

Phylogenetic and functional diversity of the aldehyde-alcohol dehydrogenases in microalgae

Robert van Lis^{1,2}, Yohann Couté³, Sabine Brugière³, Nicolas J. Tourasse⁴, Benoist Laurent⁵, Wolfgang Nitschke¹, Olivier Vallon⁴ and Ariane Atteia^{1,6,*}

¹Aix Marseille Université, CNRS, BIP UMR 7281, Marseille, France

²LBE, Univ Montpellier, INRAE, Narbonne, France

³Univ Grenoble Alpes, CEA, INSERM, IRIG, BGE, Grenoble, France

⁴UMR7141 CNRS-Sorbonne Université, Institut de Biologie Physico-Chimique, Paris, France

⁵FR 550 CNRS, Institut de Biologie Physico-Chimique, Paris, France

⁶MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France.

*Corresponding author: Ariane Atteia

E-mail address: ariane.atteia@cnrs.fr

Present address : MARBEC, Station Ifremer, Avenue Jean Monnet, Sète, France

SUPPLEMENTARY FIGURES

Supplementary Fig. S1. Multiple sequence alignment of *Polytomella* sp. ADHEs with their counterparts in *E. coli* and in *C. reinhardtii*.

Supplementary Fig. S2. Protein abundance in response to anoxia.

SUPPLEMENTARY TABLES

Supplementary Table S1. Primers sequences to clone *Polytomella* sp. ADHEs.

Supplementary Table S2. Mass-spectrometry-based identification of *Polytomella* sp. ADHEs.

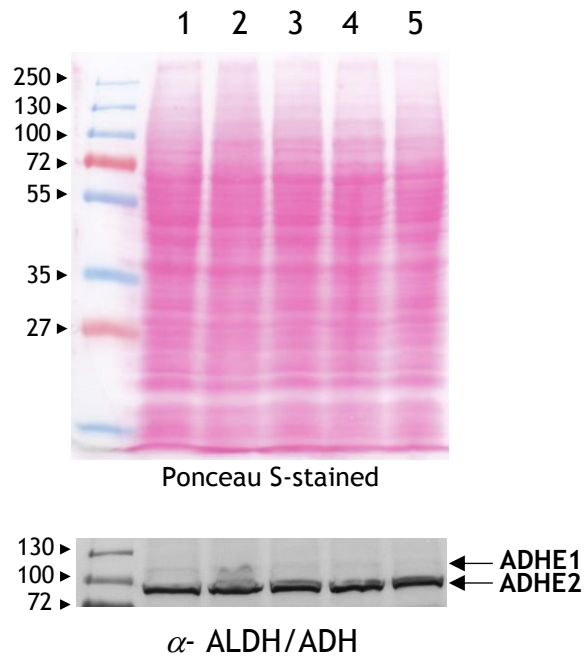
Supplementary Table S3. Mass-spectrometry-based identification and quantification of proteins in total extract from acetate-grown cells.

Supplementary Table S4. Occurrence of ADHE among free-living microalgae predicted from whole-genome sequencing projects (as of January2020).

Supplementary Fig. S1. Multiple sequence alignment of *Polytomella* sp. ADHEs with their counterparts in *E. coli* and in *C. reinhardtii*. Alignment was done using CLUSTAL Omega (1.2.4). Signatures for CoASH-dependent aldehyde dehydrogenase and iron-containing alcohol dehydrogenase are indicated. In red are shown the residues required for iron coordination in the ADH domain. Several features inferred from the cryo-EM structure of *E. coli* ADHE are indicated. The potential 8-aa linker between the ALDH and ADH domains is underlined in each algal ADHE. Key residues for enzyme auto-assembly are highlighted in black. Black arrow indicates the position of the residue required for ADHE oligomerization as determined for *E. coli*. Amino acids boxed in cyan refer to residues which ensure inter-helical interactions in the spiroosome structure. First and last tryptic peptides identified by mass spectrometry-based proteomics are in bold and underlined. The complete list of identified tryptic peptides is given in Supplementary Table S2.

<i>Ec</i> ADHE	-----	0
<i>Cr</i> ADHE	MMSSSLVSGKRV-----AVPSAAKPCAAVPLPRVAGRRTAARVVCEAAPSGAAP	49
<i>Ps</i> ADHE1	ML-TSNITGTSITASSRSKAFLAPFPIKSRPCVPSGVKKQENIRKG-TIVAKAT-----	52
<i>Ps</i> ADHE2	ML-----PTRA-----VSSL-----SRNVF	15
<i>Ec</i> ADHE	-----MAVTNVAELNALVERVKKAQREYASFTQEQVD	32
<i>Cr</i> ADHE	ASPKAEAAAPVAAAAPATPHAEVKKERAPATDEALTELKALLKRAQTAQAQYSTYTQEQVD	109
<i>Ps</i> ADHE1	-----PASAKEGDKKGI RAPIKDEELTILHETLKK AKHAQHIIYGFQSQEQVD	100
<i>Ps</i> ADHE2	ARPFANASRAYAAAPA-----AEQKSK SDEEGLSSLK STLNKAVAASKVFATYSQDQVD	69
	:: *: :::: *. :::::***	
<i>Ec</i> ADHE	KIFRAAALAAADARIPLAKMAVAESGMGIVEDKVIKNHFASEYIYNAYKDEKTCGVLS	92
<i>Cr</i> ADHE	EIFRAAAEAANAARIPLAKMAVEETRMGVAEDKVVKNHFASEFIYNKYKHTKTCGVIEHD	169
<i>Ps</i> ADHE1	HVFRCAEEAANAARLPLAKMAVAETRMGVVEDKVIKNHFASEFVYNKYKSTKTCGVIEHN	160
<i>Ps</i> ADHE2	KIFRAAAEAANAARLPLAKMAVEETRMGVVEDKVIKNHFASEFVYSQYKNTKTCGLTAHD	129
	.:**.* ** *:***** *: **:*****:*****:*. ** *****: .:	
<i>Ec</i> ADHE	DTFGTITIAEPIGIIICGIVPTTNPSTAIKSLISLKTRNAIIFSPHPRAKDATNKAADI	152
<i>Cr</i> ADHE	PAGGIQKVAEPVGVIAIVPTTNPSTAIKSLISLKTRNALVLCPPHRAAKSAIAAARI	229
<i>Ps</i> ADHE1	ENGGVTKVAEPVGVIAIVPTTNPSTAIKALLALKTRNGLIICPPHRAAKCTVQAAKI	220
<i>Ps</i> ADHE2	PISGYSKFAEPMGVIAIIPPTTNPSTAIKTLALKTRNALVICPPHRAAKCTIAAAKV	189
	* .:*****:*.*****:*****:*.*****:*****:***** .: ** :	
	-----NAD binding site 1--	
<i>Ec</i> ADHE	VLQAAIAAGAPKDLIGWIDQPSVELSNALMHPDINLILAT GGPG MVKAAYSSGKPA IGV	212
<i>Cr</i> ADHE	VRDAAVAAGAPPNIISWVETPSLPVSQLMQATEINLILAT GGP AMVRAAYSSGN PSLIGV	289
<i>Ps</i> ADHE1	IHDAAVAAGAPEGIISWVASPSMVVSQLMQAPEISLILAT GGPG MVRAAYSSGH PSLIGV	280
<i>Ps</i> ADHE2	VLDAAVAAGAPADIIITWIDSPSSLVSGALMKAPEVSLVLAT GGPG MVRAAYSSGH PSLIGV	249
	: :**:* ** .:* *: ** :* ***: :*:*****.**:*****:***:	
	-----catalytic center-----	
<i>Ec</i> ADHE	GAG NTPVVIDETADIKRAVASVLMKTFDNGVIC ASE QSVVVVDSVYDAVRERFATHGGY	272
<i>Cr</i> ADHE	GAG NTPALIDETADVAMAVSSILLKTFDNGVIC ASE QSVVVVAKAYDAVRTEFVRRGAY	349
<i>Ps</i> ADHE1	GAG NTPALIDETADIEMAVSSILLKTFDNGVIC ASE QSIVVVDSIYDEVRAEFVVRGAH	340
<i>Ps</i> ADHE2	GAG NTPALIDETADIELAVNSVLISKTFDNGVIC ASE QSIVAVDVAVYDKVRAEFVRRGAY	309
	*****.:*****: ** *:*****:*****:*. ** ** .*. :.:	
<i>Ec</i> ADHE	LLQGKELKAVQDVIKNGALNAAIVGQPAYKIAELAGFSVPENTKILIGEVTVVDESEPF	332
<i>Cr</i> ADHE	FLTEDDKVKVRAGVVVDGKLNPNIVGQSIKLAALFGIKVPQGTTVLIGEVEKIGPEEAL	409
<i>Ps</i> ADHE1	FLDEAQKEKVRGGILKDGHLNADIVGQISIRLTEIFGISVPAGSKVLIGEIERIGKDEPF	400
<i>Ps</i> ADHE2	FLNEEEKVKVRKGIKDGKLNADIVGQPVAKLAQMFGITVPAGAKVLIGEVSVVGKEEPL	369
	:* : *: : * ** **** :: : *.** .:*****: .: .* :	
<i>Ec</i> ADHE	AHEKLSPTLAMIRAKDFEDAVEKAIEKLVAMGGIGHTSCLYTDQDNQPARVSYFGQRMKTA	392
<i>Cr</i> ADHE	SQEKLCPILAMYRAPDYDHGVKMACELIMYGAGHTSVLYTNPLNNA-HIQYQS AVKTV	468
<i>Ps</i> ADHE1	SQEKLSPILGMIRAKDYDDALQKTYKLIMGGAGHTSVLYSSPLNKE-NIGRFQ MAKTV	459
<i>Ps</i> ADHE2	SQEKLSPILAMIRSSDFNTALATTHKLIMFGGAGHTSVIYTHPDNRQ-NIEAFQ MAKTV	428
	:.***.* ***: *: : :*: ** **** :*: *. .: : **	

Supplementary Fig. S2. Protein abundance in response to anoxia. *Polytomella* cells were kept under anoxia in 50 mM potassium phosphate medium (pH 7.0) at a cell concentration of $5 \cdot 10^6$ cells ml^{-1} . Proteins in cell extracts (50 μg) were separated on a 8% SDS-PAGE and transferred to nitrocellulose membrane. Lane 1, exponentially grown cells; lane 2, 2 h of anoxia; lane 3, 4 h of anoxia; lane 4, 8 h of anoxia; lane 5, 8h in aerobic potassium phosphate medium. ADHE proteins were detected by immunoblotting experiment with anti-ALDH/ADH.



Supplementary Table S1. Primers sequences to clone *Polytomella* sp. ADHEs. Sequences are given from 5' to 3'.

To clone <i>Polytomella</i> ADHE full-length coding sequences. Start and stop codons are indicated.		
	Forward primer	Reverse primer
ADHE1	cagaa <u>ATG</u> CTTACCTCTAATATCACTGGC	catcat <u>CTA</u> AGCGATGGGAGTCACGAC
ADHE2	taaa <u>ATG</u> CTTCCCACTCGTGCCGTC	accaat <u>TTA</u> CAAACCCTGGTGGGCGGCGACG
To clone ADHE coding sequences into pET24a expression vector. Restrictions sites used for PCR product cloning are underlined.		
Protein (aa position) ¹	Forward primer	Reverse primer
rADHE1 (T ⁵² -A ⁹⁹⁰)	gagggcgCATATG <u>AC</u> CCCCGCTTCTGCAAAGGAG	gtc <u>CTCGAG</u> AGCGATGGGAGTCACGACGTT
rADHE1_t (T ⁵² -V ⁹⁴⁹)	gagggcgCATATG <u>AC</u> CCCCGCTTCTGCAAAGGAG	gtc <u>CTCGAG</u> GACAGTGCTGCCTCCGAAGCT
rADHE2 (A ²⁸ -L ⁸⁸⁶)	gagggcgCATATGGCTGCTCCCGCTGCCGAGCAG	gtc <u>CTCGAG</u> CAAACCCTGGTGGGCGGCGAC

Supplementary Table S2. Mass spectrometry-based identification of *Polytomella* sp. ADHEs. Stacking analysis of total protein extract from acetate-grown cells.

ADHE1					
Start	Stop	Peptide	PTM	Score	Protein Sets
67	83	APIKDEELTILHETLKK		51,82	1 (ADHE1)
94	104	FSQEQVDHVFR		85,64	1 (ADHE1)
137	147	NHFASEFVYNK		49,31	1 (ADHE1)
168	191	VAEPVGVIAGIVPTTNPSTAIFK		60,28	1 (ADHE1)
269	306	AAYSSGHPSLGVGAGNTPALIDETADIEMAVSSILIS		90,45	1 (ADHE1)
307	331	TFDNGVICASEQSIVVVDSIYDEVR	Carbamidomethyl (C8)	50,53	1 (ADHE1)
357	371	DGHLNADIVGQSISR		74,01	1 (ADHE1)
372	385	LTEIFGISVPAGSK		73,15	1 (ADHE1)
386	393	VLIGEIER		69,42	1 (ADHE1)
405	413	LSPILGMYR		70,71	1 (ADHE1)
427	446	LIMNGGAGHTSVLYSSPLNK		45,64	1 (ADHE1)
516	522	EGMLWFR		61,92	1 (ADHE1)
531	542	SGCLEPALMDLK	Carbamidomethyl (C3)	46,08	1 (ADHE1)
547	559	AFIITDKPLFDLK		66,23	1 (ADHE1)
599	618	AFNPDVIALGGGSPMDAAK		69,69	1 (ADHE1)
599	618	AFNPDVIALGGGSPMDAAK	Oxidation (M16)	39,35	1 (ADHE1)
619	630	VMWLMYCEPETR	Carbamidomethyl (C8)	56,69	1 (ADHE1)
619	630	VMWLMYCEPETR	Oxidation (M2); Carbamidomethyl (C8)	50,34	1 (ADHE1)
619	630	VMWLMYCEPETR	Oxidation (M5); Carbamidomethyl (C8)	48,22	1 (ADHE1)
631	637	FDGLAMR		31,73	1 (ADHE1)
645	653	IYSVPELGR		67,1	1 (ADHE1)
654	684	KAILVAIPTTSGTGSEVTPFVSVTNEVTGQK		76,13	1 (ADHE1)
655	684	AILVAIPTTSGTGSEVTPFVSVTNEVTGQK		64,82	1 (ADHE1)
685	709	YPIADYALTPSMAIIDPQLVLTMPK		84,28	1 (ADHE1)
710	737	RLTAWSGIDALTHALESYVSVCSDDYNR	Carbamidomethyl (C22)	72,63	1 (ADHE1)
711	737	LTAWSGIDALTHALESYVSVCSDDYNR	Carbamidomethyl (C21)	91,34	1 (ADHE1)
742	749	EAISLLFK		34,91	1 (ADHE1)
768	793	VHYAATIAGMSFANAFLGICHSMAMK	Carbamidomethyl (C20)	41,07	1 (ADHE1)
794	815	LGAAYHVPGLANAALISHVIR		68,07	1 (ADHE1)
816	824	YNATDAPFK		38,54	2 (ADHE1,ADHE2)
836	861	AKNDYAQIADV LGLGGSPGAMTDDEK		103,16	1 (ADHE1)
838	861	NDYAQIADV LGLGGSPGAMTDDEK		95,18	1 (ADHE1)
865	874	LIEAIEELKR		63	1 (ADHE1)
875	883	ELEIPATLK		31,29	1 (ADHE1)
884	917	DIFNDPAKDAEFLANIDHLAEEAFDDQCTGANPR	Carbamidomethyl (C28)	95,32	1 (ADHE1)
892	917	DAEFLANIDHLAEEAFDDQCTGANPR	Carbamidomethyl (C20)	78,33	1 (ADHE1)
926	940	QILLDAHSPITPLK		62,74	1 (ADHE1)

ADHE2					
Start	Stop	Peptide	PTM	Score	Protein Sets
38	47	SDEGLSSLK		48,26	1 (ADHE2)
59	70	VFATYSQDQVDK		73,54	1 (ADHE2)
59	73	VFATYSQDQVDKIFR		54,27	1 (ADHE2)
106	118	NHFASEFVYSQYK		89,08	1 (ADHE2)
122	136	TCGLIAHDPISGYSK	Carbamidomethyl (C2)	74,27	1 (ADHE2)
137	160	FAEPMGVIAGIIPPTNPTSTAIFK		59,17	1 (ADHE2)
137	160	FAEPMGVIAGIIPPTNPTSTAIFK	Oxidation (M5)	43	1 (ADHE2)
189	220	VVLDAAVAAGAPADIITWIDSPSSLVSGALMK		100,33	1 (ADHE2)
189	220	VVLDAAVAAGAPADIITWIDSPSSLVSGALMK	Oxidation (M31)	67,42	1 (ADHE2)
221	237	APEVSLVLATGGPGMVR		107,22	1 (ADHE2)
221	237	APEVSLVLATGGPGMVR	Oxidation (M15)	48,5	1 (ADHE2)
238	275	AAYSHGHPSLGVGAGNTPALIDETADIELAVNSVLIS		127,1	1 (ADHE2)
276	300	TFDNGVICASEQSIVAVDAVYDKVR	Carbamidomethyl (C8)	92,34	1 (ADHE2)
276	298	TFDNGVICASEQSIVAVDAVYDK	Carbamidomethyl (C8)	46,54	1 (ADHE2)
306	316	RGAYFLNEEEK		50,04	1 (ADHE2)
306	318	RGAYFLNEEEKVK		43,88	1 (ADHE2)
307	318	GAYFLNEEEKVK		49,23	1 (ADHE2)
326	340	DGKLNADIVGQPVAK		34,82	1 (ADHE2)
329	340	LNADIVGQPVAK		61,06	1 (ADHE2)
341	354	LAQMFGITVPAGAK		73,19	1 (ADHE2)
341	354	LAQMFGITVPAGAK	Oxidation (M4)	57,49	1 (ADHE2)
355	373	VLIGESVVGKEEPLSQEK		124,41	1 (ADHE2)
355	365	VLIGESVVGK		55,16	1 (ADHE2)
374	382	LSPILAMYR		64,7	1 (ADHE2)
383	395	SSDFNTALATTHK		86,75	1 (ADHE2)
396	415	LIMFGGAGHTSVIYTHPDNR		165,88	1 (ADHE2)
396	415	LIMFGGAGHTSVIYTHPDNR	Oxidation (M3)	45,17	1 (ADHE2)
416	426	QNIQAFQAMAK		69,79	1 (ADHE2)
416	426	QNIQAFQAMAK	Oxidation (M9)	39,45	1 (ADHE2)
430	478	ILVNSPSSQGAIGGIYNQILAPSMGLGCGSWGGNIV	Carbamidomethyl (C28)	79,02	1 (ADHE2)
485	513	HGAGGFNVPATVLLQSGGAASTFAQLSGK		161,06	1 (ADHE2)
514	534	KVLVVADKNADVAAVSAALAK		151,62	1 (ADHE2)
515	534	VLVVADKNADVAAVSAALAK		78,08	1 (ADHE2)
522	534	NADVAAVSAALAK		69,1	1 (ADHE2)
535	559	AGAEATVFNQVGAASLAAVQAGLK		91,38	1 (ADHE2)
560	582	ATVAFHPDTIVAVGGGAIDAIAK		139,49	1 (ADHE2)
583	594	LIWLGYPETR		63,8	1 (ADHE2)
608	618	RVYNVPALGAK		31,26	1 (ADHE2)
609	618	VYNVPALGAK		59,12	1 (ADHE2)
619	647	ASLIAIPTAIGSGSEVTPFSSLVDEATGK		125,93	1 (ADHE2)
648	673	SYTIADAAFFPSAAIVDPSLIAALPK		55,31	1 (ADHE2)
674	699	AAVAAGAFEAIHAVESFVIAASDR		143,22	1 (ADHE2)
706	722	EALTQIFDALPSADAEK		76,87	1 (ADHE2)
730	748	AGMAYANAF LGVTQSLANK		125,49	1 (ADHE2)
730	748	AGMAYANAF LGVTQSLANK	Oxidation (M3)	58,36	1 (ADHE2)
749	770	VAVACDIPVGVAAAVLLPYVIR	Carbamidomethyl (C5)	110,7	1 (ADHE2)
771	779	YNATDAPFK		38,54	2 (ADHE2,ADHE1)
780	790	QAIFPSYHSPR		75,03	1 (ADHE2)
791	803	AVADYAELANALK		66,31	1 (ADHE2)
804	824	LGGSTPVEKAENLAAAIEGLR		119,03	1 (ADHE2)
804	812	LGGSTPVEK		28,8	1 (ADHE2)
813	824	AENLAAAIEGLR		72,77	1 (ADHE2)
846	868	FLAVVDKLAEEAFDDQCQLANPR	Carbamidomethyl (C17)	147,3	1 (ADHE2)
853	868	LAEEAFDDQCQLANPR	Carbamidomethyl (C10)	121,93	1 (ADHE2)
869	876	YPLIEDLK		42,61	1 (ADHE2)
877	886	AILVAAHQGL		28,13	1 (ADHE2)

Supplementary Table S3. Mass-spectrometry-based identification and quantification of proteins in total extract from acetate-grown cells. A total of 2307 different proteins were identified and quantified. This table presents the 100 proteins with the highest iBAQ values.

Accession (Algotblast)	Protein name	Molecular Weigth (Da)	Coverage (%)	Identified peptides	iBAQ
UTR_g1330.t1	Glyceraldehyde 3-phosphate dehydrogenase 3	39611	68,39	36	824603884
UTR_g7939.t1	Fructose-1,6-bisphosphate aldolase 2	41130	76,39	42	778106880
UTR_g5155.t1	Alcohol dehydrogenase 3	43662	72,57	51	740860245
UTR_g853.t1	Phosphoenolpyruvate carboxykinase	61328	85,74	52	737455250
UTR_g7824.t1	Glyceraldehyde 3-phosphate dehydrogenase 1	39427	67,49	35	735308958
UTR_g6404.t1	Isocitrate lyase 1	45260	69,47	28	731956019
UTR_g4113.t1	Citrate synthase 1	50678	78,62	38	695130717
UTR_g4737.t1	Malate dehydrogenase 2	35204	74,93	31	646246520
UTR_g7469.t1	Phosphoglycerate kinase 1	46946	69,39	32	588986095
UTR_g3058.t1	Fructose-1,6-bisphosphate aldolase 1	41570	73,82	39	533024469
UTR_g6806.t1	ATP synthase subunit beta	61874	74,91	37	528178417
UTR_g449.t1	Enolase 1	51700	75,99	38	520586057
UTR_g3623.t1	Isocitrate lyase 2	45057	85,82	32	485585822
UTR_g5195.t1	Tubulin beta 1	49576	74,94	45	469720390
UTR_g1942.t1	ATP synthase subunit alpha	60759	58,9	33	416877541
UTR_g8560.t1	ADHE2	91667	79,8	56	416738323
UTR_g8362.t1	Malate dehydrogenase 3	36376	75,36	24	394260570
UTR_g8202.t1	Malate dehydrogenase 4	36343	76,5	24	379971192
UTR_g5793.t1	Pyruvate decarboxylase 1	63127	68,25	34	346631050
g1843.t1	Predicted protein	50771	70,63	38	345088465
UTR_g4027.t1	Tubulin alpha-2 chain	38536	78,75	16	344867145
UTR_g2978.t1	Mitochondrial ADP/ATP translocator protein (CBN20772.1)	33307	51,3	22	343043852
UTR_g6397.t1	Triose phosphate isomerase	30585	68,66	20	342057643
UTR_g3129.t1	Tubulin alpha-1 chain	49629	84,48	34	341843768
UTR_g6936.t1	2-cys peroxiredoxin	21573	83,42	11	312452166
UTR_g1449.t1	[PTA]	61554	81,99	45	262787824
g4183.t1	Mitochondrial dicarboxylate/tricarboxylate transporter	31891	77,26	26	246430866
UTR_g5093.t1	[ACK]	43795	78,64	26	241553920
UTR_g7511.t1	Mitochondrial phosphate carrier protein	36091	72,65	21	205279091
UTR_g5448.t1	Aconitase 1	86205	60,38	47	192932096
UTR_g3037.t1	voltage-dependent anion-selective channel protein, partial (CBU30419.1)	28557	76,64	21	190493172
UTR_g7451.t1	ATP synthase associated protein ASA2	50501	76,77	27	189880798
UTR_g4659.t1	[12-oxophytodienoate reductase]	45119	74,76	23	189165696
UTR_g2915.t1	Phosphoserine aminotransferase 1	42946	70,41	32	188306193
UTR_g8790.t1	Aconitase 2	86254	60,75	48	186415006
UTR_g6540.t1	Predicted protein	31534	76,69	16	174650214
UTR_g9052.t1	Predicted protein	25904	39,43	9	173362266
UTR_g6298.t1	Predicted protein	29396	82,24	24	172957979
UTR_g1630.t1	Malate synthase 1	60347	76,38	37	165269832
UTR_g3661.t1	Predicted protein	21243	76,14	13	161867894
UTR_g6499.t1	Isocitrate dehydrogenase [NAD] subunit 1	38692	54,87	21	160093277
UTR_g2167.t1	Predicted protein	17855	75,15	13	158071401
UTR_g6318.t1	Succinyl-CoA ligase (GDP-forming), beta subunit	44728	66,35	30	156680268
UTR_g1109.t1	ATP synthase subunit gamma	34569	58,68	17	155240512
UTR_g2242.t1	60S ribosomal protein L27	15246	40	7	154137760
UTR_g6170.t1	Uncoupling protein	32650	77,85	23	154088359
UTR_g8544.t1	ATP synthase associated protein ASA4	33945	70,94	21	153259833
UTR_g8345.t1	ATP synthase associated protein ASA1 (CAD90158.2)	68587	65,7	41	152459895
UTR_g4933.t1	Peptidyl-prolyl cis-trans isomerase 1	18307	68,64	10	146207860
UTR_g4458.t1	Succinate-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	33792	72,09	18	144074692
UTR_g1561.t1	Alpha-1,4 glucan phosphorylase	98924	65,42	50	139884875
UTR_g2703.t1	Heat shock protein 70-1	70631	63,26	38	137061388
UTR_g2379.t1	ATP synthase associated protein ASA3	34814	56	15	135720348
UTR_g6672.t1	D-3-phosphoglycerate dehydrogenase	61031	71,63	33	133048943
UTR_g7528.t1	Malate dehydrogenase 5	37132	12,29	4	132290624
UTR_g1156.t1	Predicted protein	28643	59,3	17	130479242
UTR_g3590.t1	Alcohol dehydrogenase 1	44217	63,17	21	128842228
UTR_g4764.t1	H(+)-translocating inorganic pyrophosphatase	79488	38,29	34	118707880
UTR_g1063.t1	ATP synthase associated protein ASA5_1	13984	34,96	5	117868376
g8053.t1	Glutamine synthetase	42154	66,32	17	113302105
UTR_g3311.t1	Nucleoside diphosphate kinase 2	16497	50	15	110033489
UTR_g2346.t1	ATP synthase subunit OSCP	25497	56,33	14	107072753
UTR_g6765.t1	40S ribosomal protein S11	17671	33,33	6	104333965
UTR_g6201.t1	Adenylate kinase 1	26283	60,08	12	104236386
UTR_g8511.t1	Predicted protein	34083	25,87	7	103983872
UTR_g1520.t1	COX2B (AAK30366.1)	17219	74,51	9	103532896
UTR_g5069.t1	Uncoupling protein 2	32254	71,2	21	97767703
UTR_g4497.t1	Ubiquinol:cytochrome c oxidoreductase 50 kDa core 1 subunit	55189	68,42	27	95457333
UTR_g656.t1	Predicted protein	24151	61,33	10	94956237
UTR_g2501.t1	40S ribosomal protein S8	23637	45,89	10	94594789
UTR_g7612.t1	Predicted protein	43136	67,16	18	88519588
UTR_g6680.t1	cytochrome c oxidase subunit COXI (AAC24895.1)	42689	9,16	2	86014925
UTR_g2509.t1	Phosphofructokinase 1	53916	62,06	21	85475376
UTR_g8218.t1	60S ribosomal protein L12	17695	38,55	5	84760083
UTR_g4689.t1	60S ribosomal protein L6	19138	56,9	10	83903127
UTR_g1084.t1	Inorganic pyrophosphatase 1	31353	60,14	13	83199670
UTR_g8886.t1	40S ribosomal protein S9	21908	42,33	9	83061989
g4548.t1	40S ribosomal protein S18	17369	40,52	8	81845675
UTR_g4079.t1	60S ribosomal protein L7 (CBV76073.1)	27328	58,16	16	80891816
UTR_g6449.t1	S-adenosylhomocysteine hydrolase protein SAH1 (CBN20774.1)	52852	56,61	24	80766493
UTR_g4111.t1	60S ribosomal protein L26	16140	48,28	7	80387192
UTR_g125.t1	Glucose-6-phosphate isomerase 1	69761	56,51	27	79806983
UTR_g5791.t1	40S ribosomal protein S17	14866	61,54	9	78771995
UTR_g2444.t1	Histone H4	11443	41,75	5	76449776
UTR_g7024.t1	40S ribosomal protein S5a	21765	55,1	11	76396009
UTR_g4822.t1	60S ribosomal protein L23a	16694	48,67	8	75895374
UTR_g6901.t1	60S ribosomal protein L4	43342	47,49	18	75627436
UTR_g241.t1	Predicted protein	20860	62,31	7	75245446
UTR_g3652.t1	S-adenosylmethionine synthetase 1	42522	79,95	25	74459075
UTR_g4745.t1	ADHE1	106378	52,02	37	74219220
UTR_g7943.t1	Plastid triose phosphate/phosphate translocator (id to <i>P. parva</i> ABH10984)	25,06	41790	5	72579355
UTR_g1716.t1	40S ribosomal S5b	21650	49,23	10	71788376
UTR_g7187.t1	Phosphoglucomutase	65567	59,01	29	69686090
UTR_g6446.t1	40S ribosomal protein S13	17017	52,32	8	68213596
UTR_g5991.t1	Ubiquinol:cytochrome c oxidoreductase, 14kDa subunit	14058	58,2	7	68802387
g6593.t1	Tryptophan synthase, alpha chain	41847	63,13	19	68535313
UTR_g8808.t1	40S ribosomal protein SA (CBX25070)	30000	56,51	13	68130064
UTR_g1433.t1	40S ribosomal protein S15a	14783	53,85	8	68025992
UTR_g8847.t1	plastid ATP/ADP transporter (id to <i>P. parva</i> ABH10983)	63647	37,44	19	67739684
UTR_g4516.t1	Predicted protein	93898	63,67	53	67356239

Supplementary Table S4. Occurrence of ADHE among free-living microalgae predicted from whole-genome sequencing projects (as of January 2020). Species are sorted following supergroups and clades as defined by Adl et al (2012; 2019).

¹, Habitats: F, freshwater; S, salty (brackish, marine);

², Total sequence length of the nuclear genome in megabases (Mb);

³, Searches were done using public web-servers:

A, Algoblast (<http://lobosphaera.ibpc.fr/algoblast>);

B, JGI genome Portal (<https://genome.jgi.doe.gov/portal/pages/tree-of-life.jsf>);

C, Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>);

D, US National Library of Medicine National Institutes of Health (<https://www.ncbi.nlm.nih.gov/pubmed>),

E, CryptoDB (<https://cryptodb.org/cryptodb/showApplication.do>)

⁴, ADHE occurrence predicted by BlastP searches; p, partial sequence covers 70% of the protein;

⁵, Sequence number in phylogenetic tree in Figure 6.

	Habitats ¹	Size (Mb) ²	BlastP ³	ADHE ⁴	Tree ⁵
Archaeplastida supergroup					
Chlorophytes					
Chlorophyceae					
<i>Chlamydomonas reinhardtii</i>	F	111	C	1	3
<i>Chromochloris zofingiensis</i>	F	58	C	0	
<i>Dunaliella tertiolecta</i>	M		A	0	
<i>Dunaliella salina</i> CCAP19-18	M	343	C	0	
<i>Gonium pectorale</i>	M	148.8	D	2	4,5
<i>Polytomella</i> sp. SAG 198.90	M	80	This work	2	1,2
<i>Monoraphidium neglectum</i>		69.7	D	1(p)	
<i>Raphidocelis subcapitata</i>		51.1	D	1(p)	
<i>Scenedesmus</i> ARA	F	93.2	D	0	
<i>Volvox carteri</i>	F	137.7	D	2	6,7
Trebouxiophyceae					
<i>Auxenochlorella protothecoides</i>	F	22.92	B	0	
<i>Chlorella sorokiniana</i>	F	58.53	D	2	10,11
<i>Chloroidium</i> sp.	M		D	0	
<i>Coccomyxa subellipsoidea</i> C-169	F	48.83	B	0	
<i>Lobosphaera incisa</i>	F		A	0	
<i>Micractinium conductrix</i>		61.01	D	2	8,9
<i>Picochlorum soloecismus</i>	M		A	1	12
<i>Picochlorum</i> sp. SENEW3	M	15.25	B	1	13
<i>Picochlorum</i> NBRC102739	M		A	2	14,15
Mamiellophyceae					
<i>Bathycoccus prasinus</i> RCC1105	M	15.07	D	0	
<i>Micromonas pusilla</i>	M	21.95	D	0	
<i>Ostreococcus tauri</i>	M	13.89	D	0	
<i>Ostreococcus lucimarinus</i>	M	13.20	D	0	
Chlorodendrophyceae					
<i>Tetraselmis striata</i>	M	228	A	1	17
Prasinophyceae					
<i>Mesostigma viride</i>		145.9	D	0	
<i>Picocystis salinarum</i>	M	29.6	A	1	16
<i>Pyramimonas amyliifera</i> CCMP720	M		A	2	18,19
Nephroselmidophyceae					
<i>Nephroselmis pyriformis</i>	M		A	0	
Palmophyllophyceae					
<i>Prasinoderma coloniale</i>	M		D	1	20
Glaucophytes					
<i>Cyanophora paradoxa</i>	F	99.94	A	0	
Rhodophytes					
<i>Galdieria sulphuraria</i>	M	14.3	D	0	
<i>Cyanidioschyzon merolae</i>	M	14.44	D	0	
<i>Porphyridium purpureum</i>	M	19.45	D	0	
SAR supergroup					
Stramenopiles					
Bacillariophyceae					
<i>Asterionella formosa</i>	F	68.42	D	0	
<i>Fistulifera solaris</i>	M	49.7	D	2	24, 25
<i>Fragilariopsis cylindrus</i> CCMP1102	M	74.7	D	0	
<i>Phaeodactylum tricornutum</i>	M	27.4	D	0	
<i>Thalassiosira pseudonana</i>	M	32.4	D	0	
<i>Thalassiosira oceanica</i>	M	92.2	D	0	
<i>Pseudo-nitzschia multistriata</i>	M	56.7	D	0	
Eustigmatophyceae					
<i>Nannochloropsis gaditana</i>	M	29.22	D	0	
<i>Nannochloropsis oceanica</i>	M	36.45	B	0	
Pelagophyceae					
<i>Aureococcus anophagefferens</i>	M	56.66	D	0	
Synurophyceae					
<i>Ochromonadaceae</i> sp. CCMP2298	M	61.14	B	0	
Alveolata					
Chromerids					
<i>Chromera velia</i>	M	187.45	E	2	27,28
<i>Vitrella brassicaformis</i>	M	72.70	E	1	26
Rhizaria					
Chlorarachniophyte					
<i>Bigelowiella natans</i>	M	91.4	B	1	23
Haptista clade					
<i>Emiliania huxleyi</i> CCMP1516	M	167.67	B	0	
Cryptista clade					
<i>Guillardia theta</i> CCMP2712	M	87	B	2	21,22