Supplementary Information

The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry

Ellis C. O'Neill, Martin Trick, Lionel Hill, Martin Rejzek, Renata G. Dusi, Chris J. Hamilton, Paul V. Zimba, Bernard Henrissat, Robert A. Field*

Email : rob.field@jic.ac.uk

Table of Contents	
Figure S1	2
Table S1	3
Table S2	4
Figure S2	5-8
Figure S3	9
Figure S4	10
Figure S5	11
Figure S6	12
Table S3	13
Figure S7	14
References	15

Figure S1: Aromatic amino acid biosynthesis

The shikimate pathway is present in *Euglena* as the pentafunctional fungal type.¹ There are four isoforms of the DHAP synthase, one of which is only apparent in the dark transcriptome. The pathway branches at chorismate: towards folate biosynthesis, via 4-aminobenzoate formed by a bifunctional enzyme; towards tryptophan biosynthesis, via a tetrafunctional enzyme from 2-aminobenzoate; or to prephenate via chorismate mutase, which is present either as a single domain protein or as part of a trifunctional protein, with dehydratase and dehydrogenase domains. Aromatic amino acid transaminases, for the final step in tyrosine and phenylalanine synthesis, are present in the transcriptome.



Table S1: The enzymes of N-glycan biosynthesis

Transcripts were identified which encode enzymes involved in N-glycan biosynthesis by homology with known enzymes.² Highlighted in yellow are transcripts only present in the dark grown cells and highlighted in green are present only in the light grown cells. FPKM values for each cognate transcript are given in parentheses.

Activity	EC no.		OR	Fs	
dolichyldiphosphatase	3.6.1.43	lm.53772 (10.03)	dm.71782 (2.44)		
dolichol kinase	2.7.1.108	lm.96623 (1.20)			
UDP-GlcNAc-dolichol phosphate GlcNAc-1-P-transferase	2.7.8.15	lm.96341 (2.12)			
beta-1,4-N- acetylglucosaminyltransferase	2.4.1.141	lm.87840 (1.67)			
beta-1,4-mannosyltransferase	2.4.1.142	lm.92144 (1.35)			
alpha-1,3/alpha-1,6- mannosyltransferase	2.4.1.132	lm.79157 (7.10)			
alpha-1,2-mannosyltransferase	2.4.1.131	lm.67740 (8.51)			
Flippase		lm.83408 (4.92)			
alpha-1,3-mannosyltransferase	2.4.1.258	lm.68532 (7.44)			
alpha-1,2-mannosyltransferase	2.4.1.259/ 261	dm.60522 (1.56)	lm.71029 (6.67)	dm.85690 (0.80)	dm.60521 (2.61)
alpha-1,6-mannosyltransferase	2.4.1.260	lm.71029 (6.67)			
dolichol-phosphate mannosyltransferase	2.4.1.83	lm.31276 (16,28)	lm.48352 (14.30)		
dolichyl-phosphate beta- glucosyltransferase	2.4.1.117	lm.48352 (14.30)			
alpha-1,3-glucosyltransferase	2.4.1.267	lm.99883 (0.54)	dm.96080 (1.2)		
alpha-1,3-glucosyltransferase	2.4.1.265	lm.99883 (0.54)	dm.96080 (1.2)		
alpha-1,2-glucosyltransferase	2.4.1.256	lm.100691 (0.72)			
dolichyl- diphosphooligosaccharide protein glycosyltransferase	2.4.99.18	dm.41728 (5.62)	lm.68366 (11.63)	dm.17756 (44.10)	

Table S2: Annotation of enzymes involved in the biosynthesis of GPI anchors

Transcripts were identified which encode enzymes involved in GPI anchor biosynthesis by homology with known enzymes.³ Highlighted in yellow are transcripts only present in the dark grown cells and highlighted in green are present only in the light grown cells. FPKM values for each cognate transcript are given in parentheses.

EC number	Reaction	Mammali an protein	Euglena model	Closest homologue	E-value
F F 1 1	Inacital 2 phasehota		number	Prodicted inspital 2 phasehots synthese 1 like	
5.5.1.4	svnthase 1	INO1	(22.56)	ISaccoglossus kowalevskiil	0.00E+00
3.1.3.25	Myo-inositol-1- phosphatase	IMPA1	lm.49313 (5.60)	Hypothetical protein TRIADDRAFT_52515 [Trichoplax adhaerens]	1.00E-60
2.7.8.11	CDP-diacylglycerol- inositol 3-phosphatidyl transferase	PIS1	lm.37320 (12.27)	Phosphatidylinositol synthase [Spathaspora passalidarum NRRL Y-27907]	3.00E-50
			lm.79366 (4.02)	Predicted protein [Phaeodactylum tricornutum CCAP 1055/1]	1.00E-47
2.4.1.198	GPI-GIcNAc transferase	PIG-A	dm.79942 (4.57)	UDP-GIcNAc:PI a1-6 GIcNAc-transferase [Trypanosoma cruzi marinkellei]	0.00E+00
		PIG-C	None		
		PIG-H	dm.82217 (4.59)	Predicted glycosyltransferase [Ectocarpus siliculosus]	1.00E-16
		PIG-P	dm.82867 (3.40)	Predicted protein [Populus trichocarpa]	4.00E-24
		PIG-Q	dm.48080 (1.57)	Hypothetical protein BATDEDRAFT_85480 [Batrachochytrium dendrobatidis JAM81]	5.00E-36
		PIG-Y	None		
		DPM2	lm.108666 (0.26)	Unknown [Picea sitchensis]	6.00E-09
3.5.1.89	GlcNAc-PI de- <i>N</i> - acetylase	PIG-L	dm.85635 (2.13)	N-Acetyl-D-acetylglucosaminylphosphatidyl inositoldeacetylase [Leishmania major strain Friedlin]	2.00E-52
			dm.27659* (7.68)	Hypothetical protein [Paramecium tetraurelia strain d4-2] - *Also contains N-ter ManT domain	3.00E-51
2.3	Inositol acyltransferase	PIG-W	lm.93529 (2.32)	Predicted protein At4g17910-like [Brachypodium distachyon]	2.00E-64
2.4.1	α-(1-4)- Mannosyltransferase	PIG-M	lm.94033 (1.57)	GPI mannosyltransferase 1 [Dicentrarchus labrax]	2.00E-100
		PIG-X	None		
2.7	EtNP transferase	PIG-N	lm.94615 (2.25)	Predicted GPI ethanolamine phosphate transferase 1-like [Brachypodium distachyon]	3.00E-177
2.4.1	α-(1-6)- Mannosvltransferase II	PIG-V	lm.98056 (1.15)	Dolichol-P-mannose mannosyltransferase ISelaginella moellendorffiil	4.00E-65
2.4.1	α-(1-2)- Mannosyltransferase III	PIG-B	dm.85690 (0.79)	Predicted protein [Physcomitrella patens subsp.	3.00E-86
	GPI transamidase	PIG-K	dm.51731 (10.94)	Predicted protein [Physcomitrella patens subsp.	3.00E-122
		GAA1	lm.88011 (2.90)	Predicted protein [Naegleria gruberi]	4.00E-25
		PIG-S	dm.90214 (1.83)	Unnamed protein product [Vitis vinifera]	8.00E-21
		PIG-T	dm.67348 (3.66)	Hypothetical protein BATDEDRAFT_35820 [Batrachochytrium dendrobatidis JAM81]	8.00E-52
		PIG-U	lm.71943 (4.97)	Gpi transamidase component pig-u [Colletotrichum gloeosporioides Nara qc5]	3.00E-31
	Glycosylphosphatidylino sitol deacylase	PGAP1	lm.73955 (5.40)	GPI inositol-deacylase [Metarhizium anisopliae ARSEF 23]	2.00E-36

Figure S2. Alignments of didomain CAZys (Im.71174 and dm.47703) with well characterised single domain proteins.

Alignments were performed using AlignX (Invitrogen) and confirmed by modelling proteins using SWISS-MODEL in Automated mode.⁴ The key catalytic residues are highlighted in red boxes. **A**. Alignment of Im.71174 with GT11s⁵ (1) and GT15s⁶ (2). **B**. Alignment of dm.47703 with GT1s⁷ (1) and GH78s⁸ (2). FPKM values are 0.90 and 3.61 respectively.

											- Section 1
	(1)	1	,10		20		,30		40		55
light_m.71174	(1)		-IFFRT	AARMA	AWCAPI	WAVLL	LTGLGI	GPR	LRVVI	SHWS	GNIGNNL
E.coli GT11	(1)	MIWUOMD	ECEDMA	YCCLS	GGLGN		AAAYI				YFNQPQ
Vibrio GT11	(1)	MPAA MPA		VMKTS	GGLGN		AVCRA	ATON		LDVS	AYKN
1010 0111	(1)			v 11	00201	X <mark>M</mark> + X+		11 × ×			- Section 2
	(56)	56		,70		80		90		,100	110
light_m.71174	(48)	LQYAVGL	SVA <mark>V</mark> KL	NRSFH	VPHH <mark>E</mark>	LQAA	F <mark>L</mark> L <mark>SS</mark> S	5		SM	APTDVPP
E.coli GT11	(45)	KDTIR	HLE <mark>LDQ</mark>	FKIIF	DRFSS	KDEKVI	KINRLE	R		KH	KK I PL <mark>I</mark> N
Human GT11	(56)	GPSQLRG	MWTINA	IGRLC	NQMGE	YATLY	ALAKM	GRPAI	TIPAQN	1HS <mark>TL</mark>	APIFRIT
Vibrio GT11	(41)	YKLHN	G <mark>M</mark> RLDQ	FNINA	DIANE	DEIFH	LK <mark>GSSI</mark>	N		R	SRILRRL Continu 2
	(111)	111	120		130		140		150		- Section 3
light m 71174	(91)	ROTAGL	PAGADP	LAPGH		RNPSPI	NIPRTI	CDECE	VIE		KSTPPTT
E.coli GT11	(86)	SELQETA	IKLCNK	YSLND	ASYYN	PESI-	ND	-VACLE	SFYOI	SKLL	NEHRDLI
Human GT11	(111)	LPVLHSA	TASRIP	WQN <mark>Y</mark> H	LNDWM	EEEY-	RHIPGI	EYVRF	GY PCS	WTFY	HH <mark>LR</mark> QE <mark>I</mark>
Vibrio GT11	(82)	G <mark>WL</mark> KKNT	YY <mark>a</mark> ekq	RTIYD	VSVFM	QA		- PRY <mark>I</mark> I	GY <mark>W</mark> Q1	EQ <mark>YE</mark>	SQI <mark>R</mark> A <mark>VL</mark>
		00000		22020		2.5.5		Loton (0200025	- Section 4
	(166)	166		180		190		200	•	210	220
light_m.71174	(146)	LDIFFID	TPMLHY	CHNIC	QR RG.	AFGPA	ATEVA	HARL	DRALT	PN	WLAPT
E.COIL GT11	(158)	LPLEELR LOFFTIH		AKEL		NGSPP	TEVG			MPKU	WKGVUAD
Vibrio GT11	(103) (128)	LOELWP-	NOPL	SINAO	AHOLK	IOOTH	AVS	HARP	D D N-	H	PEIGVLD
	(~ _						U		- Section 5
	(221)	221	230		240		250		260		275
light_m.71174	(197)	AEYVVAA	VEHVRR	LAGNA	V <mark>F</mark> ILF	DCAV	CS <mark>R</mark> IM(2I <mark>R</mark> RS-	TPHV	WHLH	S <mark>GT</mark> A <mark>M</mark> QA
E.coli GT11	(191)	MDYYISA	MEYIES	ECGSQ	TEIIF	DDVI	WAKEKI	SKYSN		ADEN	KFS-VID
Human GT11	(220)	RRYLQQA	LDWERA	RYSSL VTEND		SNGMAI	WCREN.	NETD	SDVEF	AGDGI	OTE-TDD
VIDRIO GTII	(1/4)	L L L L KK	V DITE	TEMP		DVA		NEID-	-SP V I	TERI	- Section 6
	(276)	276		290		300		310		320	330
light m.71174	(251)	LALMVHC	THFVLT	GG <mark>SF</mark> G	WWAAM	LGETA	ASTVIT	TCGEYT	DAR	MAIR	IHKMVPA
E.coli GT11	(245)	MYLMSLC	NNNIIA	NSTYS	WWGAW	LN]	RSEDKI	VIAPE	QWYIS	GNE-	CSLKNEN
Human GT11	(275)	FALLTQC	NHTIMT	IG <mark>IF</mark> G	I <mark>WA</mark> AY	LTGGD	IYLAN	NYTLPI	SPFLF	IFKP	EAAF <mark>LP</mark> E
Vibrio GT11	(226)	IMIMC <mark>QC</mark>	Q <mark>HNI</mark> VA	NSSFS	W <mark>W</mark> AAW	LN	SNVDK	VIAPE	(TWMAE	PKG	YKW <mark>V</mark> PD <mark>S</mark>
		004	210						170		- Section 7
1	(331)	331	340		350	VEADD	360		,370		385
E coli CT11	(306)	TWKWFHR	CHAKAS	PSERQ	AITAH	TSRD	GGSDQS	SPAVNI	SISPI	ERLL	GAHYPPK
Human GT11	(290) (330)	TGTAAD	LSPLLK	H							
Vibrio GT11	(278)	WREI									
		-									- Section 8
	(386)	386		400		410		420		430	440
light_m.71174	(361)	LKSECNF	AGWSVG	GRSPY	CECRP	GWTAE	TCTNRM	NPLPGE	RPRGVI	AYLL	YGAAHYA
E.coli GT11	(300)										
Human GT11	(344)										
VIDRIO GI 11	(202)										

A1

	(261)	261	270	280		290	,300	312
light_m.71174	(261)	VLTGGSE	GWWAAYL	GETAAST	VITCGEY	FTDARRN	AIRIHKM <mark>V</mark>	PATWK <mark>v</mark> fhr
Penecillium GT15	(1)					<mark>MS</mark> F	IQ <mark>RITK</mark> RL	PSAPSLPLE
Magnoporthe GT15	(1)					MAR	PVRVLAT	ALFMWCVFL
Saccaromyces G115	(1)					MADFL	KKULKET	Section 7
	(212)	313	320	330	34	0	350	364
light m 71174	(313)	CAAKASE	520 SE			GSDOSP	NYSYSP	VERTINAHY
Penecillium GT15	(21)	DAPNEKG	RLHPREA	FFERRIRI	LKGNSSI	SIPLGLV	LFPCLVI	VFILLEVR
Magnoporthe GT15	(21)	YMIFRPS	<mark>S</mark>	PLLVA	D <mark>ees</mark> nfqi	RD <mark>PM</mark> HDP'	IG <mark>E</mark>	
Saccaromyces GT15	(23)	LLTLNSN	SRTQQMI	PSS <mark>IS</mark> A <mark>A</mark> H	F <mark>dfts</mark> gs:	IS <mark>P</mark> E <mark>Q</mark> Q <mark>V</mark>	SEENDAK	KLEQS <mark>A</mark> LNS
					1.1.1.1			Section 8
	(365)	365 370		,380	,390		400	416
light_m./11/4	(358)	PPKLKSL	CNFAGWS	DESTRUCT	CRPGW	FAETCIN.	KNPLPG	RPRGVIAYL
Magnoporthe GT15	(51)	nr Sree		FFSIRRIA		TLRRV <mark>SP</mark>	RYAPDANP	TERTNATIT
Saccaromyces GT15	(75)	EASEDSE	AMDEESK	ALKAAAEI	KADAPID	KTTMDY	ITPSFANK	AGKPKACYV
	()							- Section 9
	(417)	417	43	30	440	450		468
light m.71174	(408)	LYGAAHY	ALQLAEV	LPILDRY	NDRYHYI	ILIFHS	NMAKETM	PCTGKSYLD
Penecillium GT15	(125)	VLARNKE	LDG <mark>VI</mark> ES	LKSIERH	N R WWHYI	YVFLND	GD <mark>F</mark> DE <mark>EFK</mark>	AT <mark>V</mark> KNY <mark>TSA</mark>
Magnoporthe GT15	(78)	ALVRNEE	LDG <mark>MLQ</mark> A	MGD <mark>ler</mark> T	N S K F N Y I	PWTFFND	VPF <mark>SK</mark> EFK	QKTQAM <mark>T</mark> K <mark>A</mark>
Saccaromyces GT15	(127)	TLVRNKE	LKGLLSS	IKY <mark>ve</mark> nki	I <mark>nkkf</mark> p <mark>y</mark> j	PWVF <mark>L</mark> ND	PF <mark>T</mark> E <mark>E</mark> FK	EAVTKAVSS
		100	100		00	500	510	- Section 10
links 71174	(469)	469	480	4	1110 KMD	500	510	
Penecillium GT15	(400) (177)	EVERGET	DNTMMGE	PDWWDHE	AKEGTE	SWQRCI	AATMYCC	
Magnoporthe GT15	(130)	KCNYEII	PKEHWDM	PSWINKD	YDESVK	ILK	NKIQYAD	KINYHOMCR
Saccaromyces GT15	(179)	EVKFGII	PKEHWSY	PEWIN-Q1	TK <mark>A</mark> AEIRA	A D <mark>A</mark>	ATKY <mark>IYGG</mark>	SEEYRHMCR
								- Section 11
	(521)	521	530	,540		550	,560	572
light_m.71174	(512)	FTYMF	LRPELQP	YE YV <mark>WR</mark> MI	NLGLTH	ra <mark>v</mark> pc <mark>dv</mark>	QVMRRTN.	A <mark>VEGE</mark> YA <mark>A</mark> Q
Penecillium GT15	(225)	FYSGHEY	KHPLLMK	YEMYWRI H		DITYDP.	LKMALAN	KTYGFTIA-
Saccaromyces GT15	(226)	VOSCEEW	RHELLEE	VOLIWEVE			KENQDNN KWMOENE	
	(220)							-Section 12
	(573)	573	580	590	60	0	610	624
light_m.71174	(564)	QNEAAQC	TGPIGGE	AAEYAKT-				– <mark>FR</mark> FQPNHL
Penecillium GT15	(276)	VKELRET	VPNIFRY	<mark>a a</mark> a <mark>y k</mark> r k i	NLKSKGI	LWEMFLE	PAQPET <mark>P</mark>	EENK <mark>Q</mark> DKLP
Magnoporthe GT15	(229)	LYDAPES	IPTLWPE	TEKFLAEI			<mark>P</mark>	Q <mark>Y K</mark> H P N N A L
Saccaromyces GT15	(277)	THEYEV	IPT <mark>LW</mark> QT	SMUTIKKI			<mark>P</mark>	EYLDENNLM Section 13
	(())	625 630	5	640	650		0.00	- Section 13
light m 71174	(593)	DRIGPGR				ANCAM		RSKT MOHMA
Penecillium GT15	(328)	DEILOTE	PGDNNLK	DVDPEAME	GESYNM	CHEWSNEI	IARLDWE	RSKEYEDFF
Magnoporthe GT15	(261)	DWLTDKE	KRPE	HNRE	K <mark>ANGY</mark> ST (CHEWSNEI	EVADMNFW.	RSKTYEDYF
Saccaromyces GT15	(309)	SFLSNDN			GK <mark>T</mark> YNL(CHEWSNEI	E <mark>IANLN</mark> LW	<mark>rs</mark> pa <mark>y</mark> r <mark>eyf</mark>
								- Section 14
	(677)	677	,69	90	,700	,710		728
light_m.71174	(625)	QAVRQSG	MVYSHRL	GEQTMLLA	GISLLV	PPEAIHQ	GGLSPLF.	HNRKQLWRY
Magnoporthe GT15	(306)	NHLDRAG	CFEVEDW	CDIPVHSI	LA LOL FEI	DASKTHWI	RDIGICH	T PEFNCPNS
Saccaromyces GT15	(346)	DTLDHOG	GFFYERW	GDAPVHSI	TAAALFL	KDKIHY	SDIGYHH	PPYDNCPLD
	(0.0)							-Section 15
	(729)	729	,740	7	750	760	,770	780
light_m.71174	(677)	DQPEMFI	RRDDPVC	RAVAADN-	P <mark>W</mark>	DPPLPY	FVPQPM <mark>P</mark>	N <mark>Q</mark> KHQLN <mark>E</mark> G
Penecillium GT15	(432)	APARQLA	RIPYLEM	TTDDE <mark>k</mark> ki	RIEEDE	WANP <mark>DP</mark> V	E <mark>ngvg</mark> cr	CRCDTDIVD
Magnoporthe GT15	(358)	FKCKGC	AGRETOG	EKWLHR	El	CRENWE	MHGMG	VNWVVED
Saccaromyces G115	(248)	REVINSI	INC CDQC	ND <mark>e</mark> lfQG-		GVUIT		Section 16
	(721)	781	790	80	2			- 360001 10
light m.71174	(724)				-			
Penecillium GT15	(484)	VEGKQGS	CLNEWVD	VAGGWASH	P			
Magnapartha CT1E	(202)							



								Section 1
	(1)	1	,10	20	-	30	40	52
dark_m.47703	(1)	- DLFFFA	I <mark>AA</mark> R <mark>PM</mark>	RALILAY	LP <mark>PD/E</mark> B	PY <mark>VAL</mark> CA	ALAAAGH <mark>W</mark> C	TVCTHALLR
Arabidopsis GT1	(1)	MEESKTP	HV <mark>AI</mark> IP	SPG <mark>M</mark> GHLI	IPLVEFA	RLVHLH	GLTVTFVIA	GEGPPSKAQ
Amycolatopsis GT1	(1)		M	RVLLS <mark>V</mark> C	TRCDVE	IGVALAD	RLKALGVQI	RMCAPPAAE
Human GT1	(1)	-MARAGW	TSPVPL	CVCLLLT	CGFFEAG	LVVPM	DGSHWFTMQ	SVVEKLILR
	(53)	53	60	70	,8	30	,90	104
dark_m.47703	(52)	PLVEEHG	LTFMEE	DGEDPR		DL	HERCEREGE	FIFSFVQSI
Arabidopsis GT1	(53)	RTVLDSL	PSSISS	OF T DE OT	DT	-SSTR	DSRISLIM1	RONPELEK
Amycolatopsis GT1	(41)	CHENUN		OTERCIN	UVTVO	C-V-TT	PONREQUE	MPPPPPELL
Human GT1	(52)	GHE		Q T P KS H M	VEIIS	3-1-1	DD VINKEP N	Section 3
	(105)	105 110)	120	130		140	156
dark m.47703	(91)	GSVAERW	MDGWAK	VKAAVG	DEDMVVS	T		AVGNAGI
Arabidopsis GT1	(101)	FDSFVEG	GRLPTA	LVVDLFG	DAFDVA	FHVPP		YIFMPTT
Amycolatopsis GT1	(74)	QRLAAMT	VEMQFD.	AVPGAAE	GCAAVVA	G		D <mark>LAAAT</mark> G
Human GT1	(102)	AQSIFSL	LMSS <mark>SS</mark>	GF <mark>LDLF</mark> F	HCRSLFI	D RK L VE	YLKESSFDA	VFLDPFDTC
				-			et ::	
	(157)	157		170	,180	,19	90	208
dark_m.47703	(127)	HLAEASG	PVCLA	HLFAGE		MHFFHP	MAPAPLERS	TARPSFVNL
Arabidopsis GT1	(142)	ANVLSFE	LHLPKL		FREL	EPLMLP	GCVPVAGKI	FLDPAQDRK
Amycolatopsis GT1	(110)	VRSVAEK	LGLPEP	SVPS		VALASE NOCEADI	ALPPAYDEP	FORMER
Human GT1	(154)		ra <mark>ur</mark> a <mark>v</mark>	VERGIE		ACCE <mark>M</mark> ET	SI <mark>VE</mark> NDLLG	Section 5
	(200)	209	220		230	240	25	260
dark m.47703	(172)	FSHHMVE	YLYGIL.	ALPKANAI	RLTLGL	PIRCEH	WLAGFRAKN	DAPRIYGHS
Arabidopsis GT1	(191)	DDAYKWL	LHNTKR	KEAEGII	LVNTFFEI	LEPNAIK	ALQEPGLDE	P
Amycolatopsis GT1	(153)	VLWEERA	ARFADR	YGPTLNR	RAEIGLE	PVEDVF	GY <mark>CH</mark> GERP-	LL
Human GT1	(206)	R <mark>VW</mark> NHIV	HLEDHL	CQY <mark>L</mark> -F	NALEIAS	SEILQTP	VT <mark>AY</mark> DLYSH	IT <mark>S</mark> IW <mark>L</mark> LRTD
		7.P200	0.000	- 10	20	12300	000.000	—— Section 6
	(261)	261	270	28	0	290	,300	—— Section 6 312
dark_m.47703	(261) (224)	261 P <mark>AV</mark> LPAP	270 Pe <mark>wp</mark> rt	28 WQ <mark>V</mark> A <mark>G</mark> YW	0 Vlpev <mark>p</mark> y:	290 PPKALA	,300 <mark>AFLAAG</mark> P <mark>A</mark> F	Section 6 312 VYIHLATEK
dark_m.47703 Arabidopsis GT1	(261) (224) (235)	261 PAVLPAP	270 PEWPRT PVYPVG	28 WQ <mark>VAG</mark> YW PL <mark>VNIG</mark> K(0 VLPEV <mark>P</mark> Y1 2EAKQTE	290 PPKALA SSECLKW	,300 AFLAAGPAF LDNQPLG <mark>S</mark> V	Section 6 312 VYIHLATEK VYISEGSG-
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Humap GT1	(261) (224) (235) (197) (257)	261 PAVLPAP AADPVLA	270 PEWPRT PVYPVG PLQPDV	28 WQVAGYW PLVNIGK DAVCTGA	0 VLPEV <mark>P</mark> Y1 2EAKQTER VLLSDER	290 FPPKALA ESECLKW	,300 AFLAAGPAF LDNQPLG <mark>S</mark> V AFLAAGSPF	Section 6 312 VYIHLATEK LYVSEGSG- VHIGEGSSS
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257)	261 PAVLPAP AADPVLA FV <mark>L</mark> DYPK	270 PEWPRT PVYPVG PLQPDV PVMPNM	28 WQ <mark>VAG</mark> YW PLVNIGK DAVGTGA IF <mark>I</mark> G <mark>G</mark> IN(0 VLPEV <mark>P</mark> YT 2EAKQTEE VLLSDER CHQGK <mark>P</mark> L	290 PPKALA SSECLKW PLPPLE M	,300 AFLAAGPA LDNQPLG <mark>S</mark> V AFLAAGSPE	Section 6 312 VYIHLATEK VSFGSG- VHIGFGSSS Section 7
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (313)	261 PAVLPAP AADPVLA FV <mark>S</mark> DYPK 313	270 PEWPRT' PUOPDV PUMPNM 320	28 WQVAGYW PLVNIGK DAVCTGA IF <mark>I</mark> GGIN 330	0 VLPEV <mark>P</mark> YT 2EAKQTER VLLSDER CHQGK <mark>P</mark> L	290 PPKALA SECLKW PPELE M	,300 AFLAAGPA LDNQPLGSV AFLAAGSPE 350	Section 6 312 VYIHIATEK VHIGEGSS SS Section 7 364
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703	(261) (224) (235) (197) (257) (313) (276)	261 PAVLPAP ADPVLA FVLDYPK 313 CATPDTL	270 PEWPRT PVYPVG PLQPDV PVMPNM 320	28 WQVAGYW PLVNIGK DAVQTGA IFIGCIN 330 CCRLAGO	0 VLPEVEYT 2EAKQTER CHQGKEL 2HQGKEL 2AVVYAPH	290 PPKALA SECLKW PPELE M	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350	Section 6 312 VVI HL AT EK V S FGS G - V H I G FGS SS Section 7 364
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1	(261) (224) (235) (197) (257) (313) (276) (279)	261 PAVLPAP AADPVLA FVEDYPK 313 GATPDTL GTLTCEQ	270 EEWPRT EWPVG EQPDV EVMPNM 320 LQAAVE LNELAL	28 WQVAGYW PLVNICK DAVOTCA IFIGCIN 330 CCRLACO CRLACSEQ	0 NLPEVEYT Deakqter Lisder Chqgkel Ravyyaph Revyyaph	290 PPKALA SSECLKW PPELE M 340 HGA SPSGIAN	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSQ	Section 6 312 VVI HL AT EK V SEGS G- VHIGESS Section 7 364
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1	(261) (224) (235) (197) (257) (313) (276) (279) (249)	261 PAVLPAP AADPVLA FVEDVPK 313 GATPDTL GTLTCEQ GRGIADA	270 EENPRT ENPRYS EQPDY EVMPNM 320 LQAAVE LNELAL AKVAVE	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGGI GLADSEQ ALRAQGR	0 VLPEVEYT 2EAKQTER LSDER CHQGKEL 3 RAVVYAPH FLWVIRS VILSRGV	290 FPEKALA SECLKW FPELE M SPG GA SPG GIAN VTE	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC	Section 6 312 VVI HL AT EK SEGS G- VHIGESS Section 7 364 TDPLTFLPP
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (313) (276) (279) (249) (286)	261 PAVLPAP AADPVLA FVEDVPK 313 GATPDTL GTLTCEQ GRGIADA	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE LNELAL AKVAVE	28 WQVAGYW DLVNIGK DAVGTGA IFIGCIN 330 GCRLAGCI GLADSEQ ALRAQGR	0 VLPEVEYT 2EAKQTER LSDER CHQGKEL 3 RAVVYAPH FLWVIRS VLLSRGV	290 FPEKALA SECLKW DPELE M 340 HGA SPSGIAN VTE	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC	Section 6 312 VVI HL AT EK SEGS G- VHIGESS Section 7 364
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (313) (276) (279) (249) (286)	261 PAVLPAP AADPVLA FVUDPVLA GATPDTL GTLTCEQ GRGIADA	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE LNELAL AKVAVE	28 WQVAGYW PLVNIGK DAVGTGA IFIGCIN 330 GCRLAGCI GLADSEQ AIRAQGR	0 VIPEVEYT 2EAKQTER Ilsder ChQGKEL 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	290 FPEKALA SECLKW D PELE M 340 HGA SPSGIAN VTE	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC	Section 6 312 VVI HLATEK CVISEGSG- VHIGEGSS Section 7 364 TDPLTFLPP Section 8
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (313) (276) (279) (249) (286) (365)	261 PAMLPAP AADPVLA FVEDVPK 313 GATPDTL GTLTCEQ GRGIADA 365 370	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE LNELAL AKVAVE	28 WQVACYW PLVNIGK DAVGTGA IFIGCIN 330 GCRIACOU GLADSEQ AIRAQGR 380	0 VLPEVEYT 2EAKQTER LSDER CHQGKEL 3 AVVYAPH FLWVIRS VLSRGV 390	290 FPEKALA ESECLKW CPELE M 340 IGA SPSGIAN VTE	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC	Section 6 312 VVI HL AT EK V S FGS G- VHIGESS Section 7 364 2 TDPLTFLPP Section 8 416
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703	(261) (224) (235) (197) (257) (313) (276) (279) (249) (286) (365) (306) (306)	261 PAM LPAP AA DPVLA FVD DYPK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE LNELAL AKVAVE	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN(330 GCRLAGGI GLADSEQ AIRAQGR AIRAQGR 380 RIE <mark>E</mark> VVP	0 VIPEVEYT 2EAKQTER Isder ChQGKEI 3 RAVVYAPH FIWVISS VIISRGV 390 HYWIVES	290 FPEKALA SECLKW DPELE M O SPSGIAN VTE CKVVVHH	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC 400 AGMLLAAYT	Section 6 312 VVI HL AT EK V SEGS G- VH I G EGS SS Section 7 364 TDPLTFLPP Section 8 416 V MAGVE G V
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1	(261) (224) (235) (197) (257) (313) (276) (279) (249) (286) (365) (306) (331) (331)	261 PAMLPAP AADPVLA FVDDVLA GATPDTL GATPDTL GTLTCEQ GRGIADA 365 370 VSMS GFLERTK	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE STLDVI KRGFVI	28 WQVAGYW PLVNIGK DAVCTGA IFIGCIN 330 GCRLAGO GLAISEO AIRAQGR 380 RIEEVVPH PFWAPQA	0 VIPEVEYT 2EAKQTER LSDER CHQGKEL 3 AVWYAPH FIWVIRS VIISRGV 390 HYWLVPE 2VLAHPS	290 FPKALA SECLKW DPFLE M SPSGIAN VTE CKVVVHH GGFLTH	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC 	Section 6 312 VYI HL AT EK Y S F G S G - Y H I G F G S G - Y H I G F G S S S
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (286) (306) (331) (279) (286)	261 PAM LPAP A DPVLA FVD DYPK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE STLDVI KRGFVI DDRDDC	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGOI AIRAQGR AIRAQGR 380 RIEEVVPH PFWAPQA FAI E VNI	0 VIPEVEYT 2EAKQTER LSDER CHQGKEI 3 RAVWYAPH FIWVIRS VIISRGV 4 VIISRGV 2 VIASRGV 2 VIASRGV 2 VIASRGV	290 FPEKALA SECLKW DPELE M OPELE SP GA SP GA SP GA SP GA SP GA SP GA SP GA SP C SP GA SP SP SP SP SP SP SP SP SP SP	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSC 400 AC0 AC0 SSYFDSHSC GNNSTLES GSAGE EHVP	Section 6 312 VYI HL AT EK Y S FGS G- VH I G FGS SS Section 7 Section 7 364 TDPLTFLPP Section 8 416 VKAGVE GVH VS G E LLA TRAGVE QLV
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (286) (306) (331) (279) (286)	261 PAMLPAP AADPVLA FVDDVLA GATPDTL GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE STLDVI KRGFVI DDRDDC	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGOI GLAISEOI AIRAQGR 380 RIEEVVPH PFWAPQA FAI E VNI	0 VIPEVPYT 2EAKQTER LSDER CHQGKEI 3 AVWYAPH FIWVIRS VIISRGV 4 VIISRGV 390 HYWIVPE 2 VLAHPS FQAIFREN	290 FPKALA SECLKW DPFLE M SPGIAN VTE CKVVVHH GGFLTH AAVIHH	,300 AFLAAG PAF LDNQPLGSV AFLAAG SPF ,350 SSYFDSHSC 	Section 6 312 VVI HL AT EK C VV SEGS G- VH I G EGS SS Section 7 364 TDPLTFLPP Section 8 416 VKAGVE G VH V SG ELLA TRAGVE QUV Section 9
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (331) (279) (286) (331) (279) (286) (311)	261 PAM LPAP AA DPVLA FVD DYPK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE STLDVI KRGFVI DDRDDC	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGGIN GLAISEQ AIRAQGR 380 RIEEVVPH PFWAPQA FAITEVNH 	0 VIPEVEYT 2EAKQTER LSDER CHQGKEI 3 RAVWYAPH FIWVIRS VIISRGV 2VIAPE 2VIAPE 2VIAPE 2VIAPE 440	290 FPKALA SECLKW DPELE M SPGIAN VTE CKVVVHH TGGIAN VTE CKVVVHH TGGIAN 44	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSC 400 ACO ACO ACO MIGTLES GSAG2 EHVP 50	Section 6 312 VVI HL AT EK C VW SEGS G- VH I G EGS SS Section 7 364 TDPLTFLPP Section 8 416 VM AGVE GVH V SG PLLA TRAGVE QUV Section 9 468
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703	(261) (224) (235) (197) (257) (257) (279) (249) (286) (279) (286) (306) (331) (279) (286) (331) (279) (286) (335) (355)	261 PAVLPAP ARDPVLA FVDDVLA GATPDTL GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE 	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGGIN GLADSEQ AIRAQGR 380 RIEEVVPH PFWAPQA FAILEVNI 430 LYDLGCS	0 VIPEVPYT 2EAKQTER ILSDER CHQGKEI 3 AVWYAPH FIWVIRS VIISRGV 4 VIISRGV 390 HYWIVPE 2 VLAHPS FQAIFRE 440 PP	290 IPEKALA SECLKW IEEEE M SPSGIAN VTE CKVVVHH IGGIIN VTE CKVVVHH AVIHH 44 44 	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSC 400 AGMLAAYT GSNGTLES SAGZEHVA 	Section 6 312 VYI HL AT EK T YV SEGS G- VH I G EGS SS Section 7 364 TDPLTFLPP Section 8 416 VMAGVE GVH VVS C ELLA TRAGVE QUV Section 9 468 417 QS P
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (306) (331) (279) (286) (305) (306) (331) (279) (286) (355) (383)	261 PAVLPAP ARDPVLA FVEDVPK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PIQPDV PMPNM 320 LQAAVE AKVAVE 	28 WQVAGYW PLVNICK DAVCTCA IFIGCIN 330 GCRLAGCI AIRAQCR AIRAQCR 380 RIECVVPH PFWAPQA FAICEVN FAICEV 430 LYDLGCSH LSEDIRA	0 VIPEVPYT 2EAKQTEE ILSDER CHQGKEI REUVTRS VIISRGV VIISRGV VIISRGV VIISRGV VIISRGV VIISRGV VIISRGV PP	290 IPPKALA SECLKW IPPKLE SPGIAN VTE CKVVVHH CGGIITH AAVIHH 44 DGLVR	300 AFLAAGPAF LDNQPLGSV AFLAAGSPF 350 SSYFDSHSC 400 AGAMLAAYT CGWNSTLE GSAGTEHVP 50 AATAAALHH EFARVYKG	Section 6 312 VYIHLATEK VHIGFGSSS Section 7 364 TDPLTFLPP Section 8 416 VKACVPGVH VSGPLIA TDAGVPQLV Section 9 468 ALYTATQP
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (306) (331) (279) (286) (306) (331) (279) (286) (355) (383) (325) (383) (328)	261 PAVLPAP ARDPVLA FVDDVLA FVDDVLA GATPDTL GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PIQPDV PVMPNM 320 LQAAVE AKVAVE 	28 WQVAGYW PLVNIGK DAVGTGA IFIGGIN 330 GCRLAGO GLAISEO AIRAQGR AIRAQGR 380 RIEEVWPH PFWAPQA FAILEVNI 430 LYDLGCSH LSEDIRA VAALGIGV	0 VIPEVPYT 2EAKQTEE ILSDER CHQGKEI RAVWYAPH FIWVIRS VIISRGV UVPE QVLAHPS FQAIFREV ALRPRAG	290 TPEKALA SECLKW LEPELE M 440 HGA SPSGIAN VTE CKVVVVHH TGFLTH AVIHH A4 CKVVVRH AVIHH A5 CKVVVVHH CKVVVRH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH A5 CKVVVVHH 	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSC 400 AGAMLAAYI CGNNSTLE GSNGTEHVP 50 AATAAALHH EFARVYKG SSAALTT	Section 6 312 YI HLATEK YES FGS G- YH G FGS SS Section 7 364 TDPLTFLPP Section 8 416 VEACVE GVH VSG ELLA TRACVE QUV Section 9 468 NEGEEGKG YLAPETRAR
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (306) (331) (279) (286) (365) (306) (331) (279) (286) (417) (355) (383) (328) (286)	261 PAVLPAP ADPVLA FVDPVLA GATPDTL GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PMPNM 320 LQAAVE AKVAVE 	28 WQVACYW PLVNICK DAVCTCA IFICCIN 330 GCRLACCI AIRAQCR AIRAQCR 380 RIECVVPH PFWAPQA FAICCI 430 LYDLCCS LSEDIRA VAALCIG	0 VIPEVPYT 2EAKQTEE ILSDER CHQGKEI 3 RAVWYAPE FIWVIES VIISRGV 4 VVIISRGV 2VLAHPST FQAIFREV 440 PP	290 TPEKALA SECLKW LPPELE M 440 HGA SPSGIAN VTE CKVVVVHH TGFLTH AVIHH AVIHH DGLVRE GETPTE	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSC 400 AGAMLAAYI CGNNSTLE SAGTEHVP 50 AATAAALHH EFARVYKG SSAALTT	Section 6 312 YI HL AT EK Y I S PGS G- YH I G FGS SS Section 7 364 TDPLTFLPP Section 8 416 VEACVP GVH VSG PLIA TRACVP QV Section 9 468 NA YTA QP LAPETRAR Section 10
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (306) (331) (279) (286) (305) (306) (331) (279) (286) (417) (355) (383) (328) (286)	261 PAM LPAP A DPVLA FVD DVVK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVYPVG PLQPDV PVMPNM 320 LQAAVE LNELAL AKVAVE STLDVI KRGFVI DDRDDC HLWASR KMNAVL PYEAGR	28 WQVACYW PLVNIGK DAVGTGA IFIGIN(330 GCRLAGE AIRAQGR AIRAQGR 380 RIEEVVPH PFWAPQA(FAIEVNH 	0 VIPEVEYT 2EAKQTEE ILSDER CHQGKEI 3 RAVWYAPE FIWVIRS VIISRGV 4 VVIISRGV 2VLAHEST FQAIFREV 440 PP ALRPRAG VAH	290 TPEKALA SECLKW LPPELE M SPGIAN VTE CKVVVVHH TGEFLTH AVIHH AVIHH CKVVVVH CKVVVVHH CKVVVV CKVVVVH CKVVVVH CKVVVVH CKVVVVH CKVVVVH CKVVVVH CKVVVVH CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVVVH CKVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVVV CKVVVVVV CKVVVVVV CKVVVVV CKVVVVVVV CKVVVVVVV CKVVVVVV CKVVVVVV CKVVVVVVVVVV	,300 AFLAAGPAF LDNQPLGSV AFLAAGSPF ,350 SSYFDSHSC SSYFDSHSC GSSGTEHVP GSAGAMLAAYT CGNNSTLE SAGTEHVP 50 AATAAALHH EEMARVYKG ESISAALTT	Section 6 312 YI B FGS G- YH C FGS SS Section 7 364 TDPLTFLPP Section 8 416 VKACVP GVH VS C PLIA TRACVP QV Section 9 468 KA YTA QF KAGVE GKG VLAPET RAR Section 10
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (279) (249) (286) (306) (306) (331) (279) (286) (305) (306) (331) (279) (286) (417) (355) (383) (328) (286) (469) (400)	261 PAVLPAP A DPVLA FV DPVLA FV DPVLA GATPDTL GTLTCEQ GRGIADA 	270 PEMPRT PMPVG PVMPVG STLQPDV STLDVI KRGFVI DDRDDC HLWASR KMNAVL PYEAGR 480		0	290 PPEKALA SECLKW PPELE M SPSGIAN VTE CKVVVVHH CGFLTH AVIH CKVVVVHH CGFLTH AVIH CKVVVVH CKVVVVH CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVVV CKVVVV CKVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVVV CKVVVVVV CKVVVVVV CKVVVVVVV CKVVVVVVVVVV	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSQ 400 AGAMLAAYI CGNNSTLES SNGTEHVP 50 AATAAALHH EEMARVVKG ESSAALTT 510	Section 6 312 YI I S PGS G- YI I S PGS G- YI I S PGS SS Section 7 364 TDPLTFLPP Section 8 416 VEACVE GVH VEGEELIA Section 9 468 VI AGVE QUV Section 9 468 VI AGVE C SC VI AP DT R R Section 10 5200
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (249) (249) (286) (306) (331) (279) (286) (331) (279) (286) (417) (355) (383) (328) (286) (459) (409) (435)	261 PAVLPAP A DPVLA FV DPVL 313 GATPDTL GTLTCEQ GRGIADA 	270 PEMPRT PMPVG PMPVM 320 LOAAVE LNELAL AKVAVE STLDVI KRGFVI DDRDDC HLMASR KNNAVL PYEAGR 480 ALSERV		0	290 PPEKALA SECLKW PPELE M SPGIAN VTE CKVVVHH GGFLTH AVIH CKVVVHH AVIH CKVVVHH CKVVVVHH CKVVVHH CKVVV	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSQ CGNNSTLES SAGEHVP 50 AFTAAALHH EEVARVWKG ESSAALTT 510 LPREDASLW	Section 6 312 YI HL ATEK S FGS G- YH G FGS SS Section 7 364 TDPLTFLPP Section 8 416 VEACVEGH VGG PLIA Section 9 468 ATERGVEQUV Section 9 468 ATERGVEQUV Section 10 0 200 VGADLWTRL GNH
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1	(261) (224) (235) (197) (257) (257) (249) (249) (286) (331) (279) (286) (331) (279) (286) (331) (279) (383) (328) (286) (417) (355) (328) (286) (417) (355) (328) (286) (459) (400) (435) (336)	261 PAVLPAP A DPVLA FV DPVLA FV DPVLA GATPDTL GTLTCEQ GRGIADA 	270 PEMPRT PMPVG PQPU PMPNM 320 LOAAVE LNELAL AKVAVE STLDVI KRGFVI DDRDDC HLWASR KMNAVL PYEAGR 480 ALSERV		0 V PEVEY Z AKQTEH LSDER LSDER S AVYYAPH FLWVIRS VI SRGV 390 HYWLVPE QVLAHPS FQALFRE 440 PP ALRPRAG 7AH 490 AVKSLEA 5 TKALEA AAVGEE	290 PPEKALA SECLKW PPEEE M 440 4GA SPSGIAN VTE CKVVVHH FGGFITH AVIH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVV CKVVV CKVVV CKVVV CKVVVH CKVVVV CKVVVV CKVVV CKVVV CKVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVVVVVVV	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSQ SSYFDSHSQ CGNNSTLES SAG EHVA SAG EHVA SAG EHVA SAG ENVKG SSG ENVA SAG ENV	Section 6 312 Y I H L AT EK Y I S P G S G Y I G F G S S S Section 7 364 PTD PL T F L P P Section 8 416 V I G P L I A Y I G P L I A Section 9 468 AT R AG V P Q I V Section 9 468 AT R AG V P Q I V Section 10 0 520 V T AL V T AL Section 10 520 V T AL V T AL Section 10 520 Section 10 Section 10
dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Amycolatopsis GT1 Human GT1 dark_m.47703 Arabidopsis GT1 Human GT1	(261) (224) (235) (197) (257) (257) (249) (249) (286) (306) (331) (279) (286) (331) (279) (286) (417) (355) (383) (328) (286) (469) (400) (435) (376) (286)	261 PAVLPAP AADPVLA FVEDVPK 313 GATPDTL GTLTCEQ GRGIADA 	270 PEWPRT PVMPVG PVMPNM 320 LOAAVE LNELAL AKVAVE STLDVI KRGFVI DDRDDC HLWASR KMNAVL PYEAGR 480 ALSERV LKEARCI	28 WQVAEYW PLVNICK DAVGICA IFIGIN GLADSEQ AIRAGE AIRAGE ,380 RIEEVVP FAILEVNP FAILEVNP 430 LYDIGCE AAEDGLQA RVLKDGC AAADLVI	0 VIPEVEYT 2EAKQTEH CHQGKEL 3 RAVYYAPH FIWVIRS VIISROV 390 HYWIVPE 2VLAHPS1 FQALFREV 440 PP	290 PPKALA SECLKW PPFLE M 440 4GA SPSGIAN VTE 45 CKVVVHH FGFLTH AVIHH 45 CKVVVHH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVV CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVVH CKVVV CHH CKVVV CKVVV CHH CKVVV CHH CKVVV CHH CKVVV CKVVV CHH CKVVV CKVVV CHH CKVVV CKVVV CHH CKVVV CKVVV CKVVV CHH CKVVVV CKVVV CKVVV CKVVV CKVVV CKVVV CKVVVV CKVVV CKVVV CKVVVV CKVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVV CKVVVVVVVVVV	300 AFLAAG PAF LDNQPLGSV AFLAAG SPF 350 SSYFDSHSQ SSYFDSHSQ 400 AGMLAAY CGNNSTLES SAG EHVA SAG EHVA SAG EHVA 50 AATAAALHH EENARVWKG SLSAALTT 510 LPREDASLW	Section 6 312 VY I H L AT EK S F G S G - VH I G F G S S S Section 7 364 PTD PL T F L P P Section 8 416 V G P L I A V G P L I A Section 9 468 AT RAGVE Q V Section 9 468 A P T A Q P T R Section 10 0 520 VQ T A L WT R L

B2

								Section 9
dark m 47703	(417)	417 GLOAAVES	430	4		450		468
Aspergillus GH78	(1)	MLWS	SWILTPAL	LAIGSHAV	PEDYIL!	FOS	RTLNPS	LV <mark>Y</mark> Q <mark>V</mark> N
Bacillus GH78 Lactobacillus GH78	(321) (1)	EAAFT	PAGDNGVP	LATIGTED MAFTEQ	QSEYEDH Inn dy QF	GR H <mark>N</mark> Q	ALLDKS	AS <mark>YP</mark> P-
	(469)	469	480	,490	-	00	,510	520
dark_m.47703 Aspergillus GH78	(469)	AGLCLAAS	FLKRVPYR	LAGACAA P Guvdn pea	LISET AP	TALLGLP	DGTRD	RNTALE
Bacillus GH78	(366)		ALP	EARPTAAA	LEAFASWI	/KPF <mark>E</mark> PSLY	TEENVE	GSNVWR
Lactobacillus GH78	(29)		I <mark>LK</mark>	ETQVK <mark>AA</mark> S	IVALELD	QY <mark>LEG</mark> WG\	/KQIAPI	Section 11
4-4- 47700	(521)	521	530	540	,550	5	60	572
Aspergillus GH78	(80)	VSLOVTSV	SSKAROLG	TFTESS	WISSPACE	ATSDAGL	SP	LWES
Bacillus GH78	(405)	TLAERRAV	PR <mark>S</mark> VI <mark>N</mark> AI	PVPEPG <mark>v</mark>	LPVFEDGI	CELVIDLO	AERSG-	-F <mark>I</mark> GFE
Lactobacillus GH78	(68)	ELKRDDQ	IID <mark>EG</mark> DHQ	QFSIN	NAMGSPM	PUCFK	(FA	Section 12
4-4 47702	(573)	573 5	BO	590	600	610	DID	624
Aspergillus GH78	(126)	GHGPGTY	GADKKHLR	AFRALTL	VNNSTATI	SLDGLSIN	IYTAAPT	D-RG
Bacillus GH78	(455)	LE <mark>APAGT</mark> I	I D	AYG <mark>V</mark> EYM <mark>P</mark>	EGY <mark>T</mark> QHT	(G <mark>lp</mark> ntfry	CRECR	QSYVSP
Lactobacillus GH78	(111)	PRELARK	SE	DEMUSK	SWIQEE	HLOVLPTI	· · · · ·	Section 13
1.1	(625)	625 630	64	0	650	660	Henen	676
Aspergillus GH78	(025)	YKCYFHSS	DELINRIW	AGAYTLQ	LCTIDPT	GDSLIWL	VISSSI	NITLPQ
Bacillus GH78	(501)	VRRG <mark>FRY</mark> L	FLTVRGNS	<mark>A</mark> PVKI	HEIYIRQ	TYPVAEQ	SFRCSD	ALLNAT
Lactobacillus GH78	(149)	REYSIRYA	EITVVDTS.	PKWR <mark>av</mark> fs	NPVMA	AVDTAT	PELA	Section 14
	(677)	677	690		00	710	_	728
dark_m.47703	(677)	YAGANTL	ELCTCGTP:	SVE LOSK	WWTO	FVEHL	AYVSNG	DVASAL
Bacillus GH78	(550)	NEISRHTT	RLCMEDT		YEDMEWW	ESBNEAL1	NYYVFG	ETEIVE
Lactobacillus GH78	(201)	TENG LK TL	AD <mark>CMQD</mark> V-		RWI	ELR <mark>L</mark> Q <mark>AL</mark> A	NYATEK	
	(729)	729	,740	,750	,7	60	770	780
dark_m.47703	(729)	NS <mark>LAILA</mark> E	AQD P D G Y V	THLSPPKD	DPLRNLSN	IF <mark>A</mark> K <mark>E</mark> LYB	TEYCA	YVVYTW
Aspergillus GH78 Bacillus GH78	(279)	RELNEVPS	ADETP		QLPYAC	PSAWS	PNWTF	SLIGAS MILACR
Lactobacillus GH78	(250)	RC <mark>L</mark> YLEGA	PTTACRI	PAN	VETKP1	AVPDDEF	FDYSLF	FISILA
	(781)	781	790	800	.810	8	20	Section 16 832
dark_m.47703	(781)	1 <mark>H</mark> YL <mark>HTGD</mark>	D <mark>tfa</mark> refe	<mark>evi</mark> d <mark>ram</mark> g	WLCGLVA	D <mark>GLL</mark> K <mark>y</mark> ge	ET-AST	WHVPEV
Aspergillus GH78 Bacillus GH78	(319) (639)	SYYQYTGD SYAAHTGN	EAFAARIN	GQYKKGLQ PANKHTLT	HYLEHIDI	ISGLANITE GLINMA	SA WNLLDW	APTOP
Lactobacillus GH78	(295)	DYEAFSSD	KTVLND LY	RMAKNQMD	LALAQVIS	SE <mark>gkl</mark> k <mark>le</mark> e	E N P	VEILWS
	(833)	833 8	40	850	860	870	1	Sectior 17 884
dark_m.47703	(832)	VL <mark>G</mark> AP <mark>ID</mark> I	NCLMAHA F	RCANALR	AVG-KPP	TRDER	SRMVEA	NAR <mark>IW</mark> D
Aspergillus GH78	(367)	MGCHNIE	NALLYVL	NDAISLAS POSPALAI	DDRAN	GN STAA	SKIKAAA	ANAR <mark>IWD</mark>
Lactobacillus GH78	(344)	FDKET	GQALIIYT	LKQFITL.	ELVNDTS	LETYTEL	RKLNQY	KTQ <mark>1FD</mark>
	(995)	885 890	9	0	910	920		Section 18
dark_m.47703	(883)	SLGVYLN	SDNEDSIP	<mark>1</mark> 005	TTTAVWT	EVAPLORA	RALDYI	ERENGV
Aspergillus GH78	(419)	AONSLYRI	NETTTLHE	SDUV	SNAWATKA Cuvayte	NLTLSSNQ	SEALSSA	ALAARWG
Lactobacillus GH78	(396)	SQSGLFVS	GDOREVN-	VA	QVWMTLA	HVLDPEQT	ALMQTI	IVIK1
	(937)	937	950	q	60	970		Sectior 19 988
dark_m.47703	(931)	RECVLTAR	EHH <mark>PSMIS</mark>	YLSPFASE	RELLAFG	MGDAAGA		LWVHNAE
Aspergillus GH78	(466)	P <mark>XG</mark>	APAPEAGS	T <mark>VSPE</mark> GO	GELQ <mark>A</mark> H	LANEPDRA	LDLLRL	WGFLD
Lactobacillus GH78	(440)		FP	TA <mark>TPY</mark> MYH	HITEALE	EAGLKQEA	VQL MK DY	INGK <mark>NI</mark> T
	(989)	989	1000	1010)	1020	1030	Section 20 1040
dark_m.47703	(983)	TPEDVEN	EKVSRDGE	IEPYNQLO	GMC <mark>P</mark> TFT <mark>S</mark>	TCHCWSAG	TYALT	MLLGWT
Aspergillus GH78	(513)	DPRMINST	FIEGIST	GSLAMAP	RNTPRVS -DMETRS	AHGWSTG	SALTE	HYTAGLE STICK
Lactobacillus GH78	(481)	LG-ADTYN	EAFDPN	-QPDYSP1	GSPILNS	YCHAWSCT	VYLINH	(YLV
	(1041)	1041	1050	1060	107	0 .	1080	Section 21 1092
dark_m.47703	(1035)	PVTPGYAD	AVRPALA	DLRWAEG	VPTPHGG	IAVAWET	FGPAECI	LRVAVP
Aspergillus GH78	(565)	LTGPAGS	WLFKPOPC	NLTEVQAG	FETQLCL	FATQYQK <mark>S</mark> i	ATGTEQO	L TETAP
Lactobacillus GH78	(526)	KG DONK		DETWALG			V DAG K LI	
	(1002)	1093 1	100	1110	1120	142	n	Sectior 22
dark_m.47703	(1086)	AGURATVE	FEVAGVPV	RRLGAEI	DEPAVVA	DAEAIPLV	GPATAAV	NATLSCP
Aspergillus GH78	(617)	NGTS GSVE	EGATGQL	ISKR QAV	/KLVNGK <mark>A</mark>	RELQGGTW	TLKGL	
Lactobacillus GH78	(526)	SDE SVNV1	LAIFGE	A VALUE				

Figure S3: Tocopherol biosynthesis

Strong candidates could be found for each gene. There are many amino transferases in the transcriptome which may act on tyrosine and there are many methyltransferases whose substrates are difficult to conclusively assign. The genus of the closest homologue of each Euglena isoform is shown in blue.



Figure S4: Thiol biosynthetic pathways

Trypanothione is synthesised by joining one molecule of glutathione to each end of spermidine. The proposed biosynthetic pathway for the novel *nor*-trypanothione is shown, whereby aminopropane is transferred to 1,3-diaminopropane, derived from aspartate semialdehyde, to form *nor*-spermidine. Glutathione is then attached to this, either sequentially by two trypanothione synthases or by the separate isoforms adding one molecule each. Ovothiol biosynthesis is shown in the inset. Only the first two steps are known. The genus of the closest homologue is shown in blue.



Figure S5: Analysis of the thiol content of *E. gracilis*

Thiols were labelled with monobromobimane and analysed by HPLC. Collected fractions were analysed by LC-MS (Figure S6). There are peaks that have retention time and masses matching cysteine (Peak 1) and glutathione (Peak 4). There is a small peak (5) that coelutes with trypanothione and contains masses, which match the diprotonated mass, as well as a 14 Da smaller analogue, namely *nor*-trypanothione. Additionally there is a small peak (2) matching glutathionyl spermidine, which also contains a 14 Da smaller analogue. Together these data, along with the MS2 fragmentation (Figure S6) indicate the presence of a novel analogue of trypanothione that has one fewer carbons in the spermidine chain, namely *nor*-trypanothione. Peak 3 has a retention time that matches the reported relative retention times of ovothiol,⁹ though we were unable to obtain a standard. This peak has an exact mass matching mono-protonated ovothiol and its dimer, together with sodiated adducts, which fragment to give the expected monomers (see figure S6C). mB = monobromobimane. CyS = cysteine. GspdS = glutathionyl spermidine. GS = glutathione. TryS = trypanothione.



Peak	Retention time (min)	Compounds (bimane derivative)	Formula	Calculated m/z (Da)	Observed m/z (Da)
1	14.517	Cysteine	$[C_{13}H_{17}N_{3}O_{4}S+H]^{+}$	312.10124	312.10077
2	16.017	Glutathionyl spermidine/	[C ₂₇ H ₄₄ N ₈ O ₇ S+H] ⁺ /	625.31262/	625.31239/
2	10.017	Glutathionyl nor-spermidine	$[C_{26}H_{42}N_8O_7S+H]^+$	611.29697	611.29658
3	16.517	Ovothiol	$[C_{17}H_{21}N_5O_4S+H]^+$	392.13869	392.13870
4	17.875	Glutathione	$[C_{20}H_{26}N_5O_8S+H]^+$	498.16529	498.16536
5	26.008	Trypanothione/	[C ₄₆ H ₆₇ N ₁₃ O ₁₄ S ₂ +2H] ²⁺ /	552.73364/	552.73344/
5	20.300	nor-Trypanthione	$[C_{46}H_{67}N_{13}O_{14}S_2+2H]^{2+}$	545.72585	545.72543

!

Figure S6: Mass spectrometric analysis of bimane derivatives of thiol-containing fractions from HPLC

A. Peak 1. B. Peak 2. The glutathionyl spermidine containing masses were not selected from fragmentation. C. Peak 3. Ovothiol did not successfully fragment but the cluster ions fragmented to the monomers. D. Peak 4. E. Trypanothione in Peak 5. F. *nor*-Trypanothione in Peak 5. Note the homogeneity in the MS2 signals and deviation compared to the trypanothione masses indicates the 14 Da deviation cannot be in the glutamate or cysteine residues.



Table S3: Transcripts for polyketide and non-ribosomal peptide synthases.

Predicted proteins were identified using BLASTP to search for keto synthase (for PKSs), condensation, and adenylation (for NRPSs) domains. There is a high degree of uncertainty using this technique. The Kozak sequence for the predicted start codon is listed and the presence of any upstream stop codons is described, indicating the probability that this is the complete cognate transcript. FPKM values for each cognate transcript are given in parentheses.

A – Amino acid adenylation. AA-syn – Amino acid synthesis. ANK – Ankyrin domain. AT – Acyl transfer. C – Condensation. CoAL – Acyl-CoA ligase. DH – Dehydratase. AmT – Aspartate amino transferase. EH – Enoyl CoA hydratase ER – Enoyl reductase. HCS – HMGCoA synthase. KR – keto reductase. KS – Polyketide synthase. TE – thioesterase.

PKS		Role	Domains	Kozak sequence (Flagellate Kozak consensus is AnnATGnC ¹⁰)
1	lm.8157 (34.0)	PKS	KS-AT-DH-ER-KR- ACP-EH-EH-TE-HCS	ACGATGAT, also stop 4 codons
2	lm.60697 (2.79)	PKS	DH-KS-ACP-AmT	CAGATGGC
3	lm.53854 (2.62)	PKS	KS-AT-KR-ACP-KS	CCCATGCC
4	lm.82030 (3.47)	PKS	KR-ACP-KS	GGCATGGC
5	lm.42557 (1.77)	PKS	KR-ACP-KS-DH-KR	GGG <u>ATG</u> GC
6	lm.91532 (2.16)	PKS	A-ACP-KS	GTG <u>ATG</u> CA
7	lm.95952 (1.04)	PKS	ACP-KS	CCA <u>ATG</u> TT
8	lm.88225 (2.10)	PKS	KS-AT	GGC <u>ATG</u> G <u>C</u>
9	lm.23151 (2.17)	PKS	KS	CAC <u>ATG</u> CT
10	lm.94376 (1.70)	PKS	C-ACP	
11	lm.88941 (1.50)	PKS	KS	GTC <u>ATG</u> CT
12	lm.110121 (0.15)	PKS	KS	GGC <u>ATG</u> G <u>C</u>
13	lm.102218 (0.45)	PKS	KS	GGG <u>ATG</u> G <u>C</u>
14	lm.97081 (1.83)	PKS	ACP-KS	
NRPS		Activity		
1	lm.66007 (6.73)	NRPS	C-A-ACP	<u>A</u> CC <u>ATG</u> GA
2	lm.9669 (8.76)	NRPS	C-C-A-ACP-TE	<u>A</u> AC <u>ATG</u> G <u>C</u>
3	lm.32232 (3.16)	NRPS	C-C-A-ACP-C-A	<u>A</u> GG <u>ATG</u> CT
4	lm.96272 (1.73)	NRPS	C-A-A	<u>A</u> AC <u>ATG</u> A <u>C</u>
5	lm.21957 (19.80)	AA-Syn	C-A-ACP-TE	<u>A</u> CG <u>ATG</u> G <u>C</u>
6	lm.23118 (33.25)	AA-syn	A-ACP-TE	GCC <u>ATG</u> G <u>C</u>
7	lm.78138 (3.06)	А	A-ANK	TGT <u>ATG</u> TG
8	lm.54590 (2.55)	А	A-ANK	<u>A</u> GC <u>ATG</u> G <u>C</u>
9	lm.87820 (2.50)	А	A-ACP-TE	GAC <u>ATG</u> GC
10	lm.77877 (2.39)	А	A-ACP-ANK	CCG <u>ATG</u> GA, also stop 11 codons upstream
11	lm.98982 (1.30)	А	A	CCG <u>ATG</u> CT
12	lm.89785 (1.29)	А	A-ACP	GAA <u>ATG</u> CA
13	lm.47668 (4.62)	А	A	TAC <u>ATG</u> AT
14	lm.94698 (1.87)	A	A	GTG <u>ATG</u> CG
15	lm.44795 (2.15)	А	А	CGC <u>ATG</u> GG, also stop 6 codons
40		•	0-41	upstream
10	Im.44327 (5.36)	A	COAL	
17	Im.11010(127.33)	A	COAL	TICATOOC also stop 7 codops
18	lm.81072 (6.55)	A	CoAL	upstream
19	lm.97175 (0.69)	C	CoAL	
20	lm.26470 (20.15)	С	CoAL	CGG <u>ATG</u> C <u>C</u>
21	lm.22119 (3.22)	C	CoAL	GCC <u>ATG</u> A <u>C</u>
22	Im.37810 (6.84)	C	COAL	CCC <u>AIG</u> G <u>C</u>
23	Im.3119 (549.70)	C	CoAL	01017010
24	Im.28443 (11.01)	C	CoAL	CAG <u>ATG</u> AC
25	Im.9400 (16.40)	C	CoAL	AICATGAT
26	Im.26875 (17.61)	C	CoAL	CGC <u>ATG</u> G <u>C</u>
27	Im.12408 (12.41)	C	CoAL	

28	lm.3346 (17.23)	С	CoAL	
29	lm.44112 (13.17)	С	CoAL	CCA <u>ATG</u> CG
30	lm.17087 (13.39)	С	CoAL	<u>A</u> AC <u>ATG</u> G <u>C</u>

Figure S7: Potential siderophore production by *E. gracilis*

The chrome azurol S (CAS) assay was used to show the production of siderophores.¹¹ High nutrient media agar was prepared containing CAS (60 mg/l), hexadecyltrimethylammonium bromide (73 mg/l) and additional FeCl₃ (2.7 mg/l); cells were diluted such that approximately 10 cells were plated per petri dish. After two weeks a colourless halo around these colonies, particularly vivid when viewed with an orange filter, indicates the uptake of iron, potentially via a siderophore-based mechanism.



References

- 1. K. Duncan, R. M. Edwards and J. R. Coggins, *Biochem. J.*, 1987, 246, 375-386.
- 2. A. Varki and R. Schauer, *The Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, 2009.
- 3. P. Orlean and A. K. Menon, J. Lipid Res., 2007, 48, 993-1011.
- 4. K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, 22, 195-201.
- 5. T. Takahashi, Y. Ikeda, A. Tateishi, Y. Yamaguchi, M. Ishikawa and N. Taniguchi, *Glycobiology*, 2000, 10, 503-510.
- 6. Y. D. Lobsanov, P. A. Romero, B. Sleno, B. Yu, P. Yip, A. Herscovics and P. L. Howell, *J. Biol. Chem.*, 2004, 279, 17921-17931.
- 7. A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh and R. M. Garavito, *Biochemistry*, 2004, 43, 5170-5180.
- 8. Z. Cui, Y. Maruyama, B. Mikami, W. Hashimoto and K. Murata, J. Mol. Biol., 2007, 374, 384-398.
- 9. M. R. Ariyanayagam and A. H. Fairlamb, *Mol. Biochem. Parasitol.*, 2001, 115, 189-198.
- 10. K. Yamauchi, *Nucleic Acids Res.*, 1991, 19, 2715-2720.
- 11. B. Schwyn and J. B. Neilands, *Anal. Biochem.*, 1987, 160, 47-56.