

Supporting Information

Figure S1. Genes associated with Huc and Hhy in *Mycobacterium smegmatis*. (a) Organisation of the genes associated with the group 1h [NiFe]-hydrogenase, Hhy. (b) Organisation of the genes associated with the group 2a [NiFe]-hydrogenase, Huc. Abbreviations: HMP = hydrogenase maturation protease; NHL = NHL repeat protein; GmhA = phosphoheptose isomerase homolog; AraC = arabinose regulatory protein homolog. The length of the genes is shown to scale. The red line is used to denote genes that are transcribed in operons, as verified by previous RT-PCR studies (Berney et al., 2014b).

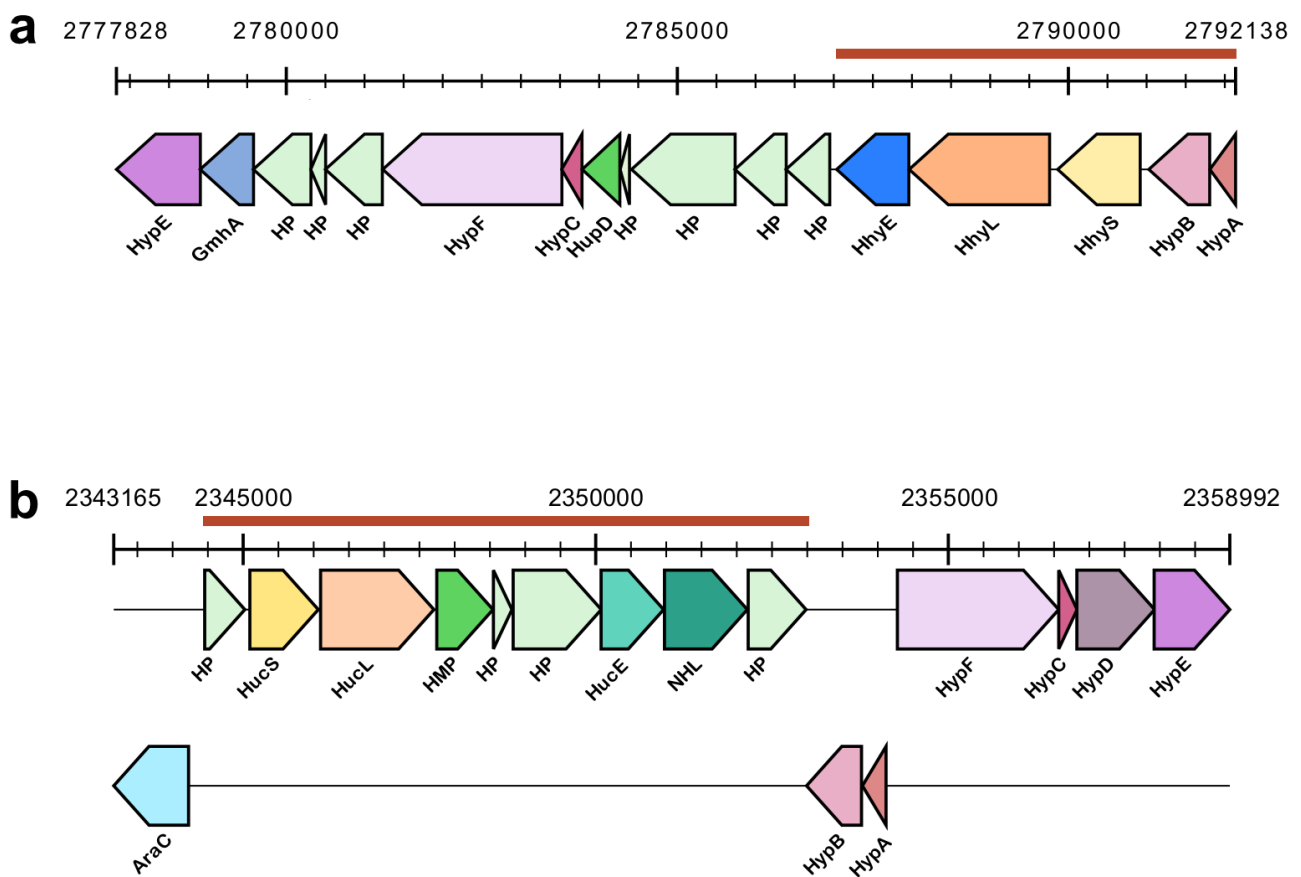


Figure S2. Deletion of the *hucE* and *hhyE* genes in *Mycobacterium smegmatis*. The schematic shows the four main steps that led to the production of knockouts of (a) *hucE* (MSMEG_2268) and (b) *hhyE* (MSMEG_2718). (i) Construction of the pX33-*hucE* / pX33-*hhyE* vectors containing a fused left flank (LF) and right flank (RF) of each gene. (ii) Temperature-mediated integration of the vector into the *M. smegmatis* chromosome via either the left flank or right flank of the chromosomal *hucE* / *hhyE* genes. (iii) Chromosomal excision of the vector due to *sacB*-mediated sucrose toxicity to either wild-type revertants or

ΔhucE / *ΔhhyE* mutants. (iv) PCR-based screening through primers (2267_chrom_fwd, 2269_chrom_rvs, 2717_chrom_fwd, and 2719_chrom_rvs) targeting the flanks to confirm whether colonies are (a) wild-type revertant (2774 bp product) or *ΔhucE* mutant (1890 bp product) and (b) wild-type revertant (4110 bp product) or *ΔhhyE* mutant (3183 bp product). Both primer sets were used to confirm the double mutant.

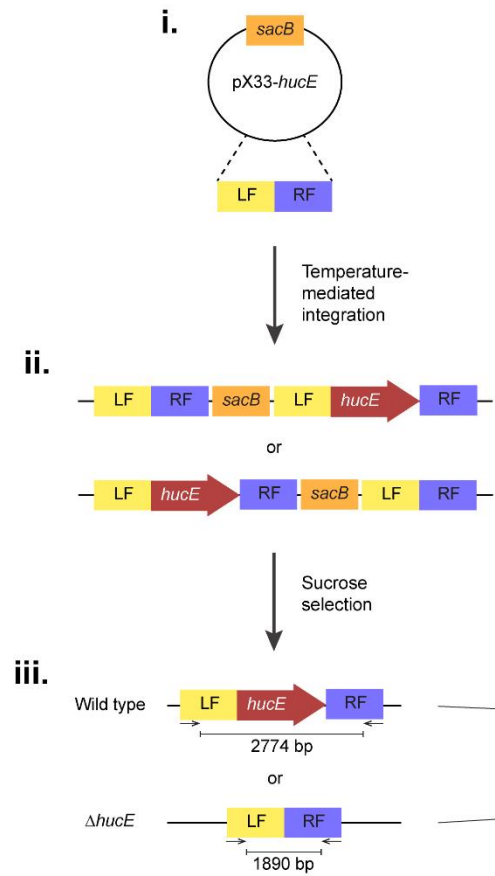
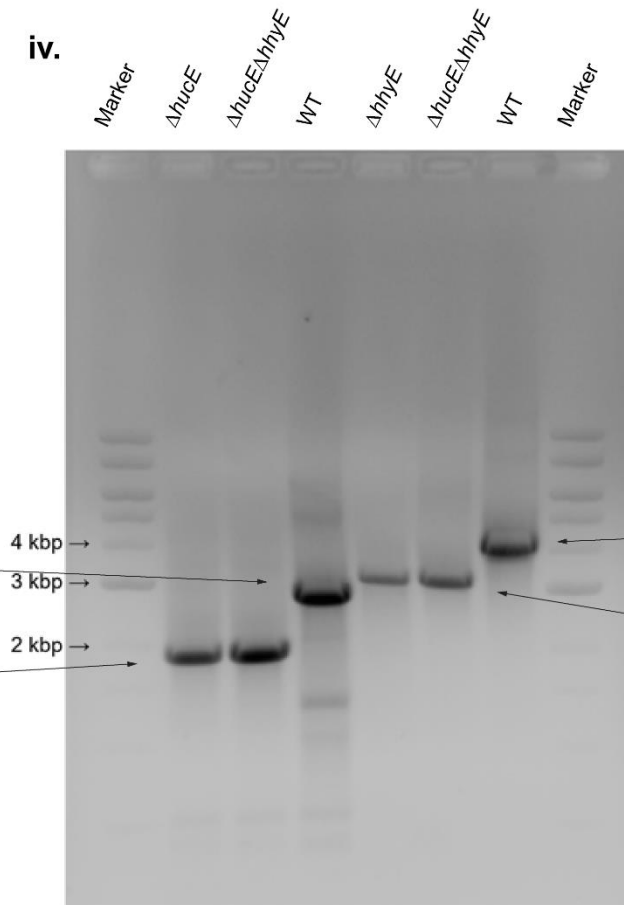
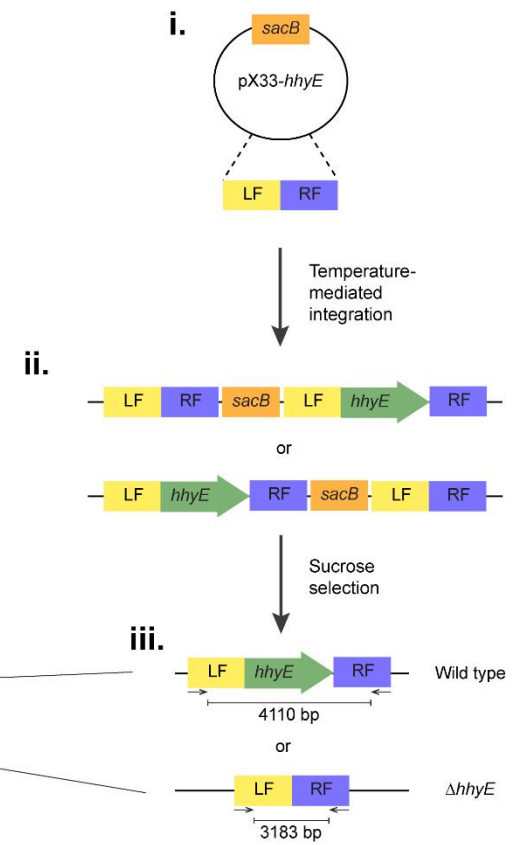
a**iv.****b**

Figure S3. Construction of complementation vectors. (a) pMV*hucE* digested with PstI and HindIII (left side) produced two fragments (893 bp and 4480 bp) consistent with *hucE* insertion to the plasmid. Map of the plasmid is shown on the right side. (b) pMV*hhyE* digested with BamHI and HindIII produced two fragments (932 bp and 4468 bp) consistent with *hhyE* insertion to the plasmid. Map of the plasmid is shown on the right side. The plasmids have an *Escherichia coli* pBR322 origin of replication, a mycobacterial OriM derived from pAL5000, a kanamycin resistance cassette (KanR), and a promoter P_{hsp60}.

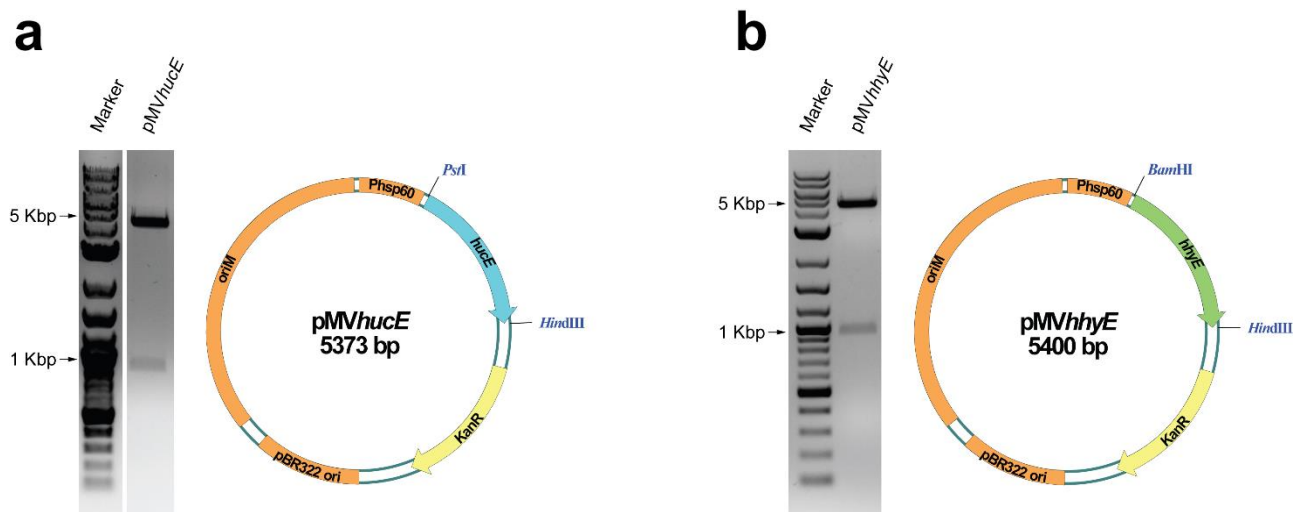


Figure S4. Distribution of HucE and HhyE proteins in hydrogenase-encoding microorganisms. (a) Distribution of HucE proteins in genomes of bacteria that encode group 2a [NiFe]-hydrogenases. (b) Distribution of HhyE proteins in genomes of bacteria and archaea that encode group 1h [NiFe]-hydrogenases. Genomes are coloured depending on whether a full-length protein is present, a truncated protein is present, or the protein is absent.

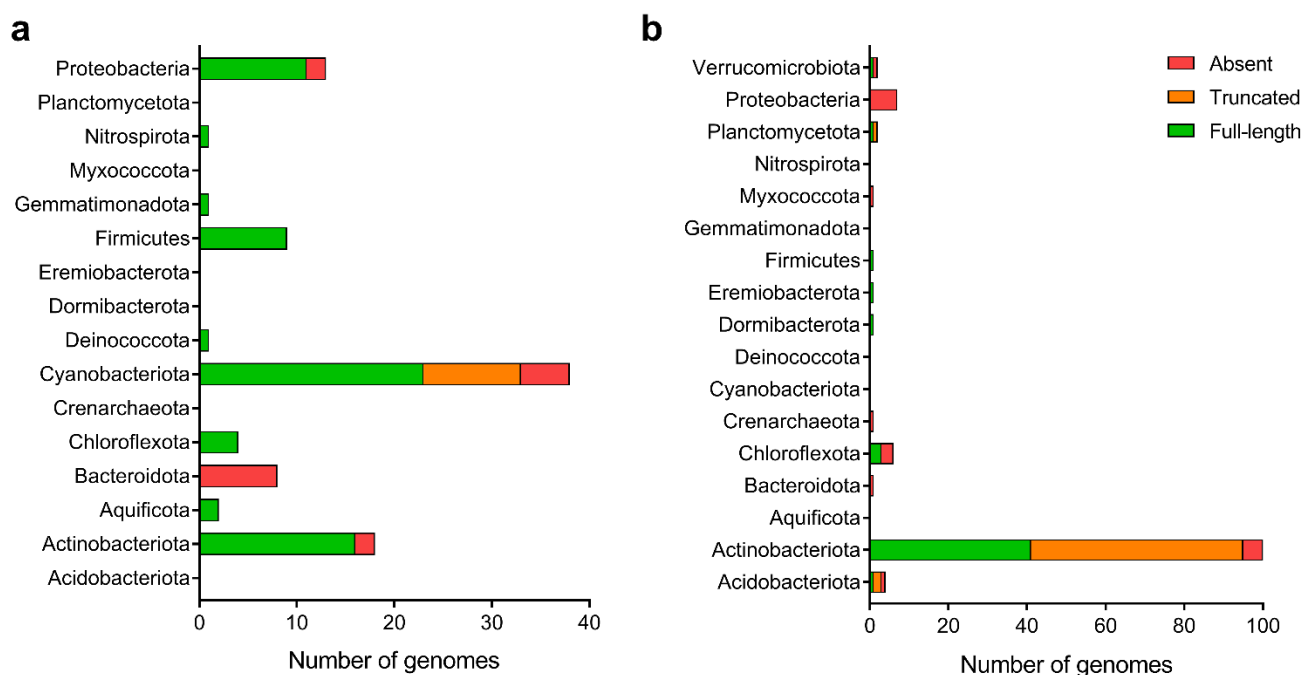


Figure S5. Multiple sequence alignment of the HucE and HhyE family proteins from bacteria that have been shown or predicted to mediate aerobic H₂ respiration. The proteins shown are: WP_011728268.1 (*Mycobacterium smegmatis* HucE), WP_013678347.1 (*Pseudonocardia dioxanivorans* HucE), WP_053382025.1 (*Nitrospira moscoviensis* HucE), WP_012963098.1 (*Hydrogenobacter thermophilus* HucE), WP_015213936.1 (*Anabaena cylindrica* HucE), WP_015115404.1 (*Nostoc* sp. PCC 7107 HucE), WP_016860574.1 (*Fischerella muscicola* HucE truncated), WP_003894102.1 (*Mycobacterium smegmatis* HhyE), WP_041978571.1 (*Pyrinomonas methylaliphatogenes* HhyE), WP_084922766.1 (*Rhodococcus equi* HhyE), WP_010988763.1 (*Streptomyces avermitilis* HhyE truncated), PZR71543.1 (*Candidatus* Dormibacterota bacterium HhyE), PZR61736.1 (*Candidatus* Eremiobacterota bacterium HhyE). The conserved cysteine and histidine residues predicted to bind iron-sulfur clusters are coloured in red and the motifs are highlighted.

WP_011728268.1	-----MTAAVEAPELEKLAQRVDNAVRA--LVDLDPEARKVAE	36
WP_013678347.1	-----MTASTKSTPAASSGPAATEAEPSFEDLAGRVDRAVEA--ASALDGAAGEAAA	50
WP_053382025.1	-----MEKQESTVLETQPTLASLVKSIKSLEVI--VEGWNDSQRLTVQ	41
WP_012963098.1	-----MTEELSFEKLAERIDELLAK--VKNFEDEKRETVG	33
WP_015213936.1	-----MNQLEELVEDINRFEAI--IAEWDESQRCVAV	30
WP_015115404.1	-----MTNLEEFVVEINRFEAI--ISEWDESQRCVAV	30
WP_016860574.1	-----MTNLEELVQEISRFEAI--ISEWEESQRCVAI	30
WP_003894102.1	---MAPAVTPLADVP--VDT---Q-SDDEARWRTAGERIESLLDA---AAGPTARARAE	47
WP_041978571.1	-----MEQ---MQQACARRIEELVQRIESLTD-----ARARDVAV	32
WP_084922766.1	-----MASAHDE--TRY---E-TPESDRWRQAGDRIEALLDA--TSAGGPVARERAE	44
WP_010988763.1	-----MSAVS-----ATNRVTAEQTGRRIEDVLDRLAA--TGDPAAGAAAE	39
PZR71543.1	-----MAIL-----QRK---EFDRRLQQTESLINEVEALAD-----EEVKRRRAV	36
PZR61736.1	MDERTPALTEGHDRRSAPPD---HLARRPEDPRALGDRVDALLQQLATSPNPAHARAKAE	57

..

WP_011728268.1	ELKAALEEIHRVGLVTIVKMRSGE-----ATREALFELVDDPVVRMLFTLHGIIRP	88
WP_013678347.1	ELRAAIEEIHRVGLRAIVRAMRER-----ETRDLLFTLVDDPVVHLLSLHGIVRP	102
WP_053382025.1	ALRKAIEDLHKEAFSRLIRGIKADP-----AAIPALRQAVGDEIVYAVLRHLGLLKP	93
WP_012963098.1	ELIKSIEEFTRMALVKLVKLMKEDS-----AGKDILLKAVREPEIYSLFLKHGIIRE	85
WP_015213936.1	GLKRSIEALHKAALTNLIKSLKQN-----NMSGLRDAVSDEIVYAVLLYHELVKP	80
WP_015115404.1	GLKRAIEALHKEALTRLIKSLKQE-----SISALRQAADDEVVYAVLLYHDLVKP	80
WP_016860574.1	GLKRAIEDLHKEALTRLIKSVKQE-----SISALRNAVQDEVVYGVLLYHELVKS	80
WP_003894102.1	QLVREVTDLYGAAALERMMLRAAV-----TAAPELAETFAADDLVASLLLHGLHPL	97
WP_041978571.1	ELVQELMRMHGAALERLLEIVWG-----RNEPELIEAFARDDLVGPLLLLYDLHPL	83
WP_084922766.1	QLVREVVDLYGEALGRVLTIAA-----RAPGLVDELARDELVSSMLLVGGLHPL	93
WP_010988763.1	ELVRTLMFEFYGSGLARILHLLSSAPGG---PRPPGDPLAALLGDELVASLLVLHDLHPE	95
PZR71543.1	EAIQALLNLHGEGLERTLELISDS-S-----NQQQALIDELGKDPVVGGLLLHGVHPL	89
PZR61736.1	ELVSTLVRLYGAGLTRILEIVDENAG-----DSGGRIFRHLCDDDLVASLLILHGLHPL	111

: : .: : : : : :

WP_011728268.1	DT---TTLAEQALAQVRPQLRSHGGDVTLVRI---ES-GTAFVRL-EGA	CNGC	SMSAVTL	140
WP_013678347.1	DP---ATHAQRVLTTRVPQLQSHGGDVAFAHV---AD-GVAYVRL-EGA	CNGC	SMSAVTL	154
WP_053382025.1	S---LNERIERALASVRPLLQGHGGNVELVRV---DPPDSVEIKL-LGA	CDGC	PASGLTL	146
WP_012963098.1	DD---RTKAIKALELIRPYIRSHGGDVEFVDL---KE-KTLYVRL-RGT	CTGC	SQVSFTL	137
WP_015213936.1	P---LAERIQTALAEVRPGLQSHNGDIELVAI---KPPDTVEVKL-IGS	CSSC	PSSTLTL	133
WP_015115404.1	PLPPILQRIQAALEEVRPGLKSHNGDIELVAF---KPPDTVEVKL-IGT	CSSC	PASHLTL	136
WP_016860574.1	PTPPLQQRIQQALEEVRPGLKSHNGDVELVAI---KAPDTVEVKL-IGT	CGNC	PASTLTL	136
WP_003894102.1	P---VHRIADALDRVRPYLGSHGDDVRLLDVPEADGAVARLRF-SGS	CKSC	PPSSAATL	153
WP_041978571.1	D---QETRVREAIERVRPYLRSHGGDVELLRI---SDGVVHLRL-RGS	CHGC	ASSAQT	135
WP_084922766.1	D---VETRVRTALDSVRPYLGSHGDDVELVDVT---DDGVVRLRL-LGS	CHGC	PSSAVTL	146
WP_010988763.1	D---RDVRIARALDSVRDQAL---EVVAFD---ETSGTLRLRTRAGAG	C-GC	AGTGTA	144
PZR71543.1	G---LEDRLVQALDKVRPYLGSHGDNVQLVGL---DEGVVRLRL-EGT	CQGC	ASSAVTL	141
PZR61736.1	S---LEERIVQALDRVRVYLKSHGDDVQLLRI---EGETVYLRM-AGS	CHGC	PSSASTM	163

.: :* : : : : *

WP_011728268.1	RQIVETALLEGVQGVSKVEVLPNE-----PTPTVISIDAVLAGSTVRRRAEEGWVRAGSV	195
WP_013678347.1	RNLVEEALVAQVPAIRSVEVLPNE-----PTPTLISVESMFPRST---GLPPGWVDGLGT	206
WP_053382025.1	AEGIEKAIREHCPEIITIKKVKGI-----PPSSNVNGASVFFVSPFARSSDAGWQFAVEL	201
WP_012963098.1	QQTILEAVQAYIPHIEKVELAKDT-----PVEAFLELS-----GKEEGYIKAFDI	182
WP_015213936.1	TQGVEEAIAKKHCPEITKVIAN-----SSPNVT--ETNFNSPFPLPTDFYVVRVATV	183
WP_015115404.1	SQGVEQSIKNYCEINNVAVG-----DRSTVDNTNTGLTTFPPTITTAWVRVATV	188
WP_016860574.1	SQGIEQTIKTHCPEINHVIAS-----	158
WP_003894102.1	ELAISDAVLAAAPEVSSIEVVNAEPFEP---AAVIPAQSLARVHSDGRRHVEWHAVPEL	209
WP_041978571.1	RLAIEAAIYEAAPEIAALEVEGVAQ---PATSFVPLESGRGNANGKRKTHGVWEVEAP	191
WP_084922766.1	QLAVEGAVQAAAPETTAIEVETEDRETA-PTPGVFTVDSLMSHVREEPTRTGTWLAEEF	205
WP_010988763.1	VQAAEDALACFAPEVTAIVEMQPAAPG-EPALLQIGTGPNSGAVARPAAAKTA-----	196
PZR71543.1	KFAVEREIMEAAPDVVAIEVADDSAER--PPEGFIPLGQIKPLA--PKPASTDWEVVAGV	197
PZR61736.1	KLAIERAIHDAAPEITEIRAEGVHDGA-----VTRKPTEWVSLAAL	204

: :

WP_011728268.1	DAAAVDDVTPVTVTAPDGSEVRLLLVNLGQRMSAYRNE	CAHEALPLDGAVL-----DVG	249	
WP_013678347.1	DDVAPGALAAVQLAADDGTGADVIVNVNDRISAYRNA	CAHEGLPLDGALL-----DVR	260	
WP_053382025.1	AEIPDIGIK-----VLELAGHSLLFSRTDQVVTFCFENA	CAHMGMPDLMGLI-----A	248	
WP_012963098.1	SELKEGHVY-----RFLHEGVVDIIMLWKGRVYAYRNS	CAHQGHPLHEGEL-----T-E	230	
WP_015213936.1	DQINKSSVL-----SVRIANQDILINRQGENITCYRNS	CSHLGYPLDKGKV-----E	230	
WP_015115404.1	EQIPDLGIY-----APKVAGHALILYRQGDRTICYRNA	CSHLASPLDTGKV-----E	235	
WP_016860574.1	-----	-----	158	
WP_003894102.1	DGLSPGQVG-----GFRVDGATVMACRVGDDYFAYHDR	CAWC	GSTFAGAVL-----T	256
WP_041978571.1	LSLAPGSAR-----IEARGRILFCRLDDRFAYEDR	CPAC	GGALDAARF-----E	238
WP_084922766.1	AELAPGEVG-----GFTVGGLSMLVCRIGDELFAVRDR	CPAC	GNSMAGTSLHRRAGGHVN	260
WP_010988763.1	-----	-----	196	
PZR71543.1	GHLESGSTK-----VVRVGGSAVLFCSWGGETTFAYRDS	CPAC	GASLVGGDL-----C	244
PZR61736.1	PDLSDNGMA-----SCEVEGMPVLFRLSPDALYAYRNQ	CPRC	LVGLSRASL-----R	251

WP_011728268.1	NGTLT	CPWH	GFCYDAT-SGECMSAPGVQLEPLPLRIDDGDIWVRVTG-----	295
WP_013678347.1	AGTLT	CPWH	GFCYDAT-SGECLSAPGAQLEQLPVRVDGDRIWVRVHT-----	306
WP_053382025.1	DGVIT	CPYH	GFRYDLS-SGECLTAPQVQLVSHAVRVVVGSKVEVKLRS-----	294
WP_012963098.1	KGVLV	CPWH	RFEYSIT-SGECLTVPYVQLVSVPTKIQNGWVLLKVR-----	275
WP_015213936.1	KGIT	CSA	HEFQYDLK-TGKCLTTPDISLHSPVKIKGDKVYIKVPNSNQK---	280
WP_015115404.1	NGIIT	CPVH	GFQYQLD-TGQCLT-ADVPLQSYLVQIKGDKVFKFSR-----	280

WP_016860574.1	-----	158
WP_003894102.1	GQVLRCPHCATGFDVVHAGLSD---SAQLEPVVLMRDGVPTLALRSTSGGAGP	307
WP_041978571.1	DVMLVCPVCGRRYDVVRAGRASDGANPHLEPLPLLEHGAKIALPA-----	285
WP_084922766.1	DAVLRCPTCRTHYDVRRAGIGVEDPDGHLDPLPVLTRDGVLSVAVPAAVTG---	311
WP_010988763.1	-----	196
PZR71543.1	EAILTCAACQARFDVKLAGRGVDN-ELHLEPLPLLQTDGEVRLAGAAKS-----	292
PZR61736.1	WPLLA CVSCGQEYDVVKAGRAPDQPELHIEPFPLVVSGDKVQLAIPVVA-----	300

Figure S6. Conservation of cysteine and histidine residues predicted to bind iron-sulfur clusters in HucE and HhyE proteins. (a) WebLogo of the 52 full-length HucE sequences shown in **Figure 1**. (b) WebLogo of the 25 full-length HhyL sequences shown in **Figure 1**. The numbering refers to the residues of *M. smegmatis* HucE and HhyL respectively as per **Figure S5**.

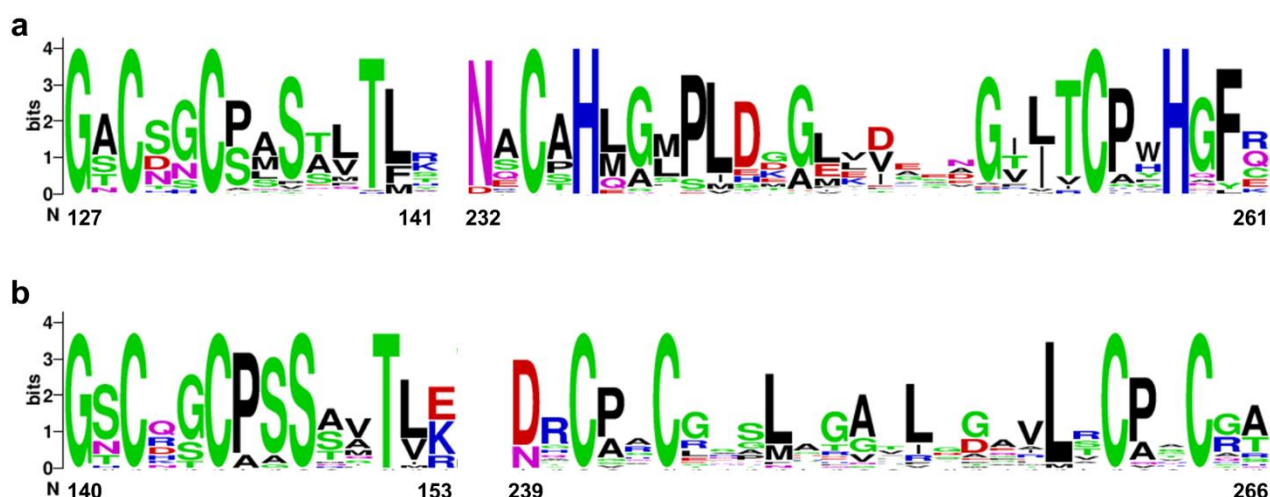
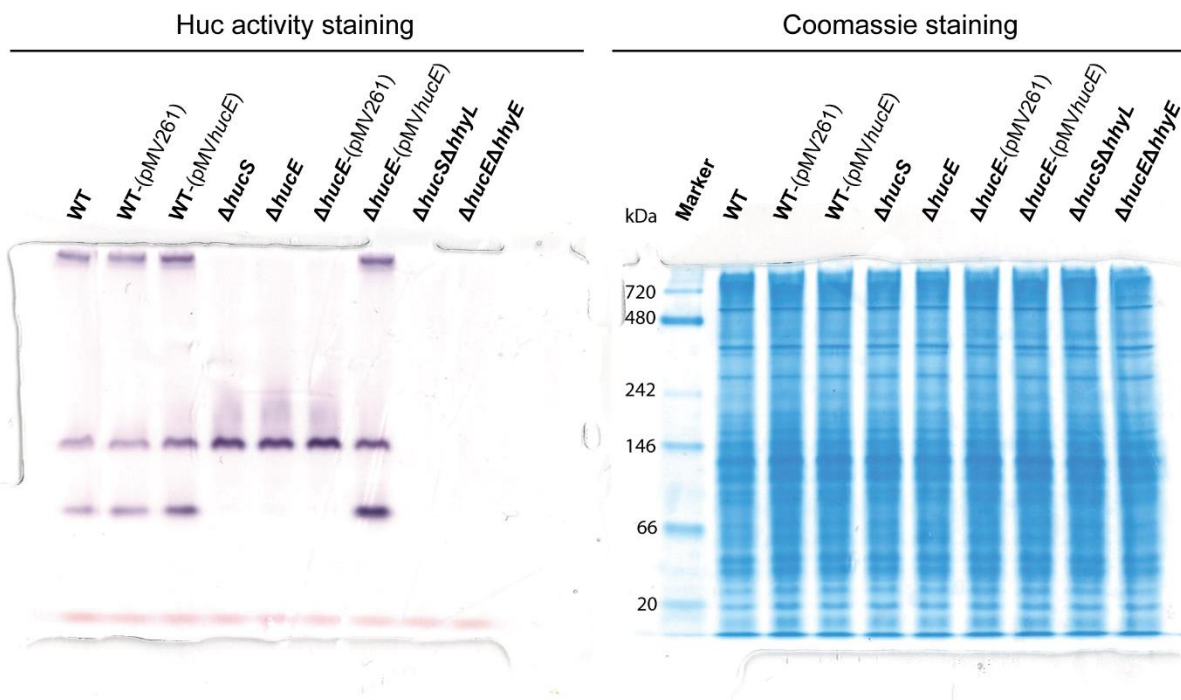
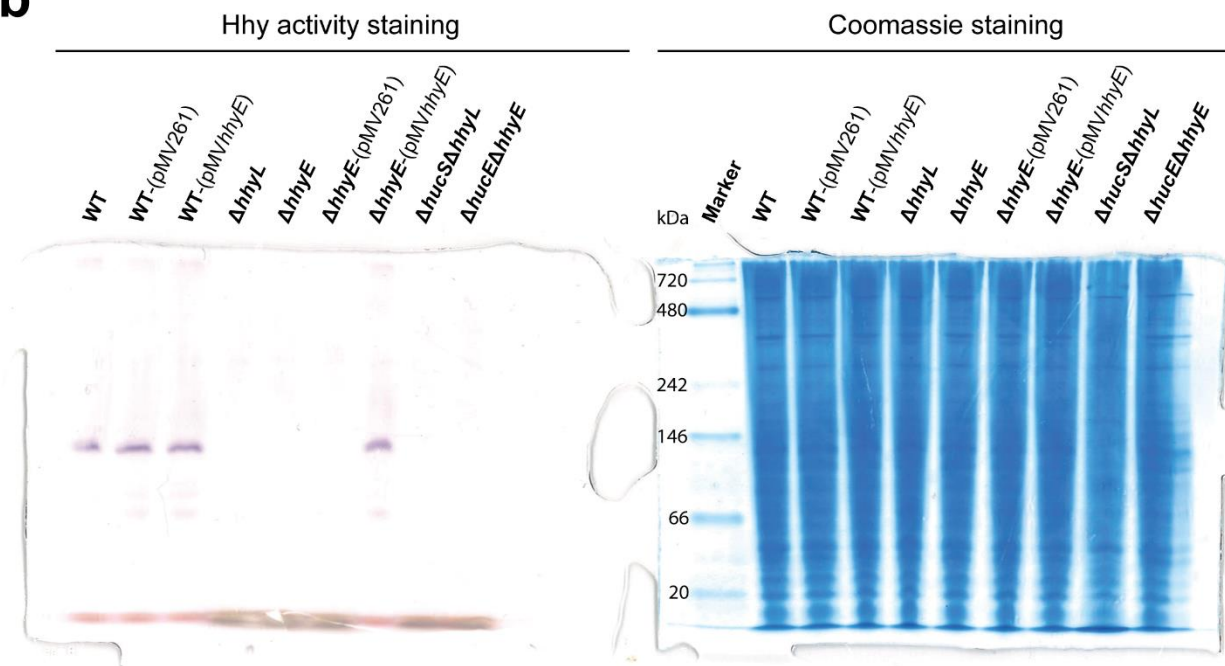


Figure S7. Original native polyacrylamide gels showing staining of hydrogenase activity. (a) Huc activity and (b) Hhy activity. The left gels show hydrogenase activity stained with the artificial electron acceptor nitroblue tetrazolium in a H₂-rich atmosphere. The right gels show protein marker and total protein stained with Coomassie Blue.

a**b****Table S1. Bacterial strains and plasmids used in this study.**

Strain	Description	Reference
mc ² 155	Electrocompetent, wild-type strain of <i>Mycobacterium smegmatis</i>	(Snapper et al., 1990)
Δ hucE	mc ² 155 with markerless deletion of MSMEG_2268	This study
Δ hhyE	mc ² 155 with markerless deletion of MSMEG_2718	This study

$\Delta hucE\Delta hhyE$	mc ² 155 with markerless deletions of MSMEG_2268 and MSMEG_2718	This study
$\Delta hucS$	mc ² 155 with markerless deletion of MSMEG_2262	(Berney et al., 2014b)
$\Delta hhyL$	mc ² 155 with markerless deletion of MSMEG_2719	(Berney and Cook, 2010)
$\Delta hucS\Delta hhyL$	mc ² 155 with markerless deletions of MSMEG_2262 and MSMEG_2719	(Berney et al., 2014b)
DH5 α	Chemically competent, wild-type strain of <i>Escherichia coli</i> F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> ϕ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169 <i>hsdR17</i> (rK ⁻ mK ⁺) λ ⁻	Thermo Fisher
TOP10	Chemically competent, wild-type strain of <i>E. coli</i> F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 araD139</i> Δ (<i>ara-leu</i>)7697 <i>galU galK</i> λ ⁻ <i>rpsL</i> (Str ^R) <i>endA1 nupG</i>	Thermo Fisher
Plasmid	Description	Reference
pX33	Gm ^r , <i>sacB</i> , mycobacterial Ts ori, p15A ori, <i>xylE</i>	(Gebhard et al., 2006)
pX33- <i>hucE</i>	pX33 harbouring the MSMEG_2268 deletion construct	This study
pX33- <i>hhyE</i>	pX33 harbouring the MSMEG_2718 deletion construct	This study
pMV261	Kan ^r , mycobacterial oriM, pBR322 ori, P _{<i>hsp60</i>} promoter	(Stover et al., 1991)
pMV <i>hucE</i>	pMV261 harbouring MSMEG_2268 gene insert	This study
pMV <i>hhyE</i>	pMV261 harbouring MSMEG_2718 gene insert	This study

Table S2. Primers used in this study.

Primer	Sequence (5' to 3')	Enzyme
del2268_fwd	aaa <u>ACTAGT</u> GTGTCAGGTTTCGCATCGAC	<u>SpeI</u>
del2268_rvs	aaa <u>ACTAGT</u> TCACAACCGATCACGACCG	<u>SpeI</u>
2267_chrom_fwd	AGACCCTGTGCACGCGATC	
2269_chrom_rvs	CGGCGACCGCGTCGGAC	
del2718_fwd	aaa <u>ACTAGT</u> ATGACCACCACAGCTCCCAA	<u>SpeI</u>
del2718_rvs	aaa <u>ACTAGT</u> TCATGGGATGTCCTCCCGC	<u>SpeI</u>
2717_chrom_fwd	GCAGCGTCCTTCACGGA	
2719_chrom_rvs	GGTAACCCTCGTAGAGCA	
PstI-2268_for	atacat <u>CTGCAG</u> AAATGACCGCGGCGGTTCGAG	<u>PstI</u>
HindIII-2268_rev	atacat <u>AAGCTT</u> CTCATCCGGTGACCCGCAC	<u>HindIII</u>
BamHI-2718_for	atacat <u>GGATCC</u> AATGGCGCCCGCCGTCACG	<u>BamHI</u>

HindIII-2718_rev

atacatAAGCTTCTCACGGTCCGGCCCCACC

HindIII

Table S3 (xlsx). Distribution and sequences of HucE and HhyE proteins in microorganisms encoding the group 2a and 1h [NiFe]-hydrogenases.