

Table 1 - Differences in functional classification between TIGR (strain 26695), AstraZeneca (Strain J99) and PyloriGene database

		Functional categories	
	TIGR	AstraZeneca	PyloriGene
Categories	Sub-categories	Sub-categories	Sub-categories
Amino acid biosynthesis	Aromatic amino acid family Aspartate family Glutamate family Pyruvate family Serine family Other	Aromatic amino acid family Aspartate family Glutamate family Pyruvate family Serine family General	1- Amino acid biosynthesis 1.1-Aromatic amino acid family 1.2-Aspartate family 1.3-Glutamate family 1.4-Pyruvate family 1.5-Serine family 1.6-Other
Purine, pyrimidines, nucleosides, and nucleotides	2-deoxyribonucleotide metabolism Purine ribonucleotide biosynthesis Pyrimidine ribonucleotide biosynthesis Salvage of nucleosides and nucleotides Sugar-nucleotide biosynthesis and conversions Other	2-deoxyribonucleotide metabolism Purine ribonucleotide biosynthesis Pyrimidine ribonucleotide biosynthesis Salvage and interconversion of nucleosides and nucleotides Sugar-nucleotide biosynthesis and conversions all members of this group	2-Purine, pyrimidines, nucleosides, and nucleotides 2.1-2-deoxyribonucleotide metabolism 2.2-Purine ribonucleotide biosynthesis 2.3-Pyrimidine ribonucleotide biosynthesis 2.4-Salvage of nucleosides and nucleotides 2.5-Sugar-nucleotide biosynthesis and conversions 3.1-General
Fatty acid and phospholipid metabolism	Biosynthesis Degradation Other	Fatty acid and phospholipid metabolism all members of this group	3-Fatty acid and phospholipid metabolism 3.1-General
Biosynthesis of cofactors, prosthetic groups, and carriers	Biotin Folic acid Heme, porphyrin and cobalamine Menquinone and ubiquinone Molybdopterin Pantothenate and coenzyme A Pyridoxine Riboflavin, FMN and FAD Thioredoxin, glutaredoxin, and glutathione Thiamine Pyrimidine nucleotides	Biotin Folic acid Heme and porphyrin Menquinone and ubiquinone Molybdopterin Pantothenate Pyridoxine Riboflavin Thioredoxin, glutaredoxin, and glutathione Thiamine Pyrimidine nucleotides	4-Biosynthesis of cofactors, prosthetic groups, and carriers 4.1-Biotin 4.2-Folic acid 4.3-Heme, porphyrin and cobalamine 4.4-Menquinone and ubiquinone 4.5-Molybdopterin 4.6-Pantothenate and coenzyme A 4.7-Pyridoxine 4.8-Riboflavin, FMN and FAD 4.9-Thioredoxin, glutaredoxin, and glutathione 4.10-Thiamine 4.11-Pyrimidine nucleotides 4.12-Other
Central intermediary metabolism	Amino sugars Phosphorus compounds Polyamine biosynthesis Other		5-Central intermediary metabolism 5.1-Amino sugars 5.2-Phosphorus compounds 5.3-Polyamine biosynthesis 5.4-Other
Energy metabolism	Aerobic Anaerobic Amino acids and amines ATP-proton motive force interconversion Electron transport Enter-Doudoroff Fermentation Glycolysis/gluconeogenesis Pentose phosphate pathway Sugars TCA cycle Other	Energy metabolism Amino acids and amines ATP-proton motive force interconversion Electron transport Enter-Doudoroff Fermentation Gluconeogenesis Pentose phosphate pathway Sugars TCA cycle Other	6-Energy metabolism 6.1-Aerobic 6.2-Amino acids and amines 6.3-Anaerobic 6.4-ATP-proton motive force interconversion 6.5-Electron transport 6.6-Enter-Doudoroff 6.7-Fermentation 6.8-Glycolysis/gluconeogenesis 6.9-Pentose phosphate pathway 6.10-Sugars 6.11-TCA cycle 6.12-Other
Transport and binding proteins	Amino acids, peptides and amines Anions Carbohydrates, organic alcohols, and acids Cations Nucleosides, purines, and pyrimidines Other Unknown substrate DNA replication, recombination, and repair	Transport and binding proteins Amino acids, peptides and amines Anions Carbohydrates, organic alcohols, and acids Cations General Nucleosides, purines, and pyrimidines all members of this group all members of this group	7-Transport and binding proteins 7.1-Amino acids, peptides and amines 7.2-Anions 7.3-Carbohydrates, organic alcohols, and acids 7.4-Cations 7.5-Nucleosides, purines, and pyrimidines 7.6-Other 7.7-Unknown substrate 8.1-DNA replication, recombination, and repair
DNA metabolism	Restriction/modification Degradation of DNA Chromosome-associated proteins Degradation of RNA DNA-dependent RNA polymerase Transcription factors	DNA replication, modification, recombination, and repair all members of this group	8-DNA metabolism 8.2-Restriction/modification 8.3-Degradation of DNA 8.4-Chromosome-associated proteins 9.1-Degradation of RNA 9.2-DNA-dependent RNA polymerase 9.3-Transcription factors
Transcription	Degradation of RNA DNA-dependent RNA polymerase Transcription factors	Degradation of RNA DNA-dependent RNA polymerase Transcription factors	9-Transcription

Table 1 - Differences in functional classification between TIGR (strain 26695), AstraZeneca (Strain J99) and PyloriGene database

Protein synthesis	RNA processing RNA aminoacylation Nucleoproteins Ribosomal proteins: synthesis and modification tRNA and rRNA base modification Translation factors Other Protein and peptide secretion, and trafficking Protein modification and repair Protein folding and stabilization Degradation of protein, peptides, and glycopeptides Other	Translation	RNA processing Amino acyl RNA synthetases Nucleoproteins Ribosomal proteins: synthesis and modification tRNA modification Translation factors Protein modification Degradation of protein, peptides, and glycopeptides General Chemotaxis and motility	10-Protein synthesis	9-4-RNA processing 10-1-RNA aminoacylation 10-2-Nucleoproteins 10-3-Ribosomal proteins: synthesis and modification 10-4-rRNA and rRNA base modification 10-5-Translation factors 10-6-Other
Protein fate	Protein and peptide secretion, and trafficking Protein modification and repair Protein folding and stabilization Degradation of protein, peptides, and glycopeptides Other	Regulatory functions	Protein modification Degradation of protein, peptides, and glycopeptides General Chemotaxis and motility	11-Protein fate	11-1-Protein and peptide secretion, and trafficking 11-2-Protein modification and repair 11-3-Protein folding and stabilization 11-4-Degradation of protein, peptides, and glycopeptides 12-1-General
Regulatory functions	Other	Regulatory functions	Chemotaxis and motility	12-Regulatory functions	12-1-General
Cell envelope	Surface structures Biosynthesis of murein sacculus and peptidoglycan Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides Other Cell division Chemotaxis and motility	Cell envelope	Surface structures Murein sacculus and peptidoglycan Surface polysaccharides, lipopolysaccharides, and antigens Membrane and porins Cell division Polyamine biosynthesis	13-Cell envelope	13-1-Lipoproteins 13-2-Surface structures 13-3-Biosynthesis of murein sacculus and peptidoglycan 13-4-Biosynthesis of surface polysaccharides and lipopolysaccharides 13-5-Other 14-1-Cell division 14-2-Chemotaxis and motility
Cellular processes	Cell division Chemotaxis and motility Detoxification DNA transformation Toxin production and resistance Pathogenesis Adaptation and atypical conditions Other	Cellular processes	Detoxification DNA transfer and Cag pathogenicity island Cell killing Chaperones Protein and peptide secretion General Phosphorus compounds Urea Transposable elements	14-Cellular processes	14-3-Detoxification 14-4-Transformation 14-5-Toxin production and resistance 14-6-Pathogenesis 14-7-Adaptation and atypical conditions 14-8-Other
Other categories	Plasmid-related functions Transposon-related functions General Conserved	Conserved with no known function <i>Hpilori</i> specific with no known function	Transposable elements	15-Other categories	15-1-Plasmid-related functions 15-2-Transposon-related functions 16-1-General 17-1-Conserved 17-2- <i>Hpilori</i> specific with no known function
Unknown function		Conserved with no known function		16-Unknown	
Hypothetical		<i>Hpilori</i> specific with no known function		17-Hypothetical	

Blue boxes indicate categories or sub-categories that were either merged into one single sub-category or fragmented into two or several (sub-) categories by comparison to the other functional classifications.

Gray boxes indicate sub-categories classified in a different category but kept intact compared to the other annotation efforts. Red boxes indicate missing categories or sub-categories from one of the functional classifications.

Table 2 - CDS with new assignments or re-assignments of function

CDS	Biological function	Proposed gene name	Original functional category with <i>AstraZeneca</i> ^a	Reassigned functional category	Basis for the proposed function ^b
HP0037.1/JHP0033.1	DNA transformation competence protein	<i>comb7</i>	n.a.	14.4	<i>H. pylori</i>
HP0045/JHP0039	GDP-fucose synthetase (previously annotated as nodulation protein Nolk)	<i>fel</i>	13.4	2.5	<i>H. pylori</i>
HP0049/JHP0042	Non-functional type II restriction endonuclease	-	17.1	8.2	<i>H. pylori</i>
HP0066/JHP0061	Predicted ATP-binding protein	-	17.1	16.1	by similarity
HP0086/JHP0079	Malate:quinone oxidoreductase	<i>mgo</i>	17.1	6.11	<i>H. pylori</i>
HP0089/JHP0082	Predicted 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	<i>min</i>	17.1	5.4	by similarity
HP0102/JHP0094	Predicted glycosyl transferase	-	17.1	13.4	by similarity
HP0105/JHP0097	Autoinducer-2 synthase	<i>luxS</i>	17.1	12.1	<i>H. pylori</i>
HP0111/JHP0103	Predicted heat-inducible transcription repressor	<i>hrcA</i>	17.1	12.1	by similarity
HP0112/JHP0104	Predicted L-fucose-1-phosphate aldolase	<i>fucA</i>	17.2	13.4	by similarity
HP0160/JHP0148	Cysteine-rich protein D	<i>hepD</i>	17.1	16.1	<i>H. pylori</i>
HP0211/JHP0197	Cysteine-rich protein A	<i>hepA</i>	17.1	16.1	<i>H. pylori</i>
HP0216/JHP0202	Predicted 1-deoxyxylulose-5-phosphate reductoisomerase	<i>dxr</i>	17.1	4.12	by similarity
HP0217/JHP0203	Predicted β-1,4-N-acetylglactosamyltransferase	<i>cglA</i>	17.2	13.4	by similarity
HP0228/JHP0213	Predicted sulfate permease	-	17.1	7.2	by similarity
HP0233/JHP0218	Predicted glutathioneperoxidase synthetase	<i>gsp</i>	17.1	5.3	by similarity
HP0235/JHP0220	Cysteine-rich protein E	<i>hepE</i>	17.1	16.1	<i>H. pylori</i>
HP0262/JHP0246/JHP0247	Non-functional type II restriction endonuclease	-	17.1	8.2	<i>H. pylori</i>
HP0291/JHP0276	Predicted chorismate mutase	<i>pheA</i>	17.1	1.1	by similarity
HP0309/JHP0294	Predicted N-carbamoyl-D-amino acid amidohydrolase	-	17.1	1.6	by similarity
HP0320/JHP0303	Sec-independent protein translocase protein	<i>tatA</i>	17.1	11.1	<i>H. pylori</i>
HP0335/JHP0336	Cystein-rich protein B	<i>hepB</i>	17.1	16.1	by similarity
HP0347/JHP0321	Predicted ribosomal large subunit pseudouridine synthase D	<i>rluDI</i>	17.1	10.4	by similarity
HP0368/JHP1013	Non-functional type II restriction endonuclease	-	17.2	8.2	<i>H. pylori</i>
HP0396/JHP0985	Predicted 3-o-octaprenyl-4-hydroxybenzoate carboxy-lyase	<i>ubiD</i>	17.1	4.4	by similarity
HP0400/JHP0981	Predicted ADP hydrolase (previously annotated as lysis tolerance protein LytB)	<i>ispH</i>	12.1	4.12	by similarity
HP0404/JHP0977	Predicted S-adenosylmethionine-dependent methyltransferase	-	14.8	2.4	by similarity
HP0419/JHP0965	Predicted DNA transformation competence (Comb8 homologue)	-	17.1	16.1	by similarity
HP0439	Predicted DNA helicase	-	17.2	14.4	by similarity
HP0447	Predicted DNA helicase	-	17.1	8.1	by similarity
HP0479/JHP0431	Non-functional type II restriction endonuclease	-	17.2	8.2	<i>H. pylori</i>
HP0482/JHP0434	Non-functional type II restriction endonuclease	-	17.2	8.2	<i>H. pylori</i>
HP0484/JHP0436	Non-functional type II restriction endonuclease	-	17.2	8.2	<i>H. pylori</i>
HP0485/JHP0437	Predicted catalase-like protein	-	17.1	14.3	by similarity
HP0511	Lipoprotein Lpp, urease enhancing factor	<i>lpp</i>	17.1	13.1	<i>H. pylori</i>
HP0553/JHP0500	Predicted 23S rRNA methyltransferase	<i>msrR</i>	17.1	10.4	by similarity
HP0602/JHP0549	3-methyladenine DNA glycosylase (previously annotated as an endonuclease III)	<i>magIII</i>	8.2	8.1	<i>H. pylori</i>
HP0605/JHP0552	Outer membrane protein of the HeFA3C efflux system	<i>hefA</i>	17.1	7.7	<i>H. pylori</i>
HP0625/JHP0569	Predicted 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthetase	<i>ispG</i>	17.1	4.12	by similarity
HP0627/JHP0628/JHP0571	Cysteine-rich protein F	<i>hepF</i>	17.2	16.1	<i>H. pylori</i>
HP0639/JHP0582	Predicted trans-regulatory protein	<i>exsB</i>	17.1	12.1	by similarity
HP0667/JHP0668/JHP0612	Remnant of type I restriction-modification polypeptide	<i>llaGI</i>	17.1	8.2	by similarity

Table 2 - CDS with new assignments or re-assignments of function

HP0669/JHP0613	Remnant of type I restriction-modification polypeptide	<i>llaGI</i>	17.2	8.2	by similarity
HP0691/JHP0637	Succinyl-CoA-transferase subunit A (previously annotated as involved in phospholipid synthesis: <i>scdA</i>)	<i>scdA</i>	3.1	6.11	<i>H. pylori</i>
HP0692/JHP0636	Succinyl-CoA-transferase subunit B (previously annotated as involved in phospholipid synthesis: <i>scdB</i>)	<i>scdB</i>	3.1	6.11	<i>H. pylori</i>
HP0731/JHP0668	Predicted <i>LeoA</i> required for secretion of heat-labile enterotoxin (<i>E. coli</i> H10407)	<i>leoA</i>	17.2	14.6	by similarity
HP0745/JHP0682	Predicted ribosomal large subunit pseudouridine synthase D	<i>rluD2</i>	17.1	10.4	by similarity
HP0747/JHP0684	Predicted S-adenosylmethionine-dependent methyltransferase	-	17.1	16.1	by similarity
HP0759/JHP0696	Predicted DNA-damage inducible multidrug efflux protein	<i>norM</i>	17.1	14.7	by similarity
HP0776/JHP0713	Predicted RNA polymerase omega chain	<i>rpoZ</i>	17.2	9.2	by similarity
HP0785/JHP0722	Predicted outer membrane lipoprotein carrier protein	<i>lola</i>	17.1	11.1	by similarity
HP0787/JHP0724	Predicted lipoprotein releasing system	<i>lalc</i>	17.1	11.1	by similarity
HP0792/JHP0728	Predicted DNA transformation competence (maybe be involved in recombination)	<i>comM</i>	17.1	14.4	by similarity
HP0810/JHP0746	Predicted N-6 adenine methyltransferase	-	17.1	8.2	by similarity
HP0831/JHP0770	Predicted dephospho-CoA kinase	<i>coaE</i>	17.1	4.6	by similarity
HP0834/JHP0773	Predicted GTP-binding protein	<i>engA</i>	17.1	16.1	by similarity
HP0840/JHP0778	UDP-GlcNAc C6 dehydratase/C4 reductase (previously annotated as putative sugar nucleotide biosynthesis)	<i>flaAI</i>	13.4	2.5/13.4	<i>H. pylori</i>
HP0890/JHP0823	Predicted short-chain oxidoreductase	<i>vllc</i>	17.1	6.12	by similarity
HP0891/JHP0824	Predicted acyl coenzyme A thioesterase	<i>vllD</i>	17.1	3.1	by similarity
HP0909/JHP0845	Non-functional type II restriction endonuclease	-	17.2	8.2	<i>H. pylori</i>
HP0944/JHP0879	Predicted regulator of purine biosynthesis	<i>ybbJ</i>	17.1	12.1	by similarity
HP0956/JHP0890	Predicted ribosomal large subunit pseudouridine synthase C	<i>rluC</i>	17.1	10.4	by similarity
HP0964/HP0965/JHP0898/JHP0899	Predicted ATP/GTP-binding protein (pseudogene)	-	17.2	16.1	by similarity
HP0977/JHP0911	Predicted peddyl-L-prolyl cis-isomerase D	<i>ppid</i>	17.1	11.2	by similarity
HP1020/JHP0404	Predicted bifunctional MEP cytidyltransferase/MECPs involved in isoprenoid synthesis	<i>ispDF</i>	17.1	4.12	by similarity
HP1022/JHP0402	Predicted 5'-3' exonuclease	-	17.2	8.1	by similarity
HP1053/JHP0372	Predicted septum site-directing protein	<i>minC</i>	17.2	14.1	by similarity
HP1060/JHP0365	Predicted Sec-independent protein translocase protein	<i>tafB</i>	17.2	11.1	by similarity
HP1061/JHP0364	Predicted Sec-independent protein translocase protein	<i>tafC</i>	17.1	11.1	by similarity
HP1082/JHP0343	Predicted lipid A and glycerophospholipid transporter (previously annotated as a MDR protein)	<i>msbA</i>	7.6	7.6	by similarity
HP1098/JHP1024	Cysteine-rich protein C	<i>hpcC</i>	17.1	16.1	by similarity
HP1102/JHO1028	Predicted 6-phosphogluconolactonase (previously annotated as a putative dehydrogenase)	<i>pgl</i>	6.12	6.9	by similarity
HP1117/JHP1045	Cystein-rich protein X	<i>hpcX</i>	17.1	16.1	by similarity
HP1122/JHP1051	Anti-sigma 28 factor	<i>flgM</i>	17.2	12.1	<i>H. pylori</i>
HP1149/JHP1076	Predicted 16s rRNA processing protein	<i>rimM</i>	17.1	9.4	by similarity
HP1185/JHP1111	Predicted sugar efflux transporter	<i>sofB</i>	17.1	7.3	by similarity
HP1215/HP1216/JHP1138	Predicted organic solvent tolerance protein (previously annotated as involved in outer membrane permeability)	<i>osrA</i>	13.5	14.5	by similarity
HP1221/JHP1142	Predicted undecaprenyl pyrophosphate synthase	<i>upps</i>	17.1	4.12	by similarity
HP1225/JHP1146	Predicted Crb integral membrane protein	<i>crbB</i>	17.1	16.1	by similarity
HP1228/JHP1149	Predicted diadenosine polyphosphate hydrolase (previously annotated as MutT)	<i>nuuH</i>	8.2	14.3	by similarity
HP1238/JHP1159	Formamidase (previously annotated as a aliphatic amidase)	<i>amfF</i>	6.2	6.2	<i>H. pylori</i>
HP1240/JHP1161	Predicted Maf protein (appears to be involved in septum formation)	<i>maf</i>	17.1	16.1	by similarity
HP1247/JHP1168	DNA polymerase III delta subunit	<i>holA</i>	17.1	8.1	<i>H. pylori</i>
HP1285/JHP1205	Predicted acid phosphatase lipoprotein	<i>aphA</i>	17.1	5.4	by similarity
HP1330/JHP1250	Predicted branched-chain amino acid transport protein	<i>azlD</i>	17.1	7.1	by similarity
HP1331/JHP1251	Predicted branched-chain amino acid transport protein	<i>azlC</i>	17.1	7.1	by similarity

Table 2 - CDS with new assignments or re-assignments of function

HP1337/JHP1256	Predicted nicotinate-nucleotide adenylyltransferase	<i>nadD</i>	17.1	4.11	by similarity
HP1338/JHP1257	Predicted nickel uptake regulation protein	<i>nikR</i>	17.1	12.1	by similarity
HP1351/JHP1270	Type II restriction endonuclease/Non-functional type II restriction endonuclease	<i>hpyAVR</i> / -	17.2	8.2	<i>H. pylori</i>
HP1353/HP1354/JHP1272	Predicted adenine-specific DNA methylase	-	17.2	8.2	by similarity
HP1394/JHP1433	Predicted inorganic polyphosphate/ATP-NAD kinase	<i>ppnK</i>	17.1	4.11	by similarity
HP1430/JHP1323	Predicted ATP/GTP-binding protein	-	17.1	16.1	by similarity
HP1443/JHP1336	Predicted 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase	<i>ispE</i>	17.1	4.12	by similarity
HP1444/JHP1337	Predicted SSR A-binding protein	<i>smpB</i>	17.1	10.6	by similarity
HP1459/JHP1352	Predicted ribosomal large subunit pseudouridine synthase B	<i>rluB</i>	17.1	10.4	by similarity
HP1464/JHP1357	Predicted ABC transport system substrate binding protein	-	17.1	7.7	by similarity
HP1466/JHP1359	Predicted ABC transport system permease	-	17.1	7.7	by similarity
HP1476/JHP1369	Predicted aromatic acid decarboxylase (previously annotated as UbiD)	<i>paalD</i>	4.4	16.1	by similarity
HP1482/JHP1375	Predicted exodeoxyribonuclease VII small subunit	<i>xseB</i>	17.2	8.3	by similarity
HP1490/JHP1383	Predicted TlyC-like hemolysin	<i>thyC</i>	17.1	14.5	by similarity
HP1527/JHP1416	Periplasmic protein essential for natural competence	<i>comH</i>	17.2	14.4	<i>H. pylori</i>
HP1567/JHP1475	Predicted GTP-binding protein	<i>engB</i>	17.1	14.1	by similarity
HP1573/JHP1481	Predicted DNase of the TadD family	<i>tadD</i>	17.1	8.3	by similarity
JHP0318	Cysteine-rich protein H	<i>hepH</i>	17.1	16.1	<i>H. pylori</i>
JHP0829	Predicted virulence-associated protein D	<i>vapD</i>	17.2	14.6	by similarity
JHP0921	Predicted DNA transformation competence (ComB8 homologue)	-	17.2	14.4	by similarity
JHP0922/JHP0923	Predicted DNA transformation competence (ComB9 homologue)	-	17.2	14.4	by similarity
JHP1437	Cysteine-rich protein G	<i>hepG</i>	17.1	16.1	<i>H. pylori</i>

^aThese CDS were classified in PyloRiGene categories taking only into account the original annotation from AstraZeneca

^bCDS for which the function has been characterized directly in *H. pylori* are annotated "*H. pylori*". CDS for which the proposed function is based on sequence similarity to proteins characterized in other organisms are annotated "by similarity".

Table 3 - List of CDS pairs that differ in length

CDS	gene name	Frame. Cate. in 26695	CDS size in 199	CDS size in 26695	absolut aa difference	aa difference	comments	F	IF	III
HP0009/JHP0007	<i>hopZ</i>	17.1	634	667	33	33	-33 different start codon due to slipped-strand mispairing mechanism (poly CT track)	5		
HP0021/JHP0019		13.5	190	191	1	1	-1 different stop codon due to a nucleotide deletion in JHP0019	3a		
HP0022/JHP0020	Predicted integral membrane protein	17.1	521	523	2	2	-2 insertion (natural polymorphism)	1		
HP0025/JHP0021	<i>hopD</i>	13.5	711	690	21	21	21 deletion (natural polymorphism)	1		
HP0028/JHP0024		17.1	177	175	2	2	2 different start codon with only the alternative second start codon UUG conserved (JHP0024 correct)	2a		
HP0030/JHP0026		17.2	593	565	28	28	28 deletion (natural polymorphism)	1		
HP0032/JHP0028		17.1	91	90	1	1	1 deletion (natural polymorphism)	1		
HP0033/JHP0029	<i>clpA</i>	17.1	741	742	1	1	-1 insertions and deletions (natural polymorphism)	1		
HP0035/JHP0031		17.1	97	100	3	3	-3 insertion (natural polymorphism)	1		
HP0036/JHP0032		17.1	334	335	1	1	-1 difference due to two consecutive start codon for JHP0032	2a		
HP0038/JHP0034	<i>comb8</i>	14.4	245	247	2	2	-2 insertion (natural polymorphism)	1		
HP0039/JHP0035	<i>comb9</i>	14.4	90	328	238	238	-238 sequencing error from TIGR	4		
HP0040/JHP0035	<i>comb9</i>	14.4	247	328	81	81	-81 sequencing error from TIGR	4		
HP0041/JHP0036	<i>comb10</i>	14.4	123	376	253	253	-253 sequencing error from TIGR	4		
HP0042/JHP0034	<i>comb10</i>	14.4	233	376	143	143	-143 sequencing error from TIGR	4		
HP0048/JHP0041	<i>hypF</i>	12.1	769	757	12	12	12 different start codon due to a premature stop codon in JHP0041 (a nucleotide substitution)	2b		
HP0050/JHP0043	<i>hypA/ViAM /hypY9Y/M</i>	8.2	232	230	2	2	2 different stop codon due to point mutations and nucleotide deletions	3a		
HP0058/JHP0050		17.2	0	120	120	120	-120 authentic frame-shift	4		
HP0058/JHP0051		17.2	0	120	120	120	-120 authentic frame-shift	4		
HP0059/JHP0052		17.2	284	327	43	43	-43 insertion (natural polymorphism)	1		
HP0060/JHP0053		17.2	813	143	670	670	670 frame-shift or sequencing error	4		
HP0060/JHP0054		17.2	813	522	291	291	291 frame-shift or sequencing error	4		
HP0060/JHP0055		17.2	813	78	735	735	735 frame-shift or sequencing error	4		
HP0063/JHP0058		17.2	496	500	4	4	-4 deletions and insertions (natural polymorphism)	1		
HP0066/JHP0061	Predicted ATP-binding protein	16.1	831	806	25	25	25 different start codon assignment and deletion	1		
HP0069/JHP0064	<i>ureF</i>	11.3	254	256	2	2	-2 insertion (natural polymorphism)	1		2a
HP0074/JHP0069		11.1	157	163	6	6	-6 different stop codon due to a nucleotide insertion in JHP0069 at the 3' end	3a		
HP0078/JHP0073	<i>lspA</i>	13.5	85	255	170	170	-170 duplication and insertion of <i>HorA</i> N-ter	6		
HP0079/JHP0073	<i>horA</i>	13.5	595	255	340	340	340 duplication and insertion of <i>HorA</i> N-ter	6		
HP0080/JHP0074		17.2	593	590	3	3	3 insertions and a deletion (natural polymorphism)	1		
HP0089/JHP0082	<i>min</i>	5.4	231	230	1	1	1 different start codon assignment (choice between TIGR (AUG) and Astra (GUG) due to consecutive AUG and GUG)	2a		
HP0091/JHP0084	<i>hpyA/III</i>	8.2	277	72	205	205	205 pseudogene JHP0084	6		
HP0093/JHP0086	<i>facC</i>	13.4	146	128	18	18	18 slipped-strand mispairing mechanism due to a poly (C) track	5		
HP0094/JHP0086	<i>facC</i>	13.4	155	128	27	27	27 slipped-strand mispairing mechanism due to a poly (C) track	5		
HP0102/JHP0094	Predicted glycosyl transferase	13.4	259	260	1	1	-1 insertion (natural polymorphism)	1		
HP0103/JHP0095	<i>ltpB</i>	14.2	565	564	1	1	1 difference due to two consecutive start codons for HP0103	2a		
HP0105/JHP0097	<i>lacs</i>	12.1	155	152	3	3	3 insertion (natural polymorphism)	1		
HP0107/JHP0099	<i>cysK</i>	1.5	306	305	1	1	1 difference due to two consecutive start codons for HP0107	2a		
HP0108/JHP0100		17.2	186	152	34	34	34 different start codon due to a nucleotide insertion generating a premature stop codon at the 5' end of the gene and 3' end polymorphism generating different stop codon	2b		3b
HP0110/JHP0102		11.3	189	191	2	2	-2 insertion (natural polymorphism)	1		
HP0111/JHP0103	<i>grpE</i>	12.1	276	266	10	10	10 different start codon assignment (199 start codon appears to be correct)	2a		
HP0112/JHP0104	<i>hrcA</i>	13.4	231	212	19	19	19 different start codon assignment	2a		
HP0114/JHP0106	<i>facA</i>	17.1	628	627	1	1	1 difference due to two consecutive start codons for HP0114	2a		
HP0118/JHP0110		17.2	395	412	17	17	-17 deletion (natural polymorphism)	1		
HP0129/JHP0118		17.2	141	143	2	2	-2 insertion (natural polymorphism)	1		
HP0130/JHP0119		17.2	286	285	1	1	1 deletion (natural polymorphism)	1		
HP0143/JHP0131	Predicted C(4)-dicarboxylates and tetracarboxylates/succinate antiporter	7.3	0	482	482	482	-482 regulated by slipped-strand mispairing mechanism (poly A track)	5		
HP0146/JHP0134	<i>fixQ</i>	6.5	73	72	1	1	1 difference due to two consecutive start codon for HP0146	2a		
HP0147/JHP0135	<i>fixP</i>	6.5	286	292	6	6	-6 C-ter frame-shift due to a nucleotide difference creating an extended C-ter	3a		
HP0149/JHP0137		17.2	194	197	3	3	-3 insertions (natural polymorphism)	1		
HP0150/JHP0138		17.2	196	193	3	3	3 different stop codon due to natural polymorphism of the 3' end	2b		
HP0151/JHP0139	Predicted membrane protein	17.1	270	255	15	15	15 different start codon due to a nucleotide insertion in JHP0139	3b		
HP0156/JHP0144		17.1	200	198	2	2	2 both genes have two start codons (AUG and GUG). HP0156 uses AUG and JHP0144 GUG	2a		
HP0159/JHP0147	<i>fixJ1</i>	13.4	372	377	5	5	-5 insertion (natural polymorphism)	1		

Table 3 - List of CDS pairs that differ in length

HP0160/JHP0148	<i>hcpD</i>	16.1	306	305	1	1	1 deletion (natural polymorphism)	1
HP0164/JHP0151	histidine kinase sensor	12.1	254	442	188	5	-188 slipped-strand mispairing mechanism due to a poly (C) track	5
HP0165/JHP0151	histidine kinase sensor	12.1	173	442	269	5	-269 slipped-strand mispairing mechanism due to a poly (C) track	5
HP0170/JHP0156		17.1	253	243	10	1	10 deletion (natural polymorphism)	1
HP0172/JHP0158	<i>moeA</i>	4.5	391	390	1	1	1 deletion (natural polymorphism)	1
HP0174/JHP0160		17.1	258	242	16	16	16 different start codon due to 5 nucleotides deletion	2c
HP0181/JHP0169		17.1	224	236	12	12	-12 different start codon assignment and insertion	1
HP0185/JHP0173		17.2	267	268	1	1	-1 insertion and deletion (natural polymorphism)	1
HP0186/JHP0174		17.2	404	567	163	4	-163 frame-shift or sequencing error	4
HP0187/JHP0174		17.2	95	567	472	4	-472 frame-shift or sequencing error	4
HP0188/JHP0174		17.2	33	567	534	4	-534 frame-shift or sequencing error	4
HP0201/JHP0187	<i>plaxX</i>	3.1	338	339	1	1	-1 difference due to two consecutive start codons for HP0201 and deletion (natural polymorphism)	1
HP0203/JHP0189		17.2	91	100	9	9	-9 different stop codon due to a nucleotide insertion	3a
HP0205/JHP0191		17.2	793	916	123	1	-123 extensive insertions and deletions	1
HP0206/JHP0192		17.2	108	85	23	1	23 one insertion and different stop due to TAA codon deletion and a nucleotide deletion in J99	1
HP0207/JHP0193	<i>mnp</i>	16.1	412	368	44	2a	44 different start codon assignment (corrected for HP0207 by SWISSProt)	2a
HP0208/JHP0194	<i>rfaD2</i>	13.4	0	394	394	5	-394 regulated by slipped-strand mispairing mechanism (poly AG track)	5
HP0209/JHP0195	<i>hopA</i>	13.5	450	438	12	2a	12 different start codon assignment (Astra considers J99 start codon as the correct one)	2a
HP0212/JHP0198	<i>dapE</i>	1.2	383	388	5	1	-5 deletion (natural polymorphism)	1
HP0214/JHP0200	Predicted sodium-dependent transporter	7.7	552	548	4	4	4 deletion (natural polymorphism)	1
HP0217/JHP0203	<i>gtaA</i>	13.4	380	385	5	5	-5 insertion and deletion (natural polymorphism). Also regulated by slipped-strand mispairing mechanism (two poly G tracks). JHP0203 was adjusted artificially by Astra to give 385 aa	1
HP0219/JHP0205		17.2	159	161	2	1	-2 two consecutive alternative start codons (UUG and GUG) and an insertion. HP0219 uses the second start codon and JHP0205 the first.	2a
HP0222/JHP0208		17.2	73	86	13	3a	-13 different stop codon due to a nucleotide substitution without changing the frame of lecture	3a
HP0223/JHP0209	<i>rada</i>	8.1	456	448	8	2a	8 different start codon assignment correct for HP0223 by SWISSProt	2a
HP0227/JHP0212	<i>hopM</i>	13.5	691	696	5	1	-5 insertions and deletions (natural polymorphism)	1
HP0228/JHP0213	Predicted sulfate permease	7.2	390	385	5	2a	5 difference due to a start codon not conserved in JHP0213	2a
HP0238/JHP0223	<i>proS</i>	10.1	577	576	1	1	1 deletion (natural polymorphism)	1
HP0247/JHP0232	<i>decD</i>	10.5	492	491	1	1	1 deletion (natural polymorphism)	1
HP0252/JHP0237	<i>hopF</i>	13.5	252	237	15	1	15 deletion (natural polymorphism)	1
HP0253/JHP0238	<i>hopG</i>	13.5	38	471	433	4	-433 frame-shift or sequencing error	4
HP0254/JHP0238	<i>hopG</i>	13.5	431	471	40	4	-40 frame-shift or sequencing error	4
HP0257/JHP0241		17.1	219	222	3	1	-3 insertion (natural polymorphism)	1
HP0258/JHP0242		17.1	348	351	3	1	-3 insertion (natural polymorphism) and difference due to two consecutive start codons, corrected by SWISSProt for JHP0242	1
HP0260/JHP0244	<i>hpyORF260/hpyORF1244M</i>	8.2	384	390	6	2a	-6 different stop codon due to the stop codon deletion in JHP0244	2a
HP0261/JHP0245		17.2	132	152	20	3b	-20 N-ter polymorphism (only C-ter located 38 aa are homologous)	3b
HP0262/JHP0246	Non-functional type II restriction endonuclease	8.2	200	172	28	2c	28 different start codon (corrected by PIR for HP0262) and frame-shift or sequencing error at the C-ter	2c
HP0262/JHP0247	Non-functional type II restriction endonuclease	8.2	200	69	131	3a	131 frame-shift or sequencing error	3a
HP0265/JHP0250	<i>ccfA</i>	6.5	240	239	1	4	1 difference due to two consecutive start codons for HP0265	4
HP0270/JHP0255		17.1	162	217	55	2a	-55 different start codon assignment	2a
HP0279/JHP0264	<i>wacC</i>	17.1	340	336	4	1	4 deletion (natural polymorphism)	1
HP0282/JHP0267		17.1	480	478	2	2a	2 two almost consecutive start UUG codons. HP0282 uses the first and JHP0267 the second	2a
HP0289/JHP0274	Predicted toxin-like protein	14.5	2893	2902	9	1	-9 natural polymorphism insertions	1
HP0291/JHP0276	<i>pheA</i>	1.1	96	100	4	1	-4 insertion (natural polymorphism)	1
HP0293/JHP0278	<i>pubB</i>	4.2	559	567	8	1	8-different stop codon due to natural polymorphism of 3' end	1
HP0298/JHP0283		7.1	549	546	3	3b	3 insertion and deletions (natural polymorphism)	3b
HP0304/JHP0289	<i>dppA</i>	17.2	329	331	2	1	-2 insertion (natural polymorphism)	1
HP0307/JHP0292		17.1	90	92	2	1	-2 different start codon due to a nucleotide deletion in JHP0292	1
HP0309/JHP0294	Predicted N-carboxymyl-D-amino acid amidohydrolase	1.6	292	294	2	2b	-2 insertion (natural polymorphism)	2b
HP0322/JHP0305	Predicted poly E-rich protein	16.1	505	501	4	1	4 deletions and insertions (natural polymorphism)	1
HP0324/JHP0307	<i>hopC</i>	13.5	254	245	9	2a	9 different start codon assignment (for HP0324 corrected by Astra)	2a
HP0333/JHP0316		14.4	270	266	4	2a	4 difference due to an alternative start codon GUG not conserved in JHP0316	2a
HP0337/JHP0319		17.2	101	112	11	1	-11 insertion (natural polymorphism)	1
HP0338/JHP0320		17.2	188	186	2	1	2 deletion (natural polymorphism)	1
HP0347/JHP0321	<i>hldJ</i>	10.4	298	293	5	2c	5 different stop codon due to natural polymorphism of 3' end	2c
HP0350/JHP0324		17.2	222	220	2	1	2 deletion (natural polymorphism)	1
HP0366/JHP1015	Predicted sugar nucleotide biosynthesis	2.5	375	378	3	3a	-3 different stop codon due to two nucleotide deletions and a nucleotide substitution	3a
HP0367/JHP1014		17.2	202	201	1	1	1 insertion and deletion (natural polymorphism)	1

Table 3 - List of CDS pairs that differ in length

HP0368/JHP1013	Non-functional type II restriction endonuclease	8.2	133	182	49	-49 different start codon due to a poly (A) track (7 and 8, respectively). Insertion of a A nucleotide for HP0368 gives genes of same size)	5
HP0369/JHP1012	Non-functional site-specific adenine methyltransferase	8.2	236	320	84	-84 different start codon due to extensive N-ter polymorphism (both are inactive forms of a typeII R-M system)	2c
HP0370/JHP1011	<i>accC</i>	3.1	458	455	3	3 difference due to two start codons for HP0370 (AUG and GUG) with only one AUG for JHP1011 which corresponds to the GUG of HP0370	2a
HP0371/JHP1010	<i>accB</i>	3.1	156	161	5	-5 deletions (natural polymorphism)	1
HP0372/JHP1009	<i>acd</i>	2.1	190	188	2	2 different start codon assignment (correct for HP0372 by SWISSProt)	2a
HP0373/JHP1008	<i>homC</i>	13.5	700	751	51	-51 insertion (natural polymorphism)	1
HP0376/JHP0291	<i>hemH</i>	4.3	334	335	1	-1 different stop codon due to a nucleotide deletion in JHP0291 at the 3'end	3a
HP0377/JHP1004	Predicted DsxC-like protein	16.1	221	223	2	-2 insertion (natural polymorphism)	1
HP0379/JHP1002	<i>furA</i>	13.4	425	435	10	-10 deletions and insertions with a natural polymorphism of the last 15 aa (also regulated by slipped-strand mispairing mechanism)	1 3b 5
HP0384/JHP0997		17.2	250	248	2	2 deletion (natural polymorphism)	1
HP0385/JHP0996		17.2	76	75	1	1 difference due to two consecutive start codons for HP0385	2a
HP0392/JHP0989	<i>cheA</i>	14.2	803	815	12	-12 insertion (natural polymorphism)	1
HP0393/JHP0988	<i>cheV3</i>	14.2	311	312	1	-1 insertion (natural polymorphism)	1
HP0398/JHP0983		17.2	182	170	12	12 deletion (natural polymorphism)	1
HP0399/JHP0982	<i>psfI</i>	10.3	556	552	4	4 deletion (natural polymorphism)	1
HP0408/JHP0973		17.2	162	164	2	-2 deletion and different stop codon (natural polymorphism)	1 3b
HP0427/JHP0952		17.1	111	63	48	48 different start and stop codons due to natural polymorphism	2c 3b
HP0463/JHP0415	<i>hscMI</i>	8.2	487	543	56	-56 different stop codon due to a nucleotide insertion (C-ter extension)	3a
HP0464/JHP0416	<i>hscRI</i>	8.2	1055	0	1055	-1055 regulated by slipped-strand mispairing mechanism (poly C track)	5
HP0477/JHP0429	<i>hopI</i>	13.5	367	371	4	-4 insertions (natural polymorphism of N-ter)	1
HP0481/JHP0433	Non-functional type II adenine specific DNA methyltransferase	8.2	211	309	98	-98 different stop codon due to a nucleotide insertion (both are inactive forms of a typeII R-M system)	3a
HP0482/JHP0434	Non-functional type II restriction endonuclease	8.2	170	175	5	-5 different stop codon due to a nucleotide insertion in JHP0434 at the 3'end and a nucleotide substitution	3a
HP0483/JHP0435	<i>hpy99X/M</i>	8.2	0	351	351	-351 authentic frame-shift	4
HP0484/JHP0436	Non-functional ACGT site-specific type II restriction endonuclease	8.2	250	184	66	66 different start codon due to a nucleotide substitution generating a premature stop codon	2b
HP0487/JHP0440	<i>hojD</i>	13.5	480	465	15	15 different start codon assignment (Astra considers J99 start codon the correct one)	2a
HP0488/JHP0441		17.2	957	912	45	45 different start due to natural polymorphism and deletions	1 2c
HP0492/JHP0444	Predicted neuraminylactose-binding hemagglutinin	17.1	295	181	114	114 one N-ter codon deletion and different stop codon due to C-ter polymorphism	2c 3b
HP0502/JHP0454		13.2	278	282	4	-4 insertions (natural polymorphism)	1
HP0503/JHP0454		17.2	54	343	289	-289 frame-shift or sequencing error	4
HP0504/JHP0455		17.2	246	343	97	-97 frame-shift or sequencing error	4
HP0505/JHP0455		17.1	49	251	202	-202 frame-shift or sequencing error	4
HP0506/JHP0456	Predicted outer membrane protein	13.5	154	251	97	-97 frame-shift or sequencing error	4
HP0508/JHP0458		17.1	452	450	2	-2 insertion and deletion (natural polymorphism)	1
HP0513/JHP0462		17.1	658	706	48	48 different start codon assignment with deletions and insertions	1 2a
HP0514/JHP0463	<i>mpl9</i>	10.3	150	149	1	1 difference due to two consecutive start codons for HP1313	2a
HP0517/JHP0466	<i>era</i>	14.8	302	301	1	1 difference due to two consecutive start codons for HP0517	2a
HP0519/JHP0468		17.2	276	283	7	-7 insertion (natural polymorphism)	1
HP0521/JHP0470	<i>caq2</i>	14.6	80	219	139	-139 Insertion resulting in a longer protein with a different start for JHP0470 and a frame-shift truncating the C-ter of HP0521	1 2c 3a
HP0527/JHP0476	<i>caqY</i>	14.6	1927	1819	108	108 insertions and deletions (natural polymorphism)	1
HP0534/JHP0482	<i>caq5</i>	14.6	196	199	3	-3 different stop codon due to one nucleotide substitution extending the 3' end by 3 extra codons	3a
HP0545/JHP0493	<i>caqD</i>	14.6	207	208	1	-1 insertion (natural polymorphism)	1
HP0547/JHP0495	<i>caqA</i>	14.6	1186	1167	19	19 deletion (natural polymorphism)	1
HP0552/JHP0499		17.1	287	289	2	-2 different stop codon; two codons insertion at the 3' end (natural polymorphism)	3b
HP0564/JHP0511		13.2	56	81	25	-25 different start codon assignment (incorrect reassignment by PIR for HP0564)	2a
HP0567/JHP0514	Predicted outer membrane protein	17.2	348	349	1	1 difference due to a TGT insertion before the stop in JHP0514	3a
HP0579/JHP0526		17.2	184	183	1	1 different stop codon due to a nucleotide substitution in JHP0526	3a
HP0582/JHP0529	<i>tonB1</i>	7.4	324	317	7	7 deletion (natural polymorphism)	1
HP0585/JHP0532	<i>nhl</i>	218	218	214	4	4 difference due to an alternative start codon GUG not conserved in JHP0532 and a deletion (natural polymorphism)	1 2a
HP0585.1/JHP0533		17.1	76	78	2	-2 different stop codon due to natural polymorphism of the 3'end of the genes	3b
HP0586/JHP0534		17.2	977	978	1	-1 deletion and different stop codon due to a nucleotide deletion at the 3' end of HP0586	1 3a
HP0591/JHP0539	<i>oorC</i>	6.11	186	184	2	2 different stop codon due to last three codons replaced by one single codon	3b
HP0597/JHP0544	<i>phpA</i>	13.3	659	660	1	-1 insertion (natural polymorphism)	1
HP0600/JHP0547	<i>spbB</i>	7.7	593	589	4	4 difference due to an AUG start codon not conserved in JHP0547	2a

Table 3 - List of CDS pairs that differ in length

HP0603/JHP0550	17.2	189	205	16	-16 different start codon assignment (HP0603 start is not conserved in JHP0550)	2a
HP0608/JHP0555	17.2	160	189	29	-29 different start codon assignment	2a
HP0610/JHP0556	14.5	1238	3194	1956	-1956 frame-shift or sequencing error	4
HP0611/JHP0299	14.5	1943	3194	1251	-1251 frame-shift or sequencing error	4
HP0612/JHP0299	7.7	166	251	85	-85 frame-shift or sequencing error	4
HP0614/JHP0557	7.7	79	251	172	-172 frame-shift or sequencing error	4
HP0617/JHP0560	17.2	111	110	1	1 difference due to two consecutive start codon for HP0614	2a
HP0619/JHP0563	10.1	577	579	2	-2 difference due to two codon insertion at the 3' end of JHP0560	3b
JHP0627/JHP0571	16.1	0	268	268	-268 regulated by slipped-strand mispairing (poly C track)	5
JHP0628/JHP0571	16.1	110	110 (352)	268	JHP0571 coordinates have been modified (352 aa to 110 aa). Duplicate of HP0627 with an extension	6
HP0629/JHP0572	16.1	225	110	115	115 JHP0571 coordinates have been modified (352 aa to 110 aa). Duplicate of HP0627 with an extension	1
HP0636/JHP0579	17.2	92	137	45	-45 different stop codon due to a nucleotide insertion (C-extension) and insertion	1 3a
HP0638/JHP0581	13.5	305	307	2	-2 insertion (natural polymorphism). Also regulated by slipped-strand mispairing mechanism: both genes are on:	1
HP0639/JHP0582	12.1	226	225	1	-1 different stop codon due to natural polymorphism of the 3' end of the genes	3b
HP0641/JHP0584	17.1	75	167	92	-92 different stop codon (extensive natural polymorphism)	3b
HP0650/JHP0595	17.2	196	209	13	-13 different start codon assignment	2a
HP0651/JHP0596	13.4	476	454	22	-22 deletion (natural polymorphism) (Also regulated by slipped-strand mispairing mechanism)	1
HP0655/JHP0600	13.5	916	906	10	10 deletion (natural polymorphism)	1
HP0656/JHP0601	17.1	383	353	30	30 different start codon due to a nucleotide substitution in JHP0601 that eliminates a premature stop codon	2b
HP0657/JHP0602	11.4	432	435	3	-3 insertion (natural polymorphism)	1
HP0658/JHP0603	10.6	475	474	1	1 difference due to two consecutive start codons for HP0658	1
HP0659/JHP0604	17.2	414	413	1	1 insertion and deletion (natural polymorphism)	2a
HP0660/JHP0605	17.1	338	335	3	3 difference due to two codons insertion at the 3' end	1
HP0662/JHP0607	9.4	240	239	1	1 difference due to two consecutive start codon for HP0662	3b
HP0667/JHP0612	8.2	85	450	365	-365 ancestral of a type III R-M system (HP0667/HP0668/HP0669 and JHP0612/HP0613)	2a
HP0668/JHP0612	8.2	607	450	157	157 ancestral of a type III R-M system (HP0667/HP0668/HP0669 and JHP0612/HP0613)	6
HP0669/JHP0613	8.2	933	1167	234	-234 ancestral of a type III R-M system (HP0667/HP0668/HP0669 and JHP0612/HP0613)	6
HP0675/JHP0617	8.1	362	356	6	6 deletion (natural polymorphism)	1
HP0678/JHP0620	13.4	36	315	279	-279 different stop codons between HP0679 and JHP0620 without changing the frame of lecture (nucleotide substitution)	3a
HP0679/JHP0620	13.4	289	315	26	-26 different stop codons without changing the frame of lecture (nucleotide substitution) but which splits HP0678 ; 3a	3a
HP0682/JHP0623	17.2	126	116	10	10 different start codon assignment. HP0682 codon is not conserved in JHP0623	2a
HP0684/JHP0625	13.2	105	248	143	-143 slipped-strand mispairing mechanism due to a poly (C) track	5
HP0685/JHP0625	13.2	172	248	76	-76 slipped-strand mispairing mechanism due to a poly (C) track	5
HP0688/JHP0628	17.1	166	276	110	-110 frame-shift or sequencing error	4
HP0689/JHP0628	17.1	96	276	180	-180 frame-shift or sequencing error	4
HP0694/JHP0634	13.5	257	336	79	-79 frame-shift for HP0694	4
HP0695/JHP0633	1.6	713	712	1	1 deletion (natural polymorphism)	1
HP0697/JHP0631	17.1	168	166	2	2 two almost consecutive start AUG codons HP0697 uses the first and JHP0631 the second one	2a
HP0701/JHP0641	8.1	827	828	1	-1 insertion (natural polymorphism)	1
HP0702/JHP0642	17.2	158	157	1	1 difference due to two consecutive start codons for HP0702	2a
HP0705/JHP0644	8.1	935	941	6	-6 different start codon assignment (apparent incorrect assignment of alternative UUG start codon for J99)	2a
HP0706/JHP0645	13.5	273	270	3	3 different start codon assignment (for HP0706 corrected by Astra)	2a
HP0710/JHP0649	13.5	660	657	3	3 deletion (natural polymorphism)	1
HP0712/JHP0651	17.1	124	234	110	-110 frame-shift or sequencing error	4
HP0713/JHP0651	17.1	114	234	120	-120 frame-shift or sequencing error	4
HP0717/JHP0655	8.1	578	582	4	-4 deletions and insertions (natural polymorphism)	1
HP0719/JHP0657	17.2	109	140	31	-31 frame-shift or sequencing error	4
HP0720/JHP0657	17.2	53	140	87	-87 frame-shift or sequencing error	4
HP0722/JHP0659	13.5	608	600	8	8 deletions (natural polymorphism) and phase variation (poly CT track - correcting the poly CT track the two proteins still have different sizes of 644 and 638 aa, respectively)	1 5
HP0723/JHP0661	6.2	330	332	2	-2 insertion (natural polymorphism)	1
HP0725/JHP0662	13.5	629	612	17	17 deletion, different start codon assignment and regulation by slipped-strand mispairing mechanism (poly CT track - correcting the track still gives two proteins of different size 653 and 651 aa, respectively)	1 2a 5
HP0728/JHP0665	17.1	336	338	2	-2 different start codon with only the alternative first start codon UUG conserved (JHP0665 correct)	2a
HP0731/JHP0668	14.6	573	569	4	4 different stop codon due to one nucleotide insertion	3a
HP0732/JHP0669	17.2	118	371	253	-253 HP0732/HP0733 and JHP0669/JHP0670 appear to have decayed from an ancestor protein	6
HP0733/JHP0670	17.2	521	192	329	329 HP0732/HP0733 and JHP0669/JHP0670 appear to have decayed from an ancestor protein	6
HP0744/JHP0681	17.2	0	303	303	-303 regulated by slipped-strand mispairing mechanism (poly AG track)	5

Table 3 - List of CDS pairs that differ in length

HP0747/JHP0684	Probable S-adenosylmethionine-dependent methyltransferase	16.1	393	400	7	-7 different stop codon; C-ter extension in J99 due to a G deletion	3a
HP0752/JHP0689	<i>hfd</i>	13.2	674	685	11	-11 different start codon assignment (for HP752 corrected by SWISSProt giving a 685 aa protein)	2a
HP0755/JHP0694	Predicted moeB/hif family protein	4.5/4.10	210	235	25	-22 different start codon assignment and stop codon (natural polymorphism)	2a
HP0760/JHP0697		17.1	462	529	67	-67 different start codon assignment (corrected form HP0760 by SWISSProt)	2a
HP0764/JHP0701		17.2	428	117	311	311 HP0764/HP0765 and JHP0701/JHP0702/JHP0703 appear to have decayed from an ancestral protein	6
HP0764/JHP0702		17.2	428	168	260	260 HP0764/HP0765 and JHP0701/JHP0702/JHP0703 appear to have decayed from an ancestral protein	6
HP0765/JHP0703		17.2	102	286	184	-184 HP0764/HP0765 and JHP0701/JHP0702/JHP0703 appear to have decayed from an ancestral protein	6
HP0766/JHP0704		17.2	274	182	92	92 different start and stop codons (extensive polymorphism)	3b
HP0769/JHP0706	<i>mobA</i>	4.5	201	209	8	-8 different stop codon due to a nucleotide substitution	2a
HP0772/JHP0709	<i>amiA</i>	13.3	440	469	29	-29 insertion (natural polymorphism)	3a
HP0775/JHP0712	<i>spoZ</i>	12.1	775	776	1	-1 different start codon due to one nucleotide substitution in JHP0712	1
HP0776/JHP0713	<i>spoZ</i>	9.2	84	74	10	10 different start codon assignment (J99 start codon appears the correct one)	3a
HP0779/JHP0716	<i>acrB</i>	6.11	853	852	1	1 difference due to two consecutive start codons for HP0779 corrected by SWISSProt	2a
HP0780/JHP0717		17.2	90	88	2	2 difference due to different start codons: HP0780 has almost two consecutive GUG while JHP0717 has only AUG corresponding to the second GUG	2a
HP0783/JHP0720		17.2	165	175	10	-10 different start codon assignment (J99 start codon appears the correct one)	2a
HP0784/JHP0721		17.2	44	43	1	1 difference due to two consecutive alternative start codons for HP0784 with only the second one conserved in JHP0721	2a
HP0794/JHP0730	<i>clpP</i>	11.4	196	195	1	1 difference due to two consecutive start codon for HP0974	2a
HP0801/JHP0737	<i>moaD</i>	4.5	74	73	1	1 difference due to two consecutive start codon for HP0801	2a
HP0805/JHP0741	Predicted LPS biosynthesis protein	13.4	284	292	8	-8 different start codon assignment (HP0805 start appears to be the correct one) and natural polymorphism of C-ter	3b
HP0807/JHP0743	<i>fecA2</i>	7.4	787	792	5	-5 different stop codon due to a nucleotide deletion for JHP0743	3a
HP0818/JHP0754	<i>proWX</i>	7.1	553	161	392	392 HP0818 is homologous to JHP0757 (553 aa) and JHP0754 (truncated duplication)	6
HP0833/JHP0772		17.2	292	296	4	-4 insertion (natural polymorphism of C-ter)	1
HP0834/JHP0773	<i>engA</i>	16.1	458	462	4	-4 insertion (natural polymorphism)	1
HP0836/JHP0775		17.1	119	201	82	-82 frame-shift or sequencing error	4
HP0837/JHP0775		17.1	102	201	99	-99 frame-shift or sequencing error	4
HP0843/JHP0781	<i>thiE</i>	17.1	102	201	99	2 insertion and deletion (natural polymorphism)	1
HP0844/JHP0782	<i>thiD</i>	4.10	219	217	2	1 difference due to two consecutive GUG start codons in HP0844 with only the second conserved in JHP0782	2a
HP0845/JHP0783	<i>thiM</i>	4.10	273	259	14	14 different start codon assignment (correct by SWISSProt for HP0845)	2a
HP0846/JHP0784	<i>hscR2</i>	8.2	866	875	9	-9 insertions (natural polymorphism)	1
HP0848/JHP0785	<i>hscD2</i>	8.2	298	207	91	91 HP848/HP849 are duplication of JHP785 with rearrangements	6
HP0849/JHP0785	<i>hscS2</i>	8.2	96	207	111	-111 HP848/HP849 are duplication of JHP785	6
HP0850/JHP0786	<i>hscM2</i>	8.2	527	528	1	-1 insertion at the 5' end and deletion at the 3' end	2c
HP0851/JHP0787		17.1	227	228	1	1 difference due to replacement of two last codons by three new ones in JHP787 (natural polymorphism)	3b
HP0854/JHP0790	<i>gucC</i>	2.4	327	325	2	2 different start codons almost consecutive (AUG, UUG and GUG for HP0854) and (GUG, AUG and GUG for JHP0790)	2a
HP0858/JHP0792	<i>flaE</i>	2.5/13.4	461	463	2	-2 insertion (natural polymorphism)	1
HP0859/JHP0793	<i>flaD</i>	2.5/13.4	330	329	1	1 deletion (natural polymorphism)	1
HP0860/JHP0794		17.1	173	174	1	-1 difference due to replacement of the second codon by two new ones in JHP0794 (natural polymorphism)	2c
HP0861/JHP0795		17.1	246	245	1	1 different stop codon due to 3' end natural polymorphism	3b
HP0863/JHP0797		17.2	542	541	1	1 insertion and deletion (natural polymorphism)	1
HP0874/JHP0808		17.2	291	292	1	-1 insertion (natural polymorphism)	1
HP0878/JHP0812		17.2	57	243	186	-186 frame-shift or sequencing error	4
HP0879/JHP0812		17.2	202	243	41	-41 frame-shift or sequencing error	4
HP0885/JHP0817	<i>mwjN</i>	14.6	461	460	1	1 difference due to two consecutive start codons for HP0885	2a
HP0887/JHP0819	<i>vacA</i>	14.5	1290	1288	2	2 deletion (natural polymorphism)	1
HP0890/JHP0843	<i>vilC</i>	6.12	256	271	15	-15 different start codon assignment (orthologues from other strains are 284 aa long)	2a
HP0896/JHP1164	<i>babB (hopT)</i>	13.5	708	703	5	5 different start codon due to slipped-strand mispairing mechanism (poly CT track)	5
HP0899/JHP0836	<i>hycC</i>	11.3	77	78	1	-1 difference in stop codon due to a nucleotide deletion in JHP0836	3a
HP0901/JHP0838		17.2	41	42	1	-1 different stop due to 3' end natural polymorphism	3b
HP0903/JHP0840	<i>ackA</i>	6.7	0	403	403	-403 authentic frame-shift	4
HP0904/JHP0841	<i>pta</i>	6.7	0	519	519	-519 frame-shift of <i>pta</i> gene (HP0903/HP0904) in 26695?	4
HP0905/JHP0841	<i>pta</i>	6.7	223	519	296	-296 frame-shift of <i>pta</i> gene (HP0905/HP0904) in 26695?	4
HP0906/JHP0842		17.2	527	524	3	3 several insertions and deletions (natural polymorphism)	1
HP0907/JHP0843	<i>figD</i>	13.2	301	363	62	-62 different stop codon (C-ter extension) and insertion	1
HP0909/JHP0845	Non-functional type II restriction endonuclease	8.2	201	193	8	8 insertions (natural polymorphism)	1
HP0910/JHP0846	<i>hpy8M</i>	8.2	379	381	2	-2 difference due to two close UUG start codons (HP0846 uses the first and HP0910 uses the second)	3a

Table 3 - List of CDS pairs that differ in length

HP ID	Gene	Start	End	Length	Notes	HP ID	Gene	Start	End	Length	Notes
HP0911/JHP0847	<i>pcrA</i>	8.1	675	676		HP0912/JHP0848	<i>pcrA</i>	8.1	675	676	
HP0912/JHP0848	<i>hopC</i>	13.5	515	520		HP0913/JHP0849	<i>hopB</i>	13.5	529	527	-1 insertion (natural polymorphism)
HP0913/JHP0849	<i>hopB</i>	13.5	529	527		HP0915/JHP0851	<i>hspB2</i>	7.4	562	562	2 deletion (natural polymorphism)
HP0915/JHP0851	<i>hspB2</i>	7.4	562	562		HP0916/JHP0851	<i>hspB2</i>	7.4	249	815	-253 frame-shift or sequencing error
HP0916/JHP0851	<i>hspB2</i>	7.4	249	815		HP0922/JHP0856	Predicted toxin-like protein	14.5	2529	2399	-566 frame-shift or sequencing error
HP0922/JHP0856	Predicted toxin-like protein	14.5	2529	2399		HP0923/JHP0857	<i>hopX</i>	13.5	369	366	130 deletion (natural polymorphism)
HP0923/JHP0857	<i>hopX</i>	13.5	369	366		HP0927/JHP0861	<i>hpxX</i>	11.3	326	310	3 deletion (natural polymorphism)
HP0927/JHP0861	<i>hpxX</i>	11.3	326	310		HP0928/JHP0862	<i>folE</i>	4.2	180	180	16 different start codon assignment corrected for HP0927 by SWISSProt
HP0928/JHP0862	<i>folE</i>	4.2	180	180		HP0931/JHP0866		17.2	146	146	HP0928 has two homologues, JHP0862 (truncated duplication) and JHP863 (full protein)
HP0931/JHP0866		17.2	146	146		HP0932/JHP0866	<i>prpP</i>	7.1	100	219	-119 frame-shift or sequencing error
HP0932/JHP0866	<i>prpP</i>	7.1	100	219		HP0936/JHP0871		17.2	389	452	-63 different start codon assignment corrected for HP0936 by PIR
HP0936/JHP0871		17.2	389	452		HP0937/JHP0872		17.1	264	274	-10 HP0937/JHP0938 and JHP0872/JHP0873 appear to have decayed from an ancestral protein
HP0937/JHP0872		17.1	264	274		HP0938/JHP0873		17.1	115	291	-176 HP0937/JHP0938 and JHP0872/JHP0873 appear to have decayed from an ancestral protein
HP0938/JHP0873		17.1	115	291		HP0940/JHP0875	Predicted amino acid ABC transporter, periplasmic binding protein	7.1	256	257	-1 insertion (natural polymorphism)
HP0940/JHP0875	Predicted amino acid ABC transporter, periplasmic binding protein	7.1	256	257		HP0951/JHP0885		17.1	496	493	3 different stop codon due to a nucleotide deletion
HP0951/JHP0885		17.1	496	493		HP0955/JHP0889		205	204	1 difference due to two consecutive start codon for HP0951	
HP0955/JHP0889		205	204	204		HP0958/JHP0892	<i>lgt</i>	11.1	284	283	1 different stop codon due to nucleotide substitution. JHP0889 has two consecutive stop codons
HP0958/JHP0892	<i>lgt</i>	11.1	284	283		HP0960/JHP0894	<i>glyQ</i>	17.1	254	267	-13 C-ter extension of four EESQ repeats
HP0960/JHP0894	<i>glyQ</i>	17.1	254	267		HP0963/JHP0896		10.1	303	298	5 different stop codon due to a nucleotide substitution
HP0963/JHP0896		10.1	303	298		HP0963/JHP0896		17.2	447	230	217 HP0963 and JHP0896/JHP0897 appear to have decayed from an ancestral protein. Note that JHP0896 was annotated as an orphan CDS of 199
HP0963/JHP0896		17.2	447	230		HP0963/JHP0897		17.2	447	226	221 HP0963 and JHP0896/JHP0897 appear to have decayed from an ancestral protein. Note that JHP0896 was annotated as an orphan CDS of 199
HP0963/JHP0897		17.2	447	226		HP0964/JHP0898	Predicted ATP/GTP-binding protein	17.1	363	244	119 HP0964/JHP0965 and JHP0898/JHP0899 appear to have decayed from an ancestral protein
HP0964/JHP0898	Predicted ATP/GTP-binding protein	17.1	363	244		HP0965/JHP0899	Predicted ATP/GTP-binding protein	17.1	363	500	-137 HP0964/JHP0965 and JHP0898/JHP0899 appear to have decayed from an ancestral protein
HP0965/JHP0899	Predicted ATP/GTP-binding protein	17.1	363	500		HP0966/JHP0900		17.1	443	500	-57 HP0966/JHP0965 and JHP0898/JHP0899 appear to have decayed from an ancestral protein
HP0966/JHP0900		17.1	443	500		HP0966/JHP0901		17.1	549	340	209 frame-shift or sequencing error
HP0966/JHP0901		17.1	549	340		HP0968/JHP0902		17.1	549	166	383 frame-shift or sequencing error
HP0968/JHP0902		17.1	549	166		HP0969/JHP0903	<i>ctcA1</i>	17.2	21	93	-72 frame-shift or sequencing error
HP0969/JHP0903	<i>ctcA1</i>	17.2	21	93		HP0971/JHP0905		7.4	1020	1019	1 difference due to two consecutive start codon for HP0969
HP0971/JHP0905		7.4	1020	1019		HP0973/JHP0907		17.2	413	431	-18 different start codon assignment
HP0973/JHP0907		17.2	413	431		HP0976/JHP0910	<i>bioA</i>	4.1	336	362	-9 insertions (natural polymorphism)
HP0976/JHP0910	<i>bioA</i>	4.1	336	362		HP0978/JHP0912	<i>fsxA</i>	17.2	492	493	-3 different stop codon due to a nucleotide deletion in JHP0910 at the 3' end
HP0978/JHP0912	<i>fsxA</i>	17.2	492	493		HP0990/JHP0938		17.2	222	416	-194 different start codon due to insertion creating an in-frame extended N-ter of JHP0938
HP0990/JHP0938		17.2	222	416		HP0991/JHP0939		17.2	209	337	-128 different start codon due to a poly (A) track and a nucleotide deletion
HP0991/JHP0939		17.2	209	337		HP0992/JHP0939		17.2	79	337	-258 different start codon due to a poly (A) track and a nucleotide deletion
HP0992/JHP0939		17.2	79	337		HP0995/JHP0941		8.1	355	331	24 different start codon with N-ter first 100 aa not homologous
HP0995/JHP0941		8.1	355	331		HP0996/JHP0942		17.2	600	677	-77 different start codon assignment; JHP0942 start is not conserved in 26695
HP0996/JHP0942		17.2	600	677		HP1012/JHP0411	Predicted integrase/recombinase XerCD family	17.2	444	443	1 deletion (natural polymorphism)
HP1012/JHP0411	Predicted integrase/recombinase XerCD family	17.2	444	443		HP1015/JHP0408	Predicted zinc metalloprotease	11.4	168	202	-34 different start codon assignment (correct for HP1015 by PIR)
HP1015/JHP0408	Predicted zinc metalloprotease	11.4	168	202		HP1017/JHP0406		7.1	519	475	44 different start codon assignment (HP1017 ALG is not conserved in JHP0406)
HP1017/JHP0406		7.1	519	475		HP1018/JHP0405	<i>rocE</i>	11.4	48	476	-428 HP1018 and HP1019 are split due to a nucleotide deletion
HP1018/JHP0405	<i>rocE</i>	11.4	48	476		HP1019/JHP0405	<i>htrA</i>	11.4	443	476	-33 HP1018 and HP1019 are split due to a nucleotide deletion
HP1019/JHP0405	<i>htrA</i>	11.4	443	476		HP1020/JHP0404	<i>htrA</i>	4.12	406	409	-3 insertion (natural polymorphism)
HP1020/JHP0404	<i>htrA</i>	4.12	406	409		HP1022/JHP0402	Predicted 5'-3' exonuclease	8.1	278	288	-10 different start codon assignment (J99 start codon appears to be the correct one)
HP1022/JHP0402	Predicted 5'-3' exonuclease	8.1	278	288		HP1028/JHP0396		17.2	165	167	-2 insertion (natural polymorphism)
HP1028/JHP0396		17.2	165	167		HP1030/JHP0394	<i>hly</i>	14.2	287	285	2 deletion (natural polymorphism)
HP1030/JHP0394	<i>hly</i>	14.2	287	285		HP1033/JHP0391		17.2	131	91	40 different stop codon due several nucleotide substitutions and deletions (natural polymorphism)
HP1033/JHP0391		17.2	131	91		HP1035/JHP0389	<i>hlyF</i>	13.2	459	455	4 two consecutive alternative GTG start codons and deletions (natural polymorphism)
HP1035/JHP0389	<i>hlyF</i>	13.2	459	455		HP1038/JHP0386	<i>arcD</i>	17.1	167	170	1 difference due to two consecutive start codons for HP1042
HP1038/JHP0386	<i>arcD</i>	17.1	167	170		HP1042/JHP0382		17.1	348	347	-3 insertion (natural polymorphism)
HP1042/JHP0382		17.1	348	347		HP1048/JHP0377	<i>hlyB</i>	10.5	944	949	1 difference due to two consecutive start codons for HP1042
HP1048/JHP0377	<i>hlyB</i>	10.5	944	949		HP1051/JHP0374		17.1	140	165	-5 deletion (natural polymorphism)
HP1051/JHP0374		17.1	140	165		HP1053/JHP0372		17.1	217	194	-25 different start codon assignment and insertion
HP1053/JHP0372		17.1	217	194		HP1054/JHP0371	<i>minC</i>	17.2	422	440	23 different start codon assignment (HP1053 codon was corrected by SWISSProt; O25693) and deletion
HP1054/JHP0371	<i>minC</i>	17.2	422	440		HP1055/JHP0370		17.2	314	313	-18 different start codon assignment
HP1055/JHP0370		17.2	314	313		HP1056/JHP0369		17.2	284	289	1 different start due to natural polymorphism of the 5' end
HP1056/JHP0369		17.2	284	289							-5 insertion (natural polymorphism)

Table 3 - List of CDS pairs that differ in length

HP1262/HP1183	<i>mucC</i>	6.1	266	265	1	1	1	2a
HP1266/HP1187	<i>mucG</i>	6.1	844	843	5	5	-5 insertion (natural polymorphism)	1
HP1274/HP1195	<i>pflA</i>	14.2	801	803	2	2	-2 insertion (natural polymorphism)	1
HP1276/HP1197		17.2	94	123	29	29	-29 same chromosomal location but no homology at all	6
HP1284/HP1204		17.1	346	352	6	6	-6 different stop codon due to C-ter natural polymorphism	3b
HP1285/HP1205		5.4	230	245	15	15	-15 different start codon assignment	2a
HP1288/HP1208	<i>qplA</i>	17.2	132	117	15	15	15 different start codon assignment and insertion	1
HP1301/HP1221	<i>rpl15</i>	10.3	135	133	2	2	2 insertion (natural polymorphism)	1
HP1303/HP1223	<i>rpl18</i>	10.3	119	118	1	1	1 difference due to two consecutive start codons for HP0515	2a
HP1321/HP1241		17.1	377	376	1	1	1 deletion (natural polymorphism)	1
HP1322/HP1242		17.2	201	192	9	9	9 different start codon assignment. HP1322 codon is not conserved in JHP1242	2a
HP1324/HP1244		17.2	83	64	19	19	19 different start codon due to a nucleotide insertion generating a premature stop codon in JHP1244	2b
HP1328/HP1248		7.4	338	341	3	3	-3 insertion (natural polymorphism)	1
HP1330/HP1250	<i>ccsB2</i>	7.1	115	118	3	3	-3 insertion (natural polymorphism)	3b
HP1333/HP1253	<i>acfD</i>	17.2	382	376	6	6	-6 insertion (natural polymorphism)	1
HP1339/HP1258	<i>expB2</i>	7.6	166	145	21	21	21 different start codon assignment corrected for HP1339 by SWISSProt	2a
HP1341/HP1260	<i>tonB2</i>	7.4	285	280	5	5	5 deletion (natural polymorphism)	1
HP1342/HP1261	<i>hopN</i>	13.5	691	696	5	5	-5 insertions and deletions (natural polymorphism)	1
HP1348/HP1267	<i>hopV</i>	3.1	240	237	3	3	3 difference due to 4 codons replacement by one codon just before the stop codon	3b
HP1353/HP1272	<i>placC</i>	8.2	315	304	11	11	11 regulated by slipped-strand mispairing mechanism (two poly C track). JHP1272 was artificially adjusted by Astra as a 1164 aa protein	5
HP1354/HP1272	Predicted adenine specific DNA methyltransferase	8.2	835	304	531	531	531 regulated by slipped-strand mispairing mechanism (two poly C track). JHP1272 was artificially adjusted by Astra as a 1164 aa protein	5
HP1361/HP1279	<i>comEC</i>	14.4	417	409	8	8	8 different start codon due to a nucleotide deletion in JHP1279	2b
HP1362/HP1280	<i>dhcB</i>	8.1	488	486	2	2	2 C-ter natural polymorphism due to codon deletions	3b
HP1363/HP1281		17.1	466	464	2	2	2 deletion (natural polymorphism)	1
HP1364/HP1282	histidine kinase sensor	12.1	397	344	53	53	53 different start codon (natural polymorphism)	2c
HP1366/HP1442	<i>hpyA/IR</i>	8.2	423	326	97	97	97 different start codon due to slipped-strand mispairing mechanism (poly A track- typeIIS restriction enzyme JHP1442 inactive)	5
HP1369/HP1284	Predicted type III restriction enzyme M protein	8.2	474	649	175	175	-175 HP369/HP1370 and JHP1284 have decayed from an ancestral type III R-M system	6
HP1370/HP1284	Predicted type III restriction enzyme M protein	8.2	288	649	361	361	-361 HP369/HP1370 and JHP1284 have decayed from an ancestral type III R-M system	6
HP1371/HP1285		17.1	968	972	4	4	-4 insertions (natural polymorphism)	1
HP1374/HP1328	<i>clpX</i>	11.4	446	452	6	6	-6 insertion (natural polymorphism)	1
HP1377/HP1291		17.1	146	129	17	17	17 different start codon due to a nucleotide deletion in JHP1291	2b
HP1379/HP1293	<i>lon</i>	11.4	835	831	4	4	4 deletion (natural polymorphism)	1
HP1380/HP1294	<i>lyrA</i>	1.1	265	275	10	10	-10 different start codon assignment (the 199 start codon appears to be the correct one)	2a
HP1382/HP1295	<i>mucG</i>	8.1	130	287	157	157	-157 different start codon due to extensive polymorphism of the N-ter	2c
HP1384/HP1441		17.1	68	67	1	1	1 difference due to two consecutive start codon for HP1384	2a
HP1387/HP1438	<i>dhcQ</i>	8.1	350	323	27	27	27 different stop codons with change in the reading frame	3a
HP1393/HP1434	<i>recN</i>	8.1	524	522	2	2	2 different stop codon due to point mutations (JHP1434 has 3 consecutive stops against one for HP1393)	3a
HP1394/HP1433	<i>ppnK</i>	4.11	284	307	23	23	-23 different start codon assignment (corrected for JHP1433 by SWISSProt)	2a
HP1396/HP1431		17.1	288	293	5	5	-5 1 and 5 a.a. deletion followed by a 11 a.a. C-terminal extension in 199 due to a nucleotide insertion	3a
HP1400/HP1426		17.1	271	282	11	11	-11 natural polymorphism along the entire polypeptide	1
HP1402/HP1424	<i>fccA3</i>	7.4	842	841	1	1	1 difference due to two consecutive start codons for HP1400 corrected by Astra	2a
HP1403/HP1425	<i>hscRB3</i>	8.2	993	991	2	2	2 deletion (natural polymorphism)	1
HP1408/HP1301		17.2	817	815	2	2	2 deletion (natural polymorphism)	2a
HP1409/HP1301		17.2	111	178	67	67	-67 different start and stop codons due to natural polymorphism	2c
HP1409/HP1302		17.2	578	578	310	310	310 frame-shift or sequencing error	4
HP1410/HP1303		17.2	394	410	16	16	-16 different stop codon due to a nucleotide deletion	4
HP1411/HP1304		17.2	619	309	310	310	310 frame-shift or sequencing error	3a
HP1411/HP1305		17.2	619	243	376	376	376 frame-shift or sequencing error	4
HP1412/HP1306		17.2	308	51	257	257	257 different start and stop codons due to natural polymorphism	2c
HP1412/HP1307		17.2	308	240	68	68	68 different start and stop codons due to natural polymorphism	3b
HP1415/HP1310	<i>micA</i>	10.4	266	277	11	11	-11 different stop codon (natural polymorphism of the last 5 aa); both start codons incorrect and refined by SWISSPROT to 311 and 312 aa, respectively, due to missing active site with the original CDS prediction	2a
HP1417/HP1312	Predicted ATP/GTP-binding protein	17.1	0	265	265	265	-265 regulated by slipped-strand mispairing mechanism (poly G track)	5
HP1430/HP1323	Histidine & glutamine-rich metal-binding protein	17.1	689	692	3	3	-3 insertions (natural polymorphism)	1
HP1432/HP1321		7.4	72	77	5	5	-5 insertion (natural polymorphism)	1
HP1433/HP1326		17.2	856	792	64	64	64 deletion (natural polymorphism) - regulated by slipped-strand mispairing mechanism. Both genes are on)	1

Table 3 - List of CDS pairs that differ in length

HP1436/JHP1329	17.2	82	86	4	-4 insertion (natural polymorphism)	1
HP1437/JHP1330	17.2	239	187	52	52 insertion, deletion, different start codon (natural polymorphism) and stop codon due to a nucleotide insertion	1 2c 3a
HP1438/JHP1331	17.1	338	161	177	177 different stop codon with C-ter 100 aa without homology	3b
HP1439/JHP1332	17.2	81	141	60	-60 different stop codon due to the insertion of two distinct nucleotides	3a
HP1439.1/JHP1332	17.2	74	141	67	-67 different stop codon due to the insertion of two distinct nucleotides	3a
HP1443/JHP1336	4.12	268	274	6	-6 different stop codon due to a nucleotide deletion in JHP1336	3a
HP1449/JHP1342	17.1	117	118	1	-1 insertion (natural polymorphism)	1
HP1450/JHP1343	17.1	547	549	2	-2 insertion (natural polymorphism)	1
HP1451/JHP1344	17.1	264	258	6	6 insertion (natural polymorphism)	1
HP1453/JHP1346	13.5	746	744	2	2 deletion (natural polymorphism)	1
HP1455/JHP1348	17.2	130	113	17	17 different start codon due to a nucleotide substitution generating a premature stop codon in JHP1348	2b
HP1462/JHP1355	17.1	172	195	23	-23 different start codon assignment	2a
HP1470/JHP1363	8.1	892	897	5	-5 difference due to two consecutive start codon corrected for HP1470 by SWISSProt and insertion (natural polymorphism)	1 2a
HP1471/JHP1364	8.2	397	406	9	-9 different first 70 aa of N-ter (regulated by slipped-strand mispairing mechanism poly G track)	5
HP1472/JHP1365	8.2	679	678	1	1 different stop codon due to a nucleotide substitution in JHP1365 (both are inactive forms of a type II R-M system)	3a
HP1478/JHP1371	8.1	682	681	1	1 difference due to two consecutive start codons for HP1478	2a
HP1479/JHP1372	17.1	844	841	3	3 deletion (natural polymorphism)	1
HP1490/JHP1383	14.5	449	441	8	8 different start codon assignment and deletion	1 2a
HP1499/JHP1392	17.2	272	306	34	-34 regulated slipped-strand mispairing mechanism (poly A track)	5
HP1500/JHP1393	17.2	23	26	3	-3 the two predicted proteins do not share any significant homology and are considered orthologues based on the genetic location raising the possibility that they are artifactual CDS	6
HP1507/JHP1400	17.1	385	399	14	-14 different start codon assignment	2a
HP1509/JHP1402	17.1	262	220	42	42 different start codon assignment (corrected for HP1509 by SWISSProt)	2a
HP1512/JHP1405	7.4	877	879	2	-2 insertion (natural polymorphism)	1
HP1513/JHP1406	10.6	386	390	4	4 insertion (natural polymorphism) of 4 codons at the N-ter	2c
HP1521/JHP1410	8.2	967	969	2	-2 insertion (natural polymorphism)	1
HP1522/JHP1411	8.2	0	202	202	-202 regulated by slipped-strand mispairing mechanism (poly G track)	5
HP1533/JHP1421	17.1	231	208	23	23 different start codon assignment (corrected for HP1533 by SWISSProt)	2a
HP1541/JHP1438	9.3	999	1001	2	-2 insertion (natural polymorphism)	1
HP1545/JHP1454	4.2	394	389	5	5 deletion (natural polymorphism)	1
HP1550/JHP1449	11.1	503	526	23	-23 different start codon assignment	2a
HP1551/JHP1448	17.1	127	97	30	30 different stop codon due to major deletion at 3'end	3b
HP1553/JHP1446	8.1	945	946	1	-1 insertion (natural polymorphism)	1
HP1568/JHP1476	17.1	183	185	2	-2 insertion (natural polymorphism)	1
HP1571/JHP1479	13.1	315	314	1	1 deletion (natural polymorphism)	1
HP1572/JHP1480	12.1	372	374	2	-2 insertion (natural polymorphism)	1
HP1579/JHP1486	17.2	142	140	2	2 deletion (natural polymorphism) and different stop codon (1 aa C-ter extension)	1 3b
HP1580/JHP1487	17.2	220	198	22	22 different start codon resulting in the first N-ter 14 aa without any homology	2c
HP1581/JHP1488	13.4	336	335	1	1 different stop codon due to one nucleotide substitution in JHP1488	3a
HP1587/JHP1493	17.1	155	209	54	-54 different stop codon with C-ter extension due to a nucleotide deletion in JHP1493	3a

* the code numbers correspond to the different sources of variation as classified in Table 4