SUPPLEMENTAL DATA

Biochemical characterization and essentiality of *Plasmodium* fumarate hydratase

Vijay Jayaraman¹, Arpitha Suryavanshi¹, Pavithra Kalale¹, Jyothirmai Kunala¹, Hemalatha Balaram^{1#}

¹ Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Jakkur P.O., Bengaluru, Karnataka, 560064, INDIA.Tel: 91-80-22082812 Fax:91-80-22082766. [#]E-mail: <u>hb@jncasr.ac.in</u>

Table S1. Oligonucleotide sequences used for various PCRs

Table S2. Prediction of mitochondrial targeting sequence in Plasmodial FH

Figure S1. Multiple sequence alignment of *Plasmodial*, single-subunit and two-subunit bacterial and archaeal FH proteins sequences.

Figure S2. Multiple sequence alignment of FH sequences from different *Plasmodium* species.

Methods S1. Generation of ∆fumACB strain of E. coli

Figure. S3 Genotyping of the *E. coli* strain $\Delta fumACB$.

Text S1. In vitro enzyme inhibition studies.

Figure S4. Generation of *P. berghei* FH knockout construct using recombineering.

Figure S5. Genotyping of *P. berghei* clones of knockout of fumarate hydratase

Table S1. Oligonucleotide sequences used for various PCRs

Primer Name	Primer sequence (5' to 3')	Application
PfFH FL _BamH1_FP	CGCGGATCCATAAAGTTTAAAGAAGCTTCC ATTTTG	Forward primer for cloning PfFHFL
PfFH_Sal1_RP	ACGCGTCGACTTATGATGGTAACCATTTATT ATAAAAATC	Common reverse primer for cloning PfFH
PfFH∆40_Bam H1_FP	CGCGGATCCAGTTTAAATAGTTTTATAGACA TTTTAAGCTTTAG	Forward primer for cloning PfFH∆40
PfFH∆120_Bam H1_FP	CGCGGATCCAATTATGAAAAAGAATATATA CATATCCCACC	Forward primer for cloning PfFH∆120
EcFumC_BamH I-FP	ACGCGGATCCATGAATACAGTACGCAGCGA AAAAGATTCG	Forward primer for cloning of EcFumC
EcFumC_SalI- RP	ACGCGTCGACACGCCCGGCTTTCATACTGC	Reverse primer for cloning of EcFumC
EcFumA_BamH I-FP	GATGTAGGATCCATCAAACAAACCCTTTCAT TATCAGGCTC	Forward primer for cloning of EcFumA
EcFumA_SalI- RP	GCAATAGTCGACTTATTTCACACAGCGGGTG CATTGTG	Reverse primer for cloning of EcFumA
P1	CGGAACACCCCGCCCAGAGCATAACCAAACC AGGCAGTAAGTGAGAGAACAGTGTAGGCTG GAGCTGCTTC	
P2	GCGCAGCCGCTTCGTTTGATCATTCCACGGC TGCACCTGTATGTTGCAGACTGTCAAACATG AGAATTAATTCCG	Generation and
P3	CTGAGTTAATGAGTTTTTGCATGATCAATCC CTG	strain Δ fumACB (in Figure S3 a and b)
P4	CACAGCGGGTGCATTGTGTGAGTTG	rigure 55 a and 57
P5	GGCAGATAAGCTGTGGGGGCGCAC	
P6	GCGTCTGGTACAAAGGAGATCAAAAACAAG TCC	
PfFHpGDB- Xho1-FP	GACTTACTCGAGATGATAAAGTTTAAAGAA GCTTCCATTTTGTTATC	Generation of the plasmid pGDB- PfFH
PfFHpGDB- AvrII-RP	GACTTACCTAGGTGATGGTAACCATTTATTA TAAAAATCATTGCC	Generation of the plasmid pGDB- PfFH
P1 and P2	P1:GCCATTAAAAAATAGGATAACATATATA AAATGCACAATCC P2:CATATGATCTGGGTATCTCGCAAAGCATT G	5' integration of RFA at PfFH genomic locus (Figure 1a)
P3 and P4	P3: GGAGACGGTCACAGCTTGTCTGTAAG P4:GTTGTGATATTGCACATGAATGGATCCAA TC	3' integration of RFA at PfFH genomic locus (Figure 1a)

Qcup	CATTATTCTTTTTCTTTTTGCACATATTTA			
		-		
QCdown	AATTG			
PheSR2	TCATTCTTCGAAAACGATCTGCG	Generation of <i>P. berghei fh</i> knockout construct (Figure S4)		
1 11 0	ACTTCTTAAACCTAATCTGTAGTAAGGAAGG			
ndhfr	GATTG			
	GATTTTGTGTGAAATTTATTTAATAAATTTG			
RECupR2	TAAATTTTTGAACTCAACACCGCCTACTGCG	(Figure 54)		
-	ACTATAGA			
	TTAAAACTTATGAGTTTTTTTTTTCCAACAAT			
RECdownR1	GTACATATTTTTGGAAAAAAGGCGCATAA			
	CGATACCAC			
D1	CGAACTATGGGCATGGTAATAATAATGGAA			
r I	ATAGC			
D)	GAAAAGCAGTAATGTTAGTACATCAAATGT			
Γ2	GTATG	Primers for diagnostic PCR of P.		
D3	CAAGTCCAACTATTTATGAATCATTGAAGAG			
15	ACAAC			
P4	TTCTTAAACCTAATCTGTAGTAAGGAAGGG			
1 7	ATTG			
P5	GTGGATGAAAATATTACTGGTGCTTTGAGG			
15	GGTGAGC			
P6	ATGGTGAGCAAGGGCGAGGAG	(Figure 6 and 7)		
D7	CACATATTAACAATTTGACTCTCTGCATATT			
P/	TTAGACTGATTTC			
P8	CGAGCTCTTTATGCTTAAGTTTACAATTTAA			
	TATTC			
Р9	ATGGTAGGATATCAAAAGATTAGAAAGTTT			
	AAGAAGGTTCC			
P10	CTAAGAAGGTATCCATTTATTATAAAAGTCA			
	TTTCCTTTATTATC			
Mqo_Fw	GGATAGCTATAGTTCCCTTCTTTTATTATATC	Primers for mqo		
	TCAATAATTTTTGC	gene		
Mqo_Rv	GTAAAGCAGTACCTGTAACACCACCACC (Figure 7a)			

	MitoProt	TargetP	MitoFates	PlasMIT
Organism	II	0		
	(Ref. 1)	(Ref. 2)	(Ref. 3)	(Ref. 4)
D faloinamum	CS: NP	CS:NP	CS:38	Non-mito
r. jaiciparum	(0.88)	(0.34)	(0.30)	(99%)
D noich an awi	CS:NP	CS: NP	CS:38	Non-mito
r. reichenowi	(0.97)	(0.26)	(0.29)	(99%)
D aaboni	CS:37	CS:NP	CS:28	Non-mito
r. gaboni	(0.92)	(0.40)	(0.24)	(99%)
D gallingooum	CS:NP	CS:NP	CS:57	Mito
r.gaunaceum	(0.85)	(0.18)	(0.01)	(91%)
D knowlesi	CS: 13	CS:11	CS:32	Mito
F. Knowlesi	(0.99)	(0.88)	(0.69)	(91%)
D fragila	CS:21	CS:13	CS:12	Non-mito
r. jrugue	(0.83)	(0.83)	(0.25)	(99%)
D vinakai	CS:NP	CS:NP	CS:45	Mito
F. vinckei	(0.15)	(0.09)	(0.005)	(91%)
P ahahaudi	CS:NP	CS:9	CS:10	Mito
r. chabauai	(0.89)	(0.66)	(0.17)	(91%)
D jauj	CS:13	CS:50	CS:42	Mito
F. thui	(0.97)	(0.86)	(0.32)	(91%)
P wiwar	CS:13	CS:29	CS:21	Non-mito
r. vivax	(0.96)	(0.816)	(0.630)	(99%)
D harahai	CS:46	CS:NP	CS:37	Mito
r. berghei	(0.85)	(0.262)	(0.251)	(91%)
P voelij	CS:NP	CS:NP	CS:9	Mito
Г. удени	(0.435)	(0.273)	(0.078)	(91%)
	CS:24	CS:NP	CS:23	Mito
FI DHODH	(0.50)	(0.349)	(0.996)	(91%)

Table S2. Prediction of mitochondrial targeting sequence in Plasmodial FH

Fractional values in parentheses are probability values for either export to mitochondria (MitoProt II) or presence of mitochondrial targeting sequence (TargetP and MitoFates). Percentage values obtained from PlasMIT are an indication of the confidence for a specific localization. Boxes shaded grey in column 2 are for sequences for which MitoProt II has identified a targeting sequence and those in columns 3 and 4 are for sequences with probability values greater than 0.5 for presence of targeting sequence. In column 5, sequences predicted to have mitochondrial localization are shaded in grey. CS, cleavage site with the number adjacent corresponding to the predicted site of cleavage; mito, mitochondrially localized; non-mito, not localized to mitochondria; NP, not predicted. [#]As a positive control for the analysis, *P. falciparum* dihydroorotate dehydrogenase (PfDHODH), a well characterized protein with canonical mitochondrial targeting sequence was used.

P.knowlesi P.fragile P.yoelii P.berghei P.chabaudi P.gallinaceum P.falciparum P.fal	111111111111111111111111111111111111111	MRNFARIPICRP. MRNFARIPICRP. MRNFARIPICRP. MRNFARIPICRP. MRVGQQARKFKRIHKLFPYTICTNNLVEISGKNN. MRVGQQARKFKRIHKLFPYTICTNNLVEISGKNN. MRVGQQARKFKRVPKLFPYTICTNFGVHGXKHIRNISNNNIRNINNINNINNINNINNINNINNFDIPFEFGKGNDGIEYRRIDDLSKYIEVIKFNNK MRVGQQARKFKRVPKLFPYTICTNFGVHGXKHIRNISNNNIRNINNINNINNINNINNINNINNINNFDIFFEFGKGNDGIEYRRLDLSKYIEVIKFNNK MRVGQQARKFRKVPSLFPYTICTNFGVHGXKHIRNISNNNIRNINNINNINNINNINNINNINNINNFDIFFEFGKGNDGIEYRRLDLSKYIEVIKFNNK MRVGQARKFRKVPSLFPYTICTNFGVHGXKHIRNISNNNIRNINNINNINNINNINNINNINNINNINNFDIFFFEFKGNDGIEYRRLDLSKYIEVIKFNNK MRVGYQKVRRFKKVPSLFPYTICTNFFGUHGXKHIRNISNNNIRNINNINNINNINNINNINNINNINNINNFDIFFFEFKGNDGIEYRRLDLSKYIEVIKKLONN MNNFKNIPLLFSKK. MINFKNIPLLFSKK. MINFKNIPLLFSKK. MIKFKEASILLSHNAYIVONLYFKKIR. MSNKFFIYQ MSTRFFVYZ MSTRFFVYZ MSTRFFVYZ MSTRFFVYZ MSTRFFVYZ MSTRFF
P.knowlesi P.fragile P.yoeili P.berghei P.vinckei P.dalinaceum P.falinaceum P.falinaceum P.faliparum P.reichenowi E.coli_PumB G.metallireducens B.petrii V.cholerae N.meninglidis A.vinelandii Acinetobacter Metallosphaera Fervidicoccus Pyrolobus Thermoproteus Acidilobus Caldisphaera	78 67 90 99 61 73 78 79 36 38 1 1 1 1 1 1 1 1 1 1 1	PIN.ESKYYGYNFENEDNFFHPNGELKNL.PEQVIQNEVERIKEYIHTPPFVITKICEYMPRE.TLFFLNKKHIKOLSNIKKGSSNDMYVANT. ND.DNKYYDINYDDENEFFDDNGLKKNL.PEQVIQNEGERIKEYIHYPPFVITKICEYMPRE.TLFFLNKKHIKOLSNIKKSSENDMYVANT. IND.DNKYYDINYDDENEFFDDNGLKKN.NYKESSENDMYVANT. IND.NSKYDINYDDENEFFDDNGLKKK.NNYKESSENDMYVANT. IND.NSKYDINYDDENEFFDDNGLKKK.NNYKESSENDMYVANT. IND.NSKYDINYDDENEFFDDNGLKKK.NNYKESSENDMYVANT. IND.SKYYDINYDDENEFFDDNGLKKK.NDCEKSNNWNKEYIHYPPFVITKICEYMIKE. ILFFLNKKHIKOLONILMKSSENDMFVANT. KINKDSKYYDINYENENEFFDENGLKKK.NDCEKSNNWNKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVMKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVMKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVMKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVMKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKK.NDCEKSNNKVKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNLKKI.NDCEKSNNKVKEYIHYPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNCALKIK.NDCEKSNNKVKEYIHIPPFVITKICEYMIKE INFN.STKYGYDFTUNFFDKGNCALKIK.NDCEKSNNKVKEYIHIPPFVITKICEYMIKE INFN.STKYGYDFTENFFDKGNLKY MN.ETKYYGYNFKEENNFLDEHGNIKEYIYNENKLLHKNYEKSYHHIPPFVITKICEYMIKE INFN.STKYGYNFKEENNFLDEHGNIKEYIYNENKLLHKNYEKSYHHIPPFVITKICEYMIKE INFN.STKYGYNFKEENNFLDEHGNIKEYIYNENKKLLHKNYEKSYHHIPPFVITKICEYMIKE. MRTIKRODVISSVAAALOP.SSYLAAAHOVAALA. MRTIKRODVISSVAAALOP.SSYLAAAHOVAALA. MRTIKRODVISSVAAALOP.SSYLAAAHOVAALA. MTVIKODDIISSVAAALOP.SSYNDAANKAALAGANAALOO. MTTVIKODDIISSVAAALOP.SYHPPDFICKKLE.ANYERESOAANADAIAO. MTTVIKODDIISSVAAALOP.SYHPPDFICKALKE.NYERENAANAAAAALOO. MTTVIKODDIISSVAAALOP.SYHPDIFICKAANAEKAANAALAGA.AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
P.knowlesi P.fragile P.forghei P.berghei P.chabaudi P.chabaudi P.gallinaceum P.falciparum P.falc	171 161 183 192 155 161 164 173 174 90 92 51 51 52 52 52 52 52 52 52 52 52 52 52 52 55 50	LIKNAVISAEONLJGCODTGTALILGKKDE EL. LITYEHKYLNLGVYNAY, RNNNFRYSOLSPLSMFSEETKN LPCOLETYSSVRGK LIKNAVISAEONLJGCODTGTALILGKKDE DI. LITSEHKYLNLGVYNAY, RNNNFRYSOLSPLSMFSEETKN LPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LVTYEHKYLLGVYNAY, RNNNFRYSOLSPLSMFDEETKN LPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LVTYEHKYLLGVYNAY, KNNPFYSOLSPLSMFDEETKN LPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LVTYEHKYLLGVYNAY, KNNPFYSOLSPLSMFDEETKN LPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LVTYEHKYLLGVYNAY, KNNPFYSOLSPLSMFDEETKN LPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LTYENKYLLGVYNAY, KNNPFYSOLSPLSMFDEVTKNSLPCOLETYSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE EL. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSNONLJGCODTGTALLGKKDE, EL. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, EL. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, EL. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLGVYNAY, KNNPFYSOLSPLSSVRGKUNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLGVYNAY, KNNPFYSOLSPLSSVRGK LIKNALSSKOLJGCODTGTALLGKKDE, DI. LTYENKYLGVYNAY, KNNPFYSOLSPLSSVRGKKVYTKKNEY GTELPACIDTGTALVAKKGO, CV. W. TGGGDEALSKOVYNT, IEDNLRYSONALDMYKEYTGTTLPACIDLY LLVSSRCALSKRWVCODTGTALVYKKGO, CV. W. TGGGDEALSKOVYNAT, SUDAPKSONTKKNT, TENLRYSONALDMYKEYTGTTLPACIDLY LLVSSRCALSKRENTCODTGTALVYNKKGO, CV. W. TGGGDEALSKOVYNT, IEDNLRYSONALDMYKEYTGTTLPACIDLY LLNSSRCALSKRENTCODTGTALVYNKKGO, CV. W. TGGGDEALSKOVYNT, IEDNLRYSONALDMYKEYTGTTLPACIDLY LLNSSRCALSKRENTGTOTTVAKKGO, CV. W. TGGKDEALSKOVYNT, IEDNLRYSONALDMYKEYTGTTLPACIDLY LLNSSRCALSKRENTGTOTTVAKKGO, CV. M. TGGKDEALSKOVYNL, SUDAPAGSKRCTTROTTPACIDLY LLNSSRCALSKRENTGTOTTVAKKGO, CV. M. TGGKDEALSKOVYNL, KYNNEYSOLSSNYKRASTRAKSSONALSSNYKRAN LLNSSRCALSKRENTGTOTTVAKKSGO, CV. M. TGGKDEALSKOVYNL, KYNNEYSONALDPAGSKRCTTROTTPAVINK LLNSSRCALSKRENTYN
P.knowlesi P.fragile P.yoejli P.berghei P.otabaudi P.galinaceum P.falciparum P.reichenowi E.coli_PumB G.metallireducens B.petrii V.cholerae N.meninglidis A.vinelandii Acinetobacter Metallosphaera Fervidicoccus Pyrolobus Thermoproteus Acidiobus Caldisphaera	259 249 271 280 269 252 261 262 172 174 141 137 137 137 137 138 135 130 131 132 132	AHGVKTPPGVQLSASSTAPIDSSPRYEIIFIAKGGSANKTPLPOOTKSILMEKLYDFLLEKI AHGCVTTPGVRASPPSSPPTCAASTASIARYEIIFIAKGGSANKTPLPOOTKSILMEKLYDFLLEKI NLNDK

P.knowlesi P.fragile P.fragile P.operghel P.chabaudi P.chabaudi P.falciparum P.falc	323 318 321 330 293 319 316 355 3217 219 182 178 178 178 178 178 178 178 178 178 178	K ETTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDNEKKEGGGYGKAPRDTKSERIT LEKKAGAGGTAAG. PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEKKEGGGYGKAPRDMXSERIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEKKEGGGYGKAPRDMXSERIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEKKEGGGYGKAPRDMXSERIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEKKEGGYGKAPRDMXSERIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEFKEGSIYGKAPRDIKSEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSA CP KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEFKEGSIYGKAPRDIKSEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSAC KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEFKEGSIYGKAPRDIKSEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSAC KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEFKEGSIYGKAPRDIKSEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSAC KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDDEFKGNIYGKAPRDIKSEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSAC KEGTSACTOPH. AVVIGGLSAENNLKVVG. TASCRYLDNEKKSGANGGAARDOVERSEKIT KEGTSACTOPH. AVVIGGSAENNLKVVG. TASCRYLDNEKTGGIYGKAPRDIESEKIT LLDKSGANGGAG, PGCKYFNHDVFVIREPBSAC KEGTSACTOPH. AVVIGGSAENNLKVVG. TASCRYLDNEKTGGIYGGAAPRDVERSKINGGANGGARGAC KEGTSACTOPH. AVVIGGSAENNLKVVG. TASCRYLDNEKTGGIYGGAAPRDVERSKINGGANGGARGAC KEGTSACTOPH. AVVIGGISAENNLKVVG. TASCRYLDNEKTGGIYGGAAPRDVERSKINGGANGGARGAC KEGTSACTOPH. AVVIGGISAENNLKVVG. TASCRYLDAENTERGEINGGAARDOVELEEKLONGKAGGAC KYGGAAGGCOCH. GUGGGGAAANAVAKAKAKAMAMAKANAMENTERGIYGGAAPRDVERSKER KYGGAAGGCOCH. GUGGGGAAANAVAKAKAKAMAMAKAKAMANAMANA KYGGAAGGCOCH. GUGGGGAAANAVAKAKAKAMANAKAKAMANANAKALANAMENTERGIYGGAARDOVELEEKLONGKAGGAACAC KYGGAAGGCOCH. GUGGGGAAANAVAKAKAKAMANAKAKAMANANAYA. AAANAYAKAKAKAMANANAYA TASCRYLDAAANAYA KYGGAAGGCOCH. GUGGGGAAANAVAKAKAKAMANAKAKAMANANAKAKAMANANAYA KAAAANA YAYANAYAYAYAYAYAYAYAYAYAYAYAYAYAY
P.knowlesi P.fragile P.yoelii P.berghei P.vinčkei P.galinaceum P.falciparum P.falci	418 413 416 425 388 414 451 312 275 271 275 271 271 271 271 271 271 271 271 271 271	T G 1 G VSG GADMO TK CLINK REGVE MEALEH DETK YLE DI TYN DLKO I GGVAT AT GGDD HA KNET GE GLP GE G V PININ DPM E SILKCLE HEP VST I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE DI TYK LKEV GGVAART GEBAG I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE TYK DI KS. GKFKAT KENSSEN EN VEN OLIND FM KILKLI NYP VST I G 1 G VSG GADMO TK CVINK EG VFFKALEH EP IK YLE TYK DI KN. GINK KFK KNSS DPN II. YEN NERESNE VAND LNOPMENIKLI I KNI I MYP VST I G 1 G VSG GADMO TK CVINK MOVFFKALEH EP IK YLE TYK DI KN. GINNK SS OPN II. YEN NERESNE VAND LNOPMENIKLI I KNI I MYP VST I G 1 G VSG GADMO TK CVINK MOVFFKALEH EP IK YLE TYK DI KN. GINNK SS OPN II. YEN NERESNE VAND LNOPMENIKLI I KNI VFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EP IK YLE TYK DI KN. GINNK SS OPN II. YEN NERESNE VAND LNOPMENIKLKI I KNI VFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFFKALEH EF IK YLE TYK DI I G 1 G VSG GADMO TK CLINK MOVFKALEH EF IK YLE TYK DI I
P.knowlesi P.fragile P.ycoli1 P.berghel P.vinckel P.raicharaum P.raichenowi E.coli_PumB G.metallireducens B.petril V.cholerae N.meningitidis A.vinelandil Acinetobacter Metallosphæra Fervidicoccus Pyrolobus Caldisphæra	513 501 515 478 502 522 383 339 335 338 339 335 334 312 329 318 312 328	LILLS GKUTVARD TAHKETVED FÜMRGIP TP:YFIKKPTVYAGBAKTPDNYASGSFGFTAGGEDAAAGELSVEMKGASLISLAFNRSSVVRSA LLLLS GKUTVARD TAHKETVEN DFVLHGKPTPYFIKKPTVYAGBAKTPYNYASGSFGFTAGGEDAAAGELSVEMKAGASLISLAFNRSSVVRSA LLLITGKUTVARD TAHKETVEN NKNTPTDFIKKPTVPTVAGBAKTPYNYASGSFGFTAGGEDAAAGELSVEMKASLISLAFNRSSVVRSA LLLITGKUTVARD TAHKETVEN NFNNTPTDFIKKPTVPTVAGBAKTPYNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVVRSA LLLITGKUTVARD TAHKETVEN NFNNTPTDFIKKPTVPTVAGBAKTPYNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVVRSA LLLITGKUTVARD TAHKETVEN NFNTPTDFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVVRSA LLITGKUTVARD TAHKETVEN NFNNTPTDFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVVRSA LLITGKUTVARD TAHKETVEN NFNNTPTDFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVRSVRSA LLITGKUTVARD TAHKETVEN NFNNTPTDFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFNRSSVRSVRSA LLITGKUTVARD TAHKETVEN NFNDTPTDFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFKNNSSLISLAFKNRSSVRSVRSA LVILTGKUVVARD GAHKETVEN OF IND NPTD EYFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDAAAGELSVENASLISLAFKNNSSLISLAFKNRSSUNGSSVRSVRSA LVILTGKUVVARD GAHKETVEN OF IND NPTD EYFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDSVDLIGSSGCGFTAGGEDOVINTALGKENSSLISLAFKNRSSLISLAFKNRSS LVILTGKUVVARD GAHKETVEN OF IND NPTD EYFIKKPTVPTVAGBAKTPNYASGSFGFTAGGEDSVDLIGSSGCGFTAGGEDSVDLIGSSGCGFTAGGEDSVDLIGSSGCGFTAGGEDSVDLIGSSGCGFTAGGEDSVDLIGGSSVDLIGGSNESTENASLENNSSVVRNA LVILTGKUVVARD GAHKETVEN OF IND NPTD EYFIKKPTVPTVAGBAKTPAGYPGSSLGFGFTAGGEDSVDLIGGSGGGVGGGVGGGVGGGVGGGVGGGVGGGVGGGVGGGV
P. knowlesi P. fragile P. yoejli P. berghei P. chabaudi P. chabaudi P. falciparum P. falciparum P. falciparum P. falciparum P. falciparum B. meninglidis B. petrii N. meninglidis A. vinelandii A. vinelandii A. vinelandii A. vinelandii Pervidicoccus Pyrolobus Thermoproteus Acidilobus Caldisphaera	606 596 604 601 575 614 601 575 614 475 431 427 426 400 422 406 422	CKNYHGFULGSIGGPGALLAAKNITRXWEVIDFAB.LGM. IXWHLIDYWDF. BAFIYID XGNDFY. NKWIPL. CKNYHGFULGSIGGPGALLAAKNITRXWEVIDFAB.LGM. IXWUVDFBAFILDWUDF. BAFIYID XGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWWVIDFBB.LGM. IXWUVDFBAFILDWUDF. BAFIYIDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWWVIDFBB.LGM. IXWUVDFBAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWWVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWWVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWWVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWVVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWVVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIRWVVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIKWVVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIKWVVIDFBB.LGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIKWVVIDFBBLLGM. IXWUVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIKWVVIDFBBLLGM. IXWUVDFGAFILDNVONFGAFIN. CKNYNGFULGSIGGPGALLAAKNNIKKNVVVIDFBBLLGM. IXWUVDFGAFILDVVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPGALLAAKNNIKKNVVVIDFBLLGM. IXWUVDFGAFILDNVVDFGAFILDNKGNDFY. NKWIPL. CKNYNGFULGSIGGPAALAAKNVKANDUVIDFBBLLGM. IXWUFYGNGAFILDNVVDFGAFILDNVVINGAFILDNVVVING CKNYNGFULGSIGGPAALAANNIKSNVVVIDFBLLGM. IXWISVDFY DYNKKONFY. NKWLPL. CKNYNGFULGSIGGPAALAANNIKSNVVVIDFBLLGM. IXWISVDFY DYNKKONFY. NKWLPL. CKNYNGFULGSIGGPAALAANNIKSNVVVIDFBLLGM. IXYFFYDRMN VUTVAVOTGAANAKNVLAAKN. IXWISTATUNGGAFILDNVVNNEDLLGM. IXYFFYDRMN VUTVAVOTGAANAKNVLAAKN. IXWISTATUNGGAFILDNVVNNEDLLGM. IXYFFYDRM VUTVAVOTGAANAKNVINFFYDLLGM. CKNYNGFULGSIGFAALANNIKSNVVNNEDLLGM. IYYFFYDRM VUTVAVOTGAANAAKNI IFFICATUN CKNYNGFULGSIGFAAAAANNIKSNVVNNEDLLGM. IXWISTATUNGGAALAANNIKSNVVNNEDLLGM. IXWISTATUNGGAAAAANNIKSNVVNNEDLLGM. IXWISTATUNGGAAAAANNYNNETGAAAAAANNAKNVNNEDLLGM. IXWISTATUNGGAAAAANNYNNETGAAAAAANNYNNETGAAAAAANNAKNVNNEDLLGM. IXWISTATUNGGAAAAAANNYNNETGAAAAAANNYNNETGAAAAAAANNANNAKAAANNAKAANNYNNETGAAAAAAANNYNNETGAAAAAANNYNNETGAAAAAAAANNYNNETGAAAAAAAAAAAANNYNNETGAAAAAAAAANNYNNETGAAAAA

Figure S1. Multiple sequence alignment of *Plasmodial*, **single-subunit and two-subunit bacterial and archaeal FH proteins sequences.** The black solid line shows 120 amino acid insertion at the N-terminus in *Plasmodium* FH sequences. The black arrow shows invariant cysteine residues that are involved in Fe-S cluster ligation. The green bar indicates the boundary between the N- and C-terminal domains.

P.VIVAX	1	MRNFARIPLCRPPPGKRLSRGALLG
P.inui	1	MRNEVRIPLCRPPSGKYLSKSAHLG
P.Knowles1_strainH	1	MRNETRILLCRSPSGKCLSKSAHLG
P.fragile	1	
P.yoelii	1	MVGYQQARKFKRIHKLFPYTICTNNLVEISGKNNINNINNINNINKLNINTINN
P.berghei	1	MVGYQKIRKFKKVPKLFPYTICTSNFIGVHGKKHIRNISNINNIRNINNINKINKLNINTLNN
P.Vinckel	1	
P.chabaudi	1	MVAYQKVRRFKKVPSLFSYTICTNKFMGIHGKNHISNINKINKLNINTLNN
P.gallinaceum	1	GIKTVNINSI.YKINKLNINTLNK
P.gaboni	1	MIKFKGGSILLSH.NAYFKYNLYLKKIRCVYNNIYRRDVNTLNN
P.falciparum_3D7	1	MIKFKEASILLSHKNAYLQYNLYFKKIRCVYHKIYRRHMNSLNS
P.reichenowi	1	MIKFKEASILLSHKNAYI
		N-terminal non-conserved residues
D	4.7	
P.VIVAX	43	FLOVILE SEGREDE ISTRRIDDLSKTIEVIKLKDNPIL. NOSKTYGYNFE NEENFFHPNGEDK
P.inui	43	FLOVLEFEDAKBUEIEFKRIDDESKIVEVIKIKESPI.NQSKYYGYNFENEDNFFHPNGEDK.
P.knowlesi_strainH	43	FLDVLEFEEGREDDTEYRRIDELSKYIEVVKIKDSP1.NESKYYGYNFENEDNFFHPNGELK.
P.fragile	32	FLDVLEFEEGREDDTEYRRIDELSKYVEVIKIKDSPI.NESKYYGYNFENEENFFDPNGELKK
P.yoelii	55	FLDTFEFEGRGNDGIEYRRLDDLSKYIEVIKFNN. KINDDNKYYDINYEDENEFFDDNGNDKI
P.berghei	64	FLDIFEFEEKGNDNIEYRRLNDLSKYIEVIKFDN.KINENSKYYDINYEDENEFFDDNGNLKI
P.vinckei	26	FLDTFEFEEKGNDGIEYKRLDELSKYIEVIKLDNNKINKDSKYYDINYENENEFFDENGNLKI
P.chabaudi	52	FLDIFEFEERGNDGIEYRRLDELSKYIEVIKLDNNKINKDSKYYDINYENENEFFDENGNLKI
P.gallinaceum	40	FLDVLDFEEEEDIEYTRIDELSKYIETVKINEDIF.NKIKYVGYDFIDINNFFDKNGNIKI
P.gaboni	44	FIDILSFRNDEDDDIEYKKVEDLSKYIEVIKMNKSSM.NETKYYGYNFKDENNFLDEHGNIKE
P.falciparum_3D7	45	FIDILSFRNEDDIEYKKVEDLSKYIEVIKINKSPM.NETKYYGYNFKEEYNFLDEHGNIKE
P.reichenowi	45	FIDILSFRNEE.DDIEYKKVEDLSKYIEVIKINKSPM.NETKYYGYNFKEENNFLDEHGNIK
D	104	
P.VIVAX	104	NEP 20. VNQNEGERVKETENVPPFVLAKLORDINE RETERETING GESSTADA
P.inui D. basulasi stasisu	104	KLPEQ. WMQSEGERIKETIHVPFFVLAKLCEVARREILFFLNKKHLKQLSNILNDEESKNDK
P.Knowlesi_strainh	104	NEP 20. VRQNEVERIKETIHIPPFVLIKLCETAFRETEFFLNKKHLKOLSNILKDGESSKNDK
P. fragile	94	NLPEQ. VIQNEGERIKEYIHVPPFVLIKLCEYAFKEILFFLNKKHEKQLSNILNDGESSKNDK
P.yoelii	11/	KNNYERESKKNVIKEYIHVPPFVLIKLCEYALKEILFFLNKNHLKOLONILIDKNSSENDK
P.berghei	126	KNN. YERESNKNIMKEYIHVPFFVLIKLCEYALKEILFFLNKNHLRQLONILMDKKSSKNDK
P.VINCKel	89	KNDCERKSNKNVMKEYLHIPPFVLTKLCEYALREILFFLSKKHLRQLONILMDKESSENDK
P. chabaudi	115	KND., CEKENNKNVMKEYIHVPPFVLTKLCEYALREILFFLNKKHLROLONILIDKESSDNDK
P.gallinaceum	100	YSKODTNKKDAKEYIHVPPFVLTKLCEYAFKEILFFLNKKHLNQLONILYDKESSKNDK
P.gaboni	106	YINNENKKLLDKNYEKEYIHIPPFVLTKLCEYAFKETLFFLNKKHLKQLONILODKESSKNDK
P.falciparum_3D7	105	YIYNENKKLLYKNYEKEYIHIPPFYLIKLCEYAFKEILFFLNKKHLKOLOHILODKESSKNDK
P.reichenowi	106	YIYNENKKLLHKNYEKEYIHIPPPVLIKLCEYAFKEILFFLNKKHLMQLQHILQDKESSMNDK

Figure S2. Multiple sequence alignment of FH sequences from different *Plasmodium* **species.** The initial 40 residues (Pf numbering) highlighted by a solid black line are not conserved. Many residues in the stretch of 41-120 residues, though not present in bacterial and two-subunit type FH sequences are highly conserved in all *Plasmodium* FH.

Methods S1. Generation of *∆fumACB* strain of *E. coli*

To knockout *fumA* and *fumC* genes in the *E. coli* strain JW4083-1, established protocols were followed (5). Briefly, to remove the kanamycin cassette at the *fumB* locus, JW 4083-1 cells were transformed with the plasmid pCP20 that carries the gene for the enzyme flippase. Thereafter, the cells were cured of the plasmid by incubation at 42 °C for 12 h. Subsequently, the cells were transformed with the plasmid pSC101 to introduce the high-efficiency λ -phage recombination machinery. In order to knockout the genes *fumA* and *fumC*, oligonucleotide primers P1 and P2 (Table S1) were used to amplify a kanamycin cassette using plasmid pKD13 (5) as template, such that the PCR product carries the homologous regions corresponding to the 5' flank of *fumA* and 3' flank of *fumC* (Figure S3 a). The PCR product was DpnI treated and thereafter used for transformation. Transformation was performed by electroporation and cells were selected on LB medium containing kanamycin. Kanamycin resistant colonies were screened by PCR using oligonucleotide primers P3/P4 and P5/P6, respectively (Figure S3 d) The kanamycin cassette at the *fumA/C* gene locus was removed by expressing flippase in the strain resulting in the generation of marker free fumarate hydratase null strain.



Figure S3. Genotyping of the *E. coli* **strain** Δ *fumACB.* (a) The relative orientation of *fumA* and *fumC* genes in the Δ fumB *E. coli* strain JW4083-1 and location of primers (P3-P6) used for validation of the knockout. 5'HR1 and 3'HR2 represent the 30 bp homologous regions used for gene replacement with kanamycin resistant marker. (b) fumA/C gene locus after homologous recombination and replacement with kanamycin resistance marker flanked by FRT sites. The orientation of the primers (P1 and P2) used for amplification by PCR of the kanamycin cassette is shown. (c) PCR amplified products using primers P1 and P2 and genomic DNA of 5 different colonies selected on kanamycin plate as template, showing the presence the kanamycin selection cassette integrated into the right locus. (d) Lanes 1-3 amplicons obtained on performing a PCR to check the presence of *fumA* gene using primers P3 and P4 and genomic DNA from different *E. coli* strains (as mentioned in the figure) as template. Lanes 5-7, amplicons obtained on performing a PCR to check the presence/absence of *fumC* gene using primers P5 and P6 and genomic DNA from different FH knockout strains (as mentioned in the figure) as template. The size of DNA fragments is indicated in bp.

Text S1. In vitro enzyme inhibition studies

The small molecules examined for their effect on FH activity with fumarate as substrate were DLmercaptosuccinic acid, meso-tartaric acid, succinic acid, aspartic acid, hadacidin (N-formyl-Nhydroxyglycine), alanosine, pyromellitic acid, itaconic acid, malonic acid, citric acid, trans-aconitic acid, D-malic acid, gamma-aminobutyric acid, alpha-ketoglutaric acid, oxaloacetic acid, DL-2 amino- 3 phenyl propionic acid, L-tartaric acid, sulfosuccinic acid, and ureidosuccinic acid. All molecules were tested at a concentration of 500 μ M.



Figure S4. Generation of *P. berghei* FH knockout construct using recombineering. (a) Schematic representation of the PbFH genomic clone (PbG01-2466a09) obtained from PlasmoGEM. (b) Schematic representation of the intermediate vector in which the gene of interest is replaced by the bacterial selection marker, *ble* (resistant to zeocin). The position of validation primers QCup and PheSR2 are indicated. The agarose gel shows the band of expected size (1505 bp) obtained by PCR using QCup and PheSR2 oligonucleotide primers and the intermediate vector DNA isolated from a bacterial colony (L1) from the zeocin plate, L2, molecular weight marker. (c) Schematic representation of the final knockout construct in which the Zeo-Phe cassette is replaced by *Plasmodium* positive/negative selection cassette, hDHFR-yFCU. The position of validation oligonucleotide primers QCup and hdhfr are represented by black arrows. The agarose gel on the left shows the band of expected size (1905 bp) obtained by PCR using primers QCup and hdhfr and plasmid isolated from a bacterial colony (L2) (obtained upon negative selection in an LB-agar plate containing p-chlorophenylalanine) as template. L1, molecular weight

marker. The agarose gels in the middle and right panels show the restriction mapping of the PbFHKO construct by BamHI and, NheI and HindIII digestion, respectively. All the bands of expected size are present. For transfection in *P. berghei*, this verified knockout construct was used.



Figure S5. Genotyping of *P. berghei* clones of knockout of fumarate hydratase. (a) Schematic representation of the selectable marker cassette inserted into the *fh* gene locus of *P. berghei* genome. Primers (P1-P8) used for diagnostic PCRs are indicated. (b) Schematic representation of the *fh* gene (PBANKA_0828100) flanked by 5' UTR and 3' UTR showing the location of primers P9 and P10. Agarose gel electrophoresis of PCRs with genomic DNA from (c) clones A-G (left panel) and clones H-Q (right panel) for detection of 5' integration; (d) clones J, M and Q for detection of 3' integration (other clones did not answer for this PCR); (e) clones A-G (left panel) and clones H-Q (right panel) for the

detection of *fh* gene; (f) clones A-G (left panel) and H-Q (right panel) for the presence of selectable marker cassette; (g) clones C and O using primers P3 and P8. Clones C, M and O did not answer for 5' integration while only clones J, M and Q answered for 3' integration. All clones answered for the presence of the *fh* gene (Panel d). All clones except C and O answered by PCR with primers P2 and P6 (left and right panels of f) indicating the integration of the entire selectable marker cassette into the genome. Clones C and O answered for a shorter fragment of the selectable marker cassette covered by primers P3 and P8 (panel g). hDHFR-yFCU, human DHFR-yeast cytosine and uridyl phosphoribosyltransferase, u, uncloned population, mr, molecular weight marker; wt, wild-type *P. berghei* genomic DNA; pl, pJAZZ-FH knockout construct (supplementary figure S4); nt, control PCR without template. Numbers to the right of panels d, e, f, g and h are the sizes of the marker DNA fragments in kbp.

REFERENCES

- 1. Claros, M. G., and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–86
- 2. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–16
- Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P., and Imai, K. (2015) MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol. Cell. Proteomics*. 14, 1113–1126
- 4. Bender, A., van Dooren, G. G., Ralph, S. A., McFadden, G. I., and Schneider, G. (2003) Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **132**, 59–66
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008