	Source extracted from GenBank file	Genome size (bp)	Start position of <i>cagA</i> on Genome	End position of <i>cagA</i> on Genome
82	Hpfe126	NT	NZ_CP094052	ABD	1	China	gastric biopsy	1619620	951589	955116
83	F21	NT	AP017336	ABD	1	Japan	Human stomach	1566831	748676	752206
84	F210	NT	AP017334	ABD	1	Japan	Human stomach	1634312	882758	886288
85	F211	NT	AP017335	ABD	1	Japan	Human stomach	1585641	849326	852856
86	F23	NT	AP017337	ABD	1	Japan	Human stomach	1558697	467875	471393
87	F24	NT	AP017338	ABD	1	Japan	Human stomach	1570268	832617	836150
88	F67	NT	AP017348	ABD	1	Japan	Human stomach	1587581	853825	857355
89	F70	NT	AP017349	ABD	1	Japan	Human stomach	1631634	839090	842620
90	F78	NT	AP017352	ABD	1	Japan	Human stomach	1644960	865475	869005
91	F90	NT	AP017354	ABD	1	Japan	Human stomach	1588119	564289	567819
92	MKF10	NT	AP017356	ABD	1	Japan	Human stomach	1617765	874756	878274
93	MKM1	NT	AP017359	ABD	1	Japan	Human stomach	1567867	555433	558966
94	MKM6	NT	AP017362	ABD	1	Japan	Human stomach	1577123	554759	558283
95	TN2wt	NT	NZ_AP019730	ABD	1	Japan	-	1604952	904047	907565
96	Hpbs1	NT	NZ_CP034071	ABD	1	China	gastric ulcer	1563701	922494	926024
97	DU15	NT	CP011483	ABD	1	South Korea	stomach	1614411	746100	749630
98	83	NT	NC_017375	ABD	1	-	biological product	1617426	855121	858651
99	F38	NT	AP017341	ABD	1	Japan	Human stomach	1626538	871801	875322
100	F16	SA	AP011940	ABD	1	-	-	1575399	835147	838665
101	Hpfe037	SA	CP094138	ABD	1	China	gastric biopsy	1596715	654921	658448
102	Hpfe038	SA	CP094137	ABD	1	China	gastric biopsy	1576312	655523	659056
103	F72	SA	AP017350	ABD	1	Japan	Human stomach	1625760	567002	570517
104	MKF8	SA	AP017358	ABD	1	Japan	Human stomach	1592987	842828	846346
105	Hpfe0006	SE	CP094168	ABD	1	China	gastric biopsy	1559683	649082	652633
106	Hpfe0007	SE	CP094167	ABD	1	China	gastric biopsy	1562288	930530	934075
107	#3	ST	NZ_AP024962	ABD	1	-	-	1585790	647176	650688

NO	Strain extracted from GenBank file	Two amino acid positions before the first EPIYA motif	NCBI accession number	EPIYA motif pattern	Copy number of <i>cagA</i>	Country extracted from GenBank file	Source extracted from GenBank file	Genome size (bp)	Start position of <i>cagA</i> on Genome	End position of <i>cagA</i> on Genome
108	35A	ST	NC_017360	ABD	1	-	biological product	1566655	1035429	1038659
109	Hp238	ST	CP010013	ABD	1	Taiwan	gastric biopsy	1586473	569075	572623
110	Hpfe0011	ST	CP094165	ABD	1	China	gastric biopsy	1621962	797544	801074
111	Hpfe0013	ST	CP094163	ABD	1	China	gastric biopsy	1613739	696457	699969
112	Hpfe010	ST	CP094153	ABD	1	China	gastric biopsy	1576566	685356	688904
113	Hpfe077	ST	CP094103	ABD	1	China	gastric biopsy	1624315	705890	709423
114	Hpfe093	ST	CP094087	ABD	1	China	gastric biopsy	1604124	678189	681716
115	Hpfe100	ST	CP094079	ABD	1	China	gastric biopsy	1555854	927931	931458