MOLECULAR ADAPTATIONS IN EXTREMELY HALOPHILIC PROTISTS

by

Tommy Harding

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
November 2016

Life finds a way.

- Dr. Ian Malcolm

Table of Contents

List of Tables

List of Figures

Abstract

Halophiles, organisms adapted to hypersaline habitats, represent unique cases in which to study evolutionary and ecological processes. To overcome the high level of stress in these habitats and to keep cellular components functional, halophiles have developed adaptations like amino acid content bias in proteins exposed to high salt, the use of solutes (organic or inorganic) to maintain high intracellular osmotic strength, and the salt-dependent adjustment of plasma membrane fluidity. Halophilic bacteria and archaea have been intensively studied, and substantial research has been conducted on halophilic fungi, and the green alga *Dunaliella.* By contrast, very few investigations of halophilic phagotrophic protists, *i.e.* protozoa, have been conducted and knowledge about their evolution is lacking. The goal of this thesis is to characterize the molecular adaptations of halophilic protozoa, using as the primary model the stramenopile *Halocafeteria seosinensis*, through transcriptomic and genomic analyses, including a study of gene expression as a function of salinity. Examination of the inferred cytoplasmic proteomes of two halophilic protozoa, *H. seosinensis* and the heterolobosean *Pharyngomonas kirbyi,* indicated an increased hydrophilicity compared to the proteomes of marine protists, but an absence of the acidic signature commonly detected in halophilic prokaryotes that accumulate high levels of cytosolic inorganic ions. These results implied levels of intracellular salt in halophilic protozoa higher than in marine protists, but lower than in 'salt-in' microbes. Concordantly, genes putatively involved in synthesis and transport of organic osmolytes that could contribute to balance the osmotic equilibrium in *H. seosinensis* (*e.g.* hydroxyectoine) were up-regulated at high salt. Other salt-responsive genes were involved in stress response (*e.g.* chaperones), ion homeostasis (*e.g.* Na^+/H^+ transporter), metabolism and transport of lipids (*e.g.* sterol biosynthetic genes), carbohydrate metabolism (*e.g.* glycosidases), and signal transduction pathways (*e.g.* transcription factors). Several potential gene duplication and/or lateral gene transfer events could have favored adaptability to high salt (*e.g.* involving genes coding for ion transporters and peroxidase). This study proposes that a transition toward high-salt adaptation in halophilic protozoa probably requires a shift in amino acid composition of proteins and alteration of transcriptional programs, leading to modification of cell structure properties like membrane fluidity and increased stress resistance.

List of Abbreviations Used

% ID	percentage identity
AA	amino acid
AC	adenylate cyclase
Ala	alanine
Arg	arginine
ASADH	aspartate-β-semialdehyde dehydrogenase
Ask	aspartate kinase
AT	adenine/thymine
ATCC	American Type Culture Collection
ATF2	activating transcription factor 2
AtSAT1	*Arabidopsis thaliana* sterol acyltransferase 1
bp	base pair
BZIP	basic leucine zipper
cAMP	cyclic adenosine monophosphate
CCAP	Culture Collection of Algae and Protozoa
CCMP	Culture Collection of Marine Phytoplankton
CG	cohesion group
cGMP	cyclic guanosine monophosphate
CHASE	cyclase/histidine kinase-associated sensory extracellular
CNBD	cyclic nucleotide-binding domain
COG	clusters of orthologous group of proteins
Cys	cysteine
DAT	dopamine transporter
DGAT	diacylglycerol O-acyltransferase
DGGE	denaturing gradient gel electrophoresis
DHAB	deep-sea hypersaline anoxic basin
EctA	diaminobutyrate acetyltransferase
EctB	diaminobutyrate aminotransferase
EctC	ectoine synthase
EctD	ectoine hydroxylase
ELFCS	eukaryotic long-chain fatty acid CoA synthetase
ER	endoplasmic reticulum
ETC	electron transport chain
FAD	fatty acid desaturase
FAE	fatty acid elongase
FC	posterior fold change
FeSOD	iron-dependent superoxide dismutase
FISH	fluorescent *in situ h*ybridization
GABA	Y-aminobutyric acid
GAT1	Y-aminobutyric acid transporter

GC	guanylate cyclase
Glu	glutamate
Gly	glycine
GlyT1b	glycine transporter
GPCR	G protein-coupled receptor
GRAVY	grand average of hydropathy
GRX	glutaredoxin
GT	glutathione transferase
His	histidine
HK	histidine kinase
HPLC	high-performance liquid chromatography
HpT	histidine-containing phosphotransfer
HSE	heat-shock DNA element
HSF	heat-shock factor
HSP	heat-shock protein
LeuT	leucine transporter
LGT	lateral gene transfer
MAPEG	membrane-associated proteins in eicosanoid and glutathione
MAPK	mitogen-activated protein kinase
MarProt	Marine Protists comparative dataset
MBOAT	membrane-bound O-acyltransferase
MGT	microsomal glutathione transferase
MMETSP	Marine Microbial Eukaryote Transcriptome Sequencing Project
MnSOD	manganese-dependent superoxide dismutase
NCBI	National Center for Biotechnology Information
NCMA	National Center for Marine Algae and Microbiota
NMR	nuclear magnetic resonance
NPC1	Niemann-Pick type C1
NQO	NAD(P)H : quinone oxidoreductase
NR	non-redundant
NT	nucleotide collection
OIV	2-oxoisovalerate
ORF	open reading-frame
OTU	operational taxonomic unit
P2XR	P2X receptor
PDB	Protein Data Bank
PDE	cyclic nucleotide phosphodiesterase
PEC	phosphoethanolamine cytidyltransferase
PFAM	protein family database
PG	phosphatidylglycerol
PG/PI-TP	phosphatidylglycerol/phosphatidylinositol transfer protein
PI	phosphatidylinositol
pI	isoelectric point

PIMT	protein-L-isoaspartyl/D-aspartyl O-methyltransferase
PIS	phosphatidylinositol synthase
PKA	protein kinase A
PKC	protein kinase C
PL	phospholipid
PPDE	posterior probability of being differentially expressed
PPI1	peptidyl-prolyl cis-trans isomerase
ppt	part per thousand
Prdx	peroxiredoxin
Ptc	Patched
ROS	reactive oxygen species
rRNA	ribosomal ribonucleic acid
RSA	relative solvent accessibility
SAM	S-adenosylmethionine
Ser	serine
SERT	serotonin transporter
sHSP	small heat-shock protein
SOAT	sterol O-acyltransferase
SOD	superoxide dismutase
SOS1	Salt-Overly-Sensitive 1
SSD	sterol-sensing domain
TPM	transcripts per million
tRNA	transfer ribonucleic acid
V-PPase	vacuolar H^+-transporting pyrophosphatase

Acknowledgments

On a personal note, I need to thank Scott and Sherise Jones for instigating and propagating the Don't BE Afraid message. Together with singers from VOX: A Choir for Social Change and associated friends, you have contributed to set my heart on fire and this has had a compelling impact on my work and on my life in general. You have truly awakened the human being in me. Immense thanks for this.

Thanks to my Haligonian friends for offering weekly oases of fabulousness and fictive escape. Thanks to the hosts Shaun Simpson and James Neish and to the attendees Derek Roger, Ryan Stachejczuk, Justin Chevrette, Angi Garofolo, Graham Bolton, Dustin Wiseman and Conor Meehan. Thanks for your devotion and loving company.

Finally, I also want to thank my friends from Québec for journeying on the long road that leads to the Maritimes multiple times and hand-delivering their affection: Annie Fournier, Marie-Ève Mainguy, Marc-André De La Garde, Fabien Piché and especially Dale Gilbert. Dale, you inspired me to embark on this post-graduate path. Que ferais-je sans vous? Merci infiniment, cher ami.

Not the least, thanks to my lover Ryan Koroscil for your unconditional support and open-mindedness. With you by my side, I have come to grow as a more peaceful man. Approaching the end of my doctoral study, I realize how I could have sunk into an ocean of anxiety without you. Thanks for sticking with me.

Chapter 1 - Introduction

1- Halophilic Protozoa

Hypersaline environments have a global but sporadic worldwide distribution (Javor 1989). They are found in warm or desert climates (*e.g.* hypersaline ponds of Australia) where intense evaporation lowers water content, in very cold regions (*e.g.* Antarctic lakes and sea ice) where freezing leads to salt expulsion from the ice lattice and concentration in the remaining liquid water, or in deep-sea environments (*e.g.* the Mediterranean deep-sea hypersaline basins) where dense hypersaline water originated from the dissolution of geological salt deposits. These environments represent habitats for microorganisms, some of which are adapted to even the highest salinities encountered. Due to their sparse geographical distribution, the simple food webs they harbor and their harsh conditions for biological growth, these systems are of interest for studies of microbial biogeography, ecology and evolution.

The Bacteria and Archaea that inhabit hypersaline environments have been widely studied (Oren 2002a). Comparatively, eukaryotes in these environments have received less attention and were treated simplistically. For example, despite several early studies reporting the presence of diverse unicellular eukaryotes (protists) in high-salt waters (Enz 1879; Butschinsky 1897; Florentin 1899; Entz 1904; Namyslowski 1913; Kirby 1932; Ruinen 1938; Volcani 1944), the traditional view of extremely hypersaline ecosystems designates the chlorophyte algae genus *Dunaliella* as the main, if not the only, primary producer, with these being consumed by the brine shrimp *Artemia salina,* representing the main, if not the only, predatory organism (*e.g.* Pedrós-Alió *et al.* 2000). This view implicitly envisaged no or minor ecological contribution from phagotrophic protists (i.e. protozoa). This was in contrast with marine and freshwater ecosystems where protozoa are often major agents of predation, and specifically of prokaryotic mortality (Azam *et al.* 1983; Salonen and Jokinen 1988; Sherr *et al.* 1989; Vaqué *et al.* 1994). It is only recently

that ecological surveys led to the wider recognition that several distinct evolutionary lineages of eukaryotes, mostly protozoa, are adapted to hypersaline environments and may have significant ecological roles (Hauer and Rogerson 2005).

Modern molecular investigations of microbial communities along salt gradients ranging from the salinity of seawater (~3.5%) up to saturation (>30% salt) have indicated that eukaryotic species richness is high up to around 8-15% salt, sometimes higher than for prokaryotes (Casamayor *et al.* 2002), and rapidly declines at higher salinity levels (Clavero *et al.* 2000; Casamayor *et al.* 2002; Estrada *et al.* 2004; Elloumi *et al.* 2009; Lei *et al.* 2009; Filker *et al.* 2015). This salinity threshold divides low-salt communities dominated by protists that tolerate rather than prefer high salt concentrations from protists that require high salt to grow, and that are therefore more active in habitats with higher salinities. This transition boundary is proposed as a critical physiological point above which specific adaptations are required for biological activities to be sustainable (Filker *et al.* 2015). These adaptations might include amino acid bias in proteins exposed to high salt (Frolow *et al.* 1996; Paul *et al.* 2008), mechanisms to adjust the intracellular concentration of solutes given the high extracellular osmotic strength (Borowitzka and Brown 1974; Galinski 1995; Oren 2002b), tight regulation of ion homeostasis (Niu *et al.* 1995; Gorjan and Plemenitaš 2006), salt-dependent adjustment of plasma membrane fluidity (Russell 1989; Turk *et al.* 2004; Turk *et al.* 2007; Turk *et al.* 2011) and a high capacity to manage the detriments related to oxidative stress (Panda and Das 2005; Petrovič 2006; Tammam *et al.* 2011).

Concordantly, phylogenetic analyses revealed that protists are typically evolutionarily segregated into freshwater and marine groups, and that adaptive transitions from habitats with strikingly different salinities have occurred relatively infrequently during the course of evolution (Logares *et al.* 2009). Furthermore, protists isolated from hypersaline habitats, and that grow well at salinity well above that of seawater, form distinct phylogenetic clades more closely related to marine forms and not to freshwater groups, supporting the existence of divergent protist taxa that have been adapted to hypersaline conditions for a long evolutionary period

(Park and Simpson 2010). This strong relationship between phylogeny and habitat salinity further suggests that osmotic pressure and ion gradients probably represent important factors in shaping protistan diversity and thus communities.

Halophiles and halotolerant organisms are defined based on their optimal salinity for growth and their obligate requirement for salt. One of the most widely used definitions (Kushner 1978) designates organisms that do not need salt to divide as being 'halotolerant' (extremely halotolerant if able to grow in >15% salt), while 'halophiles' require salt and grow best at salinities >3%. The latter group is further subdivided into categories depending on the optimal salinity for growth: 'moderate halophiles' growing well at 3-15% salt and 'extreme halophiles' at 15-30%. Organisms for which the salinity range for growth overlaps the latter categories (*i.e.* they grow well between 9-24% salt) are considered 'borderline extreme halophiles'. This thesis focuses especially on extreme (including borderline extreme) halophiles that are specialized phagotrophs, *i.e.* free-living protozoa. The past years have seen a substantial amount of work dedicated to characterizing halophilic fungi, which are osmotrophs (important review and research articles include: Petrovič *et al.* 2002; Butinar, Santos *et al.* 2005; Butinar, Sonjak *et al.* 2005; Cantrell *et al.* 2006; Plemenitaš *et al.* 2008; Gunde-Cimerman *et al.* 2009; Plemenitaš *et al.* 2014). The literature also contains several reports on phototrophic *Dunaliella* spp. (Cowan *et al.* 1992; Chen and Jiang 2009b; Oren 2014; Liu *et al.* 2015). While halophilic protozoa have been far less well studied, knowledge of their diversity and ecological distribution has advanced substantially in the last decade in particular, setting the stage for a broader understanding of the biology of these neglected extremophiles.

2- Hypersaline habitats and their protist communities

2.1- Solar Salterns

Solar salterns are one of the most favored hypersaline environments for studying the associated microbiota, especially since they allow for geographically-restricted investigations of controlled sites along salinity gradients. These man-made settings, designed primarily for the manufacture of sodium chloride, consist of series of successive shallow ponds in which seawater-derived waters are subject to evaporation, leading to a system of water masses with gradually incremented salinity, from seawater up to saturated brine. Since salterns worldwide use seawater as input, the salt composition of their water systems is similar. This feature has been inferred to drive similar microbial community composition in geographically remote locations (Oren 1994).

Salterns have received considerable attention from microbiologists partly because the microbiota that they contain influence salt production. For example, excretion of polysaccharide during unfavorable growth affects salt precipitation by binding certain ions during water evaporation (Javor 2002). Good saltern management also includes limiting the growth of *Fabrea salina* (a large ciliate that cannot be ingested by the brine shrimp *Artemia*) in intermediate salinity ponds, as this ultimately leads to microbial community imbalance and excessive flow of organic matter to the downstream ponds (Davis 2000). As a result of this focused research, about a third of protists observed in hypersaline environments up to 2005 were identified in salterns (Hauer and Rogerson 2005).

Microbial eukaryotes have been studied by microscopic examinations of samples from various salterns of the Iberian peninsula (Esteve *et al.* 1992), the saltpans of the Alto Guadalquivir region, Spain (Galotti *et al.* 2014), the Megalon Embolon saltworks, Greece (Dolapsakis *et al.* 2005), the saltworks of Sfax, Tunisia (Ayadi *et al.* 2002; Elloumi *et al.* 2006; Elloumi *et al.* 2009), the Port Fouad saltworks, Egypt (Taher *et al.* 1995), and the Hwasung salterns, Korea (Lei *et al.* 2009). Both molecular techniques (clone libraries or denaturing gradient gel electrophoresis

(DGGE) of the 18S ribosomal RNA (rRNA) gene or the plastid 16S rRNA gene) and microscopic investigations were used to characterize eukaryotic communities in the Bras del Port salterns, Spain (Pedrós-Alió *et al.* 2000; Casamayor *et al.* 2002; Estrada *et al.* 2004), the Cabo Rojo salterns, Puerto Rico (Casillas-Martinez *et al.* 2005; Cantrell *et al.* 2006), and the Exportadora de Sal saltworks, Mexico (Clavero *et al.* 2000; Nübel *et al.* 2000; Spear *et al.* 2003; Feazel *et al.* 2008). Attempts at amplifying the 18S rRNA gene were unsuccessful using samples from the Mara salterns, Peruvian Andes (Maturrano *et al.* 2006). Finally, high-throughput sequencing of the 18S rRNA gene V4 region was employed to reveal protistan communities of the Ria Formosa salterns, Portugal (Filker *et al.* 2015). A metagenomic analysis of samples from the Exportadora de Sal saltworks also acknowledged the presence of eukaryotes but without further details (Kunin *et al.* 2008). Several halophilic protists described using ultrastructural and/or molecular methods (see section 4) were isolated from solar salterns worldwide including in Seosin, Korea (Park *et al.* 2006; Park *et al.* 2007; Park *et al.* 2009), in San Diego, USA (Park and Simpson 2011), in Portugal (Foissner *et al.* 2014), in New Zealand and in Russia (Kirby *et al.* 2015).

For most of these studies, emphasis has been placed on ciliates or the phytoplankton (especially on diatoms and chlorophytes) with little detail about other protists (*e.g.* Dolapsakis et al. 2005; Elloumi et al. 2006; Lei et al. 2009), and ciliates were often found to dominate protist assemblages in the saltern system. Ciliates are conspicuous under the microscope, leading to a relative ease of identification. During molecular analyses, ciliates might be also advantaged relative to other protists that are typically smaller since larger cells are more likely to contain more rRNA transcripts in cases where RNA is used as input material to construct libraries. Ciliates are also famous for containing lots of DNA since they have a macronucleus as well as a micronucleus harboring rRNA gene copy numbers that can vary according to growth conditions (Prescott 1994). Acknowledging these biases, Filker *et al.* (2015) noted that 65.7% of sequence reads were assignable to Ciliophora in a 12%-salt pond of the Ria Formosa salterns, while fluorescent *in situ* hybridization (FISH) analysis of the same sample generated strikingly lower

abundance results, with ciliates accounting for 3.5% of protist cells. However, ciliates were not detected at 38% salt by FISH but represented 0.2% of total reads, indicating that deep sequencing techniques are more sensitive that microscopy.

A decrease in ciliate species richness (*e.g.* from 32 to 13 operational taxonomic units (OTUs) at 12% and 38% respectively) was encountered with increasing salinity (Filker *et al.* 2015). Moreover, distinctions in species composition were observed in space and time. For example, no stichotrichs nor phylopharyngeans were detected in January or May 2003 in Sfax (Elloumi *et al.* 2006) but were present in March 2012 in Ria Formosa (Filker *et al.* 2015) and in April 2001 in Hwasung salterns (Lei *et al.* 2009), while species of the class Karyorelictea were detected uniquely in Sfax. On the contrary, the species *Fabrea salina* occupied substantial fractions of the ciliate community at all sites with salinities around 12-15%. Ciliate species related to *Balanion*, *Cladotrichia* and *Condylostoma* were detected in ponds with highest salinities and probably represent yet undescribed halophilic protists (Filker *et al.* 2015).

Apart from the chlorophyte *Dunaliella salina,* which typically dominates the phytoplankton of saltpans with the highest salinities, other microalgae (including diatoms and phototrophic dinoflagellates and cryptophytes) tend to be major primary producers in lower salinities and to disappear above 15% salt, suggesting they are not adapted to high salt (Taher *et al.* 1995; Estrada *et al.* 2004; Elloumi *et al.* 2009). However, deep sequencing recorded sequences related to the diatom orders Bacillariaceae and Naviculaceae at 38% salt, accounting for 13.2% of total OTUs in this saltpan of the Ria Formosa salterns (Filker *et al.* 2015).

During the past years, increasing attention has been paid to heterotrophic nanoflagellates (2-20 μm in size). These were reported to represent for example between 17-25% of the total nanoplanktonic abundance of saltpans >30% salt in the Sfax saltwork (Elloumi *et al.* 2009), with abundance reaching 3×10^4 nanoflagellates/mL at salt concentration close to 30% in Bras del Port salterns

(Pedrós-Alió *et al.* 2000). Their role in the food web has been shown to be substantial, as discussed in section 3 (Park *et al.* 2003).

2.2- Salt Lakes, Lagoons and Natural Ponds

Natural water masses with extreme hypersalinity are rather rare. Similarly to investigations of low-salt ponds of salterns, studies of lakes and enclosed embayments with moderate hypersalinity (<10% salt) commonly revealed the presence of communities similar to those of coastal and marine systems. These include relatively recent investigations of the salt ponds of the Atacama Desert, Chile (Stivaletta *et al.* 2011; Farías *et al.* 2014; Rasuk *et al.* 2014), the Laguna Madre, USA (Buskey *et al.* 1998; Buskey *et al.* 2001), the Araruama Lagoon, Brazil (Debenay *et al.* 2001), and the microbial mats of the Hamelin Pool, Australia (Al-Qassab *et al.* 2002; Allen *et al.* 2009; Edgcomb and Bernhard 2013). Even samples of higher salinity (>10%) can contain organisms that grow better in seawater but that display very broad salt tolerance. For example, several different amoebae isolated from a hypersaline pond (16% salt) near the Salton Sea and established in high-salt cultures could be grown without acclimatization to low salinity (3.2%; Rogerson and Hauer 2002). Therefore, identification of extreme halophiles from descriptive surveys without a culturing component is difficult.

Nonetheless, the existence of extreme halophilic protists, and their importance in hypersaline environments, is getting increasing support from environmental inquiries. These studies include investigation of Australian hypersaline waters sampled from Lake Tyrrell (Heidelberg *et al.* 2013), the Hutt Lagoon (Post *et al.* 1983), and from various sites in the Shark Bay area (Patterson and Simpson 1996). Protists were also examined in microbial mats and water samples from Solar Lake, Egypt (Cohen *et al.* 1977; Wilbert and Kahan 1981; Minz *et al.* 1999), in the benthic area of salt lakes in Canada, Egypt and Australia (Wilbert 1995), in the water column of Lake Ursu, Romania (Máthé *et al.* 2014), and in the Fuente de Piedra saline lake, Spain (García and Niell 1993). In addition, protist communities were characterized in cold saline lakes of the Tibetan Plateau (Wu *et al.*

2009) and Antarctica (Perriss and Laybourn-Parry 1997; Laybourn-Parry *et al.* 2002; Yau *et al.* 2013), and from Antarctic sea-ice brine (Stoecker *et al.* 1997). Protists have also been anecdotally reported from the north arm of the Great Salt Lake, USA (Post 1977).

Again, samples of high salt concentration (>15% salt) are typically a source of novelty. Triadó-Margarit and Casamayor (2013) investigated the genetic diversity of microbial eukaryotes from several coastal and inland saline habitats to determine that terrestrial water bodies, especially hypersaline habitats, contained unreported taxa abundantly, with between 30-40% of sequences with identity <97% to sequences available in databases. In sites with salinity >15%, where sequences displaying the most extreme novelty were found, novel phylotypes were related to Choanoflagellida, Centroheliozoa, Bicosoecida and Colpodellidae.

The shallow water masses of Australia are interesting case examples to illustrate the presence of diverse halophilic protists in natural hypersaline environments. In the Hutt Lagoon, where salinity rarely falls under 18%, Post *et al.* (1983) reported 14 ciliates, ten flagellates and four amoebae using microscopy. Several of these species were not observed at salinity under 16% and were only reported from hypersaline habitats, for example the ciliate *Cladotrichia koltzowskii*. The authors suggested that these taxa might be unique to hypersaline niches, but they could not exclude the possibility that these species were high-salt-adapted marine protists that can tolerate a wide range of salinities (*i.e.* euryhaline). Importantly, protists were routinely identified at the surface of benthic microbial mats, suggesting they might occupy higher links of the microbial food web in these systems.

Patterson and Simpson (1996) reported the diversity of heterotrophic nanoflagellates in sediments from shallow waters of the Shark Bay area. As mentioned previously, sites with lower salinities (4.1% and 6% salt) contained species observed in marine environments, and their presence decreased with increasing salinity. By contrast, none of the flagellates observed in a saturated brine

puddle were previously reported from the ocean and two of them, *Pleurostomum flabellatum* and *Palustrimonas yorkeensis*, have been detected previously exclusively from hypersaline environments. More species had never been reported before, like *Ancyromonas melba, Bodo cygnus, Pleurostomum turgidum* and *Rhynchobodo simius*. One flagellate, identified as *Bodo saltans*, was observed in all samples suggesting it was extremely euryhaline, but it is probable that this species described on the basis of morphological features (morphospecies) was in fact several individual species (Park *et al.* 2006). This highlights the limitations related to microscopic examinations and indicates that the combination of molecular, microscopy and culture methods is required to accurately describe halophilic protists.

Finally, Lake Tyrrell, located in semi-arid northwestern Victoria, Australia, contains shallow waters (~50 cm deep), with salinity ranging from 25-30% in winter, that almost completely evaporate in summer, leaving residual brine on top of halite crust. Investigating this system, Heidelberg *et al.* (2013) determined that colpodellids (heterotrophic nanoflagellates related to Apicomplexa) dominated both 18S rRNA gene clone libraries (84% to 95% of total clones) and microscopic counts (85% of total counts) from water samples. Sequences for this group formed a single monophyletic group sister to *Alphamonas/Colpodella edax* and *Chromerida* sp.

These studies all indicate that hypersaline habitats represent rich sources of novel extreme halophilic protists that most probably play important ecological roles in these systems. On-going effort at isolating and characterizing these organisms is therefore crucial.

2.3 - Deep-Sea Hypersaline Basins

More than 3 km deep at the bottom of the Eastern Mediterranean Sea reside a series of deep hypersaline anoxic basins (DHABs) containing dense hypersaline waters that resulted from the dissolution of geological salt deposits, the Messinian evaporites (Cita 2006). Under the seawater/brine interface, salinity quickly increases in a halocline to reach >30%, and oxygen level drops to reach anoxic conditions, in a ~100 m-deep brine layer (maximum of 500 m in Bannock Basin).

Studying microbial communities in these remote environments is extremely challenging due to several factors, including i) the steep and thin haloclines (less than a couple of meters; Stock et al. 2013) that require specific high-precision sampling protocols (see Daffonchio et al. 2006), ii) the difficulty during fieldwork in establishing cultures of microbes adapted to high pressure and anoxic conditions, and iii) the presence of residual biological material that sinks and remains preserved in these basins, obscuring the validity of some molecular data (Danovaro *et al.* 2005). Despite these limitations, active microbial communities including protozoa have been detected—especially by recovery of RNA (a proxy for genetic activity) but also by imaging of apparently viable cells in a few cases—in Discovery, Urania, Medee, Tyro, Thetis, L'Atalante, and Bannock Basins, and in a basin near the 'Kryos Deep' depression (Alexander *et al.* 2009; Edgcomb *et al.* 2009; Edgcomb *et al.* 2011; Ferrer *et al.* 2012; Stock *et al.* 2012; Filker *et al.* 2013; Stock *et al.* 2013; Bernhard *et al.* 2014; Pachiadaki *et al.* 2014).

Compared to the deep oligotrophic waters of the Mediterranean Sea, the waters of the DHABs can contain higher cell numbers. Protist counts range from 6.0×10^3 cells/L in the halocline of the Thetis Basin (Stock *et al.* 2012) to 2.7×10^5 cells/L in the Urania halocline (Edgcomb *et al.* 2011), while in the water column above the basins counts range from $1.5 \sim 22.3 \times 10^4$ cells/L (Bernhard *et al.* 2014). Concordantly, metatranscriptomic analysis of the Thetis basin halocline indicated that eukaryotic actin and tubulin transcripts were four times higher in the upper halocline (7.0-16.3% salt) compared to 3.87% salinity samples taken 1,000 m above the interface (Pachiadaki *et al.* 2014). However, transcript abundance for these proteins decreased to very low levels in the lower halocline (21.4-27.6% salt), indicating restricted growth in the deeper, most saline layers. Cell counts in the sediments of the Urania, Discovery and L'Atalante basins were higher by two orders of magnitude compared to the water columns of these basins, and were also more abundant compared to the surrounding benthos of the sea (Bernhard *et al.* 2014). In the deep-sea haloclines and brines, even though the species richness is lower compared to surface seawater (Alexander *et al.* 2009), a higher richness was

observed than in the deep-sea marine waters overlying the DHABs (Edgcomb *et al.* 2009).

Analysis of 18S rRNA gene sequences indicated that populations of ciliates dominated the brine community of several DHABs. In the Thetis lower halocline (salinity of ~25%), they represented 40% of all reads (Pachiadaki *et al.* 2014). In the Medee Basin brine (salinity of 32%), 97% of all ciliate amplicons were distantly related to the genus *Anoplophrya* (80%-89% identity), while 64% and 45% of ciliate sequences were related to *Strombidinium* in the brines of Thetis (salinity of 35%) and Tyro (salinity of 32%) basins, respectively (Stock *et al.* 2013)*.* A substantial fraction of the taxonomic composition was also often attributable to ciliates, with ciliates representing 20% and 45% of phylotypes in the brines from Thetis (35% salt) and L'Atalante (37% salt) basins, respectively (Alexander *et al.* 2009; Stock *et al.* 2012). Detection of sequences of a halophilic clade of *Trimyema* in these DHABs and in inland hypersaline habitats suggests that the deep-sea communities share a historical link with surface hypersaline waters rather than have evolved *de novo in situ* (Park and Simpson 2015).

In addition, fungal sequences indicate a successful colonization of these DHABs where they represented 39% of all sequencing reads in the Thetis brine (Pachiadaki *et al.* 2014), 37% of all phylotypes in the Thetis lower halocline (Stock *et al.* 2012) and ~90% of all sequences in the sediments within the L'Atalante lower halocline (10% salt; Bernhard *et al.* 2014). In the L'Atalante lower halocline (37% salt), acanthoecid choanoflagellates also notably represented 10% of sequences (Alexander *et al.* 2009).

As for surface hypersaline environments, the DHABs contained several sequences remarkably divergent compared to deposited sequences, including some distantly affiliated with Bicosoecida, and one ciliate sequence only 86% identical to other ciliate species (Alexander *et al.* 2009; Stock *et al.* 2012).

Analysis of 18S rRNA gene sequences and canonical correspondence analysis indicated that geographic distance and temporal variation were negligible forces in

shaping community composition in these environments, but that chemical parameters had a significant influence (Alexander *et al.* 2009; Edgcomb *et al.* 2009; Filker *et al.* 2013; Stock *et al.* 2013). The chemistry of the brines is variable from basin to basin due to exhumation and submarine dissolution of varying layers of the Messinian deposits (Cita 2006). For example, in the Discovery Basin brine, concentrations of magnesium reach 5 M compared to 300-650 mM in the L'Atalante, Bannock and Urania basins, and the concentration of sodium is just ~70 mM while it ranges from 3.5 to 4.7 M in the other DHABs (van der Wielen *et al.* 2005). As a result, communities are significantly dissimilar between DHABs. The steep chemical gradient through the halocline down to the brine of individual DHABs also drives community composition. In addition to magnesium and sodium content, concentrations of sulfate, ammonium and oxygen were correlated with community structure variations (Alexander *et al.* 2009; Edgcomb *et al.* 2009; Filker *et al.* 2013; Stock *et al.* 2013). This influence of salt composition on community structure is concordant with the measured growth dependence of halophilic protists on ionic factors like the $Mg^{2+}:Ca^{2+}$ ratio (Park 2012). In the sediments of the Urania, Discovery and L'Atalante Basins, concentrations of HS^-, Mg^{2+} and SO_4^{2-} also seemed to influence protist density (Bernhard *et al.* 2014).

3- Ecological Roles of Protozoa in Hypersaline Habitats

As bacterivores, halophilic protozoa were shown to control prokaryotic abundance in two Spanish and one Korean saltern systems. In La Trinitat saltern, in a 15% salinity saltpan, prokaryotic loss by bacterivory was ~30 times higher than loss by viral lysis (Guixa-Boixareu *et al.* 1996). Bacterivory was maximal at salinities between 6% and 15%, accounting for ~100% of the prokaryotic production. Similarly, in Bras del Port saltern, highest bacterivory values were recorded for salinities between 5% and 20%, but this corresponded to <40% of prokaryotic production (Pedrós-Alió *et al.* 2000). In both systems, no bacterivory was detected

above 20% salt. As mentioned previously, earlier accounts stipulated that bacterivores were generally not observed at salinities >14% (Por 1980).

Contrastingly, Park *et al.* (2003) measured bacterivory in Korean salterns exceeding 30% salt, where grazing rates ranged from 1.4-13.0 × 10^8 prokaryotes eaten per liter per hour, corresponding to 25-358% of prokaryotic production depending on the month. In contrast to the Spanish salterns, no ciliates were observed and grazers were discovered to be nanoflagellates present in abundance (7-28 × 10^6 cells/L).

In addition to nanoflagellates and ciliates, amoebae were also observed to be grazers, for example on cyanobacteria in surface mud samples of a salt pond at Eilat (Hauer and Rogerson 2005). Moreover, some protozoa, like colpodellids and ciliates, were observed to consume other eukaryotes, especially *Dunaliella* algae, in environmental or crude samples (Post *et al.* 1983; Simpson and Patterson 1996). For example, Heidelberg *et al.* (2013) microscopically observed ingested *Dunaliella* in colpodellid cells sampled from Lake Tyrrell and inferred that these phagotrophs, together with ciliates related to *Trimyema koreanum* and *Fabrea salina* detected during the same study, probably exert considerable grazing pressure on *Dunaliella* and the prokaryotic communities in high salt systems. Since other predators like zooplankton are absent at high salt, these results suggest that phagotrophic protists might contribute to the significant ammonium recycling activity implied by Joint *et al.* (2002).

4- Described Isolates of Halophilic Protozoa

Most major groups of eukaryotes contain protozoa that have been reported from hypersaline environments (see Tables 1-3 in Hauer and Rogerson 2005 and Suppl. Table 1 in Park et al. 2009). However, only a few isolates (4 heteroloboseans, 1 bicosoecid and 2 ciliates) have been characterized using modern methods, including rRNA gene sequencing, and also confirmed to be (borderline) extreme

halophiles by growth experiments. This illustrates the need for more descriptive studies of halophilic protozoa. On the basis of morphological data and molecular phylogenies these isolates were assigned to new species and usually to new/different genera than marine forms, consistent with the idea that halophilic protozoa represent independent evolutionary lineages that adapted to hypersaline environments a long time ago.

The Heterolobosea (supergroup Excavata) comprises flagellates, amoebae and amoeboflagellates (that can transition between both stages and that commonly form cysts). They are typically found in freshwater, soil and marine environments, however several heteroloboseans are adapted to extreme conditions, including high temperature (*e.g. Oramoeba fumarolia*; De Jonckheere *et al.* 2011), and low pH (*e.g. Tetramitus thermacidophilus*; Baumgartner *et al.* 2009). Several are extreme or borderline extreme halophiles: *Pharyngomonas kirbyi* (Park and Simpson 2011), *Pleurostomum flabellatum* (Park *et al.* 2007), *Euplaesiobystra hypersalinica* and *Tulamoeba peronaphora* (Park *et al.* 2009). The latter were all isolated from solar salterns, or natural salt lakes or puddles with salinity of ~30%. All grow optimally at >12% salt in laboratory culture (remarkably at 30% in the case of *P. flabellatum*), indicating hypersaline habitats are most likely their ecological niche.

These species are not closely related to each other and, with the exception of *T. peronaphora* and *P. flabellatum*, they fall into separate phylogenetic clades, indicating that adaptation to high salt happened independently at least three times during the evolutionary history of the Heterolobosea. Interestingly, during the past years, several isolates with lower salt tolerance were discovered to be rather close relatives to the extreme halophiles. *Pharyngomonas turkanaensis* was isolated from Lake Turkana in Africa (0.4% salt), and can grow *in vitro* at salinity between 1.5-10%, with an optimum at <3% (Park and Simpson 2016). It is therefore at most a moderate halophile, and since the common ancestor of *Pharyngomonas* spp. was probably an obligate halophile (Harding *et al.* 2013), *P. turkanaensis* likely reverted to preferred growth at lower salinity (Park and Simpson 2016). The most closely known relative of *P. turkanensis*, *Pharyngomonas* sp. strain RL, is (barely) capable of

growth at lower salinities (3.5%; Plotnikov *et al.* 2015) although it can grow up to 25% salt (Harding *et al.* 2013), indicating varying levels of salt tolerance in species of the genus. In addition, strains of *Tulamoeba bucina* were also shown to be moderate halophiles, with a remarkable salinity range (growth at 3.5-22.5% salinity), while branching inside the clade of obligate halophiles that encompasses *Tulamoeba* and *Pleurostomum* (Kirby *et al.* 2015). Greater taxon sampling and additional descriptions will provide better insights into possible evolutionary paths toward salt adaptation and the required molecular mechanisms in the Heterolobosea.

  Halocafeteria seosinensis strain EHF34 (Bicosoecida, Stramenopila) was isolated from a 30% salt Korean saltern and grows optimally at 15% salt (Park *et al.* 2006). It is a small flagellate that is related to the marine taxa *Cafeteria* and *Caecitellus*. Its closest known relatives are actually undescribed moderate halophiles with lower salt optimum (~5-7.5%) and range (1.5-15%) for growth, hinting at a progressive adaptation to high salt in the lineage leading to *Halocafeteria* (Park and Simpson 2010). Morphospecies of *Halocafeteria* and 18S rRNA gene sequences 97.5-100% identical to strain EHF34 sequence were detected from worldwide locations (Australia, North America and Europe), indicating a wide geographic distribution on a coarse scale (Park and Simpson 2015). Molecular data was generated from *H. seosinensis* and forms the main substance of this thesis (see section 5).

  Ciliates (Alveolata) have been commonly reported from hypersaline habitats using microscopic and environmental molecular techniques, as described above, but only a couple of (borderline) extremely halophilic isolates have been characterized as laboratory cultures: *Trimyema koreanum* (Cho *et al.* 2008) and *Platynematum salinarum* (Foissner *et al.* 2014). *Trimyema koreanum* belongs to the class Plagiopylea and grows over a range of salinity between 15-30% with an optimum of 22.5% salt (Cho *et al.* 2008). It is an anaerobe that can tolerate oxic conditions for three to four days. It was shown to be able to ingest *Dunaliella* as a food source, in addition to prokaryotes (Cho *et al.* 2008). Sequences of *T. koreanum* 18S rRNA genes isolated from a wide range of geographic locations grouped concordantly to their

sampling sites, suggesting the possibility for restricted biogeographical distributions for subclades of this halophile (Park and Simpson 2015). *Platynematum salinarum,* which belongs to an entirely different group of ciliates (Oligohymenophorea), was isolated from a Portuguese solar saltern of 12% salinity. The studied strain grows over a range of 14-30% in laboratory culture, with highest growth rate at 14% salt (Foissner *et al.* 2014).

5- Evolution of Halophilic Protozoa

During the past years, a picture regarding the ecology and distribution of extremely halophilic protozoa has begun to emerge. On the contrary, basically nothing is known about their evolution, and specifically about the molecular features that contributed to their adaptation to a halophilic lifestyle. Yet their patchy phylogenetic distribution in the tree of eukaryotes indicates that adaptation to hypersaline environments has happened in several independent lineages and illustrates that protist evolvability might be as great as in bacteria and archaea. This thesis is motivated by the need for acquisition of fundamental information regarding protist molecular evolution and the genetic features that allow halophilic protozoa to survive in very high salt conditions. It examines *H. seosinensis* as a case example through comparative sequence analyses (chapter 2, also including the halophilic heterolobosean *P. kirbyi*) and expression measurements of salt-responsive genes (chapter 3).

In chapter 2, I establish that the cytosols of the halophilic protists *H. seosinensis* and *P. kirbyi* probably do not contain molar concentrations of inorganic ions, but that organic solutes instead likely accumulate to equilibrate the intracellular osmotic strength with that of the medium. On one hand, this is based on the detection of an amino acid bias leading to increased hydrophilicity in the cytoplasmic proteomes of halophilic protists, combined with the absence of an acidic signature commonly detectable in proteomes of organisms that accumulate high levels of cytosolic inorganic ions. On the other hand, I infer that organic solutes

potentially build up in the cytosol based on the salt-dependent expression of genes involved in organic osmolyte synthesis and transport.

In chapter 3, I report a comprehensive analysis of genes up-regulated at high salt (30%) compared to optimal salinity for growth (15%) in *H. seosinensis.* I therein identify several aspects of cellular biology that might require modifications of the transcriptional program in order to allow protozoa to live under hypersaline conditions. For *H. seosinensis*, the transcriptional response to high salt included the differential expression of genes involved in ion homeostasis, signal transduction, stress management, lipid remodeling and carbohydrate metabolism. In this chapter, I also examine genes potentially involved in duplication and/or lateral gene transfer events in order to gain insights into the evolutionary processes that might have favored the adaptation to high salt in the lineage leading to *H. seosinensis*.

Chapter 2 - Osmoadaptative Strategy and its Molecular Signature in Obligately Halophilic Heterotrophic Protists [1]

1- Introduction

Extremely halophilic microbes are adapted to hypersaline conditions and require salt concentrations well above those of seawater in order to grow. They typically have optimal reproductive rates at ~4× the salinity of seawater (Hauer and Rogerson 2005; Oren 2008). Some halophiles can actually sustain growth in saturating salt concentrations, and some even show their optimal growth in near-saturated media (*e.g.* Park *et al.* 2007). Research on extreme halophiles has focused for decades almost entirely on Archaea and Bacteria and has yielded substantial information in terms of adaptations of their proteins and cellular metabolism and physiology (Oren 2002a).

In order to stay hydrated in hypersaline conditions, halophiles equilibrate the osmotic strength of their cytoplasm with the extracellular medium by accumulating solutes, either inorganic or organic. For example, the 'salt-in' haloarchaea accumulate molar concentrations of potassium (Oren 2002b). These have highly acidic proteomes, resulting from an abundance of negatively charged aspartate and glutamate residues and a depletion of positively charged lysine residues at the protein surface (Frolow *et al.* 1996; Paul *et al.* 2008). X-ray crystallography studies have demonstrated that this negative net charge is neutralized by water molecules and thus increases protein solubility as well as preventing protein aggregation in hypersaline conditions (Elcock and McCammon 1998; Richard *et al.* 2000). Halophilic proteins are also depleted of hydrophobic amino acids, and this is often offset by a higher content of borderline hydrophilic-hydrophobic residues like serine

---

[1] This chapter was published as a research article in Genome Biology and Evolution and reproduced by permission of Oxford University Press: Harding, T., Brown, M.W., Simpson, A.G.B. and Roger, A.J. (2016) Osmoadaptive strategy and its molecular signature in halophilic heterotrophic protists. Genome Biol Evol, 8(7): 2241-2258.

and threonine (Lanyi 1974). High salt increases the hydrophobic effect, and the low hydrophobicity of halophilic proteins possibly allows them to avoid overly rigid folded conformations.

On the other hand, the proteomes of halophiles that use organic solutes as their main osmolytes ('salt-out' organisms) are not enriched in highly acidic proteins, although they typically produce extracellular proteins that are very acidic and less hydrophobic compared to mesophilic counterparts (*e.g.* Coronado *et al.* 2000; Oren *et al.* 2005). Organic osmolytes are strong 'water structure-formers' that are excluded from protein hydration shells, and thus prevent cell dehydration while being compatible with metabolism (Bolen and Baskakov 2001). They are uncharged or zwitterionic low-molecular weight molecules and include some amino acids and their derivatives (glutamate, aspartate, proline, ectoine), polyols (myo-inositol, mannitol, glycerol), sugars (trehalose, sucrose) and betaines (Galinski 1995). To save energy, osmolytes are preferentially imported from the extracellular environment if available, otherwise they are synthesized *de novo* (Oren 1999). Accumulation of organic osmolytes is not limited to halophiles but rather represents a widespread long-term osmotic balance response in most organisms. By contrast, massive accumulation of inorganic osmolytes is restricted to a few groups, including the *Halobacteria* (*i.e.* haloarchaea), the *Halanaeorobiales* (Rengpipat *et al.* 1988), *Salinibacter ruber* (Oren *et al.* 2002) and *Halorhodospira halophila* (Deole *et al.* 2013).

Recently, the view that intracellular salt content is absolutely coupled with proteome acidity has been challenged (Oren 2013). For example, the proteobacterium *Halorhodospira halophila* employs K^+ as its main osmolyte and expresses a highly acidic proteome, but can grow at salinity as low as 3.5%, with intracellular K^+ concentrations varying from 0.4-2.1M as salinity ranges from 5-35% (Deole *et al.* 2013). This shows that acidic proteins can function in cytosolic K^+ concentrations experienced by mesophiles (*e.g. Escherichia coli*), and contradicts the hypothesis that halophilic proteins contain an excess of negative surface charges to establish stabilizing interactions with K^+. It is suggested instead that the high

hydration state of ionized glutamate and aspartate side chains contributes to protein solubility (Deole *et al.* 2013). The use of a mixture of inorganic and organic osmolytes has also been observed, as in the halophilic archaeon *Haladaptatus paucihalophilus* which accumulates trehalose and glycine betaine as compatible solutes while keeping a stable intracellular K^+ content that is 17× higher than *E. coli*, but 0.3-0.6× that of *Halobacterium salinarum* (Youssef *et al.* 2014). As expected, *H. paucihalophilus* possesses a highly acidic proteome. More puzzling, however, are the cases of the Halanaerobiales *Halanaerobium praevalens*, '*Halanaerobium hydrogeniformans*' and *Halothermothrix orenii* that accumulate high levels of K^+ but apparently do not have an acidic signature in their proteomes (Mavromatis *et al.* 2009; Elevi Bardavid and Oren 2012). These studies show that proteome acidity is, at best, strongly suggestive of osmoadaptive strategy, rather than a perfect diagnostic.

Knowledge regarding salt adaptation in halophilic and halotolerant eukaryotic microbes is restricted to species of the chlorophycean alga *Dunaliella,* and certain yeasts, like *Hortaea werneckii, Debaryomyces hansenii, Wallemia ichthyophaga* and *Aureobasidium pullulans* (Chen and Jiang 2009a; Gunde-Cimerman *et al.* 2009). The yeasts *D. hansenii* and *A. pullulans* are considered halotolerant and can grow without salt. In fact, the former grows optimally at salinity around 3-6%, and the latter grows better without added salt (Kogej *et al.* 2005; Gunde-Cimerman *et al.* 2009). In contrast, *W. ichthyophaga* is an extreme halophile that cannot grow without salt and grows optimally at ~21-27% salinity (Gunde-Cimerman *et al.* 2009). These organisms are 'salt-out' strategists that accumulate glycerol as the main compatible solute. Glycerol is one of the cheapest and simplest osmolytes to produce, but is rarely found in other halophiles studied to date, possibly because of its high diffusion rate through standard cell membranes (Oren 1999). Yeasts and *Dunaliella* have developed mechanisms to increase glycerol retention, potentially lowering permeability by adjusting the membrane sterol content and lipid composition, or cell wall melanin level (Sheffer *et al.* 1986; Kogej *et al.* 2006; Gunde-Cimerman *et al.* 2009). Yeasts also use other polyols like erythritol, arabitol and mannitol, with the

composition of the compatible solute mix depending on salinity (Hohmann 2002). Nonetheless, the halotolerant yeast *D. hansenii* has a higher intracellular sodium content than *Saccharomyces cerevisiae* (Prista *et al.* 1997), *A. pullulans* and *H. werneckii* (Kogej *et al.* 2005), suggesting that sodium ions can substantially contribute to osmotic adjustment along with glycerol in this species.

Compared to halophilic archaea, bacteria, the alga *Dunaliella* and fungi, virtually nothing is known about halophilic heterotrophic protists, *i.e.* the ecological guild of 'protozoa'. Several evolutionarily distinct groups of halophilic protozoa have been recorded in hypersaline environments (Hauer and Rogerson 2005; Foissner *et al.* 2014; Stoeck *et al.* 2014; Park and Simpson 2015), and they represent potentially important grazers of prokaryotes in these habitats (Park *et al.* 2003). Several species show high minimum and optimum salinities for growth under laboratory conditions (Park *et al.* 2006; Cho *et al.* 2008; Fiore-Donno *et al.* 2008; Foissner *et al.* 2014), and based on this, the 'salt-in' strategy has been speculatively suggested for at least one species (Foissner *et al.* 2014). Here we present a molecular examination of two protozoa: *Halocafeteria seosinensis* strain EHF34 (Park *et al.* 2006) and *Pharyngomonas kirbyi* strain AS12B (Park and Simpson 2011). Both are obligate halophiles that cannot grow at salinities <7.5% under laboratory conditions. The AS12B strain of *P. kirbyi* grows optimally at a slightly lower salt concentration than *H. seosinensis* EHF34 (12% *vs.* 15% respectively). Concordantly, *H. seosinensis* EHF34 can still divide in media close to saturation (at 30% salt) while *P. kirbyi* AS12B has a maximum salt concentration for growth around 25% (Park *et al.* 2006; Park and Simpson 2011).

As a first investigation of the adaptations of these protozoa to high salt conditions, we aimed to gather evidence about their molecular adaptations, and strategy for osmotic adjustment. Since they grow in mixed cultures with food prokaryotes, direct measurement of the intracellular content in these protists, although required, is technically challenging. Instead, as a first investigation, we sequenced their transcriptomes and studied the molecular features of their predicted proteomes, assuming that a high level of intracellular salt would lead to

the classical acidic signature of 'salt-in' microbes (for example). We also investigated the expression of putative organic osmolyte synthesizers and transporters as a function of salinity in *H. seosinensis*. We find that the cytoplasmic proteomes of *H. seosinensis* and *P. kirbyi* are not highly acidic, although they are significantly more hydrophilic than eukaryotic microbes inhabiting marine environments. At high salt, *H. seosinensis* up-regulated genes related to ectoine hydroxylase, amino acid transporters and myo-inositol carriers. Collectively, these observations suggest that these halophilic protists exploit organic solutes as major osmolytes, although they potentially also have higher intracellular salt content relative to mesophilic protists.

2- Materials and Methods

2.1- RNA and DNA Extraction and Sequence Generation

RNA was extracted from mid-exponential-phase cultures of *H. seosinensis* strain EHF34 (Park *et al.* 2006) and *P. kirbyi* strain AS12B (Park and Simpson 2011) grown at 37 °C on a shaker at 50 rpm in salt medium adjusted to desired concentrations after dilution of 30% salt medium (4.7 M NaCl, 102 mM KCl, 133 mM MgCl$_2$, 12 mM CaCl$_2$, 13 mM MgSO$_4$, medium #5 in Park 2012). *Halocafeteria seosinensis* was grown in triplicate at 15% and 30% salt, and was fed with *Haloferax* sp. isolated from one of our cultures. *Pharyngomonas kirbyi* was grown as flagellated cells at 15% salt and as amoebae at 10% salt, and fed in both cases with *Citrobacter* sp. As growth controls, cultures grown in parallel with the RNA experiment replicates were not sacrificed for RNA extraction, but kept alive until they reached stationary phase, to ensure proper growth. RNA was extracted using TRIzol following the manufacturer's instructions (Ambion, Carlsbad, USA). RNA extracts were treated with Turbo DNAse (Ambion) prior to cDNA library preparation using the TruSeq RNA sample preparation kit version 2 (Illumina, San Diego, USA).

Halocafeteria seosinensis samples were sequenced on an Illumina HiSeq 2000 platform by Génome Québec generating a total of 188,229,640 100-bp paired-end

reads. *Pharyngomonas kirbyi* samples were sequenced on a MiSeq platform generating 26,040,129 250-bp paired-end reads. Reads were trimmed to remove low-quality sequences and adapter sequences using Trimmomatic 0.30 with a sliding window of 10 nucleotides and a PHRED33 quality threshold of 25 (Bolger *et al.* 2014), and mapped to genomes of prokaryote species known to be in the culture (*Haloferax volcanii* DS2 - GenBank assembly accession GCA_000025685.1, *Citrobacter freundii* 4_7_47CFAA – GCA_000238735.1, *Salinivibrio costicola* subsp. *costicola* ATCC 33508 – GCA_000390145.1) in order to discard contaminant sequences (~1% of total reads) using Stampy 1.0.23 (Lunter and Goodson 2011). After decontamination, reads with k-mer coverage between 6× and 50× were selected using BBNorm (from the BBMap package: http://sourceforge.net/projects/bbmap/) and assembled using Trinity 2.0.2 (Grabherr *et al.* 2011). Open-reading frames (ORFs) were predicted using TransDecoder (from the Trinity package) and translated to protein sequences. Trinity typically generates many putative isoforms and polymorphic sequences. In order to reduce the redundancy in the dataset, ORFs were compared to each other by BLASTP searches (Altschul *et al.* 1990). For highly similar ORF pairs, we discarded the smallest sequence if the alignment covered >90% of its length and if <5 mismatches were observed. Finally, to remove sequences belonging to unknown prokaryotic contaminants present in the cultures, the nucleotide sequences of ORFs were compared to sequences in the NT database using BLASTN. Sequences having 100 bp-long fragments >90% identical to a prokaryotic sequence were discarded. The generated transcriptomes contained 16,852 ORFs for *H. seosinensis* and 15,521 ORFs for *P. kirbyi*.

DNA from *H. seosinensis* was purified using a salt extraction protocol (Aljanabi and Martinez 1997). Briefly, cells were disrupted by vortexing in lysis buffer (750 mM sucrose, 50 mM Tris-HCl pH 8.3, 40 mM EDTA pH 8.0) and digested with proteinase K (0.2 mg/mL) and 0.01% sodium dodecyl sulfate for 1 hour at 55 °C. DNA was separated from the other organic phases by centrifugation in a supersaturated NaCl solution (3M) and precipitated with 70% ethanol. A paired-end

DNA library (250 bp) was prepared using the Nextera XT kit (Illumina) prior to high-throughput sequencing on a MiSeq platform. 19,726,040 reads were generated and 'cleaned' as described for the RNA-derived sequences. Genomic contigs were assembled with MIRA 4.9.5_2 (Chevreux *et al.* 2004). Genes, including intron/exon boundaries, were delimited by mapping the read sequences obtained from RNA extracts using TopHat2 2.0.13 (Kim *et al.* 2013) and predicted by Braker 1.1 (Hoff *et al.* 2016).

2.2- Protein Localization Prediction and Calculations of pI and GRAVY

Sequences were assumed to encode soluble cytoplasmic proteins if no mitochondrial targeting peptide, signal peptide, or chloroplast transit peptide was predicted by TargetP 1.1 given a reliability class of 1 or 2 (Emanuelsson *et al.* 2000), and if no transmembrane domain was predicted by TMHMM 2.0 (Krogh *et al.* 2001). Protein sequences were considered to be secreted if predictions by Phobius 1.01 (Käll *et al.* 2004), WoLF PSORT 0.2 (Horton *et al.* 2007), TargetP and SignalP 4.0 (Petersen *et al.* 2011) all agreed on the extracellular localization of the protein and if no transmembrane domain was predicted by TMHMM. Furthermore, predicted secreted proteins were excluded if they contained the ER retention signals KK, KxK, KDEL or HDEL at the C-terminus or RR at the N-terminus. Signal peptides of predicted secreted proteins were cleaved prior to further analysis using the cleavage sites predicted by SignalP (Nielsen *et al.* 1997).

The isoelectric points of protein sequences were computed by iteratively calculating protein charge at given pH values, using side chain pKa values of charged amino acids (Nelson and Cox 2005), until a neutral charge was obtained. Hydropathicity of protein sequences was determined by calculating the GRand AVerage of hydropathY (GRAVY) score using the Kyte and Doolittle hydrophobicity scale (Kyte and Doolittle 1982). In order to investigate which amino acid had the most impact on the hydrophilicity of the halophilic protist proteomes, we iteratively calculated GRAVY scores by bringing the frequency of amino acids to the average frequency of mesophilic sequences at fastest-evolving sites of the Mandor dataset

(see below). At each round, the amino acid that contributed the most to bringing the GRAVY scores closest to the mesophilic level was identified and its frequency was kept at this level for further calculation rounds.

2.3- Homology Searches

Homologs of *H. seosinensis* and *P. kirbyi* proteins were searched for, using the BLAST algorithm, in the predicted proteomes of 24 protists sequenced during the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Table 2.1; Keeling et al, 2014). Taxa were selected if the salinity of the habitat from which they were sampled, or the salinity of the medium in which they were cultured, was close to seawater (*i.e.* ~3.5%). Extremophiles (*e.g.* psychrophiles or acidophiles) were avoided, as were organisms fed with other eukaryotes (due to cross-contamination issues). As a result, these taxa were not 'salt-in' extreme halophiles, and their proteomes did not contain a molecular signature for halophilicity. Homologous sequences were kept if alignments covered >60% of the smallest sequence examined with identity >25%. Harvested homologs were screened out if they were not predicted to be soluble cytoplasmic proteins given the criteria described above. We refer to these two datasets (one for each halophile) as the Marine Protists comparative datasets (MarProt).

In order to compare halophilic protists to 'salt-in' Archaea, we identified orthologous groups from the eggNOG 4.1 database (Powell *et al.* 2014) common to 16 species of *Halobacteria* and *H. seosinensis* and *P. kirbyi* through hidden Markov model searches using the hmmscan program of the HMMER package (Eddy 1998). Orthologues with E value < 0.00001 were chosen for further comparisons using BLAST. The same coverage and identity thresholds (*i.e.* >60% of smallest sequence and >25% identity respectively) used during the MMETSP homologue searches were applied to select sequences for statistical comparison. Transmembrane proteins were excluded as described previously. As a control, sequences in MMETSP taxa homologous to halobacterial genes were also identified and compared.

Table 2.1. Salinity of habitats from which taxa included in the Marine Protists (MarProt) datasets were sampled

Species	Strain	MMETSP ID	[salt] %
Amphiprora sp.	CCMP467	MMETSP0724	3.0
Amphora coffeaeformis	CCMP127	MMETSP0316	3.5
Aplanochytrium sp.	PBS07	MMETSP0954	2.5
Aristerstoma sp.	ATCC50986	MMETSP0125	3.2
Aurantiochytrium limacinum	ATCCMYA-1381	MMETSP0958	2.5
Aureococcus anophagefferens	CCMP1850	MMETSP0914	3.2
Aureoumbra lagunensis	CCMP1509	MMETSP0890	3.2
Bolidomonas pacifica	CCMP1866	MMETSP0785	2.1
Cafeteria roenbergensis	E4-10	MMETSP0942	3.2
Chaetoceros debilis	MM31A-1	MMETSP0149	3.4
Corethron hystrix	308	MMETSP0010	2.1
Grammatophora oceanica	CCMP410	MMETSP0009	2.1
Heterosigma akashiwo	CCMP2393	MMETSP0292	3.2
Ochromonas sp.	CCMP1393	MMETSP0004	3.5
Odontella aurita	Isolate 1302-5	MMETSP0015	2.1
Odontella sinensis	Grunow 1884	MMETSP0160	3.6
Paraphysomonas imperforata	PA2	MMETSP0103	3.4
Pelagococcus subviridis	CCMP1429	MMETSP0882	3.5
Pelagomonas calceolata	CCMP1756	MMETSP0886	3.2
Percolomonas cosmopolitus	AE-1	MMETSP0758	3.2
Pteridomonas danica	PT	MMETSP0101	3.4
Skeletonema costatum	1726	MMETSP0013	2.1
Thalassionema nitzschioides	L26-B	MMETSP0156	3.4
Thraustochytrium sp.	LLF1b	MMETSP0198-0199	3.4

Salinity values were obtained from MMETSP metadata available at http://data.imicrobe.us/project/view/104.

Putative genes for ectoine/hydroxyectoine synthesis in a wide range of protists were recovered from the complete MMETSP nucleotide dataset using TBLASTN, with *H. seosinensis* EctABCD protein sequences as queries. Subject sequences >15% identical to *H. seosinensis* sequences were compared to nucleotide sequences in the NT database using BLASTN to identify putative prokaryotic contaminants. Sequences >50% identical (without alignment length threshold) to bacterial or archaeal homologs were removed from further analysis. Protein sequences corresponding to the remaining nucleotide sequences (*i.e.* confirmed to be from protists) were phylogenetically examined as described below.

2.4- Phylogenetic Analysis and Ancestral Sequence Inference

We also used a curated 'phylogenomic' dataset containing 252 house-keeping genes from a broad range of eukaryotes (Table 2.2; Brown *et al.* 2012; Burki *et al.* 2012), referred as the Mandor dataset. Using this dataset provided a phylogenetic context to correct for non-independence among organisms during statistical analyses (see section 2.5). For comparison purposes, we only included free-living mesophilic organisms (*i.e.* no multicellular, symbiotic, parasitic or thermophilic organisms), and discarded organisms that had <50% of genes represented, with the exception of *Dunaliella salina* (45% of genes present), since this was the only halophile in the original dataset. Sequences orthologous to the 252 genes were obtained from *H. seosinensis*, *P. kirbyi*, *Wallemia ichthyophaga* EXF-994 (GCA_000400465.1), *Debaryomyces hansenii* CBS767 (GCA_000006445.2), *Aureobasidium pullulans* EXF-150 (GCA_000721785.1) and the MarProt taxa using methods described by Brown *et al.* (2012). Concatenated sequences from each organism were automatically aligned using MAFFT 7.205 (Katoh *et al.* 2002) and trimmed using BMGE 1.1 (Criscuolo and Gribaldo 2010), resulting in a super-alignment containing 110 taxa and 61,522 sites. A maximum-likelihood phylogenetic tree was generated in RAxML 8.1.22 (Stamatakis *et al.* 2005) using the PROTCAT-LGF model of amino acid substitution and 12 independent starting trees. Support values were calculated from 100 replicates using a rapid bootstrap analysis. Evolutionary rates were computed at each site in the alignment using dist_est

Table 2.2. Salinity of habitats or culture media of taxa included in the Mandor dataset

Species	Strain	MMETSP ID	Medium	[salt] %	Reference[a]
Acanthamoeba castellanii	-	-	NN (CCAP): PAS + agar	0.04	CCAP 1501/1A
Achlya hypogyna	-	-	50% LB	0.5	(Misner *et al.* 2015)
Allomyces macrogynus	-	-	PYG	0.05	(Ji *et al.* 2014)
Amoebidium parasiticum	-	-	ATCC Medium 890	0.07	ATCC 32709
Amphiprora	CCMP467	MMETSP0724	-	3.0	(MMETSP)
Aplanochytrium kerguelense	-	-	1/2 artificial seawater	1.6	(Nagano *et al.* 2011)
Aristerstoma	ATCC50986	MMETSP0125	-	3.2	(MMETSP)
Aspergillus fumigatus	Af293	-	ATCC Medium 312	0.4	ATCC 1022
Aurantiochytrium limacinum	ATCCMYA-1381	MMETSP0958	-	2.5	(MMETSP)
Aureobasidium pullulans	EXF-150	-	-	0.02	(Kogej *et al.* 2005)
Aureococcus anophagefferens	CCMP1850	MMETSP0914	-	3.2	(MMETSP)
Aureoumbra lagunensis	CCMP1509	MMETSP0890	-	3.2	(MMETSP)
Batrachochytrium dendrobatidis	-	-	Artificial pond water	0.02	(Webb 2010)
Bigelowiella natans	-	-	media in reference	3.0	(Moestrup and Sengco 2001)
Blepharisma japonicum	R1072	MMETSP1395	Standard balanced salt solution	0.2	(Giese 1973)
Bolidomonas pacifica	CCMP1866	MMETSP0785	-	2.1	(MMETSP)
Cafeteria roenbergensis	E4-10	MMETSP0942	-	3.2	(MMETSP)
Cafeteria sp.	Caron Lab	MMETSP1104	-	3.6	(MMETSP)
Chaetoceros debilis	MM31A-1	MMETSP0149	-	3.4	(MMETSP)
Chattonella subsalsa	CCMP2191	MMETSP0947	-	2.0	(MMETSP)
Chlamydomonas leiostraca	SAG11-49	MMETSP1391	Standard Volvox Medium	0.04	(Andersen *et al.* 2005)
Chlamydomonas reinhardtii	-	-	Sueoka's High Salt medium	0.27	(Chaudhari *et al.* 2015)

Species	Strain	MMETSP ID	Medium	[salt] %	Reference[a]
Chlorarachnion reptans	CCM449	MMETSP0109	-	3.5	(MMETSP)
Chromulina nebulosa	UTEXLB2642	MMETSP1095	DY-V	0.01	(NCMA)
Chrysochromulina rotalis	UIO044	MMETSP0287	-	2.5	(MMETSP)
Chrysocystis_fragilis	CCMP3189	MMETSP1165	-	3.2	(MMETSP)
Compsopogon coeruleus	SAG 36.94	MMETSP0312	-	0.2	(MMETSP)
Corethron hystrix	308	MMETSP0010	-	2.1	(MMETSP)
Cryptomonas paramecium	CCAP977/2a	MMETSP0038	-	0.06	(MMETSP)
Cyanophora paradoxa	-	-	Waris medium	0.01	(Betsche et al. 1992)
Cyanoptyche gloeocystis	SAG4.97	MMETSP1086	Standard Volvox Medium	0.04	(Andersen et al. 2005)
Debaryomyces hansenii	CBS767	-	-	6.0	(Prista et al. 2005)
Dictyocha speculum	CCMP1381	MMETSP1174	-	3.2	(MMETSP)
Dictyostelium discoideum	-	-	HL5 axenic medium	0.05	(Phillips and Gomer 2015)
Dinobryon sp.	UTEXLB2267	MMETSP0019	DY-V	0.01	(NCMA)
Dunaliella salina	-	-	-	15	(Arun and Singh 2013)
Ectocarpus siliculosus	-	-	seawater	3.5	(Cock et al. 2010)
Emiliania huxleyi	-	-	-	3.2	(Rokitta et al. 2014)
Fibrocapsa japonica	CCMP1661	MMETSP1339	-	3.2	(MMETSP)
Filamoeba nolandi	NC-AS-23-1	MMETSP0413	ATCC medium 802	0.008	(MMETSP)
Florenciella sp.	RCC1587	MMETSP1324	-	3.5	(MMETSP)
Fonticula alba	-	-	nutrient agar plates	0.5	(Brown et al. 2009)
Gloeochaete witrockiana	SAG46.84	MMETSP1089	DY-V	0.01	(NCMA)
Goniomonas sp.	m	MMETSP0114	-	3.5	(MMETSP)
Guillardia theta	-	-	seawater enriched media	3.2	CCMP327
Halocafeteria seosinensis	EHF34	-	-	15	(MMETSP)
Heterosigma akashiwo	CCMP2393	MMETSP0292	-	3.2	(MMETSP)

Species	Strain	MMETSP ID	Medium	[salt] %	Reference[a]
Isochrysis sp.	CCMP1324	MMETSP1129	-	3.4	(MMETSP)
Karenia brevis	SP1	MMETSP0573	-	3.5	(MMETSP)
Lotharella amoebiformis	CCMP2058	MMETSP0042	-	3.5	(MMETSP)
Mallomonas sp.	CCMP3275	MMETSP1167	-	3.2	(MMETSP)
Mayorella sp.	BSH-02190019	MMETSP0417	ATCC medium 1525	3.5	(MMETSP)
Micromonas pusilla	-	-	seawater-based enriched media	3.5	CCMP 487, 489, 490
Ministeria vibrans	-	-	ATCC Medium 1525	3.5	ATCC 50519
Monosiga brevicollis	-	-	ATCC Medium 1525	3.5	ATCC PRA-258, 50154
Naegleria gruberi	-	-	ATCC Medium 710	0.004	ATCC 30224
Nannochloropsis gaditana	-	-	seawater-based enriched media	3.5	CCMP 369, 526, 536, 1894
Neoparamoeba aestuarina	SoJaBio B1-5	MMETSP0161	-	3.5	(MMETSP)
Neurospora crassa	-	-	ATCC Medium 331	0.03	ATCC 10335, 13837, 14912...
Ochromonas sp.	BG1	MMETSP1105	DY-V	0.01	(NCMA)
Odontella aurita	1302-5	MMETSP0015	-	2.1	(MMETSP)
Ostreococcus tauri	-	-	artificial seawater	3.0	(van Ooijen *et al.* 2012)
Palpitomonas bilix	NIES-2562	MMETSP0780	-	3.5	(MMETSP)
Paramecium tetraurelia	-	-	ATCC Medium 802	0.008	ATCC 50045
Paramoeba atlantica	621	MMETSP0151	-	3.5	(MMETSP)
Paraphysomonas bandaiensis	Caron Lab	MMETSP1103	-	3.6	(MMETSP)
Pavlovales sp.	CCMP2436	-	seawater-based enriched media	3.5	CCMP 2436
Pelagomonas calceolata	CCMP1756	MMETSP0886	-	3.2	(MMETSP)
Percolomonas cosmopolitus	AE-1	MMETSP0758	-	3.2	(MMETSP)
Pessonella sp.	PRA29	MMETSP0420	ATCC medium 1525	3.5	(MMETSP)

Species	Strain	MMETSP ID	Medium	[salt] %	Reference[a]
Phaeodactylum tricornutum	-	-	seawater-based enriched media	3.5	CCMP 630-633
Phaeomonas parva	CCMP2877	MMETSP1163	-	3.2	(MMETSP)
Pharyngomonas kirbyi	AS12B	-	-	12	(MMETSP)
Physarum polycephalum	-	-	ATCC Medium 1288	0.3	ATCC 96951
Picocystis salinarum	CCMP1897	MMETSP1159	-	3.2	(MMETSP)
Pinguiococcus pyrenoidosus	CCMP2078	MMETSP1160	-	3.2	(MMETSP)
Pleurochrysis carterae	CCMP645	MMETSP1136	-	3.4	(MMETSP)
Polysphondylium pallidum	-	-	Polyspondium medium	0.2	(Hohl and Raper 1963)
Porphyridium aerugineum	SAG1380-2	MMETSP0313	MiEB12	0.01	(EPSAG 2015)
Protocruzia adherens	Boccale	MMETSP0216	artificial seawater	3.2	(MMETSP)
Prymnesium parvum	Texoma1	MMETSP0008	-	1.8	(MMETSP)
Pseudonitzschia multiseries	-	-	f/2	3.5	(Pitcher et al. 2014)
Pseudopedinella elastica	CCMP716	MMETSP1068	-	2.1	(MMETSP)
Pteridomonas danica	PT	MMETSP0101	-	3.4	(MMETSP)
Pythium aphanidermatum	DAOM BR444	-	corn meal agar plate	0.5	(Al-Sheikh 2010)
Rhizochromulina marina	CCMP1243	MMETSP1173	-	3.2	(MMETSP)
Rhizopus delemar	-	-	ATCC Medium 324	0.04	ATCC MYA-4621
Rhodella maculata	CCMP736	MMETSP0314	-	2.1	(MMETSP)
Rhodosorus marinus	CCMP 769	MMETSP0011	-	2.1	(MMETSP)
Roombia truncata	-	-	f/2	3.5	ATCC PRA-313
Saccharomyces cerevisiae	S288C	-	ATCC Medium 1069	0.05	ATCC 204508
Salpingoeca rosetta	ATCC 50818	-	ATCC medium 1525	3.5	ATCC 50818
Schizochytrium aggregatum	-	-	ATCC Medium 790	3.5	ATCC 28209
Schizosaccharomyces pombe	-	-	ATCC Medium 1245, ATCC Medium 2064, ATCC Medium 1067, ATCC Medium 1793	0.03	ATCC MYA-1011, 46946, 96612

Species	Strain	MMETSP ID	Medium	[salt] %	Reference[a]
Skeletonema costatum	-	MMETSP0013	-	2.1	(MMETSP)
Spizellomyces punctatus	ATCC 48900	-	ATCC Medium 683	0.002	ATCC 48900
Spumella elongata	CCAP955/1	MMETSP1098	DY-V	0.01	(NCMA)
Stereomyxa ramosa	-	MMETSP0439	ATCC Medium 1525	3.5	ATCC 50982
Symbiodinium sp.	k8 and mf105	-	f/2	3.5	(Bayer *et al.* 2012)
Thalassionema nitzschioides	L26-B	MMETSP0156	-	3.4	(MMETSP)
Thalassiosira pseudonana	-	-	seawater-based enriched media	3.5	CCMP1015, 1988, 3367
Thecamonas trahens	ATCC 50062	-	ATCC medium 1525	3.5	ATCC 50062
Thraustochytrium sp.	LLF1b	MMETSP0198-0199	-	3.4	(MMETSP)
Tiarina fusus	TfusLIS	MMETSP0472	-	3.0	(MMETSP)
Undescribed cercomonad	D1	MMETSP0086-0087	-	3.4	(MMETSP)
Vannella sp	DIVA	MMETSP0168	-	3.5	(MMETSP)
Vaucheria litorea	CCMP2940	MMETSP0945	-	3.2	(MMETSP)
Vexillifera sp.	DIVA3	MMETSP0173	-	3.5	(MMETSP)
Volvox carteri	-	-	modified standard Volvox	0.02	(Fukada *et al.* 2006)
Wallemia ichthyophaga	EXF-994	-	-	20	(Zalar *et al.* 2005)

[a] For taxa cultured at the American Type Culture Collection (ATCC), the Culture Collection of Marine Phytoplankton (CCMP) and the Culture Collection of Algae and Protozoa (CCAP), we obtained the salinity from media used to maintain these cultures. Abbreviations: MMETSP, Marine Microbial Eukaryote Transcriptome Sequencing Project; NCMA, National Center for Marine Algae and Microbiota.

(Susko *et al.* 2003) and the 10,000 fastest-evolving sites were selected as sites that are more likely to change in response to environmental conditions like salinity.

Ancestral sequences were reconstructed with codeml in the PAML package (Yang 2007) using the WAG+Gamma model. Sites for which the probability of a particular ancestral state was >0.9 were kept for further analysis. In order to determine which types of substitutions were significantly enriched in halophiles (from hydrophobic residue to hydrophilic residue or vice versa), we identified substitutions that showed significant variations (p < 0.05) above and below the diagonal of substitution matrices (see Figure 2.7), and analyzed each half of the matrix independently. For each organism, Z-scores for these substitutions were multiplied by the difference between the GRAVY scores of the extant amino acid and of the ancestral amino acid and summed up across the considered half of the substitution matrix. Calculated scores from individual organisms were compared where high scores signify variation patterns in substitutions that led to increased hydrophobicity. Amino acid substitution matrices were generated using matrix2png (Pavlidis and Noble 2003).

Putative protistan EctABCD sequences were phylogenetically analyzed as described above, except that the PROTGAMMA-LG4X model was used for the maximum-likelihood tree searches. In two instances (EctB in *C. roenbergensis* and EctD in *A. spinosum*), we concatenated partial sequences that likely originated from the same gene. Figures including alignments were generated using AliView 1.17.1 (Larsson 2014).

2.5- Statistical Comparisons

Distributions of amino acid (AA) frequencies, GRAVY scores and pI values from *H. seosinensis* and *P. kirbyi* were compared against the equivalent data from the marine protists by Mann-Whitney tests using the scipy package 0.13.3 (Jones *et al.* 2001). Z-scores for AA enrichment analysis and AA substitution analysis were calculated as follows:

$$Z = (p_1 - p_0) / \sqrt{(\, p \times (1-p)/N_1 + p \times (1-p)/N_0 \,)}$$

where p_1 is the frequency of the amino acid considered for the halophile, p_0 is the AA frequency for the mesophiles, p is the AA frequency for all organisms, while N_1 and N_0 are the total numbers of AAs for the halophile and the mesophiles respectively. GRAVY scores computed from the Mandor dataset were compared by standard Z test. P-values were corrected for multiple testing using the Benjamini-Hochberg method.

Comparisons of the cytoplasmic proteomes and secreted proteomes of 'salt-out' halophilic bacteria were performed as described above, using the protein sequences predicted from the genomes of *Actinopolyspora halophila* (NZ_AQUI00000000.1), *Marinococcus halotolerans* (NZ_ATVM00000000.1), *Nocardiopsis halophila* (NZ_ANAD00000000.1), *Virgibacillus alimentarius* (NZ_JFBD00000000.1), *Chromohalobacter salexigens* (NC_007963.1), *Halorhodospira halochloris* (CP007268.1) and *Thiomicrospira halophila* (NZ_ARAR00000000.1).

Phylogenetic canonical correlation analysis and phylogenetic principal component analysis were performed using phytools 0.4-31 (Revell 2012) to account for non-independence of organisms in the Mandor data. Correlations between the habitat salinity and varying protein features (AA frequencies, GRAVY and pI) were tested. Salt concentrations were obtained based on the salinity of the environment from which the protists were sampled or the salinity of the medium in which they were commonly maintained (Table 2.2).

2.6- Protein Tertiary Structure Investigation

In order to investigate the nature of the molecular signature detected in halophilic protists, we modeled *in silico* the tertiary structure of nine selected proteins from *H. seosinensis*, and determined the contribution of amino acid substitutions (relative to template sequences) to the overall difference in protein GRAVY scores. Using BLASTP against the Protein Data Bank (PDB), we selected sequences that were > 60% identical to *H. seosinensis* sequences and for which the

alignment covered > 80% of the largest sequence in the pair compared. To avoid noise in the signal of surface residues, we favored soluble monomeric proteins with simple substrates over proteins interacting with nucleic acid (*e.g.* histones), with lipids (*e.g.* acyl-CoA dehydrogenase) and with several protein partners that could significantly vary between organisms (*e.g.* ubiquitin-conjugating enzyme E2 and chaperones).

Tertiary structures were modeled using SWISS-MODEL (Arnold *et al.* 2006; Biasini *et al.* 2014), and models with QMEAN4 score < -3 were discarded from the analysis. Primary sequences of each pair (*H. seosinensis* - template) were aligned using MAFFT and ambiguously aligned sites were removed manually. Homologous sites where the absolute difference in hydropathy index between substituted amino acids was > 1 and where the absolute difference in relative solvent accessibility (RSA) was < 5% were considered for further analysis.

The difference in hydrophobicity of compared proteins was examined spatially by calculating the difference in GRAVY (delta-GRAVY) of amino acids at substituted sites, given their RSA, as follows:

$$\text{delta-GRAVY} = \sum_{s=1}^{M} (Hh - Ht)/N$$

where delta-GRAVY is the cumulative difference in GRAVY for each RSA bin (range: 5%) that sums up the difference in hydropathy index (given the hydrophobicity scale of Kyte and Doolittle 1982) between *H. seosinensis* amino acid (Hh) and the corresponding amino acid in the template sequence (Ht) for each substituted site s divided by the total number of N homologous sites for a maximum of M substitutions. A positive value for delta-GRAVY implies that amino acids in *H. seosinensis* proteins for this RSA bin contributed to increase the GRAVY score (*i.e.* this class contributed to increase the hydrophobicity of the *H. seosinensis* protein) while a negative value indicates that they contributed to decrease *H. seosinensis* GRAVY score (*i.e.* this class contributed to decrease the hydrophobicity of the *H. seosinensis* protein).

2.7- Differential Expression Assessment

Gene expression at optimal and maximal salt concentrations in *H. seosinensis* was quantified using RSEM (Li and Dewey 2011). Briefly, forward sequence reads from each replicate were mapped to the Trinity assembly using Bowtie 2 v.2.2.4 (Langmead *et al.* 2009). Reads mapping to multiple isoforms were assigned proportionally to the number of reads mapping to unique regions of the said isoforms. After removal of ORFs having low read counts in all samples (75th quantile < 10 reads), differential expression was assessed using three independent softwares: the empirical Bayesian analysis tool EBSeq following 10 iterations (Leng *et al.* 2013), DESeq2 (Love *et al.* 2014) and the limma package (Ritchie *et al.* 2015) after normalization using the Voom method (Law *et al.* 2014)*.* P-values were corrected for multiple testing using the Benjamini-Hochberg method. In the text, we report values from EBSeq and consider differentially expressed ORFs with probability of being differentially expressed >0.95 and posterior fold change > 2 or <0.5.

To construct figure 2.10, RNA read sequences were mapped onto genomic contigs encoding ectoine-related genes using TopHat2 and alignments were visualized using Tablet 1.15.09.01 (Milne *et al.* 2013).

2.8- Data Deposition

Transcriptome Shotgun Assembly projects have been deposited at DDBJ/EMBL/GenBank under the accession GECG00000000 and GECH00000000. The versions described in this chapter are the first versions, GECG01000000 and GECGH01000000. *Halocafeteria seosinensis* Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LVLI00000000. The version described in this chapter is version LVLI01000000.

3- Results

3.1- Cytoplasmic Protein Set

Comparative datasets of diverse marine protistan homologues of *H. seosinensis* and *P. kirbyi* sequences were constructed from the Marine Microbial Eukaryote Sequencing Project (Keeling *et al.* 2014) dataset. These are henceforth referred to as the Marine Protists comparative datasets, or 'MarProt'. These transcriptomes were used as reference datasets since they are from organisms that thrive in seawater (and not from 'salt-in' extreme halophiles). Predicted cytoplasmic proteins in *H. seosinensis* and *P. kirbyi* (1,639 and 1,574 sequences respectively) had 10,187 and 9,192 homologs respectively in MarProt. Applying the homology criteria described in section 2.3, *H. seosinensis* and *P. kirbyi* shared 664 homologous cytoplasmic proteins. As a second approach to compare protein sequences from halophilic protists to non-halophilic ones we employed a curated 'phylogenomic' dataset of 252 highly conserved 'universal' eukaryotic proteins, 'Mandor'. Forty-two proteins from *H. seosinensis* and 41 from *P. kirbyi* were in both the MarProt and the Mandor datasets.

3.2- Cytoplasmic Proteins Without Acidic Signature in Halophilic Protists

The isoelectric point (pI) distributions of predicted cytoplasmic proteins in *H. seosinensis* and *P. kirbyi* were not enriched in acidic proteins as commonly observed in 'salt-in' halophiles (Figure 2.1A-B). The pI distributions of acidic cytoplasmic proteins (pI < 6) in *H. seosinensis* and *P. kirbyi* were actually shifted toward basic values as compared to MarProt (Mann-Whitney test, $p < 0.001$; Figure 2.1C-D). Ranked pI values calculated from amino acids at the fastest-evolving sites in the Mandor alignment also showed that cytoplasmic proteins of halophilic protists were unremarkable in their acidity amongst this taxonomically broad sample of eukaryotes (rank 31/90 for *H. seosinensis* and rank 46/90 for *P. kirbyi*).
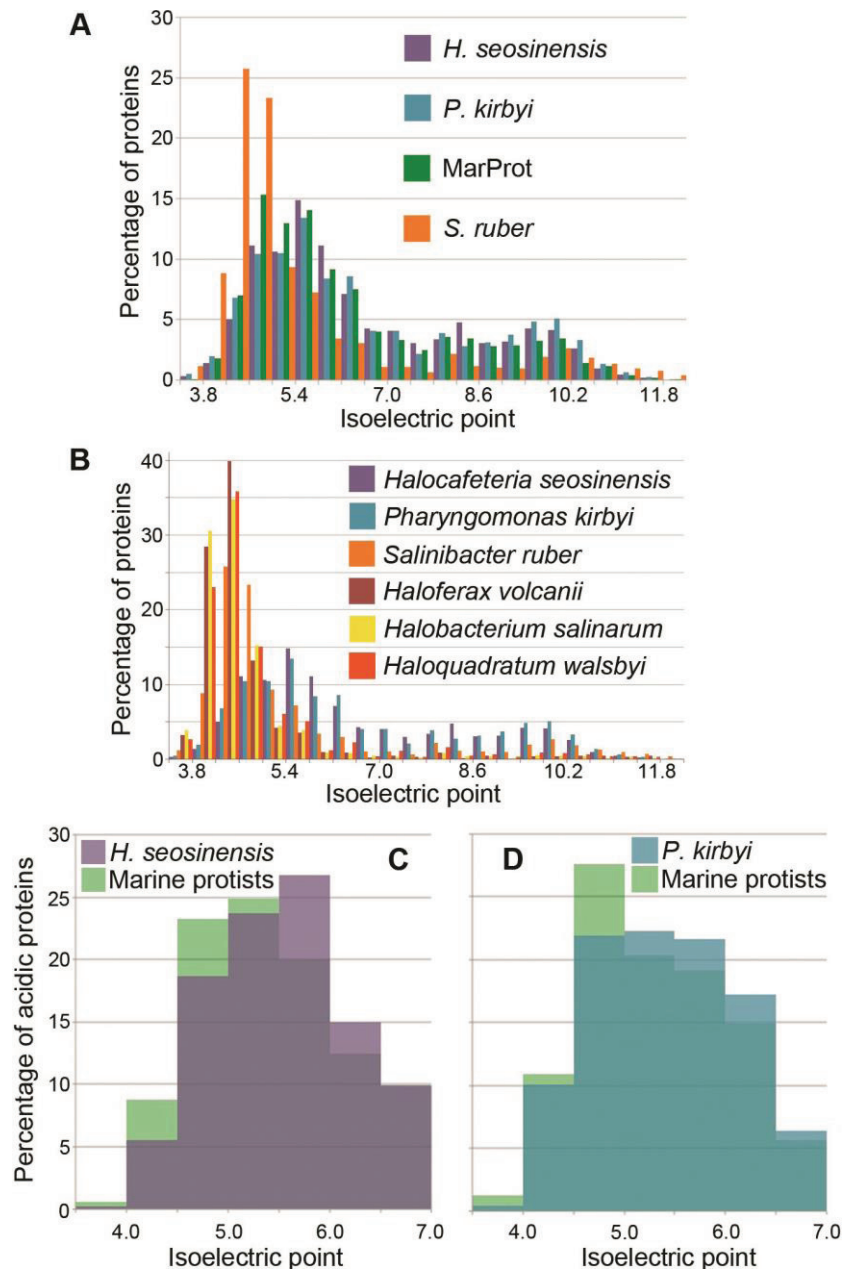
Fig. 2.1. Isoelectric point distribution of predicted cytoplasmic proteomes of halophilic protists *H. seosinensis* and *P. kirbyi,* marine protists (MarProt) and 'salt-in' microbes *Salinibacter ruber* DSM 13855 (GCA_000013045.1), *Haloferax volcanii* DS2 (GCF_000025685.1), *Halobacterium salinarum* DSM 671 (GCF_000069025.1) and *Haloquadratum walsbyi* C23 (GCF_000237865.1) showing that the proteomes of halophilic protists are not enriched in acidic proteins as in *S. ruber* (A) and halophilic archaea (B). Populations of proteins with pI < 7.0 in *H. seosinensis* (C) and *P. kirbyi* (D) are more basic than that of marine protists.

38

3.3- Increased Hydrophilicity of Cytoplasmic Proteins in Halophilic Protists

Analyses of GRAVY (GRand AVerage of hydropathY) scores showed that halophilic protists had more hydrophilic cytoplasmic proteomes than typical for marine protists, or non-extremophilic eukaryotes in general. The distributions of GRAVY scores for *H. seosinensis* and *P. kirbyi* were significantly shifted toward more hydrophilic values compared to the distributions of MarProt (Mann-Whitney, $p < 0.0001$; Figure 2.2). Concordantly, GRAVY scores calculated from the fastest-evolving sites in the Mandor alignment were significantly more hydrophilic for *H. seosinensis* and *P. kirbyi* compared to the other taxa (Z-test, $p = 0.0005$ and 0.0036 respectively). Interestingly, halophilic and halotolerant yeasts did not follow this trend ($p > 0.16$) whereas the alga *Dunaliella salina* did ($p = 0.009$). Furthermore, phylogenetic canonical correlation analysis, which corrects for species non-independence due to phylogenetic history (Revell and Harrison 2008), indicated that GRAVY scores and the salinity of an organism's habitat were significantly correlated (Mandor dataset; phylogenetic correlation analysis: canonical correlation $= 0.39$, $p = 2.4 \times 10^{-5}$; Figure 2.3). GRAVY scores obtained from the Mandor dataset were computed from amino acids at the fastest evolving sites. These theoretically represent sites in conserved proteins that are not subject to strong selective constraints based on overall protein structure or function, and could respond more easily to environmental conditions, such as salinity. The measured correlation between GRAVY scores calculated at these sites and the habitat salinity supports this assumption.

Surprisingly, the hydropathy of halophilic protist proteomes was comparable to proteomes of the *Halobacteria*, 'salt-in' Archaea that typically contain a hydrophilic signature in their proteomes. This was indicated by comparing 561 sequences from *H. seosinensis* homologous to 3,863 halobacterial sequences, and 478 sequences from *P. kirbyi* homologous to 3,418 halobacterial sequences. GRAVY distributions of halophilic protists were slightly shifted toward more hydrophilic values compared to those of the *Halobacteria* (Mann-Whitney, $p < 0.02$; Figure 2.2).
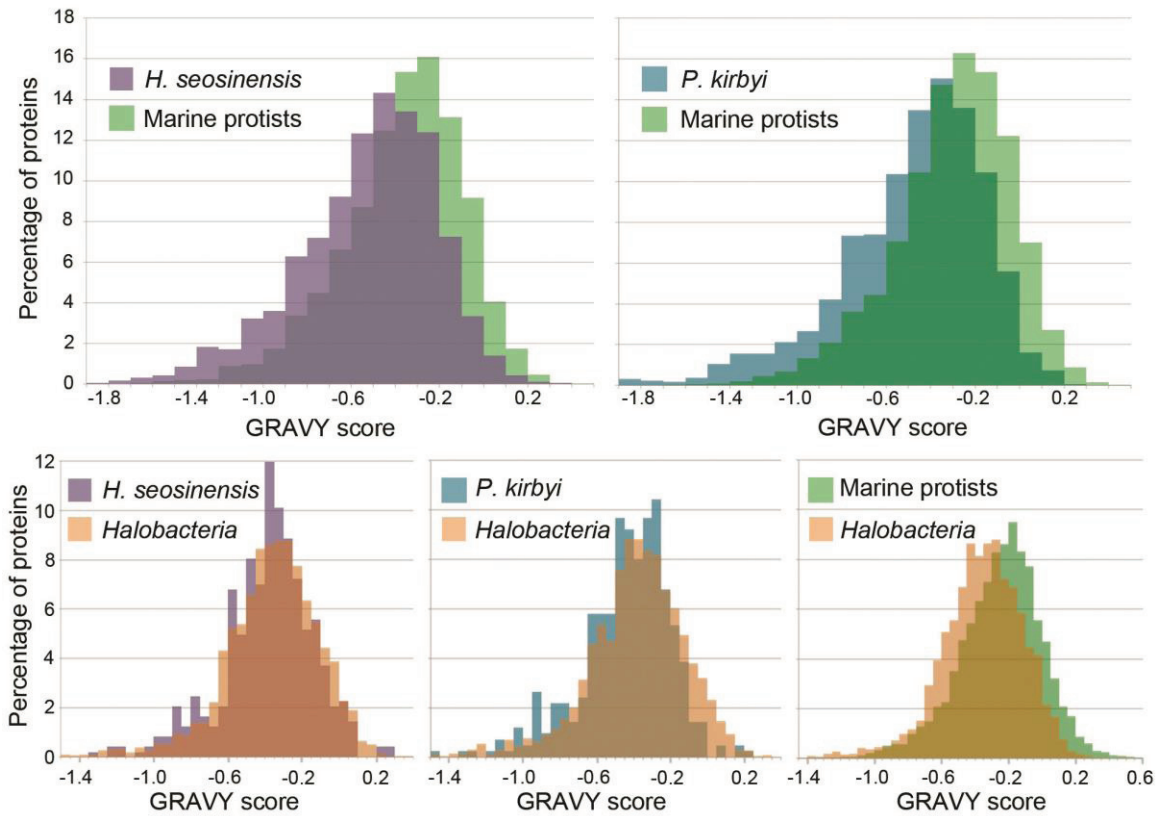
Fig. 2.2. GRAVY score distributions of predicted cytoplasmic soluble proteins of halophilic protists, marine protists and *Halobacteria*. On top, distributions of halophilic protists are significantly shifted toward hydrophilic values compared to homologous proteins from marine protists (Mann-Whitney, p < 0.0001). On the bottom, distributions of *H. seosinensis* and *P. kirbyi* are slightly shifted toward hydrophilic values compared to the ones of *Halobacteria* (Whitney, p < 0.02), while the distribution for marine protists is significantly shifted toward more hydrophobic values compared to the distribution for *Halobacteria* (Mann-Whitney, p < 0.0001).
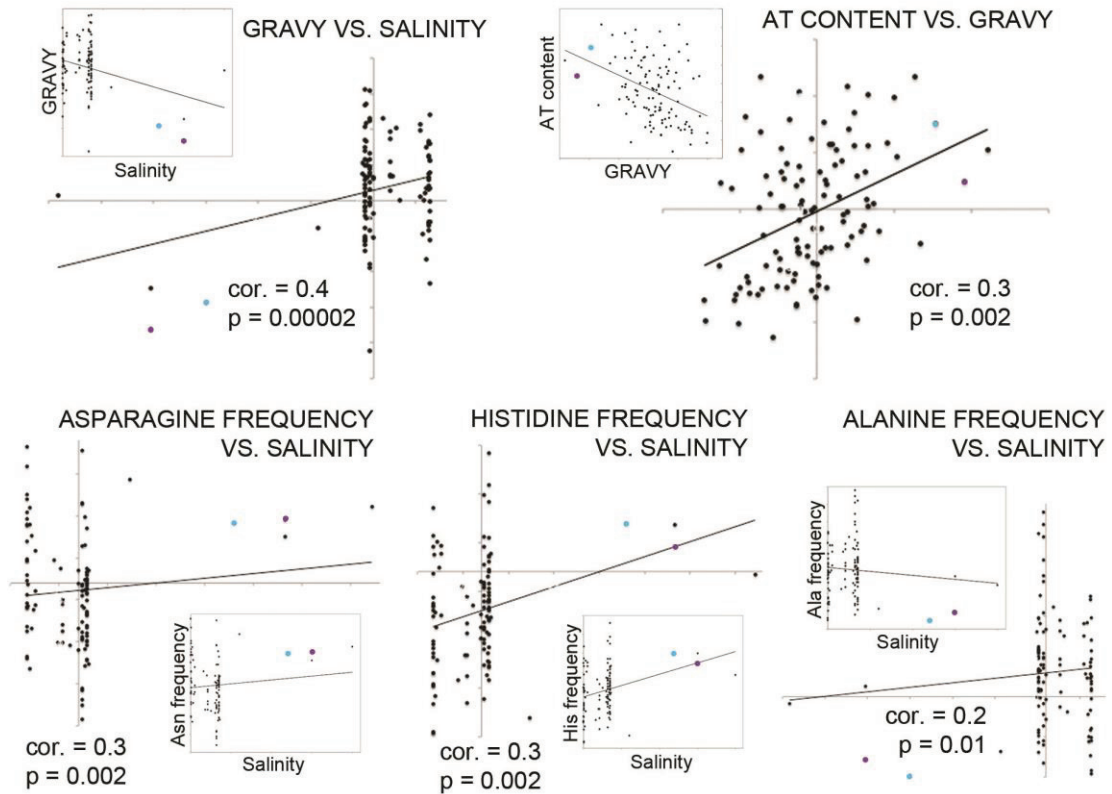
Fig. 2.3. Phylogenetic canonical correlations of GRAVY scores and habitat salinity (top-left), AT content and GRAVY scores (top-right), asparagine frequency and salinity (bottom-left), histidine frequency and salinity (bottom, in the middle), and alanine frequency and salinity (bottom-right). GRAVY scores and amino acid frequencies were calculated using fast-evolving sites of the Mandor alignment. Larger plots for each variable combination display canonical scores (*i.e.* corrected for evolutionary differences among species) while squared plots in corners contain non-transformed values. *Halocafeteria seosinensis* and *P. kirbyi* values are indicated by purple and blue dots respectively.

As expected, the distribution of GRAVY scores for 11,529 homologous sequences from the MMETSP marine protists was shifted to be more hydrophobic than the comparable distribution of GRAVY scores for 5,790 sequences from *Halobacteria* (Mann-Whitney, p < 0.0001; Figure 2.2).

The increased hydrophilicity of the cytoplasmic proteomes of the halophilic protists was due to an overall over-representation of polar residues and a general depletion of hydrophobic residues (Figure 2.4). In both *P. kirbyi*, and *H. seosinensis* alanine and leucine were significantly under-represented compared to their frequencies in mesophiles, whereas glutamate, asparagine and histidine were significantly over-represented. In addition, *H. seosinensis* cytoplasmic proteins had significantly more arginine and glutamine but less threonine and phenylalanine while *P. kirbyi* had significantly more arginine and glutamine but less cysteine, glycine and methionine. Cytoplasmic proteins from *P. kirbyi* were also enriched in phenylalanine, isoleucine and lysine, but this is potentially a result of the AT-richness of coding sequences in this organism (see below). Interestingly, frequencies of asparagine, histidine and alanine were correlated with habitat salinity (phylogenetic correlation analysis, p ≤ 0.01; Figure 2.3). Comparisons of amino acid frequency distributions of halophilic protists to those of MarProt all showed the same general trends of polar residue enrichment and hydrophobic residue depletion (Mann-Whitney, p < 0.01, except arginine for *P. kirbyi*, p = 0.48).

The contribution of each amino acid to protein GRAVY scores is a function of their frequency and hydropathicity. We therefore evaluated how adjusting each amino acid frequency to a mesophilic level would impact protein GRAVY scores. In both halophiles, these iterative calculations identified alanine as the amino acid most responsible for the low GRAVY scores. In *H. seosinensis*, the next most important amino acids, in order, were glutamate, glutamine, leucine, arginine, asparagine, valine, phenylalanine, histidine and isoleucine. In *P. kirbyi*, the order was alanine, leucine, glutamate, asparagine, lysine, arginine, valine and histidine. By contrast, when frequencies of the amino acids other than these were adjusted to the
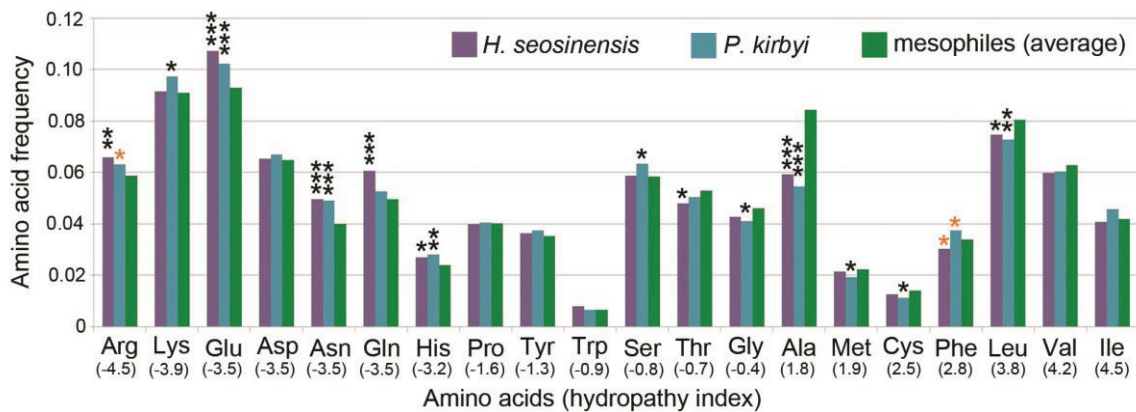
Fig. 2.4. Frequency of amino acids at fastest-evolving sites of the Mandor super-alignment. Comparison of values from halophilic protists to the average frequencies of other taxa (free-living microbial eukaryotes not adapted to extreme conditions) in the alignment revealed a general over-representation of hydrophilic residues and an under-representation of hydrophobic residues (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, in black p-values corrected for multiple testing, in orange p-values that were significant prior to correction). Amino acids are ordered on the x-axis from hydrophilic to hydrophobic residues (hydropathy indices in parentheses).

mesophilic level, the protein GRAVY scores for halophiles stayed more or less unchanged compared to the average GRAVY of mesophiles.

Phylogenetic canonical correlation analysis of the Mandor dataset showed that AT content and GRAVY scores were correlated (canonical correlation = 0.3, p = 0.002; Figure 2.3), with the proteomes of GC-rich organisms tending to be more hydrophobic than those of AT-rich organisms. In order to control for a potential influence of AT bias on GRAVY scores, we excluded organisms with AT content <45% and >55% (66 taxa removed) and recalculated the GRAVY scores at fastest-evolving sites. After this data filtering, the extremely hydrophilic GRAVY score of *H. seosinensis* stood out even more (Z = 4.05, p = 0.00003 after removal of AT-biased organisms compared to Z = 3.29, p = 0.0005 before; Figure 2.5). Since *P. kirbyi* was excluded from this previous analysis due to the high AT content of its coding sequences (59% AT on average), we also recalculated GRAVY scores by removing AT-biased sequences from MarProt, retaining 180 sequences from *P. kirbyi* with AT between 45% and 55% and comparing them to 4,352 homologous sequences from marine protists. Even when AT-biased sequences were removed, the GRAVY score distribution from *P. kirbyi* was still significantly shifted toward more hydrophilic scores compared to the distribution for MarProt (Mann-Whitney, p < 0.00001; Figure 2.5).

Next, we aimed to determine which amino acid substitutions had occurred more often in halophiles than in closely related mesophiles, or vice versa, since their divergences from common ancestors. We reconstructed ancestral sequences at internal nodes of a phylogenetic tree inferred from the Mandor dataset, and compiled the occurrence of each possible amino acid substitution in the whole alignment for each taxon (tree shown as Figure 2.6). Substitutions leading to more hydrophobic residues have occurred less often in the lineages of halophilic protists than in non-halophiles (Figure 2.7). The latter observation is supported by calculations of scores that considered the Z-scores and the difference between GRAVY scores of extant and ancestral amino acids (see section 2.4). *Halocafeteria*
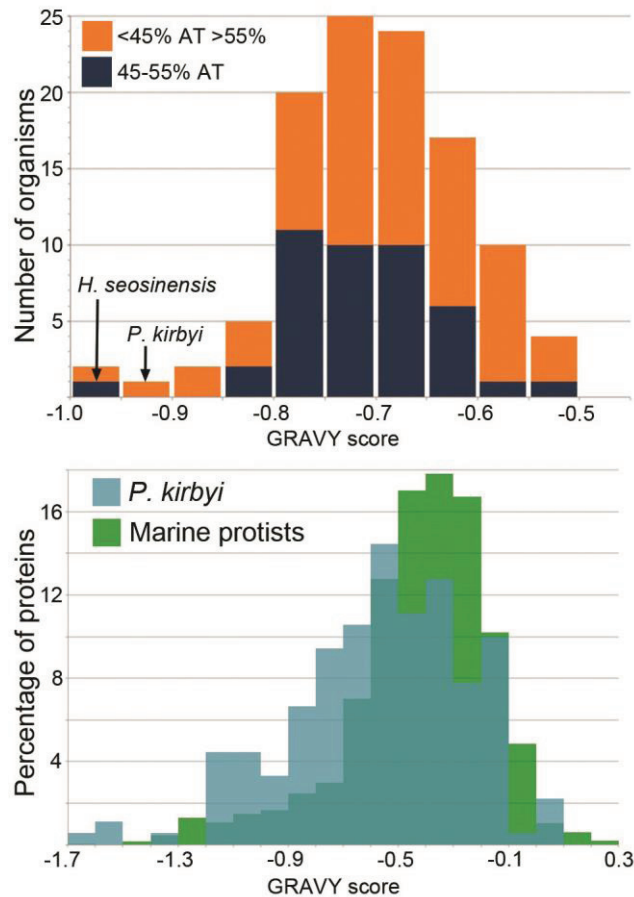
Fig. 2.5. Influence of AT content on GRAVY scores. The top panel indicates the abundance of organisms in the Mandor dataset as a function of their GRAVY score at fastest-evolving sites and AT content. When AT-rich organisms (>55% AT) are removed from the analysis, *H. seosinensis* had the most hydrophilic value of the dataset (Z score increased from 3.29 to 4.05). The bottom panel shows that the GRAVY score distribution of cytoplasmic soluble proteins of *P. kirbyi* is still shifted toward hydrophilic values compared to the one of homologous proteins from marine protists after removal of sequences with <45% and >55% AT content (Mann-Whitney, p < 0.00001).

On next page: Fig. 2.6. Maximum-likelihood phylogenetic tree generated with the Mandor super-alignment. Bootstrap values (>50% shown) from 100 replicates are indicated at branch nodes (dots indicate values of 100%). The scale bar indicates the substitution rate/site. * Sequences generated from *Tiarina fusus* were potentially contaminated with sequences the food source *Rhodomonas lens*.

Thecamonas trahens
Amoebidium parasiticum
Ministeria vibrans
74
Monosiga brevicollis
Salpingoeca rosetta
Fonticula alba
Rhizopus delemar
Wallemia ichthyophaga
Schizosaccharomyces pombe
Neurospora crassa
99
Aspergillus fumigatus
Aureobasidium pullulans
Debaryomyces hansenii
Saccharomyces cerevisiae
95
Spizellomyces punctatus
Batrachochytrium dendrobatidis
Allomyces macrogynus
80
Filamoeba nolandi
Pessonella sp.
Physarum polycephalum
Polysphondylium pallidum
Dictyostelium discoideum
Stereomyxa ramosa
Acanthamoeba castellanii
99
Vannella sp.
Mayorella sp.
86
Vexillifera sp.
Paramoeba atlantica
Neoparamoeba aestuarina
Pharyngomonas kirbyi
Naegleria gruberi
Percolomonas cosmopolitus
Palpitomonas bilix
Roombia truncata
Tiarina fusus*
Guillardia theta
Cryptomonas paramecium
Goniomonas sp
Dunaliella salina
Chlamydomonas leiostraca
Chlamydomonas reinhardtii
Volvox carteri
54
Picocystis salinarum
Ostreococcus tauri
Micromonas pusilla
84
Rhodosorus marinus
88
Porphyridium aerugineum
95
Compsopogon coeruleus
Rhodella maculata
Cyanoptyche gloeocystis
Cyanophora paradoxa
Gloeochaete witrockiana-Worden
undescribed cercomonad
Lotharella amoebiformis
76
Chlorarachnion reptans
Bigelowiella natans
Aplanochytrium kerguelense
Schizochytrium aggregatum
Thraustochytrium sp.
68
Aurantiochytrium limacinum
Cafeteria sp.
Cafeteria roenbergensis
Halocafeteria seosinensis
Pythium aphanidermatum
Achlya hypogyna
Chrysocystis fragilis
Aureoumbra lagunensis
Aureococcus anophagefferens
Pelagomonas calceolata
Rhizochromulina_marina
Pseudopedinella elastica
Pteridomonas danica
60
Dictyocha speculum
Florenciella sp.
Bolidomonas sp.
Corethron hystrix
73
Odontella aurita
Chaetoceros debilis
Thalassiosira pseudonana
Skeletonema costatum
Pseudonitzschia multiseries
Amphiprora sp.
Phaeodactylum tricornutum
Thalassionema nitzschioides
Nannochloropsis gaditana
Ectocarpus siliculosus
76
Vaucheria litorea
Fibrocapsa japonica
Heterosigma akashiwo
Chattonella subsalsa
Phaeomonas parva
Pinguiococcus pyrenoidosus
82
Mallomonas sp.
Chromulina nebulosa
Ochromonas sp.
95
Spumella elongata
99
Dinobryon sp.
Paraphysomonas bandaiensis
Aristerstoma sp.
Paramecium tetraurelia
Protocruzia adherens
Blepharisma japonicum
Karenia brevis
Symbiodinium sp.
Pavlovales sp.
Isochrysis sp.
Emiliania huxleyi
Pleurochrysis carterae
Chrysochromulina rotalis
Prymnesium parvum
97
88
0.2

46

Fig. 2.7. Occurrence of substitutions from predicted ancestral amino acids (on the left of matrices) to extant residues (at bottom of matrices) in halophilic protists at fastest-evolving sites of the Mandor super-alignment. Amino acids are ordered from most hydrophilic to most hydrophobic. Each square is colored according to Z-score where blue represents substitutions that occurred less often in halophiles compared to the other taxa and red represents substitutions that occurred more often in halophiles. Boxed squares refers to Z-scores with $p < 0.05$ (in black p-values corrected for multiple testing, in orange p-values that were significant prior to correction).

seosinensis and *P. kirbyi* had the lowest scores for substitutions leading to more hydrophobic residues. In particular, substitutions from polar residues to alanine were observed significantly less often in the halophile lineages (adjusted p < 0.05 for substitutions from lysine in both *H. seosinensis* and *P. kirbyi*, from arginine, glutamate, asparagine, glutamine, histidine, proline, serine and threonine in *P. kirbyi*, and from aspartate in *H. seosinensis*).

3.4- Candidate Secreted Proteins in Selected Halophiles

Given the high salinity of the extracellular milieu, secreted proteins in *H. seosinensis* and *P. kirbyi* were expected to show the canonical acidic signature of halophilic proteins. However, no acidic signature was detected in proteins predicted to be secreted in *H. seosinensis* and *P. kirbyi* (N = 231 and 52 proteins respectively) when compared to those predicted from the 24 MMETSP protists used for the MarProt datasets (N = 6584 proteins). The secreted proteins from halophilic protists did show more hydrophilic GRAVY scores, as observed for cytoplasmic proteins (Mann-Whitney, p < 0.001).

In order to evaluate the accuracy of the predictions, we searched for homologs of *H. seosinensis* and *P. kirbyi* predicted secreted proteins in the NR database using BLASTP. Most queries having hits (165 and 30 proteins for *H. seosinensis* and *P. kirbyi,* respectively) could be assigned to lysosomal functions (*e.g.* proteases/cathepsin/carboxypeptidase, beta-N-acetylhexosaminidase, proteins with a saposin domain, physaropepsin) or were homologous to proteins acting at the cell membrane (*e.g.* phospholipid transfer protein, N-acylsphingosine amidohydrolase), suggesting these might employ the secretion pathway but were not exported outside the cell. Others were homologous to proteins known to work intracellularly (*e.g.* dynein). This demonstrates that predicting exoproteins in phagotrophic protists is challenging, and that our analysis probably suffered from this limitation. However, it is notable that some predicted secreted proteins with low pI values (pI < 4.3) were homologous to extracellular proteins, for example, tenascin-like proteins

(extracellular matrix proteins), growth factor-binding proteins and protocadherin (partial hit to extracellular cadherin repeats).

Secretion and retention signals have been intensively studied in yeast (*e.g.* Vonheijne and Abrahmsen 1989; Gaynor *et al.* 1994; Conibear and Stevens 1998) and, in contrast to phagotrophic protozoa, yeasts secrete various soluble exoenzymes as part of their osmotrophic lifestyle. Bearing this in mind, as a form of positive control, we also compared predicted exoproteins from the halotolerant/halophilic yeasts *D. hansenii* and *W. ichthyophaga* to those of the marine protists in MarProt. The predicted secreted proteins from *D. hansenii* (103 proteins with median pI of 4.4) and *W. ichthyophaga* (110 proteins with median pI of 4.3) were significantly more acidic than the inferred exoproteins from marine protists (6479 proteins with median pI of 4.6, Mann-Whitney, $p < 0.0008$). Therefore, these results support the existence of an acidic signature in the exoproteins of halotolerant/halophilic eukaryotes when such proteins can be confidently identified.

As another control, we compared the predicted secreted proteome to the cytoplasmic proteome in each of several 'salt-out' halophilic bacteria that had optimal salinities for growth between 9-18% (*i.e.* similar to *P. kirbyi* and *H. seosinensis*; see section 2.5). As expected, the secreted proteome was more acidic and hydrophilic than the cytoplasmic proteome in each of these species (Mann-Whitney, $p < 0.001$, Table 2.3).

3.5- Localization of the Hydrophilic Signature Inside Protein Tertiary Structure

In silico predictions of the tertiary structure of proteins in *H. seosinensis* provided insights into the localization of the hydrophilic signature. The structures of nine proteins were modeled on templates from non-halophilic organisms (Table 2.4). All examined *H. seosinensis* protein sequences were more hydrophilic (*i.e.* they had lower GRAVY scores) than the corresponding homologous template sequence, except aconitase, which was slightly more hydrophobic (difference in GRAVY score of 0.01).

Table 2.3. Intra-species comparisons of cytoplasmic and secreted proteomes of 'salt-out' bacteria

	Isoelectric point (median)			GRAVY (median)			No. of proteins	
	cyto.	secr.	p-value	cyto.	secr.	p-value	cyto.	secr.
Chromohalobacter salexigens	5.09	4.30	<0.0001	-0.228	-0.431	<0.0001	2263	241
Marinobacter halotolerans	4.91	3.90	<0.0001	-0.380	-0.584	<0.0001	2344	112
Thiomicrospira halophila	5.00	4.81	0.0008	-0.293	-0.484	<0.0001	1574	129
Virgibacillus alimentarius	5.09	4.27	<0.0001	-0.353	-0.582	<0.0001	2064	119
Nocardiopsis halophila	5.00	4.06	<0.0001	-0.202	-0.345	<0.0001	3901	400
Actinopolyspora halophila	4.95	4.63	<0.0001	-0.268	-0.431	<0.0001	3534	258
Halorhodospira halochloris	5.19	4.63	<0.0001	-0.286	-0.426	<0.0001	2199	176

Abbreviations; cyto.: cytoplasmic, secr.: secreted

Table 2.4. Information on sequences compared for the tertiary structure analysis

	Protein Data Bank templates			H. seosinensis candidates			QMEAN 4 score
Protein name	Accession number	Organism	GRAVY	Sequence name	GRAVY	% ID	
Aconitase	1ami.1.A	*Bos taurus*	-0.361	m.22652	-0.350	67	-1.029
Enolase	3otr.1.A	*Toxoplasma gondi*	-0.201	m.28365	-0.223	63	-2.460
Adenosylhomocysteinase	3ond.1.A	*Lupinus luteus*	-0.176	m.61053	-0.273	69	-2.210
Fumarate hydratase	3e04.1.D	*Homo sapiens*	-0.105	m.11962	-0.253	69	-2.693
SAM synthase	5a19.1.A	*Homo sapiens*	-0.259	m.85989	-0.415	66	-1.345
Phosphoglycerate kinase	4o3f1.A	*Mus musculus*	-0.084	m.71585	-0.243	64	-1.885
GAPDH	3sth.1.A	*Toxoplasma gondi*	0.023	m.88755	-0.195	67	-1.050
PPI1	3jb9.1.Y	*Schizosaccharomyces pombe*	-0.106	m.79418	-0.374	67	-2.891
OIV dehydrogenase	2j9f1.B	*Homo sapiens*	0.073	m.91625	-0.269	65	-2.230

Abbreviations; OIV: 2-oxoisovalerate, PPI1: Peptidyl-prolyl cis-trans isomerase 1, SAM: S-adenosylmethionine, % ID: percentage identity.

We examined the locations of substitutions that led to an absolute change in amino acid hydropathy index >1 in *H. seosinensis* proteins compared to the protein sequences of the templates used to generate the structures (Figure 2.8). Residues that contributed the most to the low hydrophobicity of *H. seosinensis* proteins had a relative solvent accessibility (RSA) between 10% and 25%. Typically, a threshold of 20% RSA is used to discriminate surface from buried residues (*e.g.* Chen and Zhou 2005). Our analysis indicated that residues that influenced GRAVY scores the most had RSA values under this threshold, *i.e.* they are not at the surface. High salt enhances the hydrophobic effect, thus halophilic proteins contain less hydrophobic interactions in order to avoid a too rigid conformation and protein aggregation. A more hydrophilic core in *H. seosinensis*, relative to template proteins, is in line with this property of halophilic proteins.
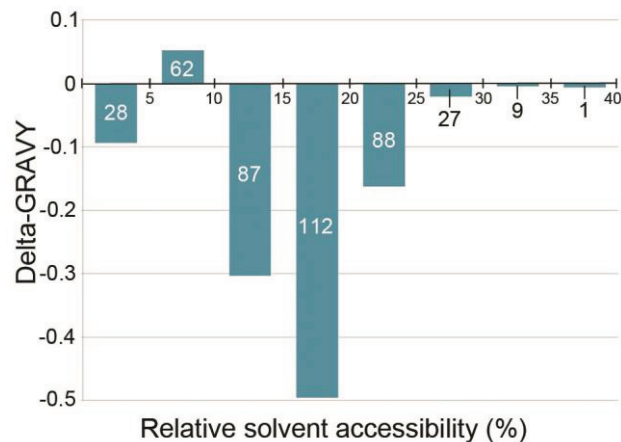
Fig. 2.8. Cumulative difference in GRAVY (delta-GRAVY) scores of amino acids at substituted sites in nine *H. seosinensis* proteins compared to the templates used to model the tertiary structures, as a function of their relative solvent accessibility. Negative y-axis values indicate that these residues contributed overall to decrease the hydrophobicity of *H. seosinensis* proteins. The total number of substitutions in each bin is indicated on respective bars.

3.6- Expression of Putative Osmolyte Synthesizers/Importers

The lack of a notably acidic proteome suggests that the halophilic protozoa under study might use the 'salt-out' strategy. That would require them to import and/or synthesize organic solutes as a response to elevated salinity. In a differential gene expression analysis of *H. seosinensis*, proteins whose transcripts were extremely up-regulated at high salt included ectoine hydroxylase and transporters for amino acids and myo-inositol. We elected to examine these interesting cases in more detail, using a purely bioinformatic approach.

5-hydroxyectoine is one of the osmolytes with the best protein-stabilizing properties (Lippert and Galinski 1992). It is synthesized from ectoine by ectoine hydroxylase (EctD), which belongs to the non-heme-containing iron(II) and 2-oxoglutarate-dependent oxygenases, a ubiquitious and large enzyme superfamily (Schofield and Zhang 1999). *Halocafeteria seosinensis* expressed transcripts annotated as EctD, and remarkably, these were 227-fold up-regulated at high salt in *H. seosinensis* (Table 2.5). The *H. seosinensis* EctD sequence contained all the conserved residues involved in binding iron, 2-oxoglurate and 5-hydroxyectoine as well as the ectoine hydroxylase consensus sequence (Figure 2.9; Höppner *et al.* 2014).

5-hydroxyectoine biosynthesis depends on a supply of ectoine, one of the most common osmolytes in halophilic bacteria (Severin *et al.* 1992). It is possible to speculate that *H. seosinensis* might import ectoine from food bacteria. This might be the case, however, *H. seosinensis* also seemed to express all the enzymes necessary for ectoine biosynthesis: diaminobutyrate aminotransferase (EctB), diaminobutyrate acetyltransferase (EctA) and ectoine synthase (EctC; Table 2.5). Interestingly, these genes are arranged in a cluster on the *H. seosinensis* genome, similarly to the way the ectoine synthesis operon is encoded on bacterial genomes, including an aspartate kinase (*ask*) gene (Widderich *et al.* 2014; Figure 2.10). A gene positioned at another locus in the *H. seosinensis* genome encoded a bifunctional

Table 2.5. Expression of genes potentially involved in organic osmolyte synthesis and transport in *H. seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	$\log_2$ FC	Adjusted p-value	$\log_2$ FC	
Ectoine biosynthesis									
m.89065	289	261	0.04	0.7	0.172	-0.5	0.233	-0.5	EctA
m.89060	238	397	1.0	1.3	1E-04	0.4	0.052	0.4	EctB
m.89066	504	239	0.7	0.4	3E-05	-1.4	0.015	-1.4	EctC
m.14216	0.7	216	1.0	220	NA	7.4	0.001	7.8	Ectoine hydroxylase (EctD)
m.89063	113	259	1.0	1.8	4E-04	0.9	0.017	0.9	Aspartate kinase
Amino acid transport									
m.15646	8	83	1.0	7.9	2E-59	3.0	2E-04	3.0	Na$^+$amino acid symporter
m.82938	4	38	1.0	7.1	2E-09	2.7	0.009	2.8	Amino acid transporter
m.16444	9	245	1.0	22.6	2E-28	4.3	4E-04	4.5	Amino acid transporter
m.16489	2	80	1.0	29.3	8E-50	4.7	3E-04	4.9	Amino acid transporter
m.62883	29	25	1.0	0.7	2E-05	-0.6	0.022	-0.6	Amino acid transporter
m.11656	16	11	1.0	0.5	2E-05	-0.9	0.012	-0.9	Amino acid transporter
m.18092	55	36	1.0	0.5	2E-11	-1.0	0.004	-1.0	Amino acid transporter
m.41609	18	14	1.0	0.6	4E-06	-0.7	0.014	-0.6	Amino acid transporter
m.1357	48	44	1.0	0.7	3E-06	-0.5	0.029	-0.5	Amino acid transporter
Myo-inositol transport									
m.89752	3	76	1.0	18.0	3E-44	4.1	2E-04	4.3	Myo-inositol transporter
m.27262	5	62	1.0	9.4	1E-26	3.2	4E-04	3.4	Myo-inositol transporter
m.25988	6	14	1.0	1.9	7E-06	0.9	0.012	0.9	Myo-inositol transporter
m.27173	2	5	0.6	2.3	0.042	1.1	0.102	1.1	Myo-inositol transporter

Abbreviations: TPM, average mRNA Transcripts Per Million; PPDE, Posterior Probability of being Differentially Expressed; Post Fold Change, Posterior Fold Change (30% over 15% salt); $\log_2$FC, $\log_2$ fold change (30% over 15% salt); EctA, 1-2,4-diaminobutyrate acetyltransferase; EctB, 1-2,4-diaminobutyrate transaminase; EctC, ectoine synthase; NA, not available due to an extreme count outlier in one of the samples.
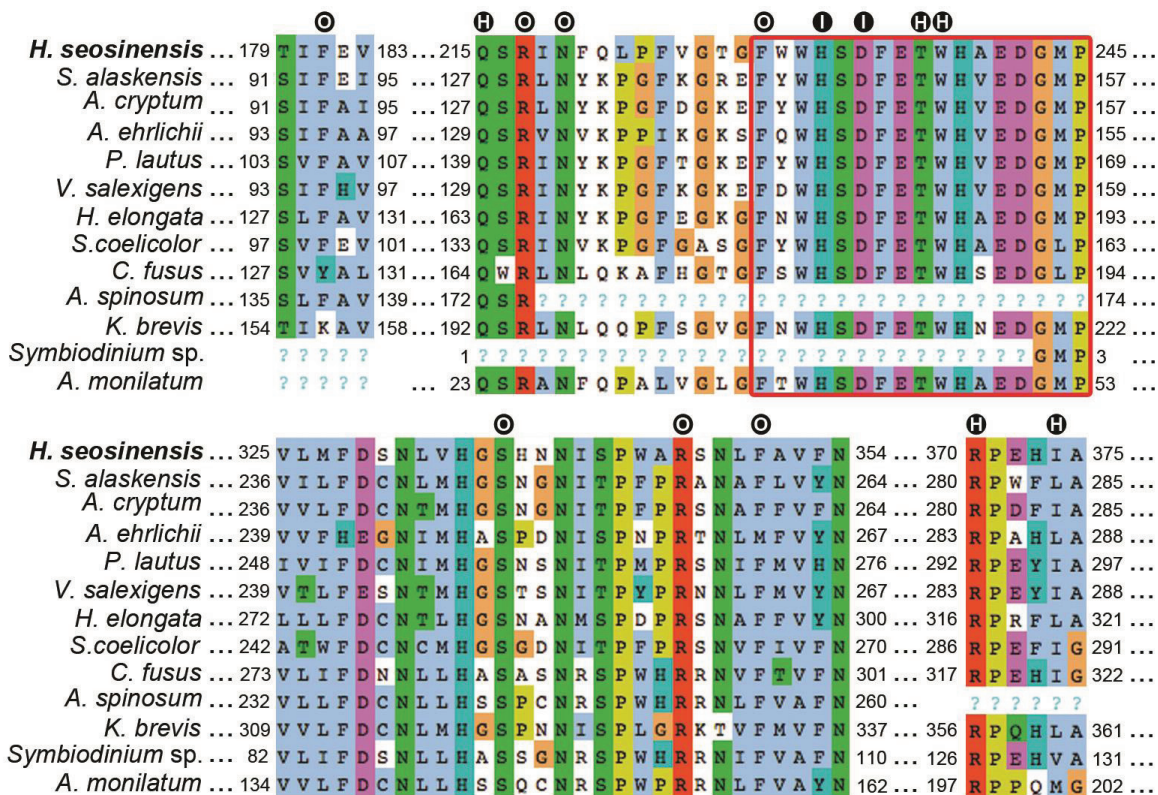
Fig. 2.9. Partial alignment of ectoine hydroxylase including sequences of characterized enzymes from *Sphingopyxis alaskensis* (WP_011543221), for which the crystal structure is available, *Acidiphilum cryptum* (AER00256), *Alkalilimnicola ehrlichii* (AER00257), *Paenibacillus lautus* (ACX67869), *Virgibacillus salexigens* (AAY29689), *Halomonas elongata* (WP_013333764) and *Streptomyces coelicolor* (Q93RV9), and sequences from the protists *H. seosinensis* (in bold)*, Ceratium fusus* (CAMPEP_0172939100), *Azadinium spinosum* (concatenation of CAMPEP_0180530970, CAMPEP_0180661784 and CAMPEP_0180535134), *Karenia brevis* (CAMPEP_0178068410), *Symbiodinium* sp. (CAMPEP_0169646080) and *Alexandrium monilatum* (CAMPEP_0175754634), showing residues involved in binding iron (circled I), 2-oxoglutarate (circled O) and 5-hydroxyectoine (circled H). The consensus sequence of ectoine hydroxylase (FxWHSDFETWHxEDG[M/L]P) is squared in red. "?" indicates missing data for partial sequences.
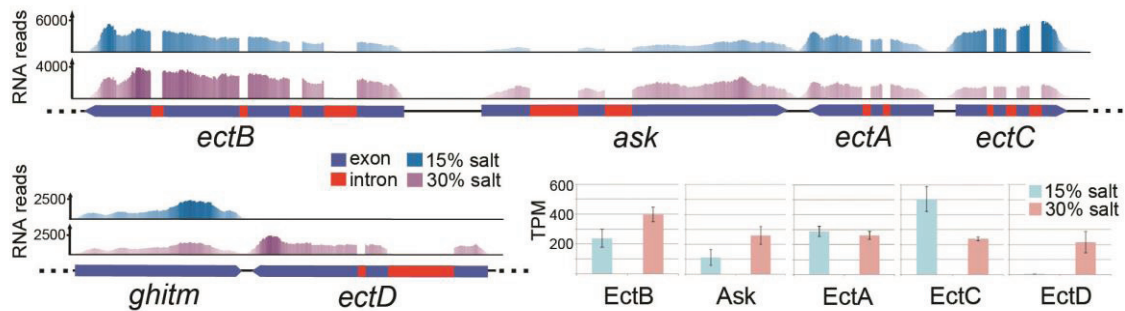
Fig. 2.10. Genomic context of ectoine/hydroxyectoine synthesis genes in *H. seosinensis* and their expression levels as a function of external salt concentration expressed in RNA read abundance mapped on genomic contigs and transcripts per million (TPM) as calculated by RSEM. On top, the region between nucleotides 48,000 and 56,100 on contig c75 contains genes coding for diaminobutyrate aminotransferase (*ectB*), aspartokinase (*ask*), diaminobutyrate acetyltransferase (*ectA*) and ectoine hydroxylase (*ectC*). On the bottom-left, the region between nucleotides 70,000 and 74,000 on contig c361 contains genes coding for growth hormone-inducible transmembrane protein (*ghitm*, shown as a reference) and ectoine hydroxylase (*ectD*). RNA read abundance is shown on top of genomic contigs on which they were mapped (15% salt in dark blue and 30% salt in purple). On genomic contigs, exons of ORFs are displayed in blue and introns in red. On the bottom-right, RNA transcript abundance (average TPM, error bars indicate 1 standard deviation) for each gene is shown in blue and pink for 15% and 30% salt conditions respectively.

aspartate kinase/diaminopimelate decarboxylase enzyme (ORF m.33370), suggesting that the *ask* gene inside the ectoine synthesis gene cluster might be specialized for ectoine synthesis, with the bifunctional enzyme instead used for lysine biosynthesis, as observed in bacteria (Stöveken *et al.* 2011). However, phylogenetic analysis of the *H. seosinensis* Ask sequence (including the aspartokinase cohesion group representative sequences; Lo *et al.* 2009) did not recover a strongly supported clade made up of Ask sequences specialized for ectoine synthesis (Figure 2.11).

Ectoine synthesis also requires the action of aspartate-β-semialdehyde dehydrogenase (ASADH), which is encoded three genes downstream of *ectC* on the *H. seosinensis* genome. Except for *asadh*, all of the ectoine-related genes are closely related to bacterial sequences (Figures 2.12-2.15). Nonetheless they all contain spliceosomal introns (Figure 2.10) and thus are truly eukaryotic sequences and do not represent bacterial contamination in our genomic assemblies.

Predicted N-terminal mitochondrial targeting signals were detected in all *H. seosinensis* ectoine synthesis-related proteins (EctABCD, Ask and ASADH), providing evidence that ectoine synthesis may occur in the mitochondria of *H. seosinensis*. Ectoine synthesis requires aspartate and glutamate as precursors; these amino acids are synthesized using intermediates of the mitochondrial Krebs cycle (Salway 1999).

Ectoine synthesis is thought to be restricted to bacteria and a few archaea (Widderich *et al.* 2014 and references therein), however no systematic search for eukaryotic homologues has been published to our knowledge. We therefore searched for homologues of EctA, B, C, and D in the MMETSP data (all taxa, not just the MarProt subset) as well as the NR and NT databases. We detected candidates for all four genes in a variety of eukaryotes. EctA and EctC were found in transcriptomic data from at least six protists: the amoebozoan *Vexillifera* sp., the dinoflagellate *Azadinium spinosum*, the cryptomonad *Goniomonas* sp., the ochrophyte stramenopile *Bolidomonas pacifica,* and in *Cafeteria roenbergensis,* which is a bicosoecid stramenopile and thus quite closely related to *H. seosinensis* (Figures 2.12 and 2.13).
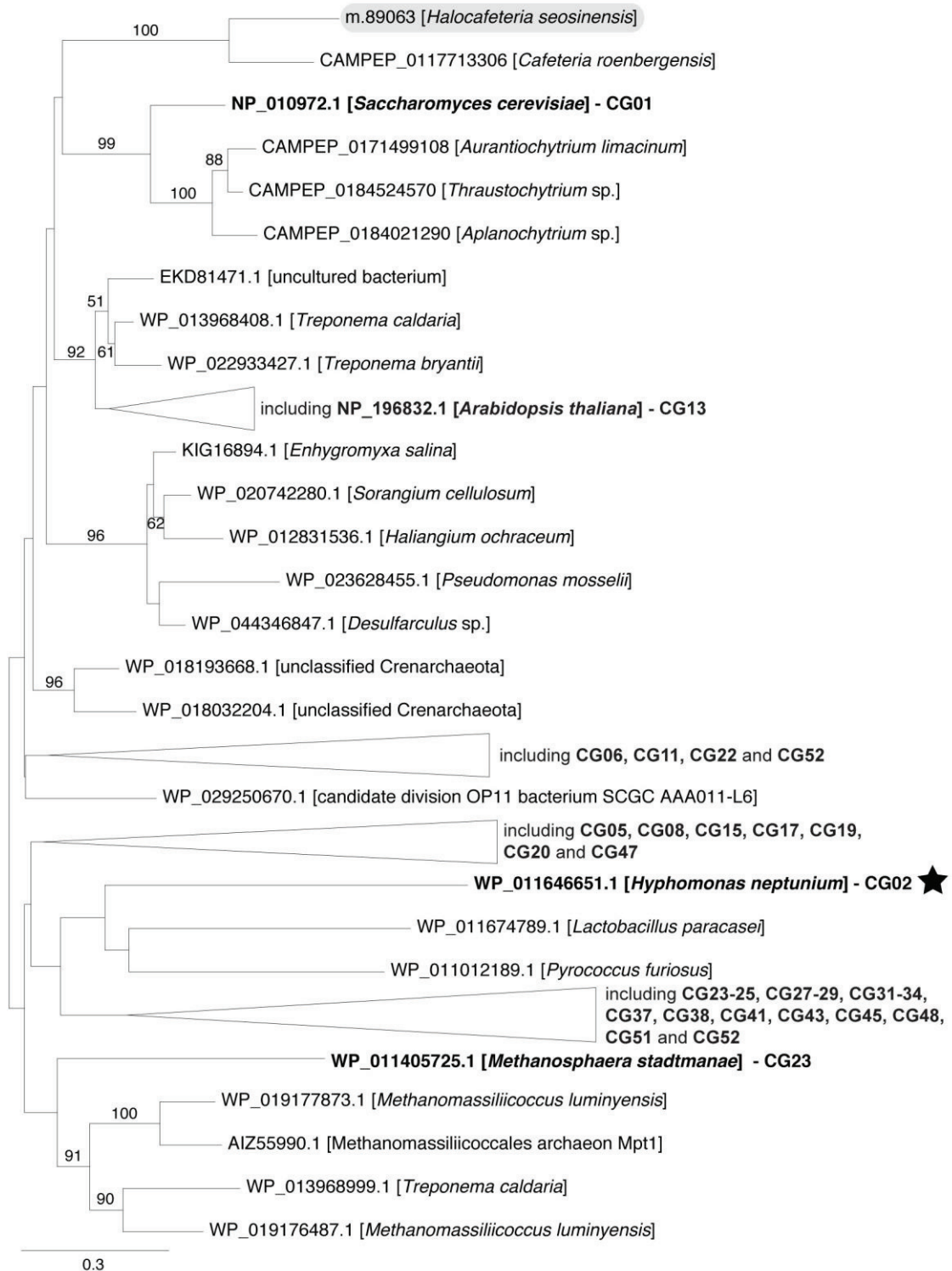
Fig. 2.11. Maximum-likelihood phylogenetic tree for aspartate kinase (Ask) including representative sequences of cohesion groups (indicated by 'CG' in boldface) as described in Lo *et al.* (2009). The Ask sequence specific for ectoine biosynthesis (*i.e.* CG02) is indicated with a star. Bootstrap values (>50%) from 100 replicates are indicated at branch nodes. The scale bar indicates the substitution rate/site.
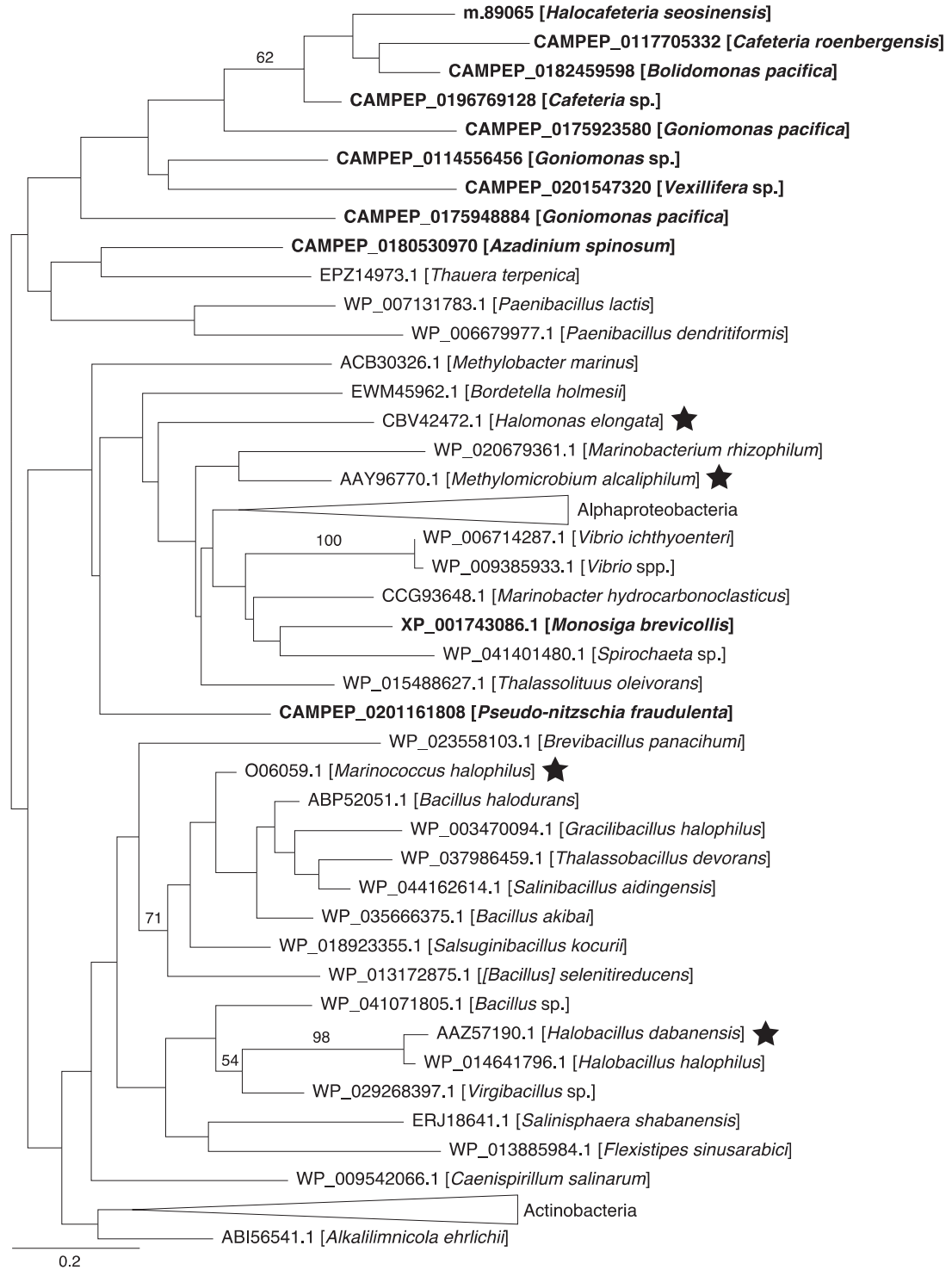
Fig. 2.12. Maximum-likelihood phylogenetic tree for diaminobutyrate acetyltransferase (EctA) showing sequences from bacteria in which the ectoine biosynthetic pathway has been enzymatically characterized (stars) and eukaryotic sequences (in bold). Bootstrap values (>50%) from 100 replicates are indicated at branch nodes. The scale bar indicates the substitution rate/site.
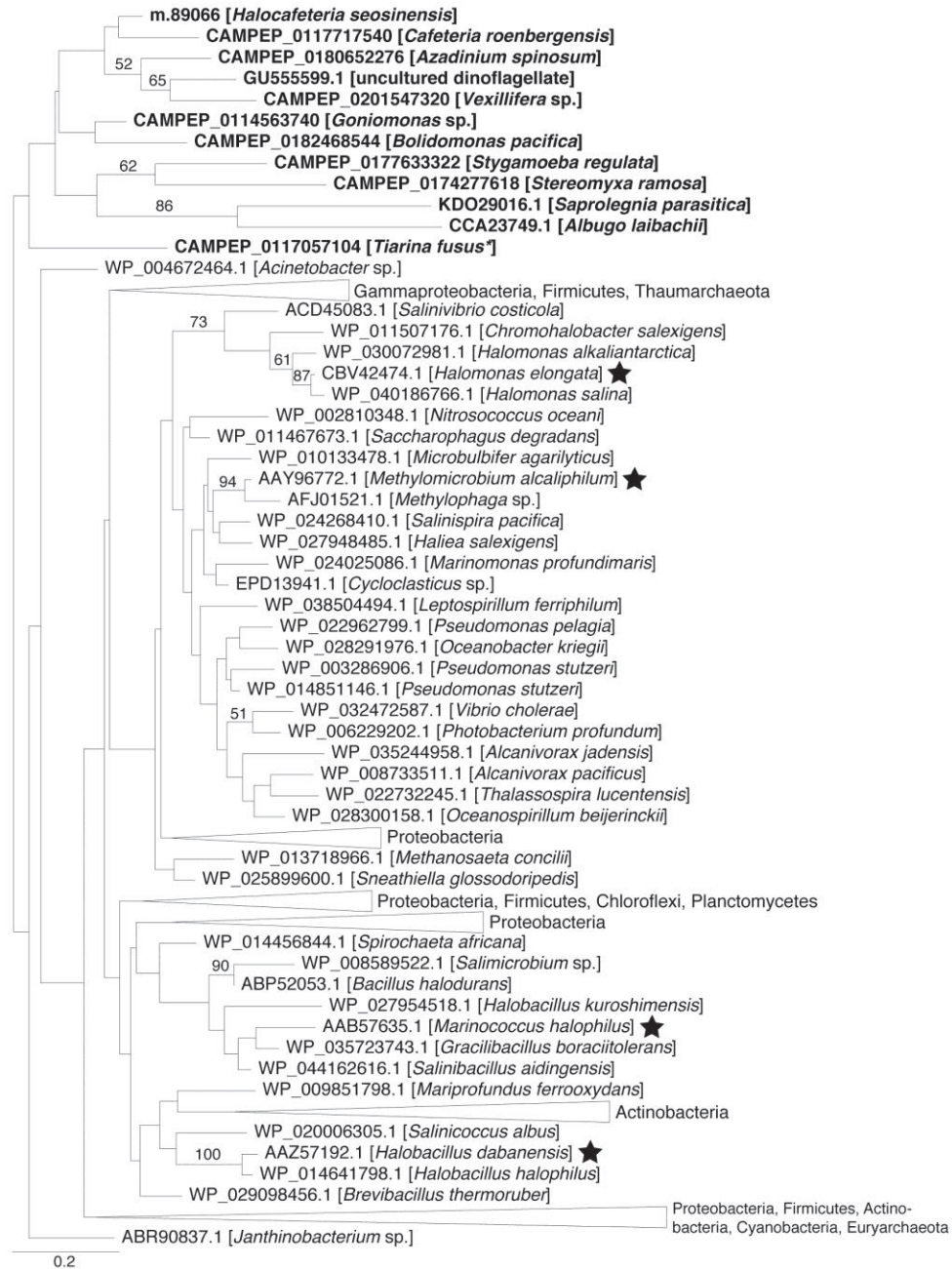
Fig. 2.13. Maximum-likelihood phylogenetic tree for ectoine synthase (EctC) showing sequences from bacteria in which the ectoine biosynthetic pathway has been enzymatically characterized (stars) and eukaryotic sequences (in bold). Bootstrap values (>50%) from 100 replicates are indicated at branch nodes. The scale bar indicates the substitution rate/site. * Sequences generated from *Tiarina fusus* were potentially contaminated with sequences from the food source *Rhodomonas lens*.

59
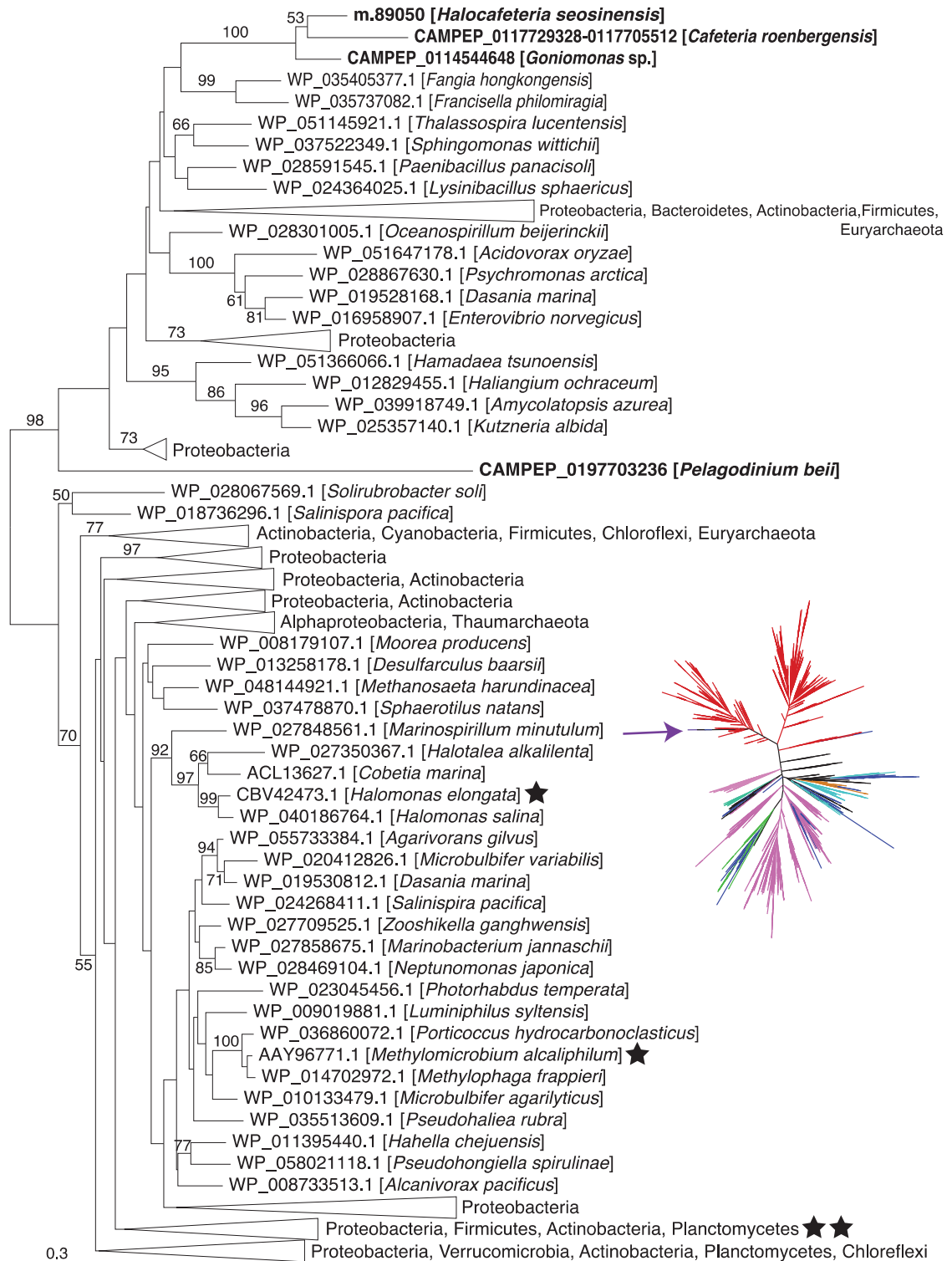
Fig. 2.14. Maximum-likelihood phylogenetic trees for diaminobutyrate aminotransferase (EctB) and related genes. The star-like tree on the right shows the diversity of genes harvested using *H. seosinensis* EctB sequence as query: EctB (red), alanine-glyoxylate aminotransferase (dark green), ornithine aminotransferase (cyan), putrescine–2-

oxoglutarate aminotransferase (brown), 2,2 dialkylglycine decarboxylase (pale green) and 4-aminobutyrate aminotransferase (pink). Sequences annotated as hypothetical proteins or as aminotransferases without further details about the substrate were left in black. The purple arrow points at *H. seosinensis* branch. EctB-related sequences (red branches) were re-aligned to generate the tree shown on the left. On the tree specific to EctB (left), eukaryotic sequences are in bold and sequences from bacteria in which the ectoine biosynthetic pathway has been enzymatically characterized are indicated with a star (the collapsed clade indicated with two stars contains sequences from *Marinococcus halophilus*, O06060.1, and *Halobacillus dabanensis*, AAZ57191.1). Bootstrap values (>50%) from 100 replicates are indicated at branch nodes. The scale bar indicates the substitution rate/site.
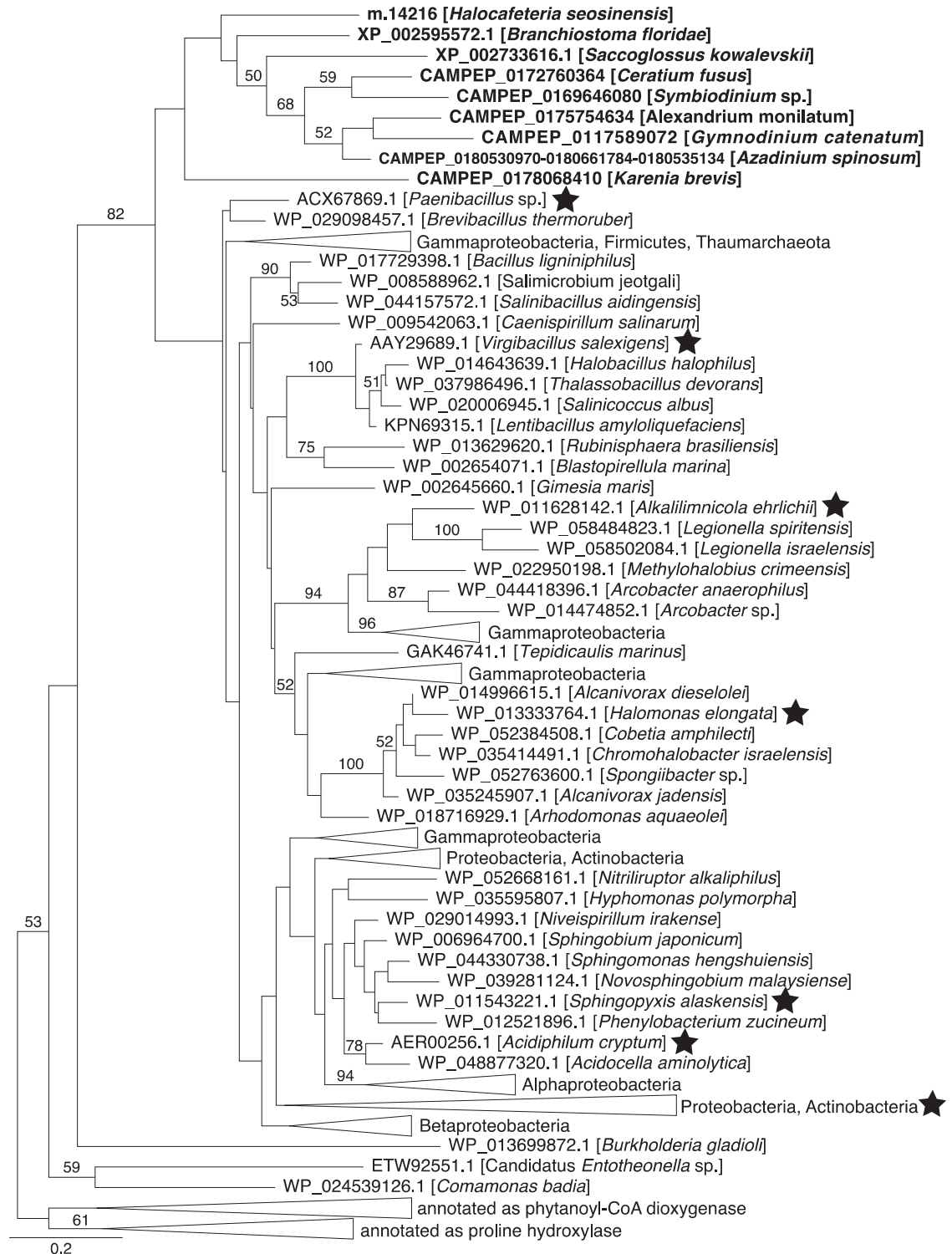
Fig. 2.15. Maximum-likelihood phylogenetic tree for ectoine hydroxylase (EctD) showing sequences from bacteria in which ectoine hydroxylase has been enzymatically characterized (stars, collapsed clade indicated with a star contains *Streptomyces coelicolor*, Q93RV9.2) and eukaryotic sequences (in bold). Bootstrap values (>50%) from 100 replicates are indicated at branch nodes. The scale bar indicates the substitution rate/site.

Interestingly, EctA and EctC were encoded on a single transcript (*i.e.* both proteins appeared to be encoded in a single ORF) in *Vexillifera* sp. (sequence CAMPEP_0201547320) and *A. spinosum* (CAMPEP_0180652276), suggesting that they were expressed as a single multi-functional polypeptide. *Azadinium spinosum* also had EctD encoded in this ORF (*i.e.* EctA, C and D encoded in a single ORF). More protists putatively expressed either EctA or EctC, like the amoebozoan *Stygamoeba regulata* and *Stereomyxa ramosa*, the ciliate *Tiarina fusus*, the choanoflagellate *Monosiga brevicollis*, the diatom *Pseudo-nitzschia fraudulenta*, and the oomycetes *Albugo candida*, *Albugo laibachii*, *Saprolegnia parasitica*, *Saprolegnia diclina*, *Phytophthora infestans*, *Phytophthora sojae* and *Plasmospora halstedii*. Interestingly, genes closely related to *ectC* were also detected in the deuterostome animals *Branchiostoma floridae* and *Saccoglossus kowalevskii*. In contrast with the wide distribution of EctA and C, only three protists, *Goniomonas* sp. (strain M), *Pelagodinium beii* and *C. roenbergensis*, expressed *bona fide* EctB candidates. These EctB candidates grouped with bacterial EctB sequences in phylogenetic trees (Figure 2.14). Other eukaryotic sequences harvested using *H. seosinensis* EctB as a query were markedly more distantly related, and were annotated as being involved in amino acid metabolism: alanine-glyoxylate aminotransferase, ornithine aminotransferase, putrescine—2-oxoglutarate aminotransferase, 2,2-dialkylglycine decarboxylase and 4-aminobutyrate aminotransferase (Figure 2.14). These latter sequences are unlikely to be orthologs of EctB.

The apparent absence of EctABC homologues in some of the protists that we have investigated must be interpreted with care; putative sequences were mostly identified in the MMETSP data where the transcriptome coverage could be too low in some cases to recover all *ectABC* genes from individual species, or, alternatively, the genes might just not be expressed in the growth conditions analyzed. Furthermore, many eukaryotic sequences from the MMETSP dataset were probably excluded from our analysis due to our stringent filtering of prokaryotic sequences that used a 50% identity cut-off in DNA sequences. However, this analysis demonstrates that ectoine synthesis might well occur in at least a few protists,

including in two marine protists that, as in *H. seosinensis*, seemed to express all *ectABC* genes. Experimental confirmation is therefore greatly needed to assess the function of these genes in protists.

Regarding ectoine hydroxylase, putative EctD sequences were detected in the dinoflagellate taxa *Ceratium fusus*, *A. spinosum, Symbiodinium* sp., *Alexandrium monilatum, Karenia brevis* and *Gymnodinium cantenatum* (Figure 2.15). Where determined, these sequences all included the consensus signature sequence of EctD and the conserved functional residues, except one residue binding 2-oxoglutarate (although some sequences were partial and were missing regions of interest; Figure 2.9). As for EctC, putative EctD sequences were detected in the animals *B. floridae* and *S. kowalevskii*, however most functional residues were substituted, including the consensus sequence that was not conserved in *S. kowalevskii*.

In *H. seosinensis*, the gene encoding EctD was greatly up-regulated at high salt while the one encoding EctA was not differentially expressed. Predictions for EctB and EctC were not consistent among the three programs used for differential expression analysis (Table 2.5). EctB seemed to be slightly up-regulated (30% increase in expression) but limma predicted this to be not significant (adjusted p-value 0.052). For EctC, EBSeq generated a relatively low posterior probability of 0.7 for a 2.5-fold repression at high salt while the other two programs returned significant p-values for over-expression (adjusted p < 0.015).

In principle, import of amino acids with osmoprotectant properties (*e.g.* glutamate, glycine, proline, aspartate) from food prokaryotes might also contribute to osmotic equilibrium in *H. seosinensis.* Several genes annotated as amino acid transporters were highly up-regulated at high salt (up to 29-fold increase; Table 2.5). In comparison, other amino acid transporter-related genes that were up-regulated at optimal salt concentration did not show more than a twofold increase in expression (Table 2.5).

A gene related to sodium-neurotransmitter symporters was 7.9-fold up-regulated at high salt (m.15646 in Table 2.5). This family of transporters is well

documented in animals. These transporters use a Na$^+$ or Cl$^-$ gradient to transport monoamines like serotonin, dopamine and norepinephrine, and the neurotransmitter Y-aminobutyric acid (GABA) and glycine (Torres *et al.* 2003). They can also potentially transport proline and taurine, two molecules reported as osmoprotectants. Alignment of the *H. seosinensis* sequence with LeuT (leucine transporter in *Aquifex*), GlyT1b (glycine transporter in humans), GAT1 (GABA transporter in humans), DAT (dopamine transporter in humans) and SERT (serotonin transporter in humans) showed conservation of functionally relevant residues, including at substrate-binding sites, and at sites involved in coordinating one of the sodium ions (Figure 2.16; Yamashita *et al.* 2005). In the *H. seosinensis* sequence, amino acids at sites known to interact with the substrate are closer in identity to Glyt1b and GAT1, consistent with the prediction that this transporter in *H. seosinensis* carries amino acids (with osmoprotectant potential) rather than biogenic amides (neurotransmitters). Expression of several tRNA synthetases and elongation factor genes was modestly repressed at high salt (up to 2.4-fold decrease) suggesting that the inferred increase in amino acid import was probably not coupled to translational activity (Table 2.6).

Myo-inositol is another well-known compatible solute (Garcia-Perez and Burg 1991; Klages *et al.* 1999; Majee *et al.* 2004). *Halocafeteria seosinensis* expressed four related genes encoding myo-inositol transporters (Table 2.5). Two were highly up-regulated at high salt (9.4- and 18-fold increases in expression) while the other two were not differentially expressed. Myo-inositol is also one of the main precursors of phosphatidylinositol (PI), a minor component of the cell membrane and an important molecule in cell signaling. However, the three enzymes that together convert myo-inositol to PI, if differentially expressed, were actually down-regulated at high salt (2.4 – 3.2-fold repression; Table 2.6), suggesting the increased gene expression of myo-inositol transporters at high salt was not directed toward lipid biosynthesis.
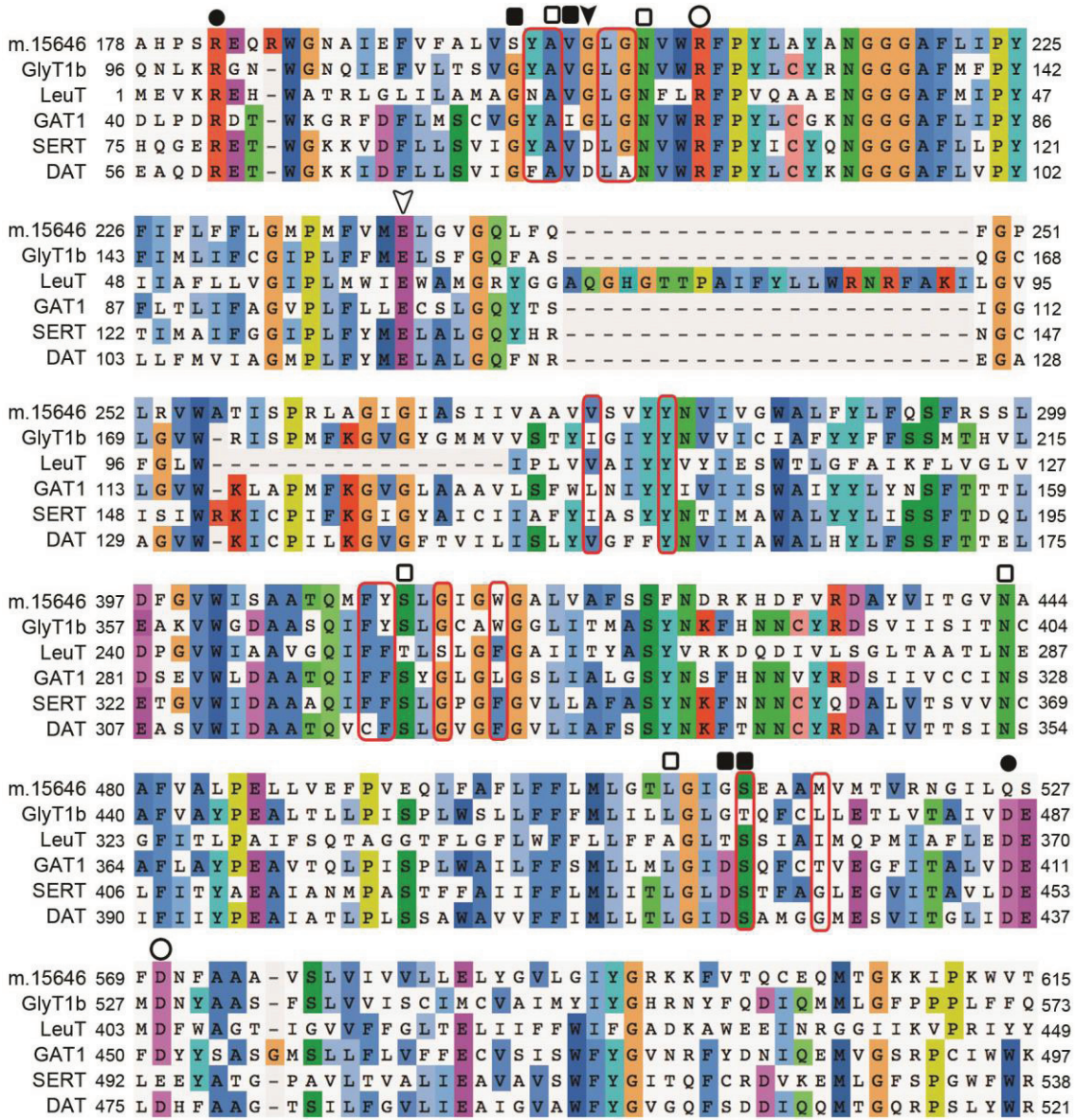
Fig. 2.16. Alignment of transporters for amino acids and derivatives: leucine (LeuT, NP_214423), glycine (GlyT1b, P48067), GABA (GAT1, P30531), dopamine (DAT, Q01959) and serotonin (SERT, P31645) based on Yamashita *et al.* (2005). The alignment, that also includes a transporter from *H. seosinensis* (m.15646), shows sites where amino acids interact with the substrate in LeuT (red boxes) and coordinate sodium ions (while box: Na1, black box: Na2). The filled arrow points at a site where aspartate indicates specificity for biogenic amide and glycine for amino acids. The empty arrow indicates a site where a mutation from Glu to Asp inactivates GAT1 (Keshet *et al.* 1995). Filled and empty circles indicate charged amino acid pairs at the extracellular and the intracellular entrances respectively. In *H. seosinensis*, amino acids at binding sites are close in identity to GlyT1b.

Table 2.6. Expression of genes involved in translation and phosphatidylinositol synthesis in *H. seosinensis*

ORF Names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$ FC	Adjusted p-value	log$_2$ FC	
tRNA loading									
m.93974	58	46	1	0.6	5E-04	-0.7	0.03	-0.7	Glutaminyl-tRNA synthetase
m.14914	32	19	1	0.5	8E-26	-1.1	0.001	-1.1	Lysyl-tRNA synthetase
m.77228	149	78	1	0.4	8E-08	-1.3	0.005	-1.3	Tyrosyl-tRNA synthetase
m.31630	53	49	0.97	0.7	0.05	-0.5	0.1	-0.5	Isoleucyl-tRNA synthetase
m.28708	41	22	0.99	0.4	9E-06	-1.3	0.01	-1.2	Tryptophanyl-tRNA synthetase (mito.)
m.60791	47	48	1	0.8	0.05	-0.3	0.1	-0.3	Tryptophanyl-tRNA synthetase (cyto.)
m.69913	96	85	1	0.7	9E-06	-0.5	0.03	-0.5	Threonyl-tRNA synthetase
m.29419	210	200	1	0.7	0.001	-0.4	0.06	-0.4	Aspartyl-tRNA synthetase
m.91051	65	36	1	0.4	5E-06	-1.2	0.01	-1.2	Prolyl-tRNA synthetase
m.76620	69	33	1	0.4	8E-09	-1.4	0.004	-1.4	Phenylalanyl-tRNA synthetase
m.38736	107	79	1	0.6	3E-05	-0.8	0.01	-0.8	Seryl-tRNA synthetase
m.54578	87	73	1	0.6	1E-06	-0.6	0.02	-0.6	Asparaginyl-tRNA synthetase
m.21630	75	69	1	0.7	0.001	-0.5	0.05	-0.4	Arginyl-tRNA synthetase
m.28234	19	13	1	0.5	1E-07	-0.9	0.006	-0.9	Leucyl-tRNA synthetase
m.59228	16	17	1	0.8	0.08	-0.3	0.2	-0.3	Methionyl-tRNA synthetase
m.8724	45	33	1	0.6	1E-08	-0.8	0.008	-0.8	Valyl-tRNA synthetase
m.28010	90	83	1	0.7	0.001	-0.5	0.05	-0.4	Alanyl-tRNA synthetase
Translation elongation factors									
m.5155	99	65	1	0.5	4E-05	-1.0	0.01	-1.0	Elongation factor EF-1α/Tu
m.78782	1299	943	1	0.6	2E-10	-0.8	0.009	-0.8	Elongation factor 2
m.66678	89	48	1	0.4	6E-11	-1.2	0.003	-1.2	Mitochondrial elongation factor

ORF Names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$ FC	Adjusted p-value	log$_2$ FC	
Translation elongation factors (continued)									
m.72653	570	272	0.99	0.4	5E-07	-1.4	0.007	-1.4	Mitoch. elongation factor Tu
m.69811	125	83	1	0.5	9E-16	-1.0	0.003	-1.0	Mitoch. elongation factor EF-Ts
Phosphatidylinositol biosynthesis									
m.22046	49	26	1	0.4	7E-11	-1.2	0.003	-1.2	diacylglycerol kinase (putative)
m.92816	34	28	0.1	0.8	0.8	-0.3	0.5	-0.6	diacylglycerol kinase (putative)
m.80525	23	32	0.1	1.1	0.7	0.1	0.6	0.2	diacylglycerol kinase (putative)
m.4043	108	90	0.4	0.6	0.02	-0.6	0.07	-0.6	phosphatidate cytidylyltransferase
m.10411	143	60	1	0.3	2E-09	-1.6	0.004	-1.6	phosphatidylinositol synthase

Abbreviations: TPM, average mRNA Transcripts Per Million; PPDE, Posterior Probability of being Differentially Expressed; Post Fold Change, Posterior Fold Change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt).

4- Discussion

4.1- Molecular Signature in Halophilic Protists

We detected an unusual signature in two halophilic protists (specifically phagotrophic protozoa), whereby cytoplasmic proteins were not highly acidic but showed an increased hydrophilicity. We interpret these observations as indicating that the cytosolic concentration of salt was lower than in typical 'salt-in' microbes, while also suggesting a higher cytosolic salt concentration than in marine organisms. In particular, the relative increase in hydrophilicity suggests that cytoplasmic proteins in halophilic protists have evolved to require fewer hydrophobic interactions to remain folded, as one would expect if the cytosolic salt concentration was higher in these organisms than in mesophiles. As mentioned earlier, although most 'salt-in' strategists have extremely acidic proteomes, while 'salt-out' strategists do not, there are several known exceptions (*e.g.* absence of an acidic signature in the proteomes of the 'salt-in' Halanaerobiales; Elevi Bardavid and Oren 2012). For this reason, the absence of an acidic signature in the predicted proteomes of *H. seosinensis* and *P. kirbyi* strongly suggests a broad 'salt-out' strategy, but is not proof of one. Further, experimental work is needed to test this hypothesis.

We also detected the up-regulation in high salt conditions of genes potentially involved in organic osmolyte metabolism and transport in *H. seosinensis.* This supports the idea that organic solutes contribute to osmotic adjustment in these organisms, although the relative importance of inorganic osmolytes remains to be determined. Obtaining direct experimental evidence for organic osmolyte accumulation would be an important avenue for further research, notwithstanding the technical challenges to working with necessarily non-axenic cultures of small bacterivorous protozoa.

The measured correlation between GRAVY scores and habitat salinity also suggests that the extracellular osmolarity might influence the hydropathicity of the cytoplasmic proteomes of 'salt-out' eukaryotic microbes (taxa in the Mandor dataset). Despite active salt expulsion, higher extracellular salt concentration likely

leads to a relatively higher salt content that the cell has to manage, at least some of the time. For instance, bacterivores presumably bring salt into the endomembrane system during phagotrophy, consequently increasing the risk of salt intrusion in the cytosol.

Contrary to expectations, most predicted exoproteins in the investigated halophilic protozoa did not show the acidic signature of halophilic proteins. This can probably be explained, to a large extent, by the difficulty of predicting secreted proteins in protists. Even when predictions from multiple programs are combined, as we did in this investigation, the accuracy of prediction is ~50% when studying protist sequences (Min 2010). Furthermore, phagotrophs like *H. seosinensis* and *P. kirbyi* do not secrete digestive enzymes outside the cell but rather engulf particles in tightly packed vacuoles. Several proteins predicted to be secreted in halophilic protozoa had functions related to lysosomal activity or to membrane biogenesis, suggesting they were following the secretory pathway but were not exported outside the cell. This is in contrast with osmotrophs like fungi and prokaryotes in which a higher number of secreted enzymes is expected. Interestingly, we actually did detect an acidic signature for the halotolerant/halophilic yeast proteins that we predicted to be secreted.

In addition, it is possible that features other than acidity could represent mechanisms of salt adaptation in proteins. The surfaces of acidic halophilic proteins are enriched in negative charges that are thought to interact with water. Protein glycosylation and phosphorylation could also lead to increased negative surface charges and increased solubility. We did not examine such post-translational modifications and hence we cannot determine if they play a role in salt adaptation.

4.2- Expression of Genes Involved in Organic Osmolyte Metabolism in *Halocafeteria seosinensis*

Analyses of genes that are differentially expressed as a function of salinity in *H. seosinensis* identified candidates for involvement in biosynthesis and transport of organic osmolytes. At high salt, we detected the up-regulation of genes related to

hydroxyectoine synthesis and to transporters of myo-inositol and amino acids, potentially glutamate, glycine, proline or taurine. Glycerol, a common osmolyte in yeasts and certain algae, probably does not accumulate in *H. seosinensis* and *P. kirbyi* since they did not express glycerol phosphatase or known glycerol transporters. This is consistent with the observation that the addition of glycerol in the medium does not improve growth of *H. seosinensis* (Park *et al.* 2006).

The presence of an ectoine hydroxylase with the consensus sequence signature and conserved functional residues, activation of its expression at high salt, and the presence of a bacteria-like ectoine biosynthesis pathway, all suggest that *H. seosinensis* uses and modifies these organic solutes for osmotic adjustment. This hypothesis is surprising *a priori*, since ectoine synthesis has not been reported in eukaryotes to our knowledge (*e.g.* Widderich *et al.* 2014). Nonetheless, we found modest evidence from available transcriptomes that ectoine biosynthesis may be found in some marine protists, including a close relative of *H. seosinensis* (*C. roenbergensis*). However, direct observations of osmolyte accumulation will be necessary to validate these conjectures.

The composition and proportion of accumulated osmolytes commonly vary in microorganisms as a function of growth conditions like external salinity (García-Estepa *et al.* 2006), and it is possible that signatures of these differences might be seen at the transcript level. Knowledge of transcriptional regulation of the ectoine operon in bacteria is still fragmented. Studies of a few bacteria showed that transcription is polycistronic, sometimes regulated by EctR, a repressor protein of the MarR family, and is complex since it involves five putative promoters, including one inside the operon (Calderón *et al.* 2004; Mustakhimov *et al.* 2010). In *H. seosinensis*, RNA read-mapping on a genomic contig indicated that four mRNA transcripts were generated from the *ectABC-ask* cluster of genes (Figure 2.10), but no EctR or MarR domains (PF01047.17, PF12802.2) were detected in any predicted protein sequences. Our transcriptomic experiment detected no differential expression of *ectA* and repression of *ectC*. Regarding *ectA*, it is possible that repression of activity at low salt occurred via allosteric regulation. This possibility

has been discussed in the case of the ectoine biosynthetic enzymes in *C. salexigens* and *Methylomicrobium alcaliphilum* (Calderón *et al.* 2004; Reshetnikov *et al.* 2005). Higher expression of *ectC* at optimal salt concentrations suggests that ectoine, the product of this gene, might be an important organic osmolyte at this salinity, while hydroxyectoine, generated by *ectD*, would become relatively more important at extremely high salinity.

Several lines of evidence suggest that hydroxyectoine may be more beneficial than ectoine when salinity gets very high (*i.e.* in our case 4.7 M NaCl). First, hydroxyectoine has superior desiccation protection properties compared to ectoine (Lippert and Galinski 1992). Furthermore, in *Brevibacterium* sp. JCM6894, accumulation of ectoine starts to plateau or decreases at salinity > 2M while hydroxyectoine starts to accumulate at salinity > 1.5 M (Nagata *et al.* 1996; Nagata *et al.* 2008). In *Brevibacterium linens*, ectoine content also decreases for salinities > 2M NaCl (Bernard *et al.* 1993). In *C. salexigens*, although ectoine content increases with salinity ranging from 0.75 M to 3 M NaCl, the magnitude of the increase is higher for hydroxyectoine compared to ectoine (13.8-fold vs. 2.8-fold increase respectively; García-Estapa *et al.* 2006). In addition, some strains of *Pseudomonas stutzeri* preferentially accumulate hydroxyectoine over ectoine (Seip *et al.* 2011; Stöveken *et al.* 2011).

Hydroxyectoine was reported to accumulate substantially only during the stationary growth phase in *Virgibacillus salexigens*, *Bacillus clarkii* (Bursy *et al.* 2007), *Marinoccocus* sp. (Schiraldi *et al.* 2006), *Streptomyces coelicolor* (Bursy *et al.* 2008) and *Halomonas elongata* (Cánovas *et al.* 1999). Our growth controls indicated that the *H. seosinensis* cultures were at mid-exponential phase at the time of RNA extraction. Therefore, the observed difference in ectoine hydroxylase expression between conditions was not a result of different growth phases. Rather, it suggests that hydroxyectoine contributed to *H. seosinensis* survival in an extreme salt condition (30% ~ 4.7 M NaCl). Furthermore, in the halophilic proteobacterium *C. salexigens*, ectoine and hydroxyectoine accumulate together during the exponential

phase (García-Estepa *et al.* 2006) and the same is observed in *S. coelicolor* when salt is added to the medium (Bursy *et al.* 2008).

The phylogenetic relationship of *H. seosinensis ectABCD* genes with bacterial sequences and the bacteria-like genomic arrangement of *ectABC-ask* suggest that these genes were acquired by lateral gene transfer from bacteria. The wide but sporadic diversity of protists that putatively express ectoine synthetic genes also suggest that these genes might have afterwards spread horizontally between protists. Explaining this distribution pattern by EctABC being ancestral to all eukaryotes and lost multiple times is less parsimonious. More sampling of eukaryotic sequences will shed light on this by either revealing further patchiness of these genes in eukaryotes or, alternatively, showing a broader distribution within eukaryotes.

5- Summary

Our results suggest that *H. seosinensis* and *P. kirbyi* most probably use organic solutes as the main osmolytes while likely experiencing higher intracellular salt content relative to organisms inhabiting marine environments. This is based on the presence of a hydrophilic signature in cytoplasmic proteins and on the expression pattern of genes potentially involved in organic osmolyte synthesis and transport, namely ectoine/hydroxyectoine, myo-inositol and undetermined amino acids. Importantly, future metabolomic investigation is required to directly measure these osmolytes using NMR, mass spectrometry or HPLC. Characterization of enzymatic activity at varying salinities could also help in determining whether cytoplasmic enzymes of halophilic protists function optimally at relatively high salt concentrations.

Chapter 3 - Transcriptomic Response to Hypersaline Stress in the Halophilic Protist *Halocafeteria seosinensis* and Contributions of Differentially Expressed, Duplicated and Laterally Transferred Genes to Survival in High Salt Habitats

1- Introduction

Hypersaline environments are habitats for a variety of micro-organisms, the so-called halophiles, that are adapted to the often-extreme conditions prevailing in these settings. True halophiles require high salt conditions to grow optimally and most of them cannot divide at salt concentrations under ~9%, which is around three times the salinity of seawater (Park *et al.* 2006; Park *et al.* 2007; Cho *et al.* 2008; Park *et al.* 2009; Park and Simpson 2011; Foissner *et al.* 2014). Challenges faced by these life forms include ionic stress (especially the toxicity of sodium and chloride ions), osmotic stress, dehydration/desiccation stress (induced by complete evaporation), and reduced solubility of metabolites including nutrients and oxygen. Halophiles have evolved to overcome these constraints by developing adaptations like amino acid bias in high-salt exposed proteins (Frolow *et al.* 1996; Paul *et al.* 2008), massive synthesis of organic osmolytes (Borowitzka and Brown 1974; Galinski 1995; Oren 2002b), tight regulation of ion homeostasis (Niu *et al.* 1995; Gorjan and Plemenitaš 2006), salt-dependent adjustment of plasma membrane fluidity (Russell 1989; Turk *et al.* 2004; Turk *et al.* 2007; Turk *et al.* 2011), and increased capability to manage oxidative stress (Panda and Das 2005; Petrovič 2006; Tammam *et al.* 2011).

Response to salt stress in eukaryotes has been investigated in various mesophilic organisms, especially in plants. Limited information is available regarding the molecular response of halophilic microbial eukaryotes, with most of

our knowledge coming from the model yeast *Hortaea werneckii, Wallemia ichthyophaga* (another yeast) and species of the chlorophycean algal genus *Dunaliella*. Optimal growth for these organisms occurs at high salt (1.7-2.6 M salt for *Dunaliella salina*, 1.5-3 M for *H. werneckii* and 3.5-4.5 M for *W. ichthyophaga*) and division is extremely slow or non-existent in the absence of salt (Ginzburg *et al.* 1990; Gunde-Cimerman *et al.* 2009; Oren 2014). Genomic and transcriptomic data are available for the yeast isolates and have revealed features of their adaptability to hypersaline environments (Lenassi *et al.* 2013; Zajc *et al.* 2013). In contrast, the *D. salina* nuclear genome-sequencing project is still ongoing (Smith *et al.* 2010) and only limited information on salt-dependent molecular response of *Dunaliella* spp. is available (Kim *et al.* 2010; Zhao *et al.* 2011).

In *H. werneckii*, higher salinities induce the up-regulation of a persistent transcriptional program for specific genes that is substantially controlled by the mitogen-activated protein kinase (MAPK) Hog1 (Vaupotic and Plemenitaš 2007). As the downstream effector of a cascade of MAPKs, Hog1 is the central regulator of the high osmolarity glycerol pathway that regulates the expression of various osmoresponsive genes including those involved in ionic homeostasis, energy metabolism and protein quality control (Plemenitaš *et al.* 2008). For instance, activated Hog1 stimulates the production of glycerol, the main osmolyte in *H. werneckii* (Petrovič *et al.* 2002; Vaupotic and Plemenitaš 2007).

Hog1 also regulates the expression of cation expulsion ATPases, the *ENA* genes, that are essential for survival under salt stress (Proft and Serrano 1999). Intracellular ion content is kept far below the extracellular salt concentration in salt-tolerant and halophilic yeasts. For example, in the halotolerant yeast *Candida salina*, intracellular sodium content is kept at ~30 mM when grown at salinity up to 2 M NaCl and increases to ~125 mM when maintained at a salinity of ~ 4M NaCl (Silva-Graça *et al.* 2003). This steep gradient is primarily maintained by two Na^+-ATPases: ENA1, the expression of which is stimulated upon acute salt exposure, and ENA2, which is overexpressed in salt-adapted cells (Gorjan and Plemenitaš 2006). A potential additional mechanism contributing to cytoplasmic sodium expulsion in

yeast is ion sequestration in vacuoles using the Na$^+$/H$^+$ exchanger NHX1 (Nass *et al.* 1997).

Since salts impact membrane fluidity, adaptation to varying salinities involves lipid composition adjustment (Russell 1989). *Hortaea werneckii* maintains a fluid membrane over a wide range of salinities by keeping a low sterol-to-phospholipid ratio and by decreasing both fatty acyl length and the saturation level of phospholipids (Turk *et al.* 2004). Key features of halotolerance in yeasts involve the preservation of membrane fluidity and a steady intracellular ion content over a broad range of extracellular salt concentrations (Kogej *et al.* 2005; Turk *et al.* 2007).

When exposed to high salt conditions, microorganisms face a significant concomitant challenge: oxidative stress. Several studies in plants indicated that acquisition of salt tolerance might be a consequence of improving resistance to oxidative stress (*e.g.* Hernández *et al.* 1995; Gossett *et al.* 1996; Gueta-Dahan *et al.* 1997; Hernández *et al.* 2000). Concordantly, the ability of *H. werneckii* to manage oxidative stress appears to be accentuated, since its capability to degrade hydrogen peroxide over a wide range of salinity is as high, or even higher, than that of *Saccharomyces cerevisiae* stressed by exposure to 3% salt (Petrovič 2006). Similarly, stress-response genes represent a major class of genes that are up-regulated during growth at 4.5 M NaCl in *Dunaliella* sp. (Kim *et al.* 2010). Therefore, the capacity to survive in extremely high salt habitats depends on the ability to manage reactive oxygen species by synthesizing antioxidants or expressing molecular chaperones like heat shock proteins Hsp70 and Hsp90 that are up-regulated at high salt in *H. werneckii* (Vaupotic and Plemenitaš 2007).

Compared to the halophilic yeasts and algae, virtually no information is available on the molecular adaptation of halophilic heterotrophic protists (*i.e.* protozoa). The feeding mode of halophilic protozoa, which phagocytose particles (typically prokaryotes), differs greatly from yeasts, which are osmotrophs. In addition, phagotrophic protozoa are typically not surrounded by a cell wall. These differences between yeasts and protozoa are likely to lead to very different salt

adaptation strategies.

Among described halophilic protozoa, the bicosoecid stramenopile
Halocafeteria seosinensis was first isolated from a 30% salt Korean saltern, and has
been frequently observed in hypersaline water samples from various geographical
locations (Park and Simpson 2015). Strain EHF34 grows optimally at 15% salt (2.5
M salt), still divides at 30% salt, although very slowly (38 h/division), but cannot
grow at salinities <7.5% (Park *et al.* 2006). Analysis of its inferred cytoplasmic
proteome revealed a molecular signature suggestive of a higher intracellular salt
content than in marine protists (see Chapter 2). This was also detected in the
halophilic heterolobosean *Pharyngomonas kirbyi*, suggesting this property might be
typical for halophilic protozoa. At high salt, *H. seosinensis* up-regulates genes whose
product are potentially involved in osmolyte synthesis and transport, namely
ectoine hydroxylase, amino acid and myo-inositol transporters, suggesting it might
use organic solutes to reach osmotic equilibrium (see Chapter 2).

Hypersaline habitats represent suitable systems to study adaptation,
especially since microbial communities were shown to have evolved faster in
extreme environments (Li *et al.* 2014). Gene duplication is an important mechanism
for creating genetic novelty. After duplication, one copy of a gene may experience
relaxed evolutionary constraints, increasing the chance for it to acquire a new
function (Long *et al.* 2003). A possible fate of duplicated genes is to retain function
while acquiring enzymatic specialization to varying ecological conditions, a concept
termed ecoparalogy (Sanchez-Perez *et al.* 2008). Several cases have been
documented in halophilic prokaryotes where ecoparalogs perform the same
function but are expressed under different salinities, for example, two glutamate
dehydrogenases in *Salinibacter ruber* (Bonete *et al.* 2003) and three chaperonins
(CCT1-3) in *Haloferax volcanii* (Kapatai *et al.* 2006). *In silico* analysis of the *S. ruber*
genome identified many ecoparalog candidates that were especially enriched in
outer membrane and periplasmic space proteins that are exposed to contrasting salt
conditions (Sanchez-Perez *et al.* 2008). This exemplifies the potential benefit of this
mechanism for the evolution of halophiles.

In halophilic yeasts, expansion of gene families for cation transporters, P-type ATPases and hydrophobins was observed (Lenassi *et al.* 2013; Zajc *et al.* 2013). Duplication of ion transporters could allow a greater potential for adapting to varying salt conditions (Lenassi *et al.* 2013). Duplication of hydrophobins in *W. ichthyophaga* might have favored adaptation to hypersaline environments, since these small proteins modulate cell wall permeability and contribute to cell aggregation into protective many-celled structures (Zajc *et al.* 2013).

In prokaryotes, lateral gene transfer (LGT) is accepted as another important mechanism that has the potential to increase the fitness of the recipient cell (Battistuzzi and Brown 2015). For example, in *S. ruber* the acquisition of genes of the K^+ uptake system from Haloarchaea might have mediated adaptation to its hyperhalophilic lifestyle (Mongodin *et al.* 2005). Notably, salt can stabilize naked DNA and protect it from thermodegradation, thus favoring environmental preservation of the raw material for LGT (Marguet and Forterre 1994, 1998; Tan and Chen 2006; Borin *et al.* 2008). Concordantly, Rhodes *et al.* (2011) noted a fundamental difference in the origin of laterally transferred genes between thermophiles and halophiles, where the former primarily had thermophilic LGT donors while LGT events into halophiles originated mostly from non-halophilic taxa. Based on the assumption that free DNA is more stable in hypersaline environments and that these habitats often occupy topographic minima where biological material accumulates over time, Rhodes *et al.* (2011) theorized that halophiles might be exposed to free DNA from a more diverse range of organisms. This suggests that hypersaline environments might represent an unusually favorable habitat for the occurrence of LGT.

Although the importance of LGT to eukaryotic evolution is still debated (Ku *et al.* 2015), its incidence in microbial eukaryotes has been increasingly documented (Keeling and Palmer 2008; Andersson 2009; Soucy *et al.* 2015). Several cases have been reported where LGT was inferred as a probable driver of niche adaptation, including adaptation to anaerobic and parasitic lifestyles, and to rumen, sea ice and soil habitats (Andersson *et al.* 2003; Richards *et al.* 2003; Eichinger *et al.* 2005;

Ricard *et al.* 2006; Andersson *et al.* 2007; Belbahri *et al.* 2008; Raymond and Kim 2012; Takishita *et al.* 2012; Schönknecht *et al.* 2014; Stairs *et al.* 2014; Nývltová *et al.* 2015; Xu *et al.* 2016). Furthermore, in the polyextremophile red alga *Galdieria sulphuraria*, adaptation to extreme environments was likely facilitated by lateral acquisition of genes coding for ion transporters, osmolyte synthesizers, and toxic metal pumps and neutralizers (Schönknecht *et al.* 2013). Interestingly, *H. seosinensis* potentially acquired via LGT all the genes necessary for production of the osmolyte ectoine (see Chapter 2). These genes were arranged on the genome as a gene cluster, similarly to the ectoine synthesis bacterial operon, and were phylogenetically closely related to the bacterial enzymes. Although production of ectoine and hydroxyectoine in *H. seosinensis* still awaits experimental confirmation, this LGT event likely contributed to salt adaptation in *H. seosinensis*, especially since the expression of some of these genes was salt-responsive. Overall, although the molecular mechanisms for acquisition and genomic insertion of foreign genes have yet to be described, these results suggest that LGT is an important mechanism for adaptation of microbial eukaryotes to new environments.

In this chapter, I present the results of a comprehensive analysis of *H. seosinensis* transcriptomes generated under contrasting salinities. I report on the long-term transcriptional program of salt-adapted cells, with a special emphasis on genes that were up-regulated at high salt. I also identify gene duplication candidates, including ecoparalogs, and probable LGT events that potentially contributed to the halophilicity of *H. seosinensis*. Given the recent explosive growth in the availability of molecular data, I also took the opportunity to describe commonalities and differences of *H. seosinensis* relative to other protists.

The ability to predict gene function solely based on sequence information is limited, especially due to the great evolutionary distance between most protists and the best-studied model organisms. In this context, differential expression studies are extremely helpful in order to flag genes with important physiological roles (*e.g.* Diray-Arce *et al.* 2015). The aim of this chapter is therefore to take advantage of this powerful approach and explore *H. seosinensis* genetic space and its expression as a

function of salinity. This technique allowed me to identify candidate systems in cell physiology (among thousands of possibilities) that were responsive to a change in extracellular salinity, and subsequently to build models for future testing.

2- Materials and Methods

2.1- RNA and DNA Extraction and Sequence Generation

Halocafeteria seosinensis transcriptomic sequences were generated as described in Chapter 2. Briefly, RNA was extracted from mid-exponential cultures of *H. seosinensis* strain EHF34 (Park *et al.* 2006) grown in triplicate at 15% and 30% salt, and fed with *Haloferax* sp. RNA was extracted using TRIzol (Rio *et al.* 2010) and treated with Turbo DNAse (Ambion) prior to cDNA library preparation using the TruSeq RNA sample preparation kit version 2 (Illumina) that included a poly-A tail purification step for eukaryotic messenger RNA (mRNA) enrichment. *S*amples were sequenced on a HiSeq platform by Génome Québec. Reads were trimmed to remove low-quality sequences using Trimmomatic v. 0.30 (Bolger *et al.* 2014) and mapped to genomes of food prokaryotes known to be in the culture in order to discard contaminant sequences, using Stampy 1.0.23 (Lunter and Goodson 2011). Reads were then assembled using Trinity 2.0.2 (Grabherr *et al.* 2011) and open-reading frames (ORFs) were predicted using TransDecoder (included in the Trinity package). Nucleotide sequences were compared to each other using BLASTN (Altschul *et al.* 1990) and ORFs sharing identical regions at least 50 nucleotides long were considered alternative spliced isoforms of the same gene. Finally, to remove sequences belonging to unknown prokaryotic contaminants present in the cultures, the nucleotide sequences of ORFs were compared to sequences in the NCBI Nucleotide collection (NT) database using BLASTN. Sequences having >100 bp-long fragments >90% identical to a prokaryotic sequence were discarded.

DNA was purified using a salt extraction protocol (Aljanabi and Martinez 1997). Briefly, cells were disrupted by vortexing in lysis buffer and digested with

proteinase K (0.2 mg/mL) and 0.01% sodium dodecyl sulfate. DNA was separated from the other organic phases by centrifugation in a supersaturated NaCl solution (3 M) and precipitated with 70% ethanol. A paired-end DNA library (250 bp) was prepared using the Nextera XT kit (Illumina) prior to high-throughput sequencing on a MiSeq platform. Reads were cleaned of low quality sequences and from prokaryotic contaminants as performed for the RNA sequences. Genomic contigs were assembled with MIRA 4.9.5_2 (Chevreux *et al.* 2004). Genes, including intron/exon boundaries, were delimited by mapping the read sequences obtained from RNA extracts using TopHat2 (Kim *et al.* 2013) and predicted by Braker 1.1 (Hoff *et al.* 2016).

2.2- Gene Annotation

Predicted proteins were annotated using the eggNOG 4.1 database (Powell *et al.* 2014) through hidden Markov model searches (E value < 0.00001) using the hmmsearch program of the HMMER package (Eddy 1998). Further protein domain characterization was done by interrogating the Pfam 27.0 (Finn *et al.* 2016) and TIGRFAMs (Haft *et al.* 2003) databases again using hmmsearch, and the NCBI conserved domain database (Marchler-Bauer *et al.* 2015) using the BLAST algorithm. Proteins were also assigned to KEGG pathways (Kanehisa *et al.* 2016) by the KEGG Automatic Annotation Server (Moriya *et al.* 2007) using the representative set for genes through the single-directional best hit method.

In specific cases, putative functions of *H. seosinensis* proteins were investigated further by inspecting multiple sequence alignments for conserved functional residues and by performing phylogenetic analyses. In these instances, searches against the NCBI non-redundant (NR) database and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Keeling *et al.* 2014) were done using BLAST to gather homologous genes. Sequences were aligned using MAFFT 7.205 (Katoh *et al.* 2002) and resulting alignments were trimmed using BMGE 1.1 (Criscuolo and Gribaldo 2010). Maximum-likelihood phylogenetic trees were built from five starting trees using RAxML 8.1.22 (Stamatakis *et al.* 2005) with

the PROTGAMMALG4X model of amino acid substitution. Bootstrap support was calculated from 100 replicates.

Protein features like transmembrane regions and signal peptides were detected to increase confidence in annotations, in cases where these characteristics were reported for the described proteins. Targeting signals were predicted using TargetP 1.1 (Emanuelsson *et al.* 2000) and mitoprot II 1.101 (Claros and Vincens 1996). Sequences positive for signal peptide detection were investigated further for the presence of the endoplasmic reticulum (ER) retention signals KK, KxK, KDEL or HDEL at the C-terminus or RR at the N-terminus. Transmembrane domains were predicted using TMHMM 2.0 (Krogh *et al.* 2001) and HMMTOP 2.0 (Tusnady and Simon 2001).

2.3- Differential Gene Expression Assessment

Gene expression at optimal and maximal salt concentrations was quantified using RSEM (Li and Dewey 2011). Briefly, forward sequence reads from each replicate were mapped to the Trinity assembly using Bowtie 2 v.2.2.4 (Langmead *et al.* 2009). Reads mapping to multiple isoforms were assigned proportionally to the number of reads mapping to unique regions of the said isoforms. After removal of ORFs having low read counts in all samples (75th quantile < 10 reads), differential expression was assessed using three independent software programs: the empirical Bayesian analysis tool EBSeq following 10 iterations (Leng *et al.* 2013), DESeq2 (Love *et al.* 2014) and the limma package (Smyth 2004; Ritchie *et al.* 2015) after normalization using the Voom method (Law *et al.* 2014)*.* P-values were corrected for multiple testing using the Benjamini-Hochberg method. In the text, I refer to the results of EBSeq, although I also report the respective predictions from limma and DESeq in tables. ORFs were considered differentially expressed if their posterior probability was above 0.95 (or adjusted p-value < 0.05) and posterior fold change (FC) <0.5 or >2 (*i.e.* $\log_2$FC <-1 or >1). ORFs were ordered according to the number of transcripts per million (TPM) averaged over replicates of the condition of interest, and ranks are reported from the highest to the lowest abundance.

2.4- Identification of Duplicated Genes and Ecoparalogs

A local database containing protist protein sequences from the MMETSP dataset (MarProt taxa listed in Table 2.1) and from published genomes (Table 3.1) was constructed in order to identify recently duplicated genes. *Halocafeteria seosinensis* sequences were added to this database after selecting the longest isoform for each gene. Following BLASTP comparison using *H. seosinensis* sequences as queries, sequences that were more similar to other *H. seosinensis* sequences than to other protist sequences were classed as candidate recent duplicates. For these, additional homologous sequences were gathered by BLASTP comparison against the NR database if the alignment covered 2/3 of the smallest sequence for sequences >30% identical. For each gene cluster, sequences were aligned and trimmed as described in section 2.2. Preliminary maximum-likelihood phylogenetic trees were inferred using FastTree 1.0.1 (Price *et al.* 2009). Trees for which *H. seosinensis* sequences clustered in a clade to the exclusion of sequences from other organisms were selected for more in-depth phylogenetic analysis using RAxML, using the PROTGAMMALG4X model of sequence evolution (50 independent starting trees for ML tree search; Bootstrap support was evaluated using 100 replicates). *Halocafeteria seosinensis* sequences that clustered exclusively in a clade with bootstrap support >50% were considered candidate gene duplicates. Since *Cafeteria roenbergensis* was the closest relative of *H. seosinensis* in the database, cases where *H. seosinensis* sequences clustered together next to a *C. roenbergensis* sequence represented examples for which I had the greatest confidence for relatively recent gene duplications. Clusters that included genes with contrasting expression values (*i.e.* one being up-regulated and the other down-regulated) were examined as ecoparalog candidates.

As controls, I performed the same analysis on proteins predicted from the genomes of *Dictyostelium discoideum* (GCF_000004695.1), *Guillardia theta* (GCF_000315625.1), *Nannochlorospis gaditana* (GCA_000240725.1), *Salpingoeca rosetta* (GCF_000188695.1), *and Thalassiosira pseudonana* (GCA_000149405.2).

Table 3.1. Genomes used to identify gene duplication candidates.

Organisms	GenBank assembly accession
Aureococcus anophageferrens	GCA_000186865.1
Blastocystis hominis	GCA_000151665.1
Ectocarpus siliculosus	GCA_000310025.1
Nannochloropsis gadita	GCA_000240725.1
Phaeodactylum tricornutum	GCA_000150955.2
Phytophthora infestans	GCA_000142945.1
Phytophthora sojae	GCA_000149755.2
Phytophtora parasitica	GCA_000247585.2
Thalassiosira pseudonana	GCA_000149405.2

2.5- Gene Enrichment Analysis

The relative abundance of genes assigned to Clusters of Orthologous Group of proteins (COG) was analyzed using STAMP v.2.1.3 (Parks *et al.* 2014) to determine if they were enriched in differentially expressed genes or duplicated genes. Significant variations in proportions were assessed with the hypergeometric test followed by multiple test correction by the Benjamini-Hochberg method. Prior to analysis, I removed the following COG classes (which contained very few genes or genes without obvious biological significance or known function) to decrease their influence on the multiple test correction: 'Cell motility', 'Defense mechanisms', 'Unknown function', 'General predictions only', 'Nuclear structures', and 'No hits found'.

3- Results and discussion

3.1- Data Overview

Halocafeteria seosinensis expressed 16,852 non-redundant ORFs corresponding to 12,020 genes. 9,026 genes had no spliced variants and 2,994 genes had alternatively spliced isoforms (median of 2 isoforms/gene). 1,657 ORFs were discarded during differential expression assessment due to low expression in all conditions (see section 2.3).

3.1.1- Gene Expression Analysis

EBSeq flagged 2,871 ORFs as being differentially expressed, of which 62% were up-regulated at high salt and 38% were down-regulated (Figure 3.1). BLASTP searches indicated that 45% of these ORFs had no homologue in the NR database (E value cutoff = 0.00001). Comparatively, DESeq2 and limma detected 3,265 and 2,882 differentially expressed ORFs respectively (adjusted p-value < 0.05). There was good agreement between the analyses: 2,418 ORFs were identified as differentially expressed by all three analyses, and the great majority of the ORFs that were called as differentially expressed by EBSeq were also identified as such by limma and DESeq2 (87% and 90% respectively). For the sake of succinctness and because predictions from EBSeq were the most conservative, the results from the latter program are given in the text, and respective predictions from limma and DESeq2 can be found in tables.

COG classes containing genes involved in metabolism and transport of inorganic ions were significantly enriched in differentially expressed genes (although with relatively low support, adjusted p-value = 0.050), while classes containing genes involved in translation, RNA processing and cytoskeleton were significantly depleted of differentially expressed genes (adjusted p-values < 7.3 × 10^{-3}, Figure 3.2). This suggested that, while core genes involved in basic cellular functions tended to not be differentially expressed, genes involved in ion homeostasis as a
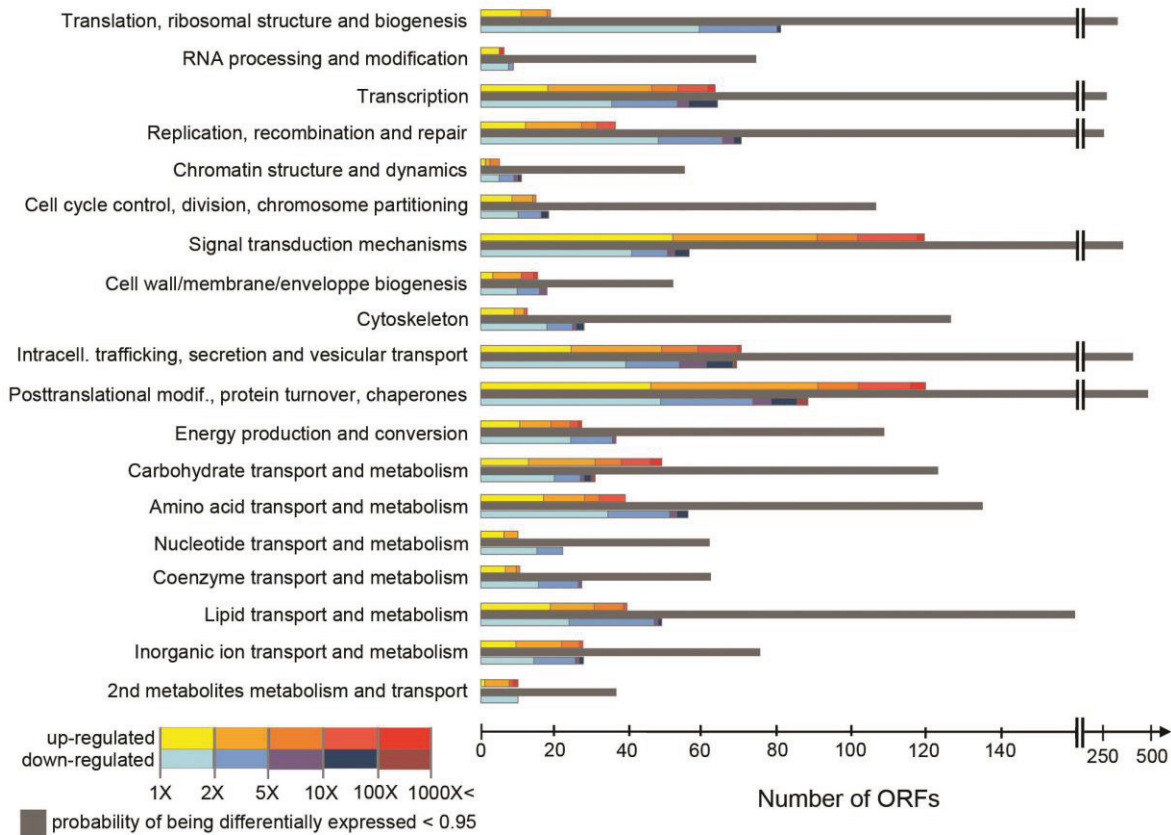
Fig. 3.1. Numbers of ORFs assigned to each COG class given their gene expression values. For each class, the first and the last bars indicate the number of ORFs up-regulated and down-regulated at high salt respectively (posterior probability of being differentially expressed (PPDE) > 0.95) given their posterior fold change (indicated by the colors), and the grey bar in the middle shows the number of ORFs with PPDE < 1.
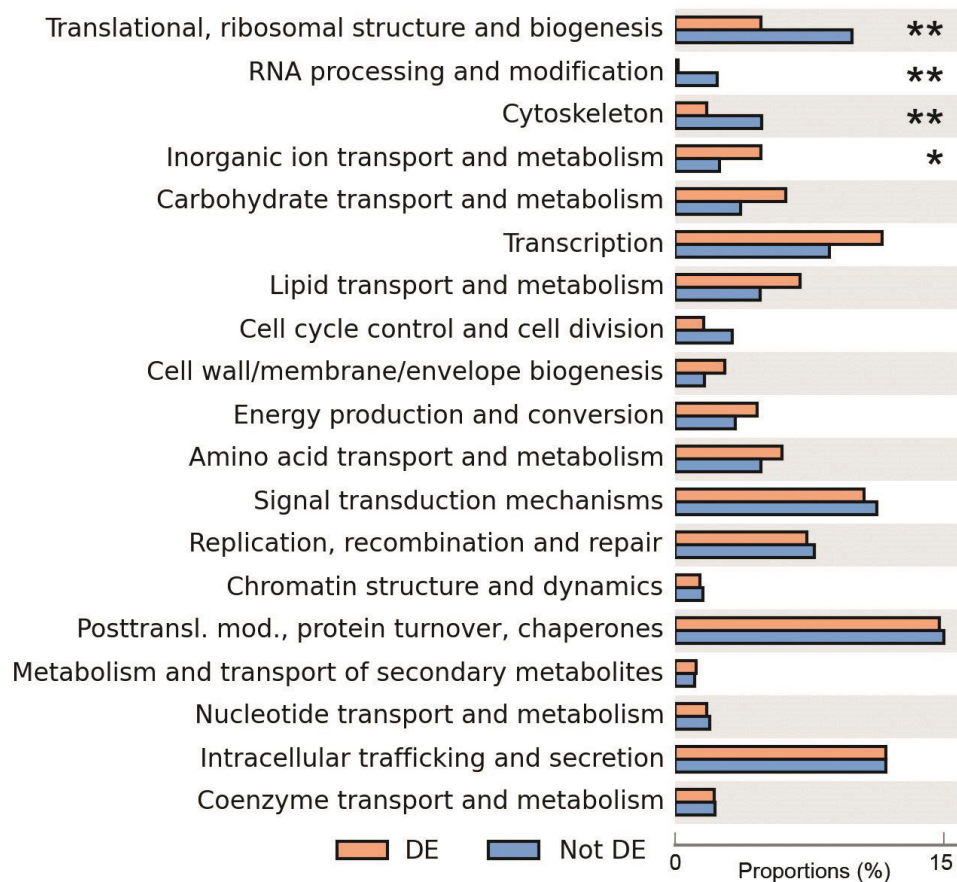
Fig. 3.2. Enrichment analysis of differentially expressed genes. The proportions of differentially expressed genes assigned to each class are shown in orange, percentages of non-differentially expressed genes are in blue. The number of asterisks indicates adjusted p-values after multiple test correction using the Benjamini-Hochberg method (***: adjusted p-value < 0.001, **: adjusted p-value < 0.01, *: adjusted p-value < 0.05).

group responded to a variation in extracellular salinity. COG categories that did not show a significant bias also contained informative genes in terms of response of *H. seosinensis* to high salt. Discussion about genes assigned to these classes is included in this chapter.

3.1.2- Gene Duplication Analysis

The gene duplication analysis revealed 494 clusters containing 1,652 genes. The bulk of these, 317 clusters encompassing 1,086 genes, contained only *H. seosinensis* sequences (*i.e.* no homologs were detected in other organisms). The remaining 153 clusters contained 444 genes where *H. seosinensis* sequences clustered together to the exclusion of all other homologous sequences gathered from the local protist database and the NR database. For the sake of stringency, I considered only these latter examples of (recent) gene duplication candidates in further analyses.

After removing uninformative COG categories for assessment of classes enriched in gene duplication (as described in section 2.5), 230 genes were detected as duplicated in *H. seosinensis*, from a total of 4,283 genes assigned to these categories (6%). Similar proportions were recovered from the genomes of *G. theta* (9%) and *D. discoideum* (6%), while the genomes of *S. rosetta*, *T. pseudonana* and *N. gaditana* had a lower percentage of duplicated genes under these criteria (2% for both)*.* Enrichment analysis showed that COG categories representing housekeeping genes ('Translational, ribosomal structure and biogenesis', 'Transcription' and 'Cytoskeleton') were depleted of duplicated genes in *H. seosinensis* (adjusted p-values < 4.5×10^{-3}, Figure 3.3). The same trend was observed in the genomes of *N. gadita*, *G. theta* and *D. discoideum*. Conversely, categories including genes involved in metabolism and transport of amino acids and inorganic ions were enriched in duplicated genes, as were genes involved in intracellular trafficking of metabolites like phospholipids (adjusted p-values < 4.3×10^{-3}). The enrichment analysis of duplicated genes indicated that core genes seemed more evolutionarily 'stable'
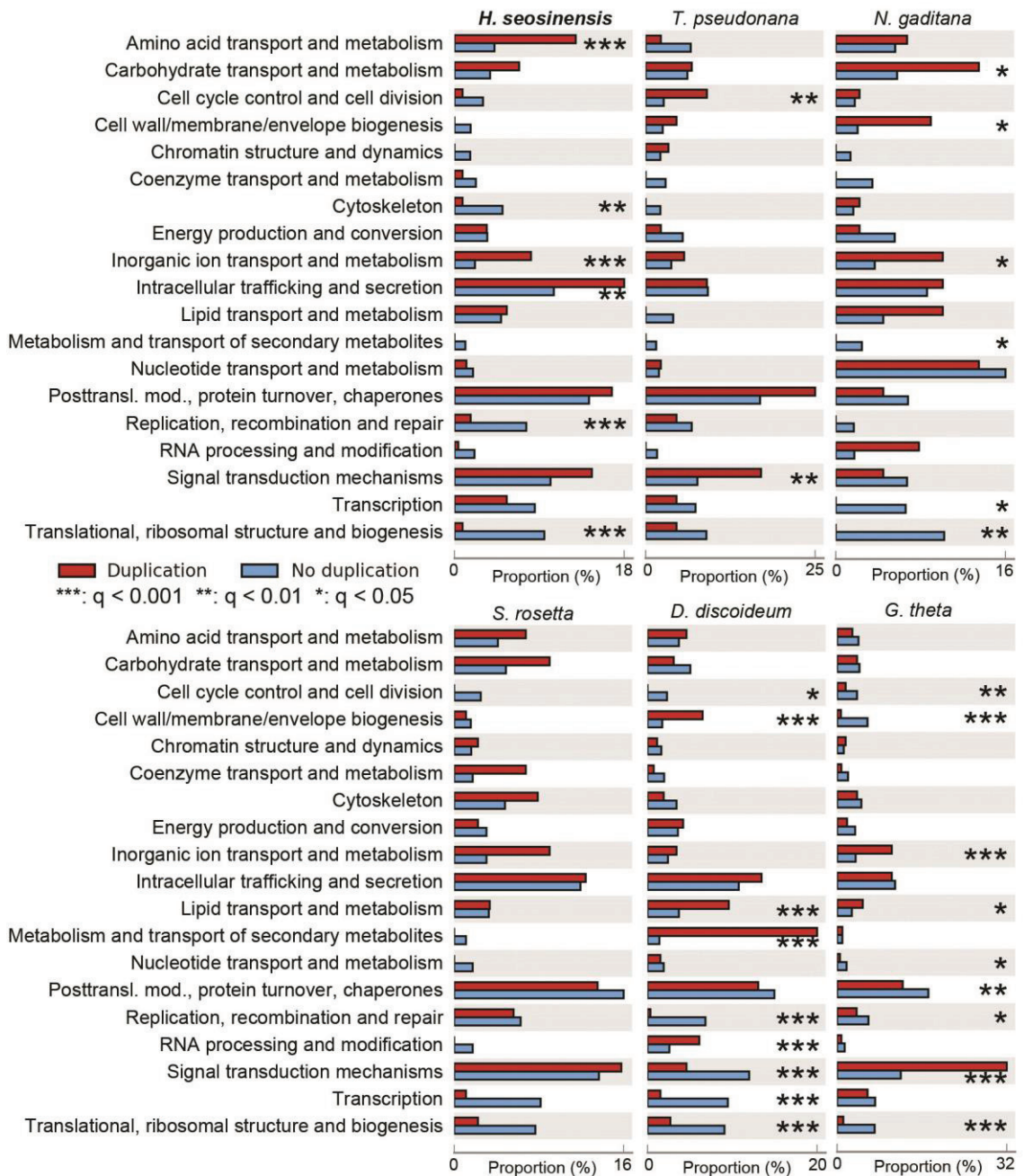
Fig. 3.3. Gene duplication enrichment analysis performed on the transcriptome of *H. seosinensis,* and the genomes of *T. pseudonana*, *N. gaditana, S. rosetta, D. discoideum* and *G. theta.* The proportions of duplicated genes assigned to each COG class are shown in red. The percentage of the remaining genes is in blue. The number of asterisks indicates the false discovery rate after multiple test correction using the Benjamini-Hochberg method (***: adjusted p-value < 0.001, **: adjusted p-value < 0.01, *: adjusted p-value < 0.05).

compared to metabolic genes, echoing the enrichment analysis of differentially expressed genes, where core genes involved in basic cellular functions were transcriptionally steadier (see section 3.1.1). The same kind of principle was described in fungi, where genes essential in growth processes had more stable copy numbers and expression, while accessory genes were more 'volatile' in this regard (Wapinski *et al.* 2007).

3.1.3- Overview of Sections 3.2-3.7

In the following sections, I first describe the potential contributors to ion homeostasis in *H. seosinensis* and discuss the enrichment of duplicated ion transporters (section 3.2). In section 3.3, I explain how oxygen level might have influenced gene expression, since salinity affects oxygen concentration. In section 3.4, I define the actors of the signal transduction pathways and types of transcription factors that may orchestrate *H. seosinensis* gene expression response. In section 3.5, I demonstrate that a major aspect of *H. seosinensis* salt response is related to stress management. In section 3.6, I illustrate how the transcriptional program for lipid metabolism was affected by high salt conditions. Finally, in section 3.7, I suggest explanations for the upregulation of carbohydrate-related metabolic genes.

3.2- Ion Homeostasis in *Halocafeteria seosinensis*

Establishing ion gradients is a universal requirement for cells. Halophiles and mesophiles keep their intracellular Na^+ content low relative to the extracellular milieu while keeping intracellular K^+ higher than outside the cell. Two molecular models have been described to establish such gradients, both using the energy generated by hydrolysis of ATP via P-type ATPases. The first one, described in animal cells, involves Na^+/K^+-ATPases that directly concentrate each ion on either side of the membrane (Morth *et al.* 2011). The second one, described in fungi, plants and algae, involves the establishment of a proton gradient that is subsequently used to generate the respective ion gradients using Na^+/H^+ and K^+/H^+ transporters (Rodríguez-Navarro 2000). Recently, evidence has accumulated favoring a third

model that combines both mechanisms, based on organisms that express both types of ATPases, Na^+/K^+ and H^+, as in protists like *Dictyostelium, Heterosigma, Chlamydomonas* and *Phytophthora* (Barrero-Gil *et al.* 2005).

As described below, in *H. seosinensis*, I obtained moderate evidence for the expression of Na^+/K^+-ATPases, but observed the differential expression of a Na^+/H^+ antiporter, indicating that proton gradients might also be important in *H. seosinensis* ion homeostasis. In this regard, *H. seosinensis* seemed to employ the third model described above, like other protists.

3.2.1- Sodium Expulsion Through Na^+/H^+ Antiporters

Na^+/H^+ antiporters are key contributors to salt resistance in plants (Blumwald *et al.* 2000). *Halocafeteria seosinensis* expressed two genes (ORFs m.11942 and m.85102) that were related to the plasma membrane Na^+/H^+ antiporter Salt-Overly-Sensitive 1 (SOS1). Importantly, one of them, m.85102, was 2.7-fold upregulated at high salt (Table 3.2). As in SOS1, these two proteins in *H. seosinensis* had a transmembrane N-terminal region homologous to the NhaP domain (PFAM00999), responsible for ion transport, and a C-terminal region distantly related to cyclic nucleotide binding domain (CNBD, PFAM00027). In plants the latter domain is involved in regulating the transporter activity via phosphorylation by serine/threonine kinase SOS2. The NhaP domain in *H. seosinensis* sequences contained aspartate and arginine residues essential for ion binding and translocation, supporting the inferred function for these transporters (Figure 3.4; Hellmer *et al.* 2003). Interestingly, *H. seosinensis* NhaP domains were more conserved than the CNBD when compared to *Arabidospis thaliana* SOS1 sequence (~40% and ~25% identity respectively). The CNBD was relatively divergent and residues shown to be involved in regulation in plants were not conserved in *H. seosinensis* (data not shown). These included the phosphorylated serine, the motif recognized by SOS2, and the auto-inhibitory sequence, suggesting that the transporters in *H. seosinensis* are regulated differently (Quintero *et al.* 2011).

Table 3.2. Expression of genes involved in ion homeostasis in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
Na$^+$/H$^+$ antiporter								
m.85102	8.7	29.5	1.00	2.67	0.00	1.40	0.00	1.44
m.11942	12.2	19.7	0.05	1.28	0.34	0.35	0.20	0.36
m.6801	2.3	7.1	0.22	2.64	0.15	1.19	0.06	1.60
m.14957	8.7	15.2	0.08	1.42	0.30	0.49	0.22	0.47
m.41235	31.7	20.6	1.00	0.50	0.00	-0.99	0.00	-0.97
m.64236	32.6	33.3	1.00	0.79	0.01	-0.34	0.04	-0.32
m.43490	61.3	119.9	1.00	1.50	0.00	0.59	0.00	0.61
P-type ATPase IIC (Na$^+$/K$^+$)								
m.3798	29.2	26.0	0.95	0.69	0.04	-0.51	0.04	-0.50
m.79586	15.8	18.8	0.04	0.91	0.60	-0.12	0.58	-0.09
P-type ATPase IB (Cu^{2+})								
m.46283	23.8	64.9	1.00	2.13	0.00	1.10	0.00	1.14
P-type ATPase IIA (Ca^{2+})								
m.52246	204.8	189.5	1.00	0.70	0.00	-0.50	0.07	-0.49
m.75748	17.1	33.9	0.05	1.59	0.18	0.65	0.19	0.73
P-type ATPase IIB (Ca^{2+})								
m.1354	0.5	0.6	0.00	0.98	1.00	-0.01	0.99	0.01
m.11968	104.6	166.8	1.00	1.23	0.01	0.30	0.11	0.32
m.58046	60.3	109.4	0.99	1.43	0.00	0.52	0.04	0.56
m.39471	25.7	46.9	0.04	1.45	0.21	0.53	0.20	0.61
m.21843	47.3	25.5	0.99	0.43	0.00	-1.20	0.01	-1.17

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC
P-type ATPase IIIA (H⁺)								
m.92338	197.5	193.5	0.03	0.79	0.40	-0.32	0.52	-0.25
P-type ATPase V (cation)								
m.10985	4.6	10.6	0.10	1.85	0.15	0.84	0.21	0.86
Sodium/calcium exchanger protein								
m.16057	16.4	17.9	0.67	0.84	0.29	-0.24	0.30	-0.24
m.92548	19.3	42.4	0.99	1.72	0.00	0.79	0.03	0.83
m.54532	32.6	25.7	1.00	0.62	0.00	-0.69	0.01	-0.67
m.58047	11.1	15.6	0.04	1.09	0.45	0.13	0.44	0.14
m.23282	9.1	7.1	1.00	0.61	0.01	-0.70	0.03	-0.74
Ca²⁺-activated potassium channel								
m.18644	9.2	20.4	0.23	1.69	0.05	0.74	0.12	0.72
m.35725	7.9	13.4	0.05	1.32	0.32	0.39	0.37	0.36
m.57786	4.2	4.9	0.13	0.89	0.72	-0.16	0.49	-0.27
Cyclic-nucleotide gated cation channels								
m.28289	5.2	15.0	1.00	2.23	0.00	1.14	0.00	1.13
m.80954	12.7	19.9	0.89	1.23	0.11	0.30	0.06	0.33
m.45596	15.0	35.6	0.26	1.93	0.04	0.91	0.03	0.96
Sodium, potassium or calcium channels								
m.32606	7.9	5.7	0.36	0.59	NA	-0.72	0.09	-0.74
m.61852	1.2	0.3	1.00	0.18	0.00	-2.33	0.00	-2.79
m.10834	6.8	10.6	0.07	1.22	0.42	0.28	0.34	0.24
m.27065	9.3	15.2	0.08	1.27	0.31	0.35	0.24	0.31
m.27029	14.6	13.9	0.23	0.75	0.31	-0.40	0.17	-0.42
m.74753	4.1	22.7	1.00	4.34	0.00	2.06	0.00	2.12

	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
ORF names	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
Sodium, potassium or calcium channels (continued)								
m.74752	7.3	8.2	0.06	0.89	0.75	-0.16	0.63	-0.17
Chloride channels								
m.1669	8.4	19.1	0.92	1.82	0.01	0.85	0.01	0.87
m.79946	10.6	39.0	0.98	3.09	0.00	1.53	0.01	1.73
m.57330	13.0	17.3	0.05	0.99	0.98	0.01	0.75	0.06
m.84325	22.7	34.7	0.05	1.19	0.35	0.26	0.20	0.27
m.65420	24.2	14.2	0.86	0.44	0.00	-1.17	0.00	-1.13
m.79944	41.2	60.6	0.06	1.12	0.75	0.17	0.56	0.22
m.16240	70.1	70.0	0.07	0.77	0.28	-0.36	0.21	-0.32
Chloride channels? (related to Tweety transmembrane proteins)								
m.48253	0.1	36.0	1.00	273.91	0.00	7.89	0.00	8.44
m.36773	2.6	33.9	1.00	10.08	0.00	3.32	0.00	3.41
m.19882	160.3	178.2	0.03	0.86	0.35	-0.22	0.40	-0.20
m.83902	58.1	63.1	0.75	0.83	0.08	-0.26	0.21	-0.24
m.92115	34.1	28.2	0.46	0.61	0.02	-0.69	0.08	-0.65
m.84735	97.7	1.0	1.00	0.01	0.00	-6.88	0.00	-6.96
m.27101	0.6	0.3	0.00	0.40	0.16	-1.15	0.04	-1.22
m.19880	0.4	0.7	0.18	1.42	0.67	0.41	0.35	0.65
m.56744	20.0	0.4	1.00	0.01	0.00	-5.85	0.00	-6.08
m.67887	14.1	19.8	0.05	1.09	0.43	0.13	0.28	0.16

Abbreviations: TPM, averaged transcripts per million (at 15% or 30% salt); PPDE, Probability of being Differentially Expressed; Post Fold Change, posterior fold change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.
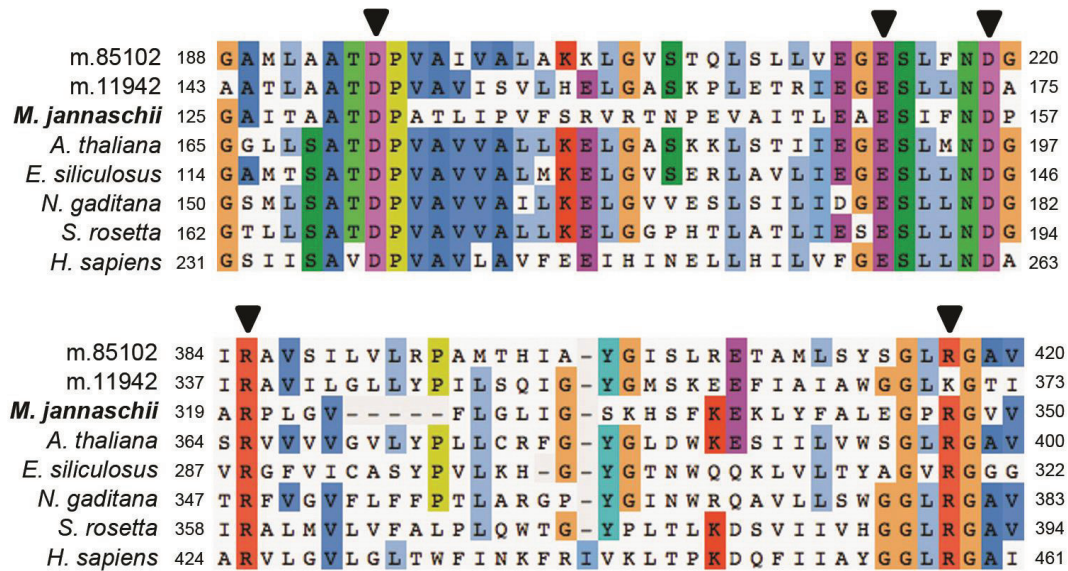
Fig. 3.4. Partial alignment of Na$^+$/H$^+$ antiporter protein sequences indicating conservation of crucial residues for catalytic activity (triangles) based on study of *Methanoccocus jannaschii* enzyme (in bold, 4CZB). The alignment also includes sequences related to *H. seosinensis* sequences (m.85102, m.11942) from *A. thaliana* (CCH26571.1), *Ectocarpus siliculosus* (CBJ26919.1), *Nannochloropsis gaditana* (EWM21970.1), *Salpingoeca rosetta* (XP_004995575.1) and from *Homo sapiens* (P19634.2) as a distant relative.

95

A search for SOS2 homologs in the *H. seosinensis* transcriptome revealed many distant candidates, as was commonly the case for protein kinases.

In plants, SOS1 contributes to Na^+ expulsion from the cytosol, including in the salt-resistant halophyte *Thellungiella salsuginea* (Oh *et al.* 2009). The halophilic alga *D. salina* also increases the expression of a Na^+/H^+ antiporter as a response to increased salinity (Katz *et al.* 1992). *Halocafeteria seosinensis* sequences encoded predicted signal peptides (predicted from genome-derived sequence for ORF m.85102) suggesting that they act somewhere along the secretory pathway, *i.e.* at the plasma membrane and/or in vacuoles (in which case they might pump sodium into vacuoles for later removal from the cell).

Interestingly, no *ENA* genes (P-type ATPase IID), which contribute to sodium expulsion in yeasts, were detected in *H. seosinensis*. Five other genes containing NhaP domains were expressed by *H. seosinensis* and are thus also potential contributors to sodium homeostasis (Na^+/H^+ antiporters in Table 3.2).

Up-regulation of a Na^+/H^+ antiporter at high salt implies that proton gradients might be of importance for ion homeostasis in *H. seosinensis*. Three enzymatic complexes that could be responsible for maintaining a proton gradient required by the Na^+/H^+ antiporters were expressed: a P-type ATPase IIIA, the vacuolar H^+-transporting ATPase and the vacuolar H^+-transporting pyrophosphatase (V-PPase). Although some of these genes had high transcript abundance (*e.g.* the gene for V-PPase, ORF m.35862, with ~1500 TPM), none of them were differentially expressed.

3.2.2- Na^+/K^+ ATPases to Maintain Ion Gradients

In order to identify P-type ATPases in *H. seosinensis*, and especially Na^+/K^+ ATPases (type IIC) that would contribute to maintain ion gradients, I used the reference sequences published by Axelsen and Palmgren (1998) to harvest homologous proteins and to generate phylogenetic trees (Figure 3.5). In addition to
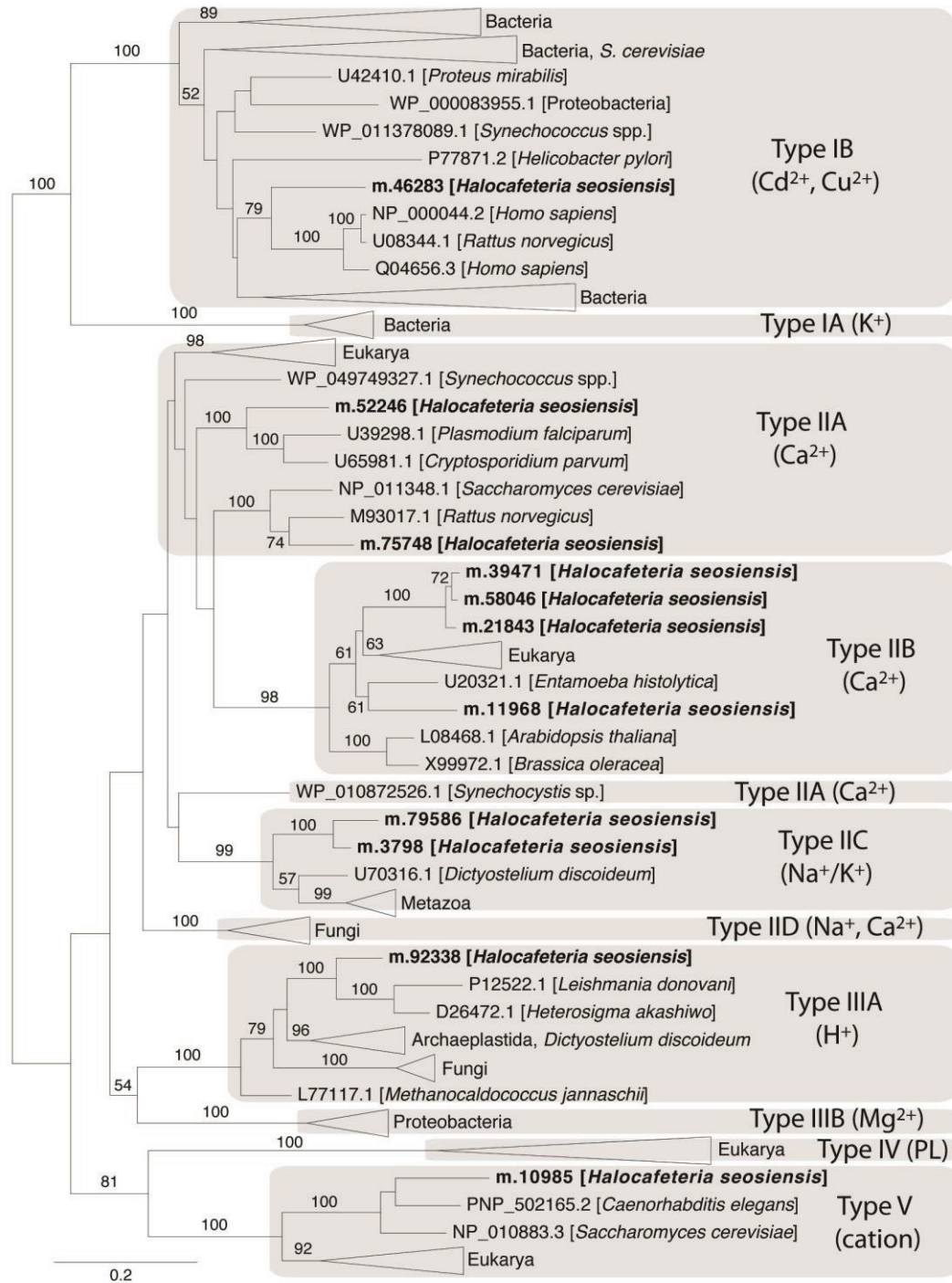
Fig. 3.5. Maximum-likelihood phylogenetic tree for P-type ATPases based on the analysis by Axelsen and Palmgren (1998) and containing *H. seosinensis* sequences (in bold). Type IV ATPases transport phospholipids (PL) and the corresponding clade included seven genes from *H. seosinensis.* Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

type IIC, *H. seosinensis* expressed genes clustering with P-type ATPase types IB (Cu^{2+}), IIA (Ca^{2+}), IIB (Ca^{2+}), IIIA (H$^+$), IV (phospholipids) and V (cation), none of them being up-regulated at high salt, with the exception of one putative Cu^{2+} transporter (Table 3.2). Interestingly, the two genes affiliated with Na$^+$/K$^+$-ATPases (ORFs m.3798 and m.79586) were closely related to each other, suggesting they were recently duplicated in *H. seosinensis* (Figure 3.6). A signal peptide was predicted for the second ORF, suggesting it might work at the plasma membrane. These ORFs grouped robustly inside a clade, sister to the animal Na$^+$/K$^+$-ATPase sequences, that contained a sequence from the oomycete *Pythium aphanidermatum* for which the corresponding protein was shown to mediate Na$^+$ expulsion and K$^+$ uptake (Barrero-Gil *et al.* 2005).

Interestingly, analysis of a multiple alignment including sequences from this clade, and from characterized animal transporters for which the crystal structure is available, showed that many residues involved in coordinating and binding Na$^+$ and K$^+$ in the animal transporters were not conserved in sequences of the *H. seosinensis*-containing clade, including the *P. aphanidermatum* Na$^+$/K$^+$-ATPase (Figure 3.7; Ogawa and Toyoshima 2002; Shinoda *et al.* 2009; Kanai *et al.* 2003; Nyblom *et al.* 2013) . Na$^+$/K$^+$-ATPases in animals require a β-subunit that is, among others, essential for K$^+$ binding (Shinoda *et al.* 2009). No such accessory protein was detected in the *H. seosinensis* transcriptome. Furthermore, residues involved in α/β subunit interaction (Nyblom *et al.* 2013) were not conserved in *H. seosinensis* sequences and related ATPases, including the one from *P. aphanidermatum* (data not shown). No β-subunit was mentioned for the transport assays of *P. aphanidermatum* Na$^+$/K$^+$-ATPase (Barrero-Gil *et al.* 2005). Absence of a β-subunit also applies to *Heterosigma akashiwo* for which the Na$^+$/K$^+$-ATPase sequence (Shono *et al.* 1996) belongs to the *H. seosinensis*-containing clade. However, the *H. akashiwo* sequence possessed all the conserved residues mentioned above. Shono *et al.* (2001) hypothesized that a 285 amino acid-long insertion in the *H. akashiwo* protein might
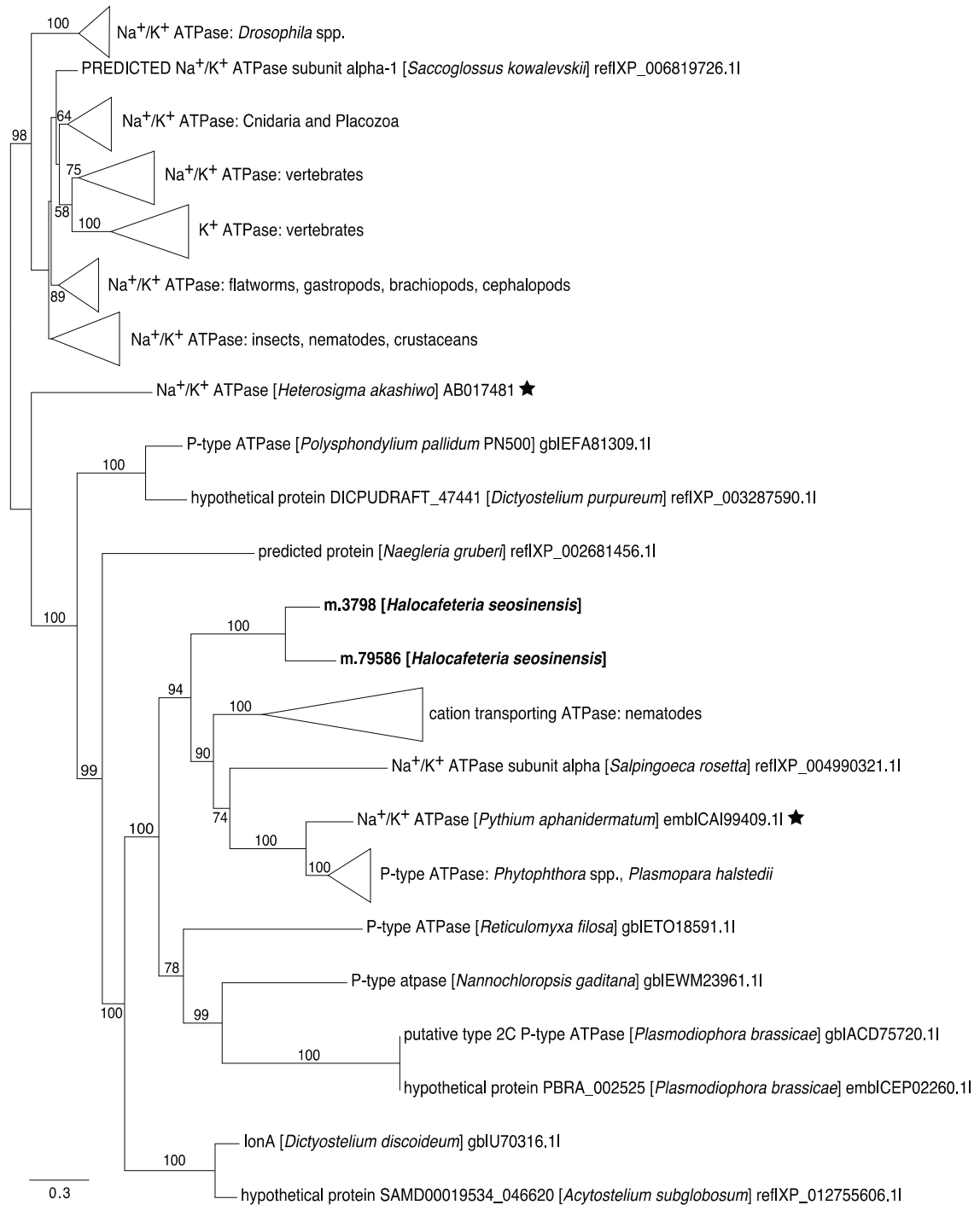
Fig. 3.6. Maximum-likelihood phylogenetic tree for sequences related to Na+/K+ ATPases. *Halocafeteria seosinensis* sequences (in bold) clusters inside a clade, sister to metazoan Na+/K+ ATPases and K+ ATPases, that contains sequences for ATPases shown to mediate Na+ expulsion and K+ uptake (stars). Bootstrap values >50% are indicated on branches. The scale bar indicates the substitution rate/site.
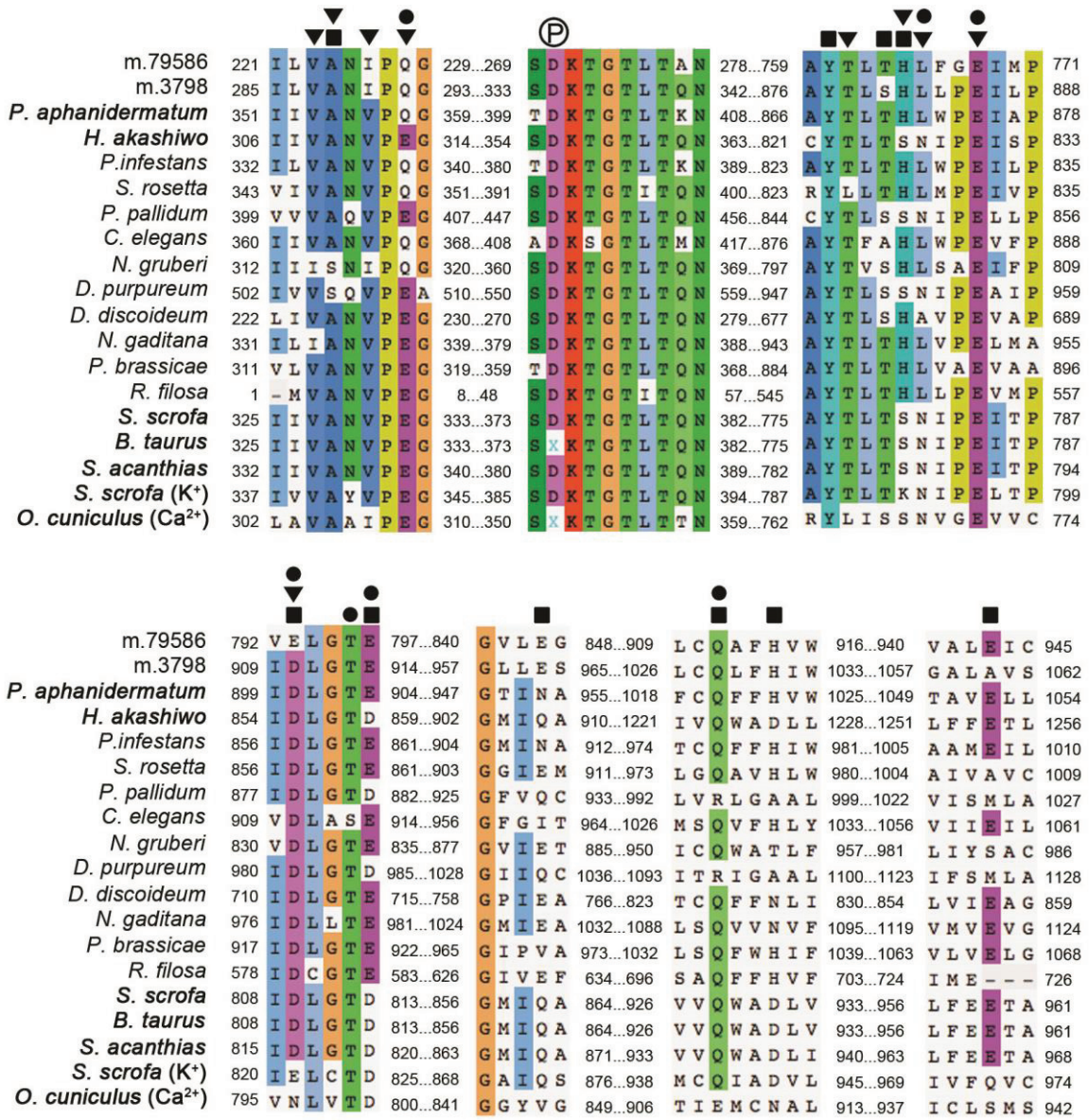
Fig. 3.7. Partial alignment of characterized Na⁺/K⁺-ATPases (in bold: *P. aphanidermatum* – 99409.1, *H. akashiwo* – AB017481, *Sus scrofa* – 4HQJ, *Bos taurus* – 4XE5 and *Squalus acanthias* – 2ZXE) and sequences clustering with *H. seosinensis* (m.79586, m.3798) in Figure 3.6 (see Figure 3.6 for accession numbers). The alignment also includes sequences for gastric H⁺/K⁺ ATPase (*Sus scrofa* (K⁺), 3IXZ_A) and Ca²⁺-translocating ATPase (*Oryctolagus cuniculus* (Ca²⁺), 3BA6_A). Residues interacting with Na⁺ (squares) and K⁺ (triangles) in Na⁺/K⁺-ATPases, and with Ca²⁺ (circles) in the Ca²⁺-translocating ATPase are indicated on top of respective positions. The phosphorylated residue involved in conformational change of the transporters and transition from the E1 to E2 states is indicated with a circled "P".

act similarly to the animal β-subunit. This insertion was unique to *H. akashiwo*. Therefore, the absence of a β-subunit in *H. seosinensis* and *P. aphanidermatum* could partly explain the divergence in functional residues compared to the characterized animal enzymes. Conservation of residues between *H. seosinensis* and *P. aphanidermatum* transporters suggest that ion homeostasis in the halophilic protist is potentially maintained partly by plasma membrane Na^+/K^+-ATPases.

These ATPases were the only candidates involved in the active transport of K^+. Despite not being differentially expressed, they might be important in ion homeostasis. Since type IIC (Na^+/K^+-ATPases) also includes K^+/H^+ ATPases, further work needs to consider them more carefully in order to determine their substrate specificity. For now, based on sequence conservation between *H. seosinensis* sequences and *P. aphanidermatum* Na^+/K^+-ATPase sequences, this alternative hypothesis for *H. seosinensis* genes coding for a K^+/H^+ ATPase is less likely.

3.2.3- Other Ion Transporters and Channels

Enrichment analysis indicated that the class 'Ion transport and metabolism' was enriched in gene duplicates (adjusted p-value = 5.8×10^{-5}, Figure 3.3). Proteins encoded by these genes were related to various ion transporters and channels. In addition to the previously mentioned duplication of Na^+/K^+-ATPase type IIC (TIGR01106, Figure 3.6), other examples included type IIB calcium-translocating ATPases (TIGR01517, Figure 3.8), chloride channels (CD03685, Figure 3.9), magnesium transporters (TIGR00400, Figure 3.10) and high affinity sulfate transporters (TIGR00815, Figure 3.11). A significant enrichment of ion transporter genes was also detected in the genome of halophilic yeasts *Wallemia ichthyophaga* and *Hortaea werneckii* (Lenassi *et al.* 2013; Zajc *et al.* 2013). However, I also detected an enrichment in duplicated genes assigned to this class in the genomes of *N. gaditana* (adjusted p-value = 0.041) and *G. theta* (adjusted p-value = 3.9×10^{-12}), implying that this situation is not unique to extreme halophiles (Figure 3.3).
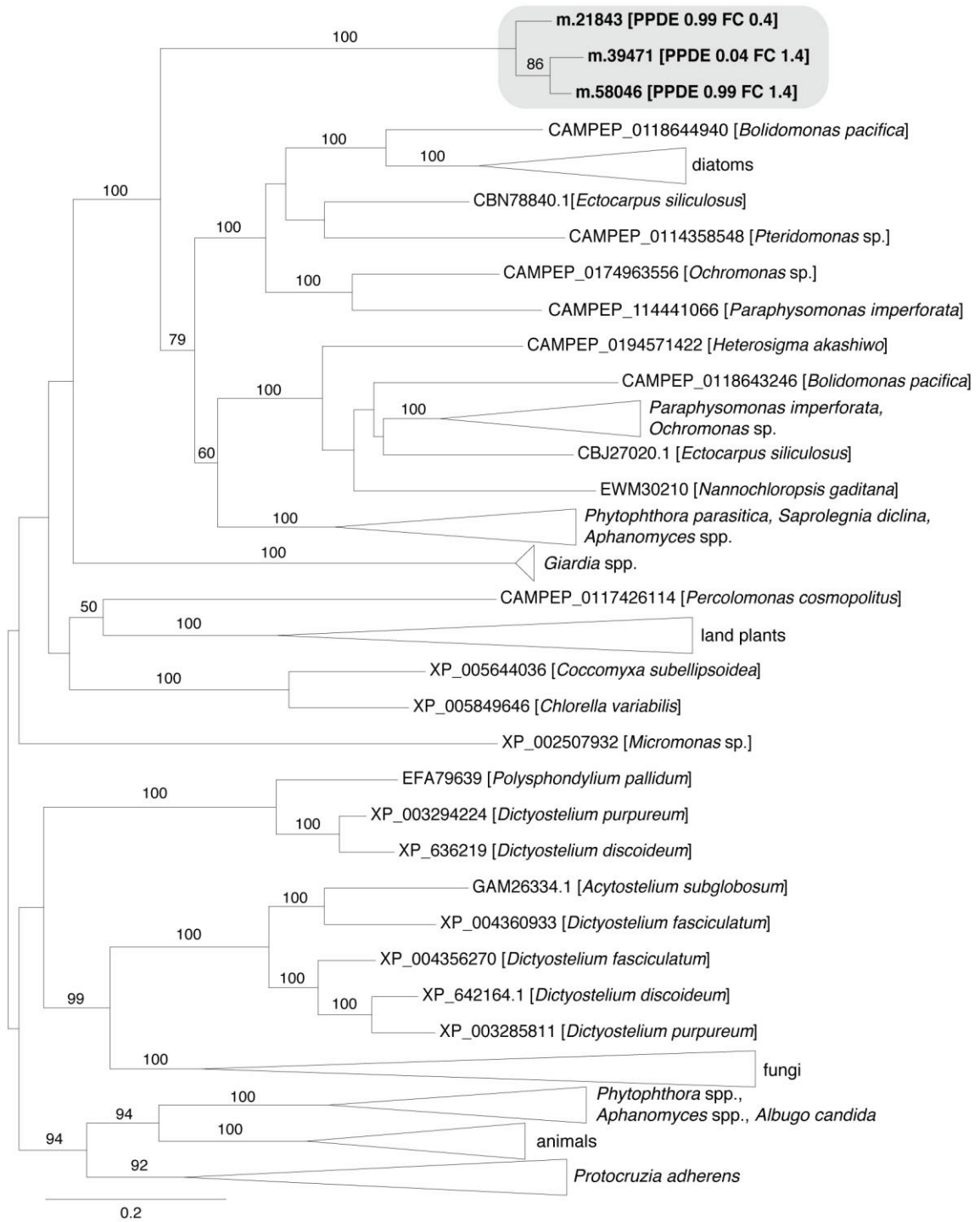
Fig. 3.8. Maximum-likelihood phylogenetic tree for gene duplication cluster 115 (grey box): calcium-transporting ATPases. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.
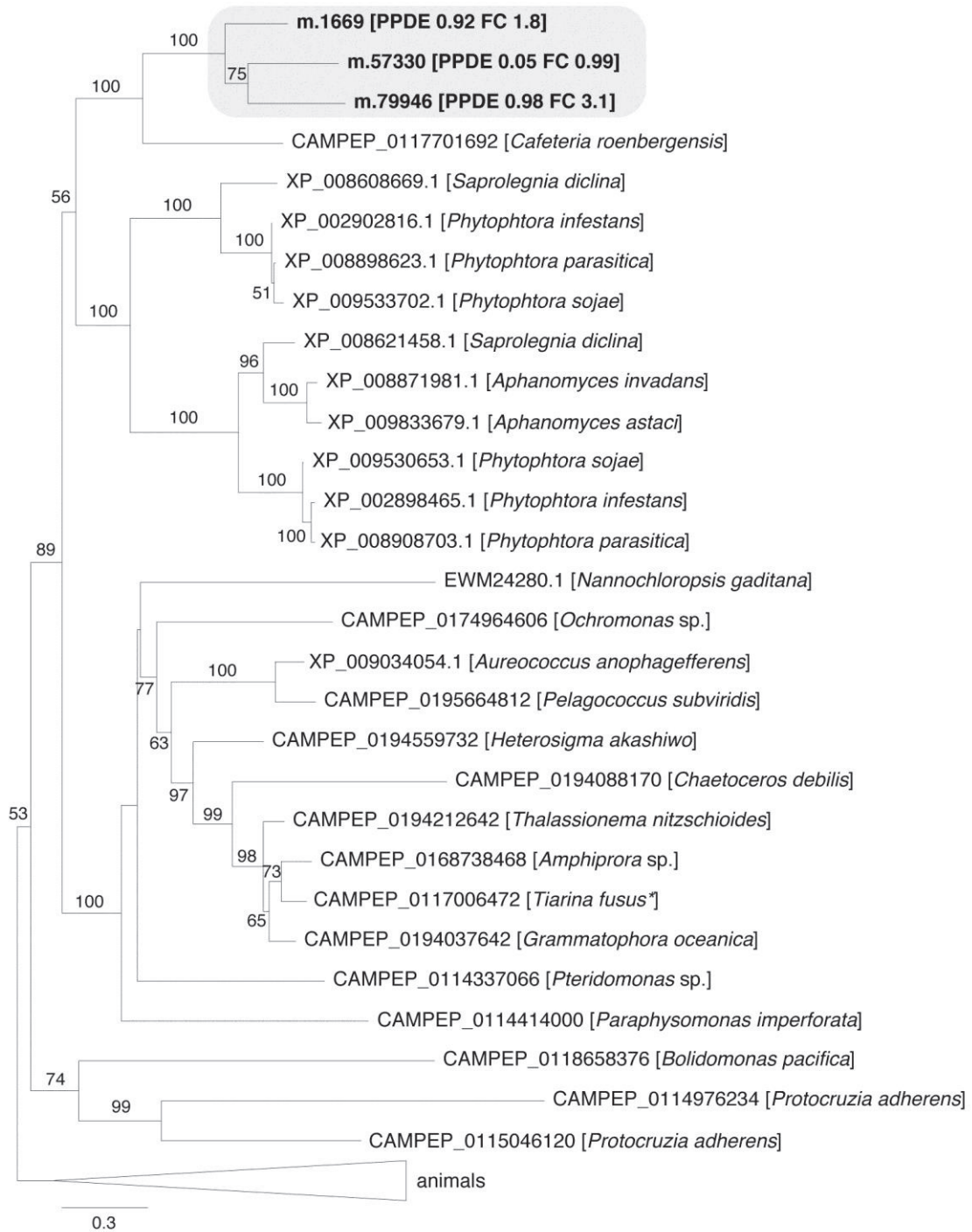
Fig. 3.9. Maximum-likelihood phylogenetic tree for gene duplication cluster 101 (grey box): chloride channels. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site. * Sequences generated from *Tiarina fusus* were potentially contaminated with sequences from the food source *Rhodomonas lens*.
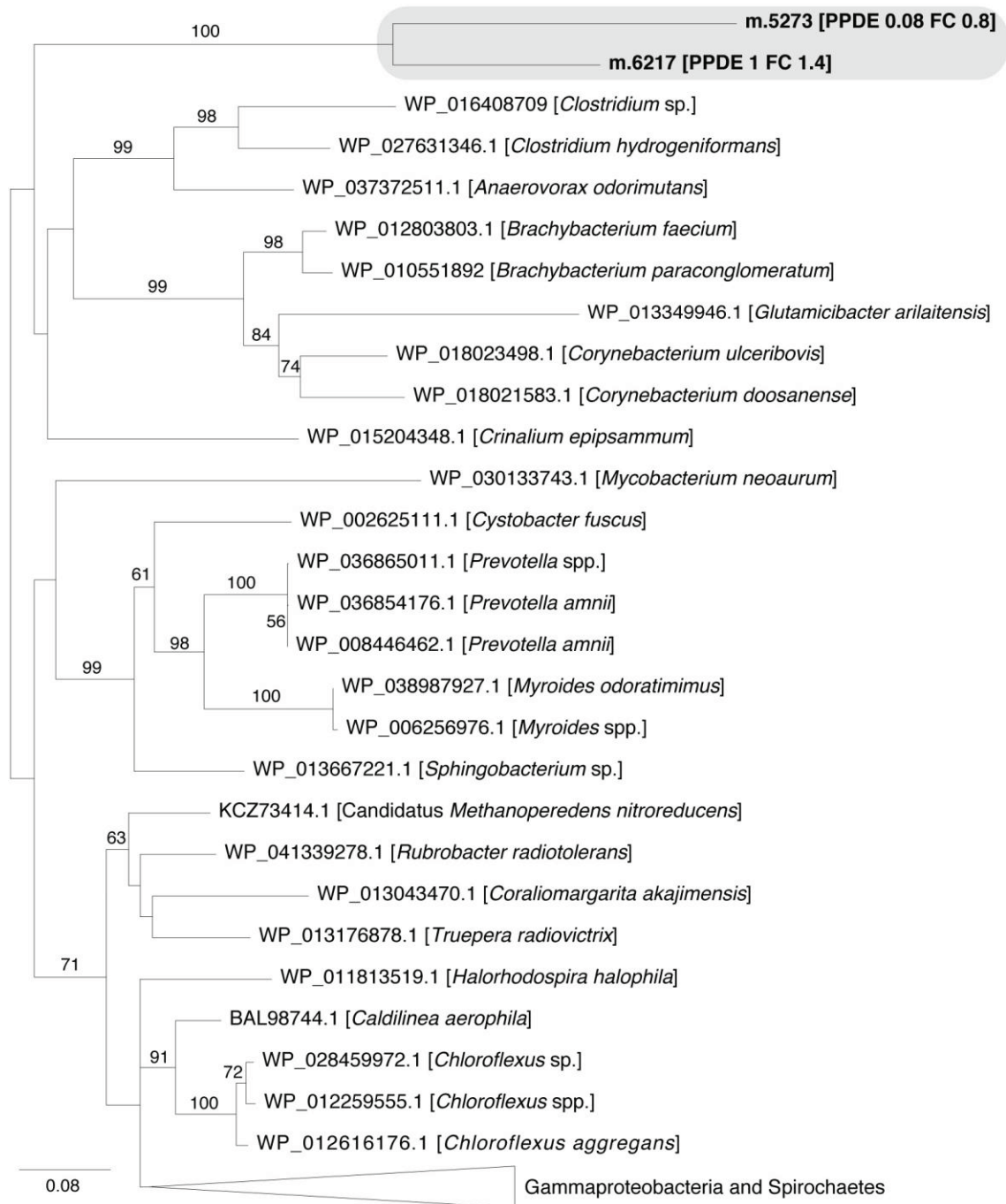
Fig. 3.10. Maximum-likelihood phylogenetic tree for gene duplication cluster 408 (grey box): magnesium transporters. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.
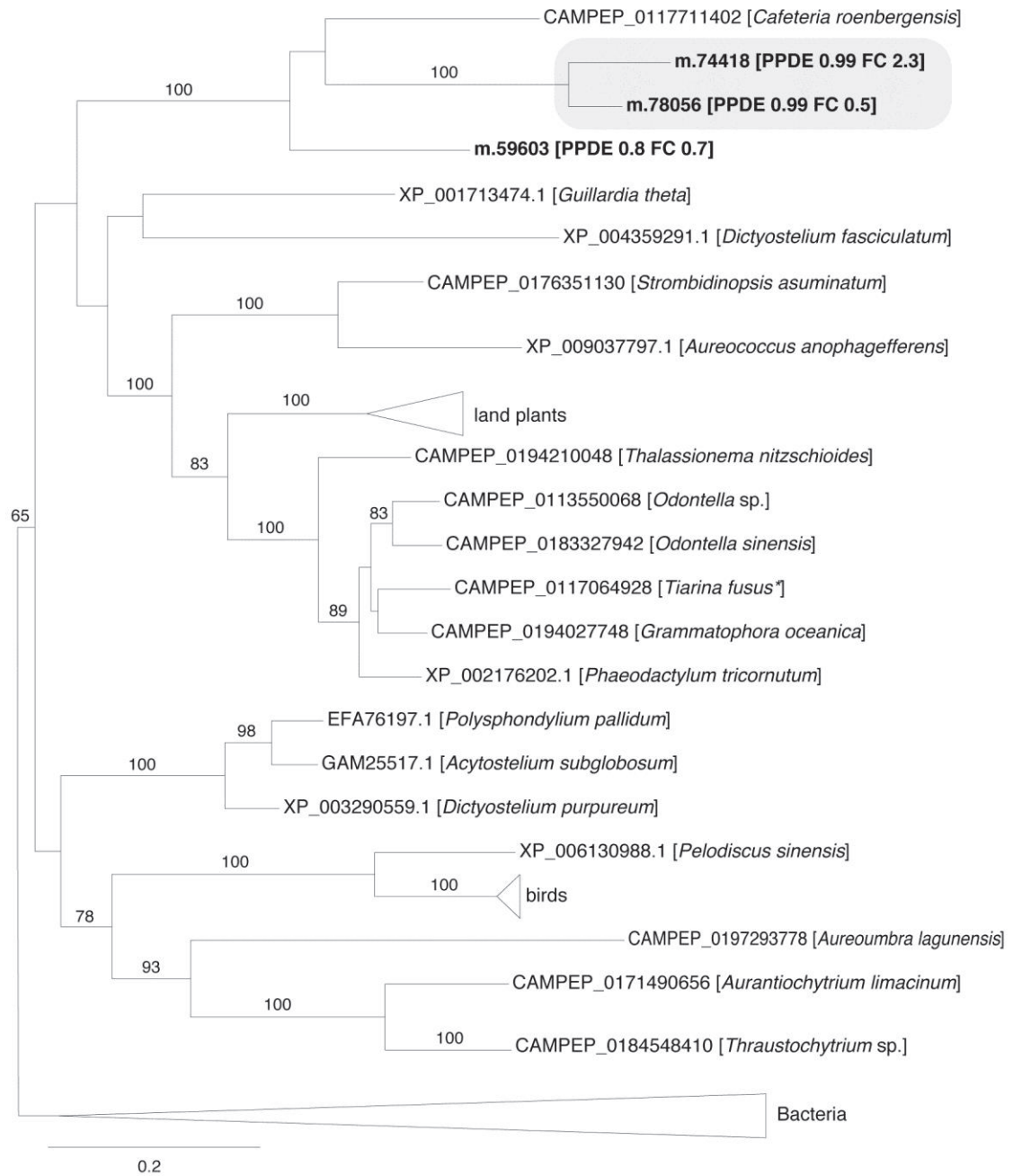
Fig. 3.11. Maximum-likelihood phylogenetic tree for gene duplication cluster 481 (grey box): sulfate transporters. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site. * Sequences generated from *Tiarina fusus* were potentially contaminated with sequences from the food source *Rhodomonas lens*.

More carriers were expressed by *H. seosinensis,* like the sodium:calcium exchanger (TIGR00927), calcium-activated potassium channels (PFAM03493), cyclic nucleotide-gated ion channels (ion transport domain PFAM00520 combined to cNMP-binding domain PFAM00027) or sodium/potassium/chloride channels (PFAM00520, Table 3.2). Interestingly a group of related chloride channels distantly affiliated to the *tweety* gene of *Drosophila* and related proteins (PFAM04906) had drastic differences in expression levels ranging from 273-fold upregulation to 127-fold downregulation (ORFs under "Chloride channels? (related to Tweety transmembrane proteins)" in Table 3.2). Given my search criteria, these sequences were too far from any homologs to be included in the gene duplication analysis but they might represent a hot spot of salt-driven adaptation.

In summary, I propose that *H. seosinensis* maintains a sodium/potassium gradient by expressing two Na^+/K^+ ATPases. Since these were the only detected transporters predicted to have potassium as substrate, further work is greatly needed in order to determine their substrate specificity. Differential expression of a Na^+/H^+ transporter indicated that establishment of a proton gradient used for subsequent Na^+ expulsion may be part of *H. seosinensis* response to high salt. Subcellular localization of this transporter (*i.e.* in vacuolar or plasma membranes) would help clarify whether vacuoles are involved in salt expulsion or if sodium is directly extruded outside cells. The high number of duplicated ion channels, which passively carry ions along their gradient, might favor rapid ionic adjustment in specific conditions. Study of the response to acute change in salinity would clarify the contribution of these channels to short-term salt adaptation in *H. seosinensis*.

3.3- Impact of O_2 availability on gene expression

As salinity increases, the solubility of oxygen decreases. Since i) oxygen is theoretically 2.6× less soluble at 30% salt compared to 15% salt (Battino *et al.* 1983; Sherwood *et al.* 1991), ii) the cell membrane is oxygen permeable (Khan *et al.* 2003), and iii) the response to low-oxygen conditions is widely regulated at the

transcriptional level (Butler 2013), I examined whether there may be an oxygen-dependent gene response to increased salinity. Inspection of the expression of metabolic and oxygen-dependent genes suggested that oxygen availability somewhat affected gene expression in *H. seosinensis*.

In aerobic organisms, oxygen is the terminal electron acceptor that drives energy metabolism through the Krebs cycle and the electron transport chain (ETC), leading to the creation of an electrochemical gradient that is used to generate ATP. When oxygen is not available, anaerobes (facultative or obligate) can either use an alternative electron acceptor, like fumarate, nitrate or sulfur, that still depends on an ETC (anaerobic respiration) or perform fermentation (Ginger *et al.* 2010). In the latter process, energy is obtained via substrate-level phosphorylation where the electron acceptor, NAD+, is regenerated by the reduction of oxidized compounds (*e.g.* pyruvate to lactate).

At high salt in *H. seosinensis*, most transcripts encoding subunits of ETC complexes I-IV, pyruvate dehydrogenase and enzymes of the Krebs cycle (isocitrate dehydrogenase, succinate dehydrogenase and malate dehydrogenase) were inferred to be differentially expressed, usually repressed up to 2.5-fold (Figure 3.12 and Table 3.3, note that transcript abundance was too low (indicated by "N/A" in Table 3.3) for genes encoded on the mitochondrial genome, for which transcripts were probably not polyadenylated thus not selected during library preparation). Furthermore, enzymes involved in generating the electron carriers tended to be more than twofold repressed at high salt. These included enzymes loading heme on apocytochrome c (cytochrome c heme lyase), synthesizing ubiquinone (Coq5 and Coq6) and the complex IV cofactor heme A (Cox15; Figures 3.13 and 3.14).

Concordantly, genes potentially involved in fermentation were up-regulated at high salt. Although glycolytic enzymes were not differentially expressed, with the exception of phosphoglyceromutase (2.4-fold repression, Figure 3.15), soluble NADH-dependent fumarate reductase (2.5-fold upregulation) and mitochondrial

pyruvate

pyruvate dehydrogenase complex
E1alpha: m.52373,
E1beta: m.87365 [PPDE 0.99 FC 0.45]
E2: m.15321
E3: m.943

acetyl-CoA   citrate synthetase
m.43475

oxaloacetate   citrate   aconitate hydratase
m.22652

malate dehydrogenase
m.75687 [PPDE 0.99 FC 0.40]
malate   isocitrate

fumarase   isocitrate dehydrogenase/
m.11962   oxalosuccinate decarboxylase
catalytic: m.54720
fumarate   **regulatory: m.29230**
[PPDE 0.98 FC 0.39]

succinate dehydrogenase
A: m.48989   oxoglutarate
[PPDE 1 FC 0.46]
B: m.48074   glutarate dehydrogenase complex
C: m.46418   succinate   E1: m.63442
D: TR3269   E2: m.13344
succinyl-CoA   E3: m.943
succinyl-CoA synthetase
alpha: m.5413
beta: m.68274

Fig. 3.12. ORFs in *H. seosinensis* coding for enzymes acting in the TCA cycle. For differentially expressed ORFs (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated.

Table 3.3. Expression of genes involved in the electron chain transport in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	$\log_2$ FC	Adjusted p-value	$\log_2$ FC	
Complex I (NADH dehydrogenase)									
f2_07	0.1	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 1
f1_05	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 2
f-3_011	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 3
f-3_07	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 4
f-3_05	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 4L
f-3_04	0.1	0.1	N/A	N/A	N/A	N/A	N/A	N/A	subunit 5
f2_02	0.1	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 6
f1_06	0.4	0.1	N/A	N/A	N/A	N/A	N/A	N/A	subunit 7
f1_04	0.1	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 9
m.78293	247.7	249.8	0.05	0.77	0.211	-0.36	0.133	-0.35	subunit 10
f-3_013	0.5	0.2	0.91	0.29	0.134	-1.50	0.112	-1.96	subunit 11
m.68372	391.9	373.2	0.04	0.71	0.312	-0.46	0.285	-0.40	1 alpha subcomplex subunit 2
m.71778	317.8	203.5	0.32	0.48	0.002	-1.02	0.038	-1.00	1 alpha subcomplex subunit 5
m.51687	182.0	119.4	0.88	0.50	0.000	-0.98	0.004	-0.97	1 alpha subcomplex subunit 6
m.58428	176.7	119.4	0.97	0.51	0.000	-0.94	0.002	-0.93	1 alpha subcomplex subunit 9
m.13012	224.8	172.3	0.17	0.58	0.021	-0.76	0.024	-0.75	1 alpha subcomplex subunit 12
TR3115	224.3	160.4	0.16	0.54	0.057	-0.84	0.174	-0.74	1 beta subcomplex subunit 7
m.80486	344.7	294.1	0.14	0.64	0.035	-0.62	0.036	-0.59	1 beta subcomplex 9
m.33187	312.3	217.5	1.00	0.54	0.000	-0.88	0.001	-0.87	flavoprotein 1
m.16087	288.1	177.7	0.20	0.46	0.005	-1.07	0.014	-1.05	flavoprotein 2
m.42249	273.7	259.3	0.04	0.72	0.165	-0.46	0.112	-0.45	iron-sulfur protein 4
m.60325	446.8	270.6	0.44	0.46	0.003	-1.08	0.010	-1.07	iron-sulfur protein 6

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$ FC	Adjusted p-value	log$_2$ FC	
Complex I (NADH dehydrogenase - continued)									
m.20620	176.0	130.9	0.39	0.57	0.008	-0.80	0.013	-0.78	iron-sulfur protein 8
m.52326	55.5	76.8	0.22	1.08	0.500	0.11	0.308	0.14	intermediate-associat. prot. 30
m.9807	22.2	20.1	0.69	0.68	0.022	-0.54	0.036	-0.49	assembly factor 6
g3835	9.5	11.2	0.07	0.92	0.742	-0.11	0.754	-0.09	assembly factor 7
m.93153	15.5	33.9	1.00	1.73	0.004	0.78	0.008	0.80	assembly factor 3
m.75939	11.1	10.9	0.99	0.76	0.158	-0.39	0.072	-0.42	assembly factor 5
TR4890	298.9	295.0	0.06	0.75	0.258	-0.39	0.350	-0.33	Nuum
TR1148	345.7	92.9	0.99	0.20	0.000	-2.19	0.01	-2.16	LYR motif-containing protein 5
m.10265	398.2	255.9	0.98	0.49	0.000	-1.00	0.002	-1.00	gamma carbonic anhydrase 1
m.79606	2653.6	1833.3	0.06	0.52	0.026	-0.91	0.038	-0.89	mitochondrial acyl carrier protein
m.74576	54.0	105.0	0.98	1.53	0.022	0.61	0.018	0.62	Ind1/MRP
Complex II (Succinate dehydrogenase)									
m.48989	331.0	198.5	1.00	0.46	0.000	-1.11	0.000	-1.10	flavoprotein subunit 1
m.48074	331.2	234.1	0.60	0.54	0.003	-0.86	0.008	-0.86	subunit 2
m.46418	488.7	334.3	0.55	0.52	0.002	-0.92	0.007	-0.91	cytochrome b subunit
TR3269	553.4	383.3	0.12	0.53	0.019	-0.89	0.082	-0.88	cytochrome b small subunit
Complex III									
m.80458	631.5	481.8	0.91	0.58	0.000	-0.76	0.005	-0.76	cytochrome c1
TR3373	987.2	564.6	0.41	0.44	0.004	-1.15	0.046	-1.13	cyt. b-c1 complex subunit 6
f-1_05	0.2	0.0	N/A	N/A	N/A	N/A	N/A	N/A	cytochrome b
m.67892	84.4	69.4	1.00	0.63	0.000	-0.64	0.003	-0.63	peptidase subunit alpha
m.89567	311.4	246.3	1.00	0.62	0.000	-0.69	0.001	-0.67	peptidase subunit beta

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	log$_2$ FC	Adjusted p-value	log$_2$ FC	
Complex III (continued)									
m.35127	779.3	747.4	0.03	0.72	0.131	-0.45	0.105	-0.43	ubiquinol-cytochrome C reductase iron-sulfur subunit
Complex IV (cytochrome oxidase)									
f2_05	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 1
f-2_06	0.2	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 2
688_f-2_02	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	subunit 3
m.93907	1927.8	1122.5	0.07	0.43	0.009	-1.14	0.024	-1.14	subunit VIb
m.58879	44.5	41.1	0.50	0.71	0.032	-0.49	0.034	-0.46	protoheme IX farnesyltransferase
m.11490	36.0	24.4	1.00	0.53	0.000	-0.90	0.002	-0.87	subunit XI assembly protein
m.74554	102.9	28.7	1.00	0.21	0.000	-2.20	0.000	-2.19	assembly protein COX15
TR1614	476.5	366.4	0.12	0.58	0.062	-0.75	0.164	-0.68	copper chaperone (COX17)
m.33952	82.7	82.3	0.20	0.77	0.213	-0.36	0.182	-0.30	assembly protein COX19
TR1765	125.0	101.0	0.84	0.62	0.015	-0.66	0.087	-0.59	assembly factor 5
m.21092	85.5	75.9	0.38	0.68	0.029	-0.55	0.031	-0.51	protein sco1
Complex V (ATP synthase)									
688_f-1_01	1.1	0.4	1.00	0.26	0.016	-1.76	0.008	-2.10	F1 subunit alpha
m.50724	784.2	721.9	1.00	0.71	0.000	-0.48	0.010	-0.46	F1 subunit beta
m.59503	555.2	482.4	0.83	0.67	0.006	-0.57	0.014	-0.56	F1 gamma subunit
m.10551	740.4	508.2	0.14	0.52	0.010	-0.92	0.019	-0.91	F1 delta subunit
m.77785	482.2	433.0	0.05	0.67	0.199	-0.55	0.322	-0.47	F1 epsilon subunit
688_f-2_03	0.0	0.0	N/A	N/A	N/A	N/A	N/A	N/A	Atp6 (subunit a)
f-3_012	0.0	0.1	N/A	N/A	N/A	N/A	N/A	N/A	subunit 8 (Atp8)

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post Fold Change	Adjusted p-value	$\log_2$ FC	Adjusted p-value	$\log_2$ FC	
Complex V (ATP synthase - continued)									
f-2_010	0.2	0.6	N/A	N/A	N/A	N/A	N/A	N/A	F0 subunit 9 (C1)
m.3341	36.8	88.6	0.99	1.78	0.055	0.81	0.025	0.99	malate dehydrogenase
m.92678	13.4	14.0	0.00	0.81	0.230	-0.30	0.197	-0.27	inner membrane protease ATP23
m.9569	16.6	18.8	0.52	0.86	0.250	-0.21	0.213	-0.19	ATP synthase regulation protein NCA2
m.9383	76.7	61.8	0.28	0.60	0.022	-0.71	0.097	-0.64	molecular chaperone Atp11
m.1513	77.2	51.6	0.99	0.51	0.000	-0.96	0.012	-0.94	F1 complex assembly factor 2

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); $\log_2$FC, $\log_2$ fold change (30% over 15% salt); N/A, transcript abundance was too low for genes encoded on the mitochondrial genome for which transcripts were probably not adenylated thus not selected during library preparation.

glycine

5-aminolevulinate synthase
m.10453

5-aminolevulinate

porphobilinogen synthase
m.10926, m.10935

porphobilinogen

hydroxymethylbilane synthase
m.79542

hydroxymethylbilane

uroporphyrinogen III synthase
m.74841

uroporphyrinogen III

uroporphyrinogen decarboxylase
m.63860 [PPDE 1 FC 0.44]

coproporphyrinogen III

coproporphyrinogen III oxidase
m.78089

protoporphyrinogen IX

oxygen-dependent
protoporphyrinogen oxidase
m.14364

protoporphyrin IX

ferrochelatase
m.60804
[PPDE 0.99 FC 0.35]

protoheme

protoheme IX
farnesyltransferase
m.58879

apocytochrome c
m.69118

heme O

cytochrome c heme lyase
m.34054 [PPDE 1 FC 0.46]

COX15
m.74554
[PPDE 1 FC 0.21]

heme A   cytochrome c

Fig. 3.13. ORFs in *H. seosinensis* coding for enzymes acting in the porphyrin biosynthetic pathway. For differentially expressed ORFs (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated.
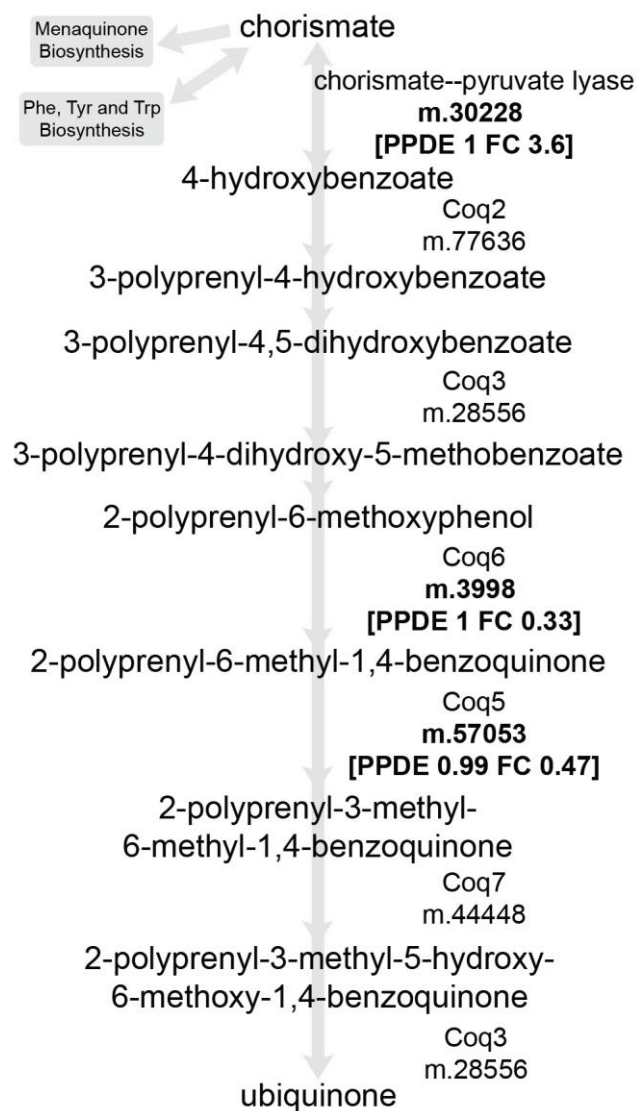
Fig. 3.14. ORFs in *H. seosinensis* coding for enzymes acting in the ubiquinone biosynthetic pathway. For differentially expressed ORFs (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated.

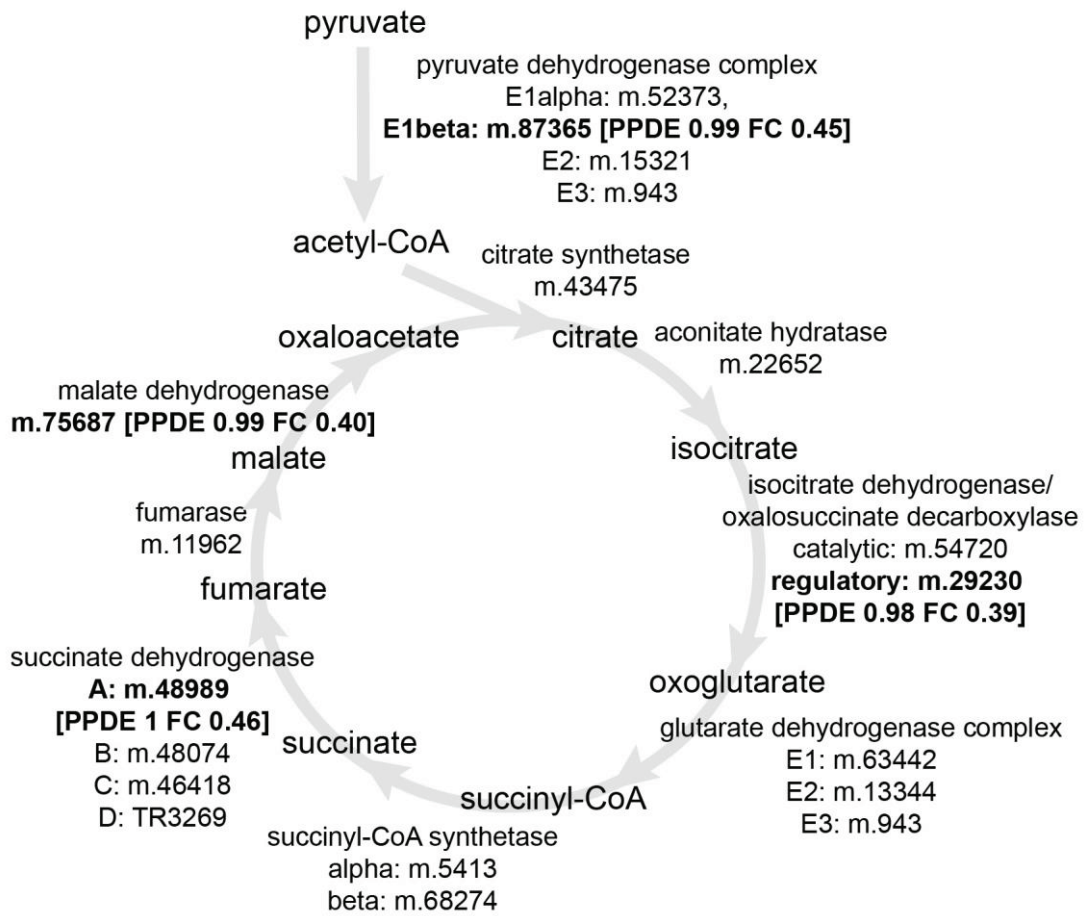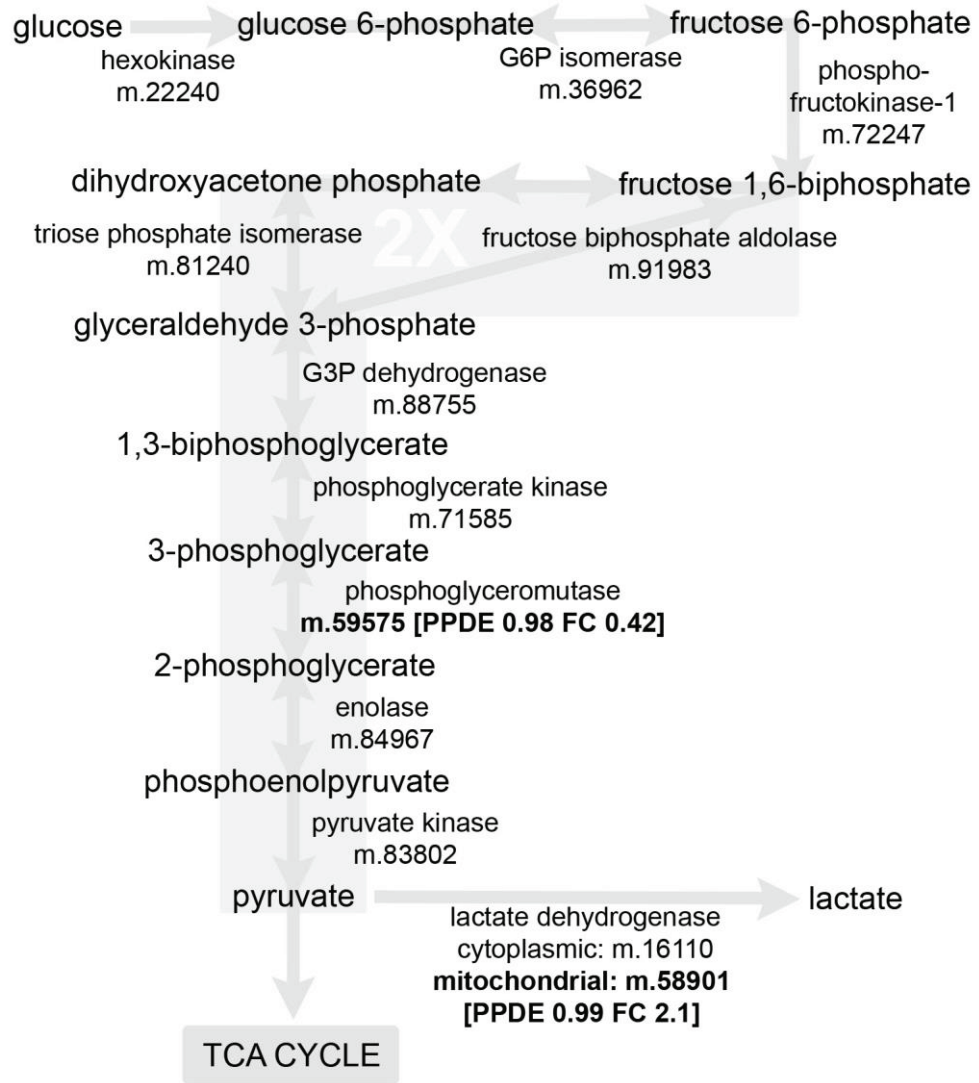**Fig. 3.15.** ORFs in *H. seosinensis* coding for enzymes acting in glycolysis. For differentially expressed ORFs (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. From fructose 1,6-biphosphate to pyruvate, two molecules of each metabolite are processed per molecule of glucose (shaded area).

lactate dehydrogenase (2.1-fold upregulation) had noticeably higher expression at high salt.

In short, oxygen limitation at high salt appears to have led to partial repression of *H. seosinensis* respiration-related genes that was compensated for by the upregulation of fermentation genes. Therefore, adaptation to lower oxygen availability is part of the *H. seosinensis* response to high salt.

3.4- Signal Transduction: From Sensing the Environment to Altering the Transcriptional Program

Grown in two different salt concentrations, *H. seosinensis* differentially expressed genes encoding proteins involved in sensing the environment and in transducing signals that ultimately lead to altering the expression of relevant genes. This transcriptional program involved genes typically acting in the G-protein pathway and in cyclic nucleotide signaling, various kinases, P2X receptors, and transcription factors especially involved in stress response, such as sirtuins and heat shock factors. Sequences for these proteins in *H. seosinensis* tended to have low similarity to known genes, lowering confidence for functional annotation. However, differential expression of many of these genes suggested that they play important roles in modulating *H. seosinensis* activity at varying salinity.

3.4.1- Cyclic Nucleotide Dependent Signaling

Using cyclic nucleotides as a second messenger is a universal signaling mechanism that regulates a wide variety of cellular activities. The first described cyclic nucleotide cascade model involved activation of a transmembrane receptor by extracellular stimuli followed by interaction with a G protein complex, leading to the dissociation of G protein subunits (Robison *et al.* 1971). Consequently, G proteins activate or inhibit adenylate cyclase, which generates the second messenger cyclic adenosine monophosphate (cAMP). The latter binds to protein kinase A (PKA),

which in turn transduces the signal to the appropriate effectors depending on the required cellular response.

As sensors of extracellular conditions, *H. seosinensis* up-regulated (by up to 350-fold) 11 genes related to G protein-coupled receptors (GPCR) with homology to domains of the slime mold cAMP receptor and membrane region of the Frizzled/Smoothened family (PFMA05462 and PFAM01534 respectively, Table 3.4). A large variety of GPCRs exist, and some are known to be transcriptionally induced by salt exposure as in *Oryza sativa* (Yadav and Tuteja 2011). GPCRs have a distinguishable structure where seven transmembrane regions separate the extracellular N-terminal tail from the cytoplasmic C-terminal tail, features that were detected by TMHMM in sequences listed in Table 3.4. Typically, GPCR sequence conservation tends to be low, with an average of <25% pairwise identity between members of the same family (Oliveira *et al.* 1999). As a result, the identity of the putative sensed stimuli could not be determined solely based on analysis of *H. seosinensis* sequences. However, based on their expression profile, many of these genes might be important in regard to long-term salt adaptation.

Ten of these genes coding for GPCRs were closely related and possibly originated from gene duplication events, however no homologs were detected in the NR and MMETSP databases given my standardized search criteria (>30% identity with 2/3 alignments coverage). When the identity threshold was relaxed to >20% identity, 20 sequences were gathered, including one expressed by *C. roenbergensis*. Phylogenetic analysis demonstrated that *H. seosinensis* sequences clustered together with moderate bootstrap support (72%) to the exclusion of the sequence from *C. roenbergensis* (Figure 3.16). These duplicated genes had contrasting expression levels (from 2.8-fold decrease for ORF m.21317 to 350-fold increase for m.82040, Table 3.4). Four were activated at high salt as their average transcript level rose from <1.2 transcript per million (TPM) at 15% salt to >15 TPM at high salt. Since these salt-responsive genes were evolutionarily dynamic, they represent attractive candidates for further studies on salt adaptation, aiming at characterizing their respective stimulus and downstream signaling partners.

Table 3.4. Expression of genes involved in signaling cascade in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
G protein-coupled receptors								
m.82040	0.05	25.14	1.00	350.34	0.000	7.71	0.000	9.18
m.28994	0.02	15.93	1.00	257.41	0.000	7.71	0.000	8.67
m.43860	13.62	37.08	1.00	2.17	0.000	1.10	0.002	1.10
m.63112	0.77	16.90	1.00	17.84	0.000	3.81	0.000	4.29
m.28285	7.83	10.09	0.09	1.06	0.924	0.08	0.967	0.02
m.77134	1.17	26.53	1.00	17.68	0.000	4.09	0.000	4.20
m.21316	5.17	29.78	1.00	4.55	0.000	2.17	0.000	2.21
m.21317	47.31	22.15	1.00	0.35	NA	-1.47	0.003	-1.44
m.56320	10.76	18.02	0.20	1.36	0.190	0.44	0.094	0.45
m.61334	0.04	0.81	1.00	8.85	0.013	2.51	0.004	3.39
G proteins								
m.12403	91.37	89.15	1.00	0.75	0.001	-0.40	0.017	-0.38
m.23018	8.64	16.25	0.71	1.42	0.112	0.51	0.074	0.49
m.73239	43.96	37.39	1.00	0.66	0.000	-0.59	0.004	-0.57
m.20288	127.70	138.83	0.04	0.84	0.334	-0.25	0.226	-0.24
Protein kinase A								
m.46128	76.30	88.58	0.03	0.88	0.514	-0.17	0.451	-0.14
m.79597	30.38	41.48	0.04	1.05	0.682	0.07	0.497	0.09
m.83951	35.29	39.27	0.06	0.84	0.372	-0.23	0.340	-0.18
Adenylate cyclases								
m.83214	8.87	9.26	0.93	0.81	0.173	-0.30	0.204	-0.30
m.60922	0.66	3.93	0.41	5.05	0.292	1.25	0.330	2.84

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC
Guanylate cyclases								
m.5216	0.34	45.22	1.00	100.71	N/A	6.53	0.000	6.72
m.72172	0.30	3.37	1.00	8.86	0.000	3.00	0.000	3.10
m.85803	2.66	4.78	0.10	1.46	0.346	0.52	0.250	0.51
m.80667	10.78	13.83	0.01	1.00	0.979	0.01	0.842	0.03
m.31327	9.50	23.13	1.00	1.90	0.000	0.93	0.000	0.95
m.65409	12.43	13.12	0.68	0.83	0.213	-0.26	0.154	-0.23
m.80299	4.54	8.89	1.00	1.54	0.001	0.62	0.006	0.64
m.24353	16.26	17.20	0.05	0.79	0.530	-0.31	0.607	-0.20
m.52473	8.19	18.96	1.00	1.79	0.000	0.84	0.003	0.85
m.84504	13.56	12.32	1.00	0.69	0.001	-0.53	0.009	-0.51
m.51628	10.37	12.79	0.04	1.00	1.000	0.00	0.904	0.04
Phosphodiesterases								
m.89581	0.49	2.86	1.00	4.64	0.000	2.06	0.002	2.11
m.43482	5.37	30.49	1.00	4.38	0.000	2.12	0.000	2.14
m.19732	13.45	28.14	1.00	1.66	0.001	0.73	0.005	0.74
m.17625	3.57	4.60	0.02	1.01	0.928	0.04	0.768	0.07
m.62195	24.04	39.84	0.06	1.30	0.200	0.38	0.094	0.42
m.1263	2.85	6.74	1.00	1.83	0.001	0.86	0.005	0.84
m.73524	5.71	12.81	0.11	1.89	0.194	0.84	0.107	0.97
m.89632	5.44	8.39	0.52	1.24	0.614	0.30	0.573	0.25
m.60926	10.72	20.64	0.96	1.48	0.005	0.56	0.010	0.58
m.35609	39.43	37.05	0.75	0.73	0.088	-0.44	0.088	-0.38
m.40399	11.68	16.65	0.05	1.14	0.644	0.19	0.514	0.19
m.7956	7.30	8.19	0.12	0.88	0.658	-0.18	0.407	-0.23

	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
ORF names	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
Sensory histidine kinases								
m.13308	0.13	4.75	1.00	32.12	0.000	3.86	0.001	5.83
m.13214	8.43	43.97	1.00	4.48	0.001	1.96	0.005	2.37
m.91175	6.58	16.07	1.00	1.94	0.000	0.95	0.001	0.98
m.29497	10.95	21.92	0.96	1.56	0.007	0.64	0.010	0.64
m.43073	30.80	47.13	1.00	1.18	0.020	0.25	0.124	0.27
m.13905	33.88	23.59	1.00	0.55	0.000	-0.85	0.002	-0.84
m.29833	34.67	28.96	1.00	0.65	0.005	-0.62	0.044	-0.61
Histidine-containing phosphotransfer proteins								
m.28961	44.90	133.11	1.00	2.29	0.000	1.20	0.002	1.23
m.2685	84.55	106.19	0.04	0.96	0.849	-0.05	0.996	0.00
Mitogen-activated protein kinases								
m.31471	0.87	22.24	1.00	18.81	0.000	4.13	0.001	4.23
m.81114	29.21	122.04	1.00	3.17	0.000	1.64	0.000	1.71
m.55540	3.09	12.11	1.00	3.23	0.001	1.60	0.004	1.68
m.51772	20.05	56.83	1.00	2.22	0.000	1.14	0.001	1.18
m.89580	2.72	8.25	1.00	2.38	0.000	1.23	0.002	1.21
m.4769	13.54	38.18	1.00	2.08	0.005	1.03	0.024	1.24
m.79532	3.30	9.20	1.00	2.07	0.014	1.02	0.031	1.26
m.24771	7.69	17.71	1.00	1.79	0.000	0.85	0.000	0.88
m.63746	38.56	28.05	1.00	0.56	0.000	-0.83	0.001	-0.81
m.4818	72.04	49.61	1.00	0.53	0.000	-0.90	0.004	-0.90
m.60901	11.84	1.53	1.00	0.11	0.000	-2.77	0.036	-3.26
m.33974	16.59	0.81	1.00	0.04	0.000	-4.43	0.000	-4.64

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	$\log_2$FC	Adjusted p-value	$\log_2$FC
Calcium-dependent protein kinases								
m.12706	0.10	3.51	1.00	19.67	0.000	3.98	0.000	4.44
m.37224	2.00	24.90	1.00	9.41	0.000	3.21	0.000	3.32
m.16106	1.15	10.84	1.00	7.08	0.000	2.80	0.000	2.92
m.24266	3.28	18.43	1.00	4.52	0.000	2.14	0.000	2.18
m.81494	5.16	14.19	1.00	2.19	0.000	1.11	0.003	1.13
m.83739	4.58	10.55	1.00	1.82	0.003	0.85	0.007	0.84
m.20526	18.17	36.24	0.96	1.55	0.132	0.63	0.042	0.72
m.25940	31.88	72.48	1.00	1.77	0.000	0.83	0.000	0.85
m.63829	21.42	51.51	1.00	1.87	0.000	0.91	0.000	0.93
m.20537	19.63	15.24	1.00	0.60	0.000	-0.73	0.005	-0.69
m.21877	35.31	26.39	1.00	0.57	0.000	-0.79	0.001	-0.77
Calmodulin-dependent kinases								
m.40160	6.14	23.10	1.00	2.98	0.000	1.56	0.000	1.58
m.20528	4.14	9.24	0.98	1.78	0.087	0.79	0.037	0.84
m.40168	22.76	38.02	0.99	1.28	0.099	0.36	0.038	0.42
Other serine/threonine protein kinases								
m.8926	0.06	19.29	1.00	185.54	0.000	7.21	0.000	7.65
m.80566	0.47	19.38	1.00	31.09	0.000	4.92	0.000	5.03
m.24240	48.70	231.70	1.00	3.67	0.000	1.87	0.000	1.92
m.88063	20.87	70.59	1.00	2.71	0.000	1.41	0.001	1.41
m.32208	36.35	95.36	1.00	2.07	0.000	1.05	0.000	1.09
m.47366	38.81	61.42	1.00	1.25	0.041	0.33	0.033	0.35
m.53058	57.56	56.21	1.00	0.75	0.001	-0.40	0.017	-0.38
m.49893	23.68	2.84	1.00	0.10	NA	-3.29	0.000	-3.33

	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
ORF names	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
Other serine/threonine protein kinases (continued)								
m.61748	31.34	22.73	1.00	0.56	0.001	-0.81	0.006	-0.76
m.51667	32.74	23.23	0.99	0.53	0.005	-0.88	0.009	-0.92
m.85991	28.86	21.37	1.00	0.55	0.000	-0.84	0.003	-0.85
P2X receptors								
m.48508	0.12	2.76	1.00	15.44	0.000	3.47	0.001	4.45
m.49662	59.70	428.55	1.00	5.52	0.000	2.44	0.000	2.49
m.23394	11.90	48.48	1.00	3.19	0.000	1.67	0.000	1.71
m.23395	2.98	11.71	1.00	3.14	0.000	1.62	0.000	1.65
m.407	3.62	0.08	1.00	0.02	0.000	-5.06	0.000	-5.47
m.85617	17.02	18.67	0.58	0.85	0.271	-0.22	0.248	-0.19
m.49618	12.97	8.85	1.00	0.52	0.000	-0.91	0.006	-0.85
m.15826	17.31	33.01	1.00	1.47	0.037	0.56	0.016	0.62
m.15807	1.48	3.20	0.15	1.84	0.449	0.72	0.308	0.89
m.35759	9.02	2.76	1.00	0.24	0.000	-2.02	0.000	-1.99
m.66849	27.48	23.85	0.71	0.66	NA	-0.59	0.024	-0.55
m.72620	15.80	31.01	1.00	1.51	0.000	0.60	0.002	0.64
m.56329	1.23	1.30	0.00	0.84	0.635	-0.26	0.413	-0.27

Abbreviations: TPM, averaged transcripts per million (at 15% or 30% salt); PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.
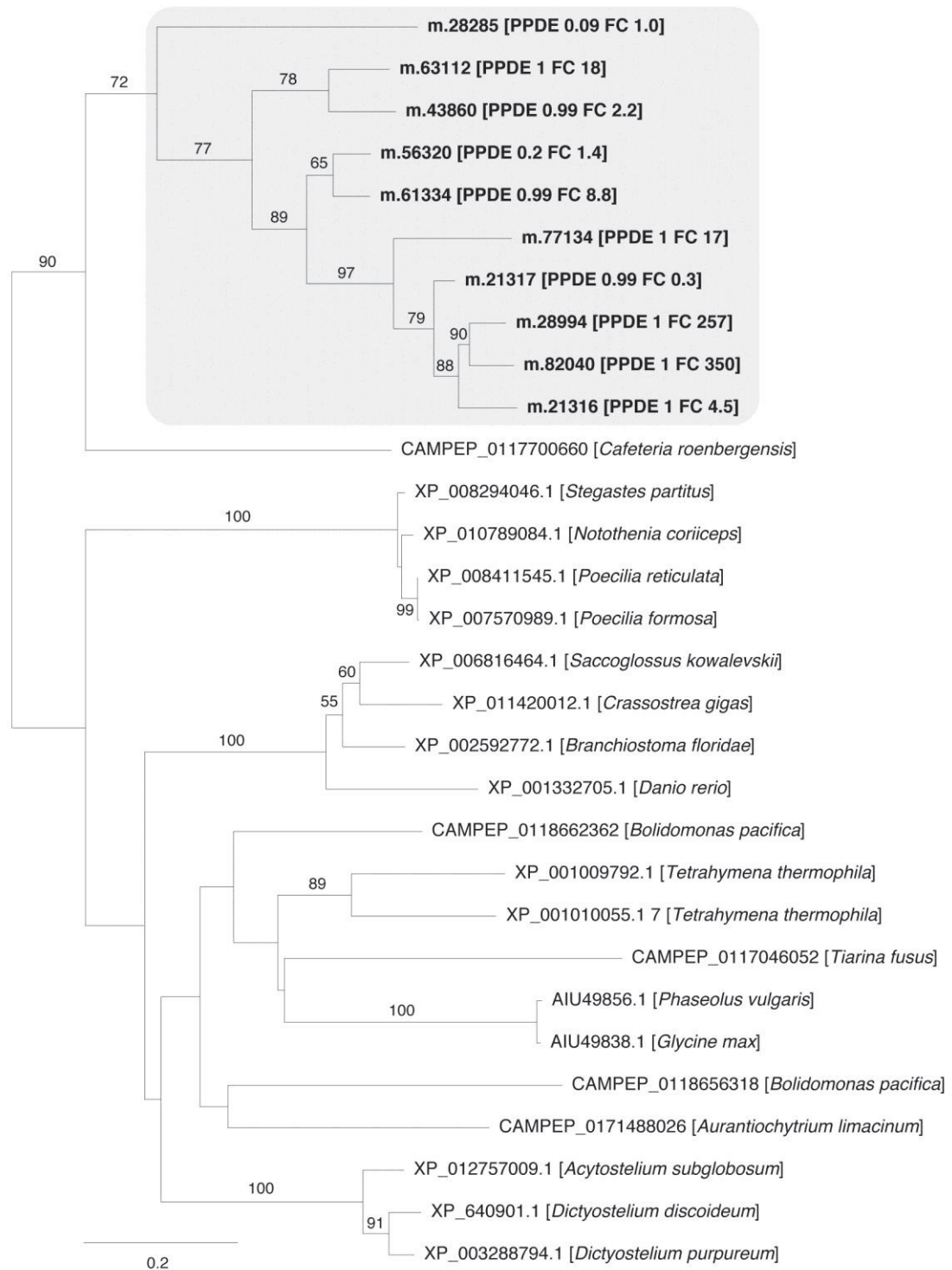
Fig. 3.16. Maximum-likelihood phylogenetic tree for gene duplication cluster 109 (grey box): G-protein coupled receptors. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

However, transducers of the canonical pathway downstream of GPCR were not differentially expressed or not convincingly identified (due to distant phylogenetic relationships). Genes related to G protein subunits alpha and beta (no gamma found) in *H. seosinensis* were not differentially expressed. Genes related to adenylate cyclase and to PKA were also not differentially expressed. Since these members of the cascade are involved in many cellular processes in parallel, it seems logical that they were not transcriptionally regulated. Nonetheless, characterization of the differentially expressed GPCR could allow clarification of how *H. seosinensis* senses the extracellular conditions.

Other evidence indicated that high salt adaptation involved cyclic nucleotide messaging, especially cyclic guanosine monophosphate (cGMP). While two genes related to adenylate cyclase (AC) were detected in *H. seosinensis*, some ten genes related to guanylate cyclase (GC) were expressed, including two genes markedly up-regulated at high salt (8.9- and 100-fold increase, Table 3.4). These GCs were all membrane-localized; they each contained two class III cyclase catalytic domains (PFAM00211) separated by two transmembrane regions each containing 5-6 membrane-spanning stretches. This topology suggested that the active site was formed by the heterodimeric association of both domains as commonly observed in some class III cyclases (Willoughby and Cooper 2007).

The specificity for *H. seosinensis* cyclase enzymes was identified based on conserved residues in the purine-binding pocket, where lysine and glutamate correspond to adenine and guanine specificity respectively (yellow boxes in Figure 3.17; Baker and Kelly 2004). Two other highly conserved residues are the asparagine and arginine thought to be required to stabilize the transition state of the enzyme (pink boxes in Figure 3.17; Yan et al. 1997). Interestingly, in the *H. seosinensis* GCs, C-terminal domains did not contain these residues, suggesting they were not catalytic domains (Yan *et al.* 1997; Linder 2005). This resembles the type I AC where the C-terminal domain is non-catalytic (it lacks the asparagine-arginine pair) but promotes the catalytic activity of the N-terminal domain (Baker and Kelly 2004). For most of the *H. seosinensis* GCs, and most importantly for the
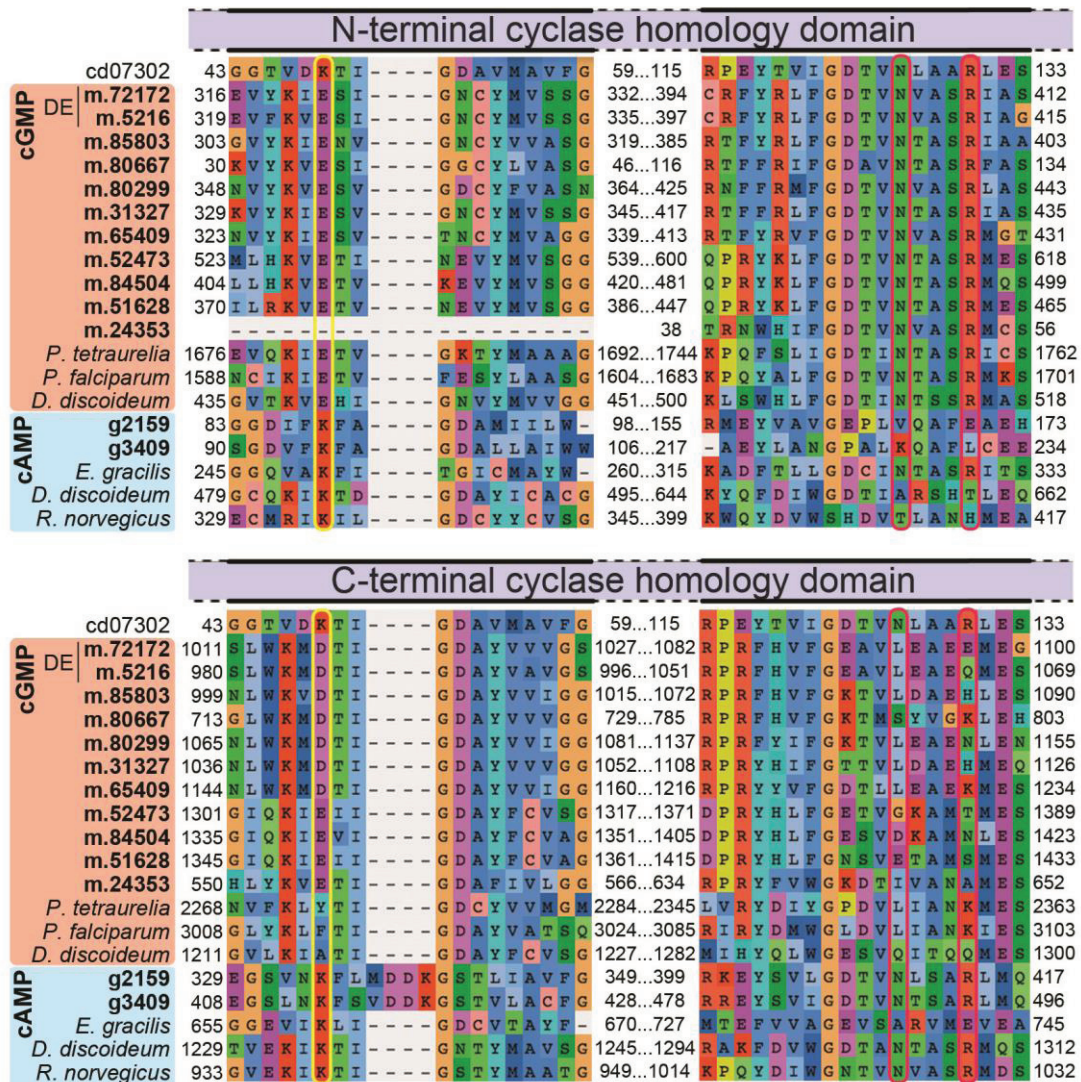
Fig. 3.17. Partial alignment of nucleotide cyclase sequences showing conservation of residues specific for adenine (K) or guanine (E, yellow boxes) and residues required for catalytic activity (N and R, red boxes). The alignment was generated using the Conserved Domain Database sequence for the cyclase homology domain (CD07302) and both N-terminal and C-terminal domains of *H. seosinensis* sequences (in bold; DE indicate sequences that were up-regulated at high salt; m.24353 is a 5' partial ORF), of guanylate cyclase from *Paramecium tetraurelia* (CAB44361.1), *Plasmodium falciparum* (CAD52725.2) and *D. discoideum* (CAB42641.1), and of adenylate cyclase from *Euglena gracilis* (BAB85619.1), *D. discoideum* (Q03100.2) and *Rattus norvegicus* (AAA40682.1).

differentially expressed ones, no evidence of regulatory domains or stimuli-derived binding domain could be detected. One possibility is that the membrane itself could trigger GC activity (Reddy *et al.* 1995; Cooper *et al.* 1998).

Antagonists of nucleotide cyclases are cyclic nucleotide phosphodiesterases (PDE) that hydrolyze nucleoside 3',5'-cyclic phosphate to nucleoside 5'-phosphate, thus terminating the associated signal. *Halocafeteria seosinensis* expressed more than ten proteins encoding a cyclic nucleotide phosphodiesterase domain (PFAM00233), including two that were up-regulated at high salt (4.4-fold and 4.6-fold, Table 3.4). These two proteins also encoded GAF domains (PFAM01590 and PFAM13492) involved in regulation and dimerization of the enzyme (Conti and Beavo 2007). PDE can be specific to cAMP, cGMP or both (dual specificity), unfortunately the characteristics leading to the recognition of these substrates with subtle differences still remain elusive (Ke *et al.* 2011). To determine whether differentially expressed PDE in *H. seosinensis* are involved in cAMP- or cGMP-dependent signaling will require biochemical characterization.

3.4.2- Sensory Histidine Kinases

Histidine kinases (HK) initiate signaling cascades by sensing varying stimuli (*e.g.* osmotic change, chemicals, nutrients or temperature) and subsequently transduce the information to a response regulator by phosphorylation, ultimately leading to gene repression/activation or allosteric regulation of target proteins (Nongpiur *et al.* 2012). The prototypical histidine kinase consists of two domains: a catalytic and ATP-binding domain (HATPase_c, PFAM02518) and a dimerization and histidine phosphotransfer domain (His kinase A, PFAM00512). In ~25% of HK, the enzyme is fused with the receiver domain to form hybrid histidine kinases in which the phosphotransfer is mediated by the histidine-containing phosphotransfer protein/domain (HpT; Gao and Stock 2009).

In *H. seosinensis,* two genes related to sensory hybrid HK were up-regulated (4.5- and 32-fold, Table 3.4). Furthermore, two genes containing the HpT module (PFAM01627) were expressed in *H. seosinensis*, one being 2.3-fold up-regulated at

high salt (ORF m.28961). The sensing domain, located at the N-terminus of HK proteins, is commonly very divergent, reflecting the idea that they detect a variety of environmental stimuli (Stock *et al.* 2000; Anantharaman *et al.* 2001; Aravind *et al.* 2002). HK genes that were the most up-regulated at high salt (m.13308, 32-fold change and m.13214, 4.5-fold change) encoded Cyclase/Histidine kinase-Associated Sensory Extracellular (CHASE) domains. These domains were surrounded by hydrophobic regions that were predicted to span the membrane by TMHMM, suggesting they might be exposed to the extracellular milieu, as predicted. Although little is known about CHASE domains and their ligands (Anantharaman and Aravind 2001; Mougel and Zhulin 2001), in plants, they bind cytokinin and regulate cell differentiation and division, but are also encoded in HK (AHK2 and AHK3) known to function in drought and osmotic stress signaling (Tran *et al.* 2007). As part of *Dictyostelium* adenylyl cyclase G, the CHASE domain is regulated by osmolarity, ensuring the cyst form stays dormant in adverse environmental conditions (van Es *et al.* 1996). These sensory proteins represent good candidates for future experimental investigation of salt-related signaling cascades in *H. seosinensis*.

Osmoadaptation in yeast is dependent on the HOG pathway, which consists of a cascade involving hybrid histidine kinases and mitogen-activated protein kinases (MAPK) that regulate the transcription factor Hog1. Several sequences from *H. seosinensis* were homologous to these genes (namely *SLN1, HHK7, SSK1, SSK2, PBS2, YPD1 and HOG1*), but they shared sequence similarity only in conserved domains (*i.e.* domains of histidine kinase, histidine kinase ATPase, response receiver, MAPK, and HpT). Whether these genes in *H. seosinensis* have the same HOG pathway-related functions as their yeast homologs is obscured by the large evolutionary distances between them (except the transcription factor Hog1, 45%ID, E value = $4e^{-104}$), and requires further experimental investigation.

Many other kinases and phosphatases were up-regulated at high salt (Table 3.4), including kinases dependent on mitogen (seven genes with 2.1- to 19-fold over-expression), calcium (five genes with 2.2- to 20-fold over-expression) or calmodulin (one gene with 3.0-fold over-expression) and other serine/threonine protein

kinases (five genes with 2.1- to 186-fold over-expression), many of which arose from gene duplication (seven and four clusters for kinases and phosphatases respectively).

3.4.3- P2X Receptors

Halocafeteria seosinensis expressed 13 genes related to P2X receptors (P2XR, PFAM00864), of which four were up-regulated (3.1- to 15-fold over-expression including one with high transcript abundance: m.49662 with 429 TPM at 30% salt) and two were down-regulated (4.2- and 42-fold repression; Table 3.4). These genes stemmed from gene duplication events (Figure 3.18). P2X receptors are known as ATP-gated cation channels involved in signaling. In vertebrates, they are trimeric receptors in which each subunit contains a short cytoplasmic N-terminal tail, a longer C-terminal tail of varying length and an extracellular domain delimited by two transmembrane domains. This overall structure was also observed in *H. seosinensis* P2XR-related sequences, although the hydrophobic signal of putative transmembrane regions was weak in a few instances. In vertebrates, P2XR are involved in varying mechanisms like synaptic transmission, cell stress, chemotransduction, inflammation and redox signaling (Khakh 2001; Burnstock and Knight 2004; Stojilkovic *et al.* 2014).

The presence of these receptors in unicellular eukaryotes was only recently established based on in-depth functional investigations of *Ostreococcus tauri* and *Dictyostelium discoideum* receptors (*e.g.* Cai 2012). Heterologous expression of *D. discoideum* and *O. tauri* P2XR in human kidney cells showed that, similarly to vertebrate P2XR, these channels were activated by ATP and permeable to calcium, although *O. tauri* P2XR was 2.5X more permeable to sodium (Fountain *et al.* 2007; Fountain *et al.* 2008). The *D. discoideum* genome encodes five P2XR genes for which the corresponding proteins all localize to contractile vacuoles, an organelle involved in osmoregulation (Ludlow *et al.* 2009). Disruption of the *p2xA* gene in strain AX4 resulted in an inability to regulate cell volume in hypotonic solution

Fig. 3.18. Maximum-likelihood phylogenetic tree for gene duplication cluster 141 (grey box): P2X receptors. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

(Fountain *et al.* 2007), although this was not observed when the gene was disrupted in a different strain, AX2 (Ludlow *et al.* 2009; Sivaramakrishnan and Fountain 2013). Although endogenous extracellular ATP released during hypotonic cell swelling seemed to play a role as a stress signal molecule (Sivaramakrishnan and Fountain 2015), the role of P2XR remained elusive in this mechanism.

Despite a low degree of amino-acid conservation in P2XR (Verkhratsky and Burnstock 2014), many functional residues are preserved between *D. discoideum*, *O. tauri* and the vertebrate sequences, including ATP-binding residues (K^{69} and K^{308} of rat P2X2), protein kinase C-interacting residues (motif Tx[K/R]) and the C-terminal motif enhancing membrane retention (YxxxK), although the latter is substituted to YxxxL in *O. tauri* (Fountain *et al.* 2007; Fountain *et al.* 2008). In *H. seosinensis* sequences, residues binding ATP were not conserved or partially conserved, suggesting these channels were insensitive to ATP (Figure 3.19). Interestingly, an essential aspartate residue (North 2002) was conserved in most of *H. seosinensis* sequences but was substituted to asparagine in four cases, as in the *O. tauri* P2XR sequence (N^{353} of *O. tauri* P2XR, circle in Figure 3.19). Fountain *et al.* (2008) suggested that this substitution could be linked to the lower calcium permeability of the *O. tauri* channel by showing that changing this site to aspartate increased calcium permeability. The protein kinase C (PKC) consensus sequence (Tx[K/R]; Wen and Evans 2009) was also conserved in all of *H. seosinensis* sequences, except in two for which the phosphorylated threonine was substituted by serine (that can potentially be phosphorylated) or alanine, suggesting that some of these channels could be regulated by cytoplasmic PKC as for the vertebrate P2XR.

The human genome encodes seven P2X proteins that assemble into homo- or hetero-trimeric receptors, multiple combinations allowing for functional versatility (North 2002). This P2XR multimeric state was also observed in *D. discoideum,* in which the genome encodes five *P2X* genes (Fountain *et al.* 2007). The *O. tauri* genome encodes four genes (Fountain *et al.* 2008). In *H. seosinensis*, 13 genes clustered into 3 independent clades, reflecting a hot spot of duplication events
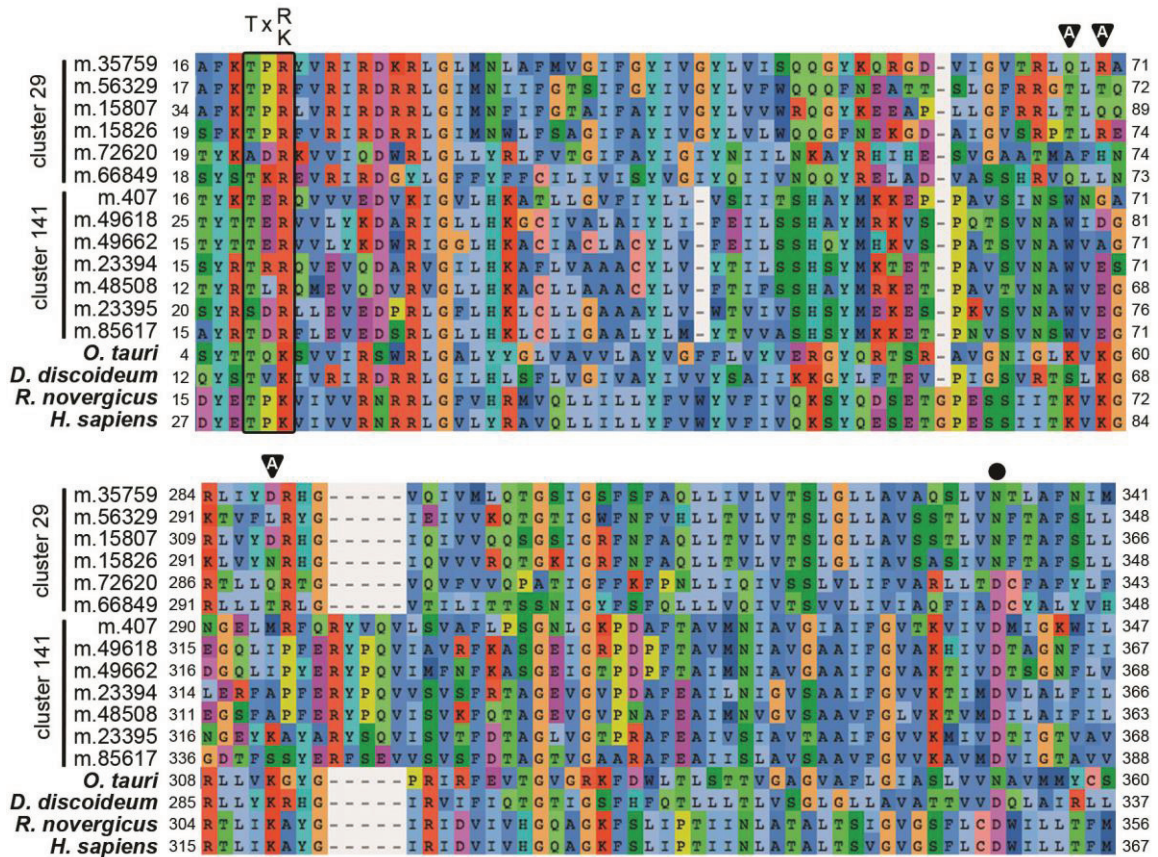
Fig. 3.19. Partial alignment of P2X receptor protein sequences indicating conservation of crucial residues for ATP binding ('A' in triangle) including reference sequences (in bold) from *O. tauri* (CEF98706.1), *D. discoideum* (XP_645378.1), *R. norvegicus* (P49653.1) and *H. sapiens* (Q9UBL9.1), and *H. seosinensis* sequences constituting gene duplication clusters 29 and 141. The protein kinase C recognition motif (Tx[K/R]) is boxed. The position indicated with a circle is essential for function and may be linked to calcium permeability.

(Figures 3.18 and 3.20, the percentage identity threshold was relaxed to 20% to gather more homologs). In EggNOG, P2XR are assigned to the class 'Intracellular trafficking, secretion and vesicular transport' that was enriched in duplicated genes in *H. seosinensis* (adjusted p-value = 4.3×10^{-3}, Figure 3.3). By representing 31% of duplicated genes in this class, P2XR contributed substantially to this enrichment. In eleven cases, genes in the same clades were located next to each other on the genome, with the exception of m.85617 and m.48508, which were separated by a single gene (encoding a *Giardia* variant-specific surface protein domain, PFAM03302).

To investigate whether this level of duplication was unique to *H. seosinensis*, I examined the genomes of 15 organisms (Table 3.5) and 359 transcriptomes sequenced during the MMETSP (excluding dinoflagellates for which genomes are known to contain highly duplicated genes) in search of sequences homologous to *H. seosinensis* P2XR-related sequences (using an E value < 0.00001 as a threshold to gather as many homologs as possible). Homologs harvested using *H. seosinensis* P2XR sequences as queries were used to interrogate the dataset from their respective species of origin in order to recover more gene duplicates (excluding sequences >90% identical). I identified P2XR-related sequences in seven (47%) of the genomes and 150 (42%) MMETSP transcriptomes. Only eight (2%) protists expressed more than ten P2XR-related proteins (*Pyramimonas parkeae, Pseudo-nitzschia fraudulenta, Pyramimonas parkeae, Chrysochromulina ericina, Chrysochromulina polylepis, Chrysochromulina rotalis, Mesodinium pulex, Dolichomastix tenuilepis*). Although duplication of P2XR-related genes happened in several organisms, high numbers of duplicates/genome seem to be rather uncommon, suggesting that these patterns of differential gene expression and gene duplication in *H. seosinensis* potentially have a role in salt adaptation, where multiple subunit combinations could lead to fine-tuned environmental stress responses.
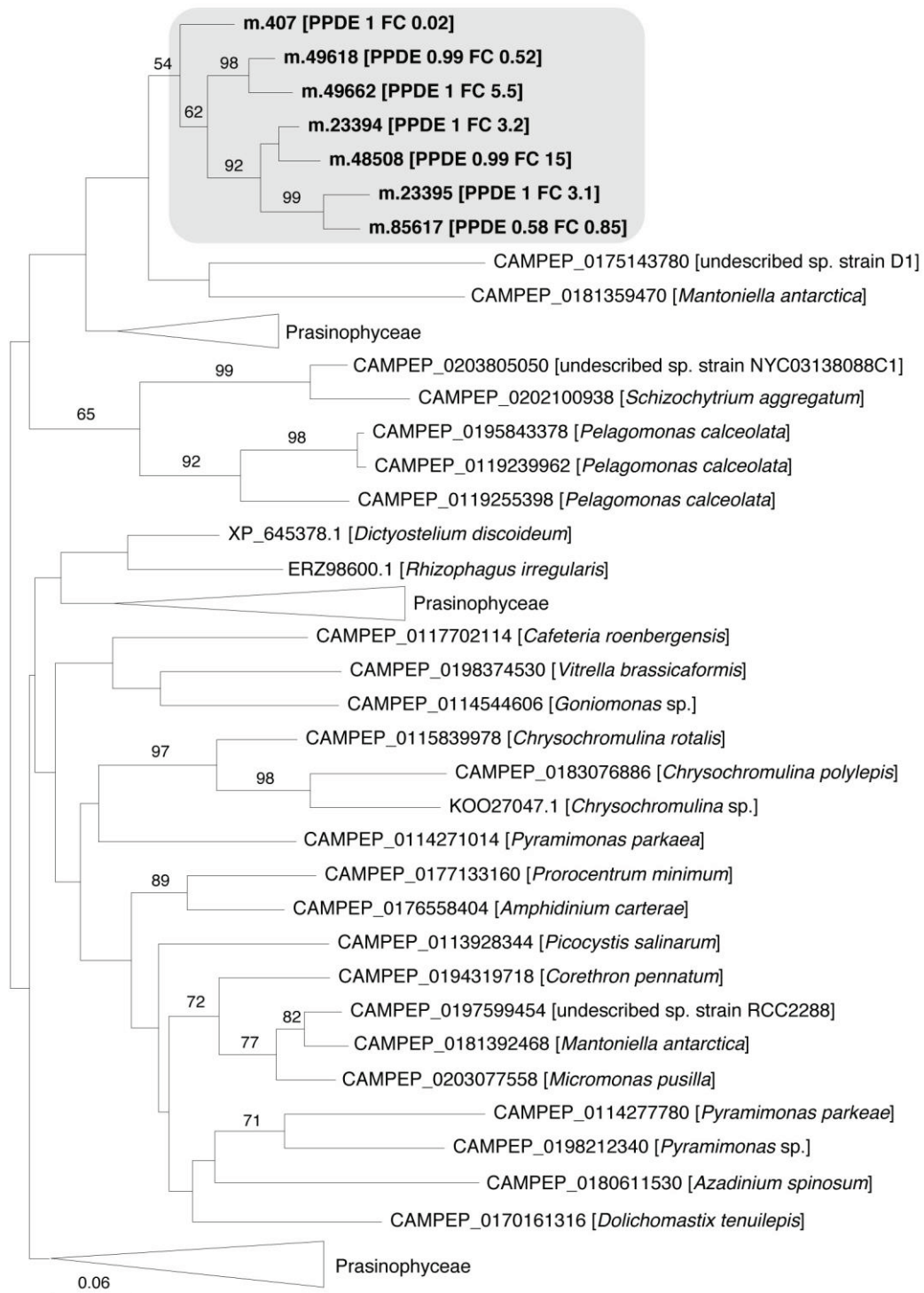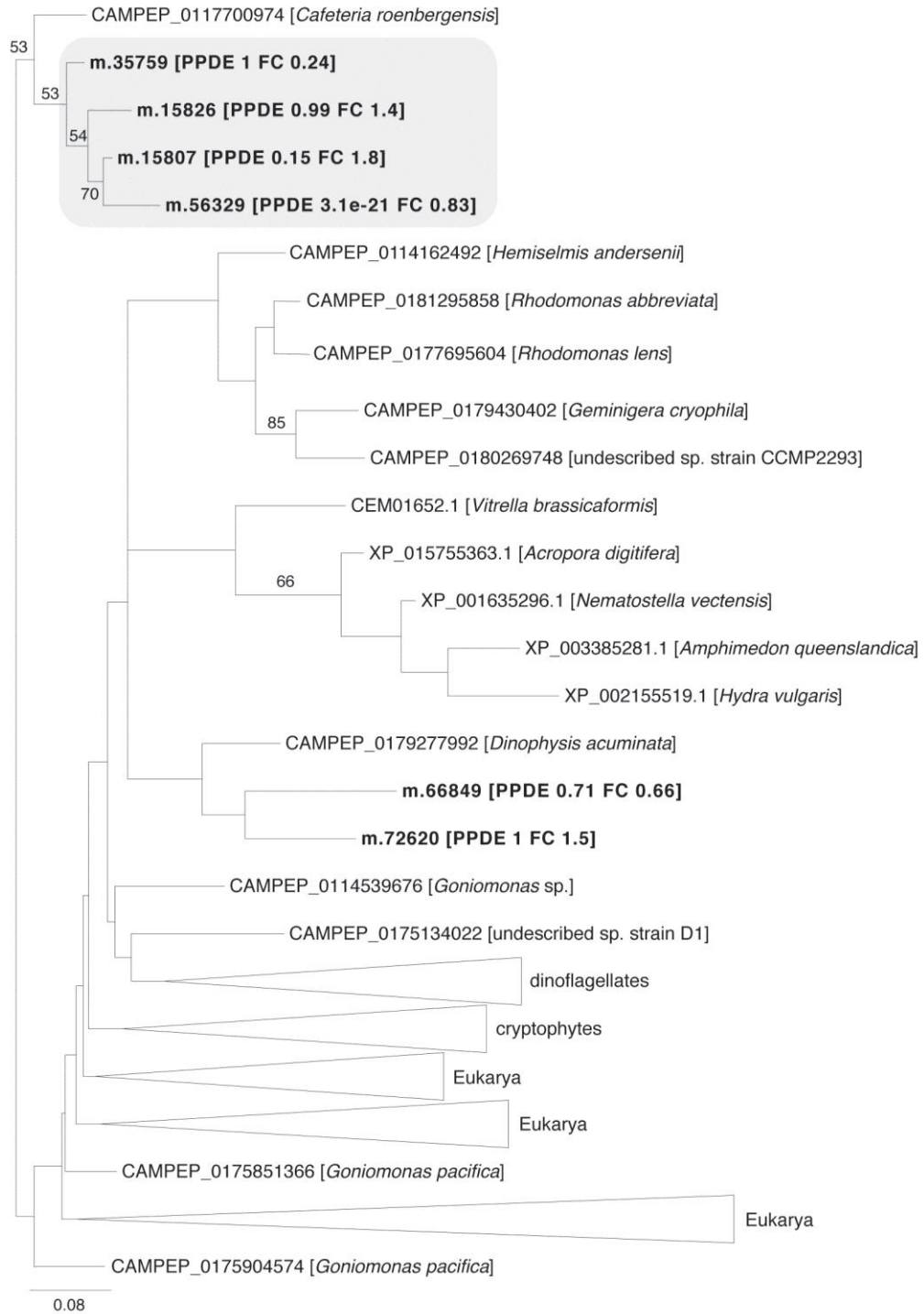
Fig. 3.20. Maximum-likelihood phylogenetic tree for gene duplication cluster 29 (grey box): P2X receptors. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

Table 3.5. Genomes used to survey gene duplication of P2X receptors

Organisms	GenBank assembly accession
*Acanthamoeba castellanii**	GCF_000313135.1
*Acytostelium subglobosum**	GCF_000787575.1
*Aureococcus anophageferrens**	GCA_000186865.1
Blastocystis hominis	GCA_000151665.1
*Capsaspora owczarzaki**	GCF_000151315.2
*Dictyostelium fasciculatum**	GCF_000203815.1
*Dictyostelium purpureum**	GCF_000190715.1
Ectocarpus siliculosus	GCA_000310025.1
Nannochloropsis gadita	GCA_000240725.1
Phaeodactylum tricornutum	GCA_000150955.2
Phytophthora infestans	GCA_000142945.1
Phytophthora sojae	GCA_000149755.2
Phytophtora parasitica	GCA_000247585.2
*Salpingoeca rosetta**	GCF_000188695.1
Thalassiosira pseudonana	GCA_000149405.2

*: Genomes encoding P2X receptors.

3.4.4- Transcription Factors

Many genes related to various transcription factors were upregulated at high salt including basic leucine zipper (BZIP) domain-containing factors, heat shock transcription factors, sirtuins, and transcription factors of the Myb superfamily.

Halocafeteria seosinensis expressed more than ten genes encoding BZIP domains (PFAM00170 and PFAM07716) and among them three were upregulated at high salt (2.1- to 40-fold, Table 3.6). The most upregulated gene (m.26350) encoded a BZIP domain related to the one of activating transcription factor 2 (ATF2, CD14687). Expression of this gene was remarkably induced at high salt, with transcript level rising from an average of 4 to 184 TPM. In mammals, ATF2 interacts with DNA at the cAMP-responsive element when phosphorylated by stress-activated protein kinases in response to varying stimuli like pro-inflammatory cytokines, UV radiation, DNA damage or variation of reactive oxygen species (van Dam *et al.* 1995). Molecular interactions between ATF2, another transcription factor c-Jun and the genomic sequence enhancer for the gene interferon-β are known (Panne *et al.* 2004). Examination of the *H. seosinensis* sequence indicated that many of the residues

Table 3.6. Expression of genes coding for transcription factors in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
BZIP transcription factors								
m.26350	3.74	184.45	1.00	40.07	0.000	5.20	0.000	5.35
m.66567	11.09	117.35	1.00	8.15	0.000	3.01	0.000	3.09
m.26504	68.52	188.84	1.00	2.14	0.000	1.08	0.002	1.09
m.8753	26.73	42.72	1.00	1.26	0.063	0.34	0.100	0.37
m.77566	205.47	337.19	1.00	1.28	0.022	0.36	0.075	0.38
m.61243	184.61	113.16	0.23	0.46	0.012	-1.07	0.075	-1.06
m.46704	98.59	126.85	0.04	0.98	0.945	-0.02	0.953	0.02
m.86230	228.47	358.31	0.48	1.23	NA	0.30	0.143	0.32
m.90423	25.14	59.77	0.74	1.88	0.008	0.89	0.043	0.89
m.31559	7.23	11.39	0.28	1.23	0.552	0.30	0.401	0.39
m.34007	35.40	74.33	1.00	1.65	0.000	0.72	0.002	0.76
Sirtuins								
m.20906	1.07	20.08	1.00	15.35	0.000	3.74	0.000	4.01
m.51236	27.68	137.90	1.00	3.76	0.000	1.89	0.000	2.00
m.8428	47.62	53.37	0.03	0.86	0.576	-0.21	0.490	-0.18
m.29062	0.60	2.36	0.00	3.00	0.256	1.24	0.056	2.23
m.29080	4.97	5.26	0.14	0.90	0.860	-0.15	0.850	-0.11
m.18553	5.64	14.49	0.92	2.02	0.067	0.98	0.038	1.06
m.18775	6.55	9.24	0.09	1.15	0.752	0.19	0.704	0.15
Heat shock transcription factor								
m.36206	7.04	29.42	1.00	3.59	0.003	1.70	0.004	1.94
m.87751	16.40	57.27	1.00	2.94	0.001	1.49	0.003	1.60
m.51685	8.75	23.30	1.00	2.11	0.001	1.06	0.004	1.05

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	$\log_2$FC	Adjusted p-value	$\log_2$FC
Heat shock transcription factor (continued)								
m.49753	24.10	51.65	1.00	1.67	0.000	0.74	0.005	0.76
m.27955	37.99	23.99	0.98	0.48	0.000	-1.02	0.003	-1.00
m.28229	43.31	35.26	1.00	0.62	0.000	-0.68	0.015	-0.65
m.77800	38.04	33.88	1.00	0.70	0.002	-0.51	0.037	-0.50
m.93231	38.79	48.36	0.04	0.95	0.698	-0.06	0.832	-0.04
m.49301	51.38	93.16	0.30	1.35	0.241	0.43	0.196	0.52
Myb superfamily of transcription factors								
m.93181	21.00	121.08	1.00	4.72	NA	2.14	0.001	2.28
m.93641	15.99	71.25	1.00	3.54	0.000	1.81	0.000	1.84
m.41207	5.87	21.75	1.00	2.82	0.000	1.48	0.001	1.62
m.83871	1.66	5.39	1.00	2.57	0.006	1.32	0.007	1.40
m.8412	0.50	1.93	1.00	2.89	0.022	1.46	0.009	1.62
m.63380	5.64	17.87	0.99	2.54	0.065	1.23	0.029	1.33
m.19672	22.56	43.88	1.00	1.50	0.000	0.59	0.002	0.62
m.45072	32.52	58.33	1.00	1.40	0.011	0.49	0.011	0.51
m.89019	44.79	100.76	1.00	1.74	0.000	0.80	0.000	0.83
m.39666	35.65	65.80	0.98	1.39	NA	0.48	0.023	0.52
m.91407	10.30	5.96	1.00	0.46	0.000	-1.12	0.000	-1.11
m.91409	2.14	2.45	0.20	0.96	0.937	-0.08	0.775	-0.18
m.26990	50.02	41.71	0.56	0.65	0.042	-0.61	0.030	-0.60
m.10692	12.08	18.58	0.08	1.17	0.454	0.23	0.378	0.19
m.94075	22.04	16.20	1.00	0.56	0.000	-0.83	0.002	-0.80
m.44643	19.23	17.59	0.20	0.71	0.149	-0.48	0.120	-0.42
m.55775	68.37	54.19	0.65	0.60	0.007	-0.71	0.013	-0.68

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC
Myb superfamily of transcription factors (continued)								
m.57095	4.14	11.38	0.28	2.30	0.123	1.08	0.076	1.20
m.42533	12.78	21.35	0.67	1.32	0.074	0.40	0.042	0.42
m.70073	18.30	26.68	0.26	1.12	0.368	0.16	0.242	0.18
m.70070	23.58	29.79	0.05	0.95	0.855	-0.07	0.898	0.03
m.32801	7.35	3.41	0.28	0.36	0.000	-1.44	0.002	-1.40
m.3350	8.60	3.49	0.99	0.31	0.000	-1.63	0.006	-1.53
m.60704	26.44	34.05	0.03	0.98	0.919	-0.03	0.955	0.01
m.92279	8.76	7.01	0.23	0.60	0.167	-0.70	0.227	-0.53
m.83876	17.40	8.53	1.00	0.38	0.000	-1.36	0.000	-1.34
m.51203	36.17	51.39	0.05	1.09	0.641	0.13	0.429	0.15
m.33860	10.39	14.30	0.04	1.05	0.734	0.07	0.570	0.09
m.42866	19.11	24.41	0.00	0.99	0.991	0.00	0.868	0.02
m.63376	87.10	132.98	0.15	1.18	0.389	0.25	0.231	0.25
m.53156	7.93	2.02	1.00	0.20	0.000	-2.23	0.000	-2.30
m.39384	19.53	21.47	0.20	0.83	0.207	-0.26	0.166	-0.24
m.23804	35.13	42.84	0.04	0.95	0.807	-0.07	0.874	-0.03
m.83256	108.97	148.71	0.05	1.05	0.704	0.08	0.507	0.10
m.83303	92.33	107.80	0.03	0.89	0.547	-0.16	0.467	-0.14
m.51232	36.39	37.81	1.00	0.80	0.015	-0.32	0.048	-0.30
m.62240	8.90	11.02	0.00	0.96	0.826	-0.05	0.801	-0.04
m.48741	47.10	49.64	1.00	0.82	0.004	-0.28	0.061	-0.26
m.80955	45.47	37.83	1.00	0.65	0.018	-0.62	0.015	-0.63
m.61959	43.65	47.35	1.00	0.85	0.092	-0.23	0.131	-0.21
m.7096	4.02	8.63	0.16	1.81	0.261	0.78	0.129	0.89

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
AP2 domain-containing protein								
m.1235	0.02	5.04	1.00	79.30	0.000	5.93	0.000	6.78
m.1227	9.01	166.75	1.00	14.10	0.000	3.77	0.000	3.90
m.52833	7.56	15.10	1.00	1.53	0.002	0.62	0.006	0.65
m.66136	79.63	71.32	1.00	0.69	0.011	-0.52	0.015	-0.50
m.74828	23.99	48.75	1.00	1.62	0.000	0.69	0.002	0.72
m.46485	21.99	20.52	1.00	0.72	0.035	-0.47	0.027	-0.47
m.66815	24.77	23.75	0.05	0.72	0.470	-0.43	0.592	-0.27
m.66817	10.62	25.54	0.37	1.98	0.039	0.94	0.027	1.01
m.63830	12.24	12.27	1.00	0.78	0.094	-0.36	0.058	-0.37
m.80289	11.20	14.84	0.03	1.00	0.983	0.01	0.901	0.02
m.51950	26.92	28.64	0.99	0.83	0.136	-0.25	0.133	-0.23
m.34779	0.55	0.28	0.02	0.42	0.252	-1.04	0.083	-1.18
m.89218	54.49	50.45	1.00	0.70	0.003	-0.50	0.014	-0.49
m.78252	155.28	182.99	0.02	0.90	0.705	-0.14	0.594	-0.14
m.68406	30.66	48.31	0.92	1.21	0.090	0.28	0.059	0.30
m.24425	29.56	36.85	0.03	0.98	0.928	-0.02	0.989	0.00

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.

involved in these interactions were conserved (Figure 3.21). However, the level of divergence was substantial outside this region, suggesting this protein is probably interacting with different partners, including the recognized activation DNA sequence. Understanding whether the *H. seosinensis* gene performs similar functions as ATF2, including the nature of the stress that led to increased expression, requires more investigation.
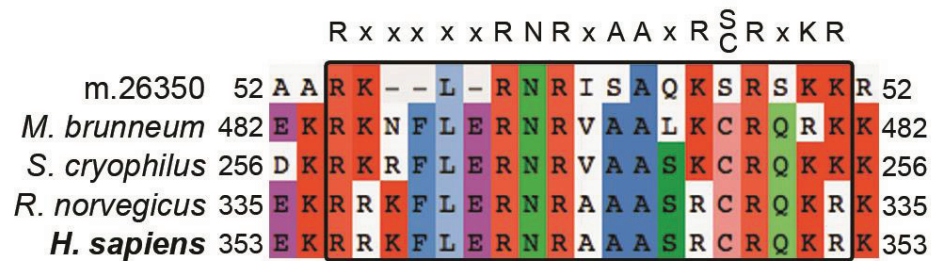
Fig. 3.21. Partial alignment of activating transcription factor 2 protein sequences showing conservation of DNA recognition site (Rx$_5$RNRxAAxR[S/C]xKR) based on study of crystal structures of the human factor (in bold, P15336; Panne et al 2004). The alignment also includes sequences from *H. seosinensis* (m.26350), *Metarhizium brunneum* (XP_014545513.1), *Schizosaccharomyces cryophilus* (XP_013025988.1) and *Rattus norvegicus* (NP_112280.1).

Silent information regulator proteins (sirtuins) were first described as de-acetylases of histones that induce a more compact protein conformation around DNA, leading to gene repression. Later they were found to also de-acetylate a wider range of proteins, such as transcription factors, co-factors and enzymes (Feige and Auwerx 2007, 2008; Haigis and Sinclair 2010). In mammals, they are implicated in a wide range of cellular processes including aging, metabolism and tolerance to oxidative stress (Feige and Auwerx 2008). Via sensing levels of their co-factor NAD$^+$, they regulate gene expression based on the energetic state of the cell. There are seven sirtuin genes encoded in the human genome (Frye 2000) and *H. seosinensis* expressed seven genes related to SIRT1, 2, 4 and 7 (Table 3.6). Two of them,

encoding a SIRT7 domain (CD01410), were highly upregulated at high salt (3.8- and 15-fold).

The cellular function of SIRT7 is the least known of the sirtuins (Liu and Chen 2015). Enriched in metabolically active tissues, SIRT7 upregulates ribosomal RNA gene transcription by de-acetylating residue K^{373} of RNA polymerase I subunit PAF53 (Ford *et al.* 2006; Chen *et al.* 2013). Under stress conditions (*e.g.* glucose deprivation, treatment with anisomycin, actinomycin D or AICAR), hyper-acetylation of PAF53 caused by SIRT7 inhibition ultimately leads to slower growth (Chen *et al.* 2013). At the homologous position to the acetylated lysine residues, the *H. seosinensis* PAF53 sequence, but also yeast sequences, instead harbor an arginine. This position is involved in DNA binding in yeast (Geiger *et al.* 2010) suggesting that no acetylation mechanism regulates DNA binding at this position in *H. seosinensis*. Despite apparent differences between the mammalian SIRT7 regulation system, my differential expression results suggest that expression of genes related to SIRT7 might also be linked to cellular stress in *H. seosinensis*.

Halocafeteria seosinensis expressed nine genes containing heat-shock factor (HSF)-type DNA-binding domains (PFAM00447); three of them were up-regulated at high salt (2.1- to 3.6-fold). HSFs regulate the expression of heat-shock proteins, chaperones that survey the protein pool to ensure proper folding, by targeting heat-shock DNA elements (HSEs) allowing for recruitment of the transcription machinery. HSEs consist of inverted repeats separated by two variable nucleotides (*e.g.* GAAxxTTC; Pelham and Bienz 1982). Sequence comparison of *H. seosinensis* HSF-related sequences to human HSF1 indicated that all sites involved in binding HSE were conserved in two of them, the most upregulated and the most downregulated genes, suggesting they might bind canonical HSE sequences (Figure 3.22; Neudegger *et al.* 2016). However, investigating the upstream genomic region of HSP genes in *H. seosinensis*, only one had the canonical 'TTC..GAA' and this gene was not differentially expressed. Nonetheless, these differentially expressed factors can potentially be linked to the over-expressed chaperones discussed in section 3.5.3.
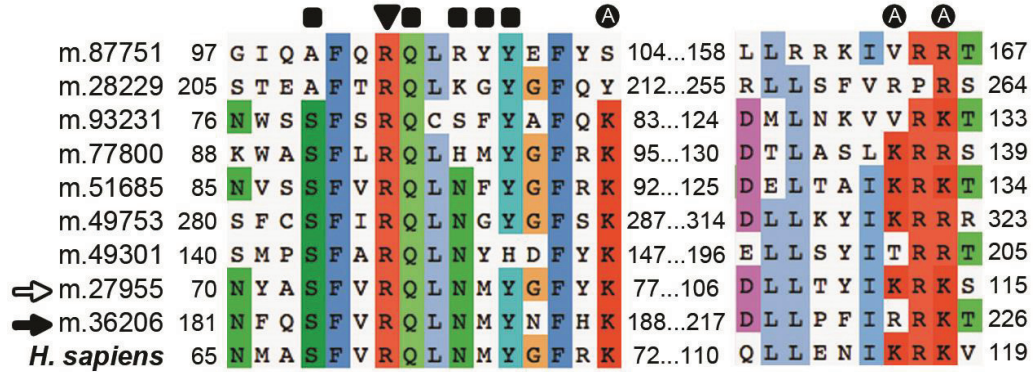
Fig. 3.22. Partial alignment of heat shock factor protein sequences indicating conservation of residues that, in human HSF1 (in bold, NP_005517.1; Neudegger et al. 2016), interact with the guanine of the heat shock DNA element (triangle), stabilize and orient the interaction (squares) and residues that, when acetylated, lead to repression of the transcription factor (circled "A"). Arrows indicate the most up-regulated (filled) and down-regulated (empty) genes at high salt in *H. seosinensis*.

Halocafeteria seosinensis expressed ~40 genes encoding domains of the MYB superfamily of transcription factors (PFAM00249 and PFAM13921), of which six were substantially upregulated at high salt (2.5- to 4.7-fold, Table 3.6). Transcription factors of the MYB superfamily are involved in the regulation of a wide range of physiological and biochemical processes, including metabolite biosynthesis, development, cell cycle control and light and hormone signaling (Chen *et al.* 2006). In plants, they also regulate abiotic stress response gene expression by modulating root growth, biosynthesis of polyamines in leaves, stomatal movement (control of pore opening/closure), cell wall, cuticle and suberin modification (to adjust barriers that protect plants from water loss) and biosynthesis of flavonoids as strong antioxidant molecules to protect cells against reactive oxygen species (Baldoni *et al.* 2015; Roy 2016). It is worth noting that MYB factors are also regulated at the post-translational level by phosphorylation, acetylation, sumoylation and proteolytic cleavage, and thus more MYB domain-containing factors in *H. seosinensis* could be involved in salt stress response (Pireyre and Burow 2015).

Sixteen genes encoding AP2 DNA-binding domains (PFAM00847) were expressed in *H. seosinensis*, including two genes that were upregulated at high salt

(14- and 79-fold, Table 3.6). ORF m.1227 notably rose from an average of 9 TPM at 15% salt to 167 TPM at high salt. First thought to be plant-specific, AP2 domain-containing proteins were identified in a wider range of organisms (Magnani *et al.* 2004; Oberstaller *et al.* 2014). Similarly to MYB factors, AP2 domain-containing factors in plants are known to be involved in a variety of key developmental steps and in environmental stress response pathways (Licausi *et al.* 2013; Dey and Vlot 2015). In the apicomplexan parasites *Plasmodium* and *Toxoplasma*, they regulate cell cycle progression and parasite virulence, amongst other functions (Behnke *et al.* 2010; Walker *et al.* 2013). However, even though highly conserved AP2-related domains between *Plasmodium falciparum* and *Cryptosporidium parvum* kept their DNA binding specificity, the overlap of targeted genes between the species is quite low (De Silva *et al.* 2008). This observation suggests that great caution is required when transferring knowledge on transcriptional regulation networks between species. As for the other transcription factors, the targeted genes in *H. seosinensis* stay undetermined and further investigation is required to specify what aspect of high salt adaptation is regulated by AP2 domain-containing proteins.

In summary, exposure to high salt stimulated the expression of many genes likely involved in modulating transcription directly or indirectly in *H. seosinensis.* I identified the importance of the cyclic nucleotide-dependent (especially cGMP) signaling pathway by detecting the differential expression of guanylate cyclases and cyclic nucleotide phosphodiesterases. These cascades were potentially initiated by G protein-coupled receptors, extracellular sensors that were involved in several gene duplication events and of which at least four were activated at high salt. Genes related to P2X receptors were massively duplicated and differentially expressed. Although these genes have yet elusive functions in *H. seosinensis*, consistent with their salt-responsive expression, a gene of this family was linked to osmoregulation in *D. discoideum*. Finally, several transcription factors known to be involved in stress response (BZIP, SIRT7, MYB, AP2 and heat shock factors) were upregulated at high salt. This is consonant with a gene response to the extreme condition to which *H. seosinensis* was exposed. Further work should characterize in greater detail these

responsive genes in order to identify the interacting partners of their respective proteins, the target genes they regulate or the factors they sense.

3.5- Stress Response: Dealing With Reactive Oxygen Species and Surveying Troublemakers

Environmental stresses lead to increased cellular levels of reactive oxygen species (ROS; Lushchak 2011; Sharma *et al.* 2012). Mitochondria are probably the dominant source of ROS, where stress conditions induce an imbalance in the electron transport chain leading to reverse electron transport and undesired oxidation of oxygen by complex I, generating ROS (Tomanek 2015). The ability to survive such stress resides in the capacity to manage these destructive ions that otherwise react with DNA, proteins and lipids (Yu 1994). For example, plant species with greater antioxidant capacities show a greater resistance to salt stress (Panda and Das 2005). As well, the capacity to repair or to degrade problematic proteins is key to survival in extreme conditions. Grown at high salt, *H. seosinensis* up-regulated an arsenal of genes, some of which were amongst the most highly transcribed in this condition, that are involved in protecting a wide diversity of molecules (especially proteins and lipids) in several cellular compartments. Proteins encoded by these genes either contributed to directly neutralize ROS (*e.g.* superoxide dismutase and peroxidase), were involved in repairing and protecting cellular components affected by ROS (*e.g.* glutathione-dependent proteins and chaperones such as heat shock proteins), or were involved in controlling potential sources of ROS (*e.g.* quinone oxidoreductase).

3.5.1- Direct Detoxification of ROS by Superoxide Dismutase and Peroxidase

Two highly expressed and up-regulated genes, superoxide dismutase (SOD) and peroxidase, indicated that the level of ROS was likely higher in *H. seosinensis* grown at high salt. SODs catalyze the dismutation of superoxide radicals into less

damaging species, namely hydrogen peroxide (H_2O_2) and molecular oxygen (Fridovich 1995). Subsequently, peroxidases reduce H_2O_2 to water.

Halocafeteria seosinensis expressed two SODs: one (ORF m.9318) was predicted to reside in the cytosol, based on absence of a targeting signal, and showed a 2.7-fold increase at high salt (Table 3.7). The second (m.23828) encoded a mitochondrial transit peptide, was concordantly closely related to mitochondrial manganese (Mn) SOD, and was only slightly up-regulated (60% increase). The expression level for these genes was notably elevated in high salt, with 1861 TPM and 824 TPM respectively (ranks 97 and 175). Although m.9318 was related to bacterial sequences, it was clearly eukaryotic since it was encoded with introns on the genome, and was also closely related to several other protist sequences.

MnSODs in metazoans, plants and yeasts commonly reside in mitochondria, while the evolutionarily unrelated cytosolic form of SOD instead depends on copper or zinc (*e.g.* Holley *et al.* 2012; Miriyala *et al.* 2012; Sarsour *et al.* 2014; Bresciani *et al.* 2015). Exceptions have been reported where MnSODs are localized to the cytoplasm in marine invertebrates (Wang *et al.* 2015), to the cytosol in the liver of primates (Steinman 1982) and in many fungi (Fréalle *et al.* 2005), or to the peroxisomes in plants (del Río *et al.* 2002). Protists are generally assumed to express cytosolic iron-dependent (Fe) SOD instead of MnSOD. This is based on studies of a few parasites (Bécuwe *et al.* 1996; Loftus *et al.* 2005; Asojo *et al.* 2006; Wilkinson *et al.* 2006). However, reports are accumulating on the occurrence of cytosolic MnSODs in protists including *Phytophthora* spp. and *Symbiodinium* spp. (Blackman *et al.* 2005; Krueger *et al.* 2015); note that chloroplast MnSODs have also been reported, in *Euglena gracilis* and *Chlamydomonas reinhardtii* (Kanematsu and Asada 1979; Allen *et al.* 2007).

FeSODs and MnSODs are evolutionarily related and, due to their wide distribution in all domains of life, they are thought to have ancient origins (Miller 2012). Concordantly, the *H. seosinensis* SODs were related to each other (44% identical), to cytosolic MnSODs (<46% identical to MnSOD from *Phytophthora*

Table 3.7. Expression of genes involved in stress response in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-values	log_2FC	Adjusted p-values	log_2FC
Superoxide dismutases								
m.9318	549.68	1860.88	1.00	2.66	5.92E-06	1.38	1.22E-03	1.42
m.23828	396.37	823.78	1.00	1.59	1.71E-02	0.66	1.77E-02	0.69
Peroxidase								
m.79082	161.47	3985.78	1.00	18.42	4.85E-27	4.04	1.49E-05	4.29
Cytosolic glutathione transferases								
m.57692	108.13	1279.90	1.00	9.06	2.49E-18	3.08	2.71E-05	3.26
m.3188	384.89	333.97	0.03	0.65	1.65E-01	-0.59	1.31E-01	-0.57
m.69131	115.33	581.10	1.00	3.91	3.48E-18	1.95	2.61E-05	1.98
m.64209	569.94	518.85	0.03	0.70	2.21E-01	-0.49	1.50E-01	-0.49
m.34325	94.88	91.27	0.26	0.73	5.61E-02	-0.44	5.53E-02	-0.40
m.75424	380.60	264.22	0.07	0.51	1.95E-02	-0.93	3.83E-02	-0.87
m.24329	74.34	4.59	1.00	0.05	7.73E-45	-4.27	6.74E-06	-4.33
m.27085	34.54	31.98	0.14	0.72	3.48E-01	-0.44	2.66E-01	-0.38
Microsomal glutathione transferases								
m.21576	306.54	1064.12	0.99	2.66	1.17E-04	1.37	2.78E-03	1.43
m.17240	266.69	118.84	0.99	0.34	1.32E-05	-1.50	1.41E-03	-1.51
m.14941	109.36	81.56	0.17	0.56	4.49E-02	-0.79	5.92E-02	-0.72
m.31322	251.52	67.34	1.00	0.21	1.52E-25	-2.24	4.89E-04	-2.23
Glutaredoxins								
m.39259	8.04	46.48	1.00	4.76	5.71E-10	2.18	1.02E-04	2.27
m.72582	222.23	821.73	1.00	2.81	5.82E-04	1.44	4.17E-03	1.56
m.84602	332.08	286.24	0.08	0.66	9.97E-02	-0.59	7.45E-02	-0.57
m.85098	186.64	164.18	0.07	0.67	1.72E-01	-0.55	1.26E-01	-0.53

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-values	log$_2$FC	Adjusted p-values	log$_2$FC
Gluthathione conjugate pumps								
m.20314	24.07	24.63	1.00	0.78	6.40E-02	-0.34	1.46E-01	-0.33
m.77447	10.03	11.10	0.05	0.83	3.76E-01	-0.27	4.75E-01	-0.21
Peroxiredoxins								
m.14635	46.81	472.62	1.00	7.78	8.83E-24	2.90	1.06E-05	3.00
m.10580	2937.56	3301.80	0.01	0.86	5.52E-01	-0.20	4.76E-01	-0.18
m.351	76.59	50.98	0.91	0.51	8.25E-04	-0.96	5.48E-03	-0.92
m.25391	985.88	392.85	0.28	0.30	2.00E-04	-1.67	5.58E-03	-1.66
m.57627	283.58	563.30	0.59	1.52	1.26E-01	0.60	7.84E-02	0.62
NADPH-dependent alcohol dehydrogenase								
m.77193	2.37	626.65	1.00	202.04	7.79E-156	7.51	8.19E-05	7.72
Small heat shock proteins								
m.64653	91.24	812.76	1.00	6.70	NA	2.54	1.17E-03	2.89
m.65830	163.13	904.27	1.00	4.24	1.00E-05	1.98	1.09E-03	2.22
m.67185	54.76	273.96	1.00	3.75	2.54E-04	1.80	1.74E-03	2.13
m.8880	20.88	2.87	0.99	0.10	8.40E-09	-3.04	3.01E-03	-2.98
ATP-dependent ClpB protease								
m.91451	12.06	100.38	1.00	6.27	1.24E-10	2.54	1.95E-04	2.74
HSP70-like proteins								
m.81151	83.27	1656.85	1.00	15.24	6.23E-20	3.75	3.32E-05	3.99
m.13417	1456.47	1290.18	0.44	0.70	2.87E-02	-0.50	4.58E-02	-0.46
m.13416	172.15	0.36	1.00	0.00	3.02E-165	-9.07	6.16E-07	-9.38
m.73633	190.27	243.81	0.01	1.00	9.96E-01	0.00	9.86E-01	0.00
m.66450	2.22	3.87	0.15	1.40	7.98E-01	0.32	2.04E-01	2.44
m.44257	996.51	1322.50	0.01	1.02	9.24E-01	0.04	8.04E-01	0.06

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-values	$\log_2$FC	Adjusted p-values	$\log_2$FC
HSP70-like proteins (continued)								
m.79811	810.87	711.59	0.92	0.68	3.10E-03	-0.55	1.40E-02	-0.54
m.1530	105.11	193.40	0.84	1.46	8.22E-03	0.54	4.32E-02	0.59
Protein-L-isoaspartate/D-aspartatyl O-methyltransferase								
m.50853	46.34	133.20	1.00	2.18	1.07E-03	1.10	3.64E-03	1.20
NAD(P)H:quinone oxidoreductases								
m.42655	10.26	160.29	1.00	11.73	3.63E-10	4.23	4.53E-06	3.70
m.35863	571.80	1581.48	0.99	2.14	3.55E-02	-0.86	3.60E-03	1.11
m.79636	167.58	435.01	1.00	1.98	9.31E-04	0.89	7.50E-04	1.02

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); $\log_2$FC, $\log_2$ fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.

nicotianae, accession AAY57577) and to FeSODs (<43% identical to FeSOD from *Trypanosoma grayi*, KEG14863.1). I therefore conducted a phylogenetic analysis to investigate the predicted localization of *H. seosinensis* SODs based on the phylogenetic context. This analysis revealed three clades: one well-supported clade containing sequences annotated as mitochondrial and, as expected, m.23828 (100% bootstrap support), another well-supported clade containing the described FeSOD sequences (93% bootstrap support), and a third clade containing m.9318 that showed weak support (bootstrap support of 56%, Figure 3.23). I investigated further the putative cellular localization of eukaryotic sequences by combining predictions from TargetP and mitoprotII. Transit peptides were predicted for 67% of sequences in the 'mitochondrial' clade (probability > 0.75). Conversely, none of the sequences in the m.9318-containing clade and only 7% of sequences in the FeSOD clade had a mitochondrial transit peptide confidently predicted. This indicates that the bifurcation between the 'mitochondrial' clade and the other two clades might be linked to varying subcellular functions.

All sequences in the alignment possessed the highly conserved residues involved in interacting with the metallic cofactors or located in the immediate environment of these, with the exception of one tryptophan residue that was substituted to phenylalanine or leucine in 26 sequences (Wintjens *et al.* 2004). Interestingly, only three sequences included in the m.9318-containing and 'mitochondrial' clades showed deviations from the MnSOD-specific residues (*i.e.* 202/205 had perfect match), while sequences clustering with FeSOD had completely different residues (N = 119 sequences; Figure 3.24). This indicated that the recovered clades probably gathered sequences according to their cofactor specificity and that both *H. seosinensis* SODs were likely Mn-specific enzymes. This contrasts with the general assumption that cytosolic SODs in eukaryotes depend on copper, zinc or iron (see above) and suggests that cytosolic MnSOD might be more common in protists than previously believed (see also Blackman *et al.* 2005; Krueger *et al.* 2015). Although more investigation is required to confirm the subcellular localization and the cofactors of *H. seosinensis* SODs, my results indicate that a
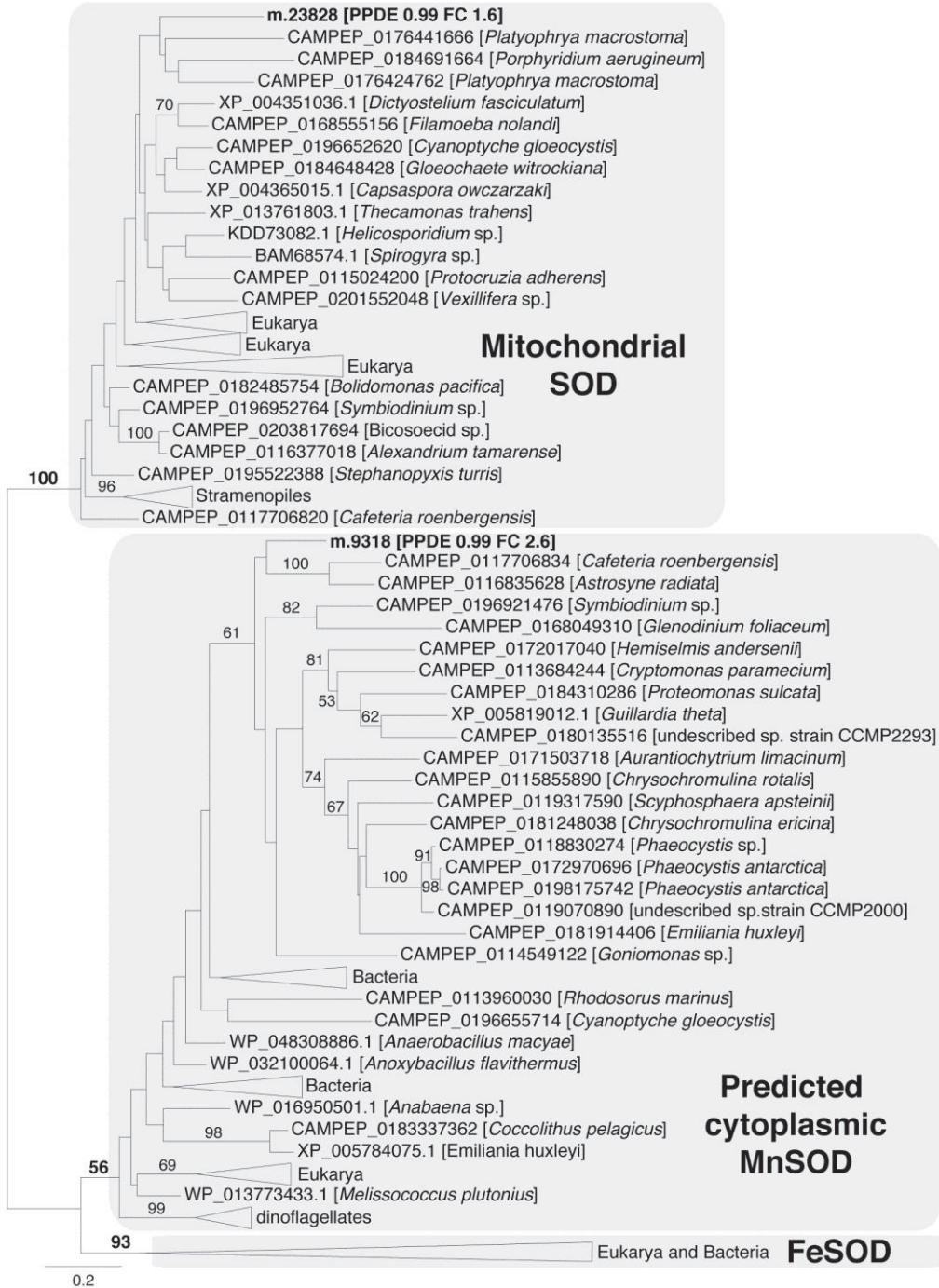
Fig. 3.23. Maximum-likelihood phylogenetic tree for superoxide dismutase (SOD) sequences showing three clades exclusively including sequences annotated as mitochondrial SOD, sequences for which no mitochondrial targeting peptide could be predicted excluding bacterial sequences (Predicted cytoplasmic MnSOD), and sequences of described iron-dependent SOD (FeSOD). For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.
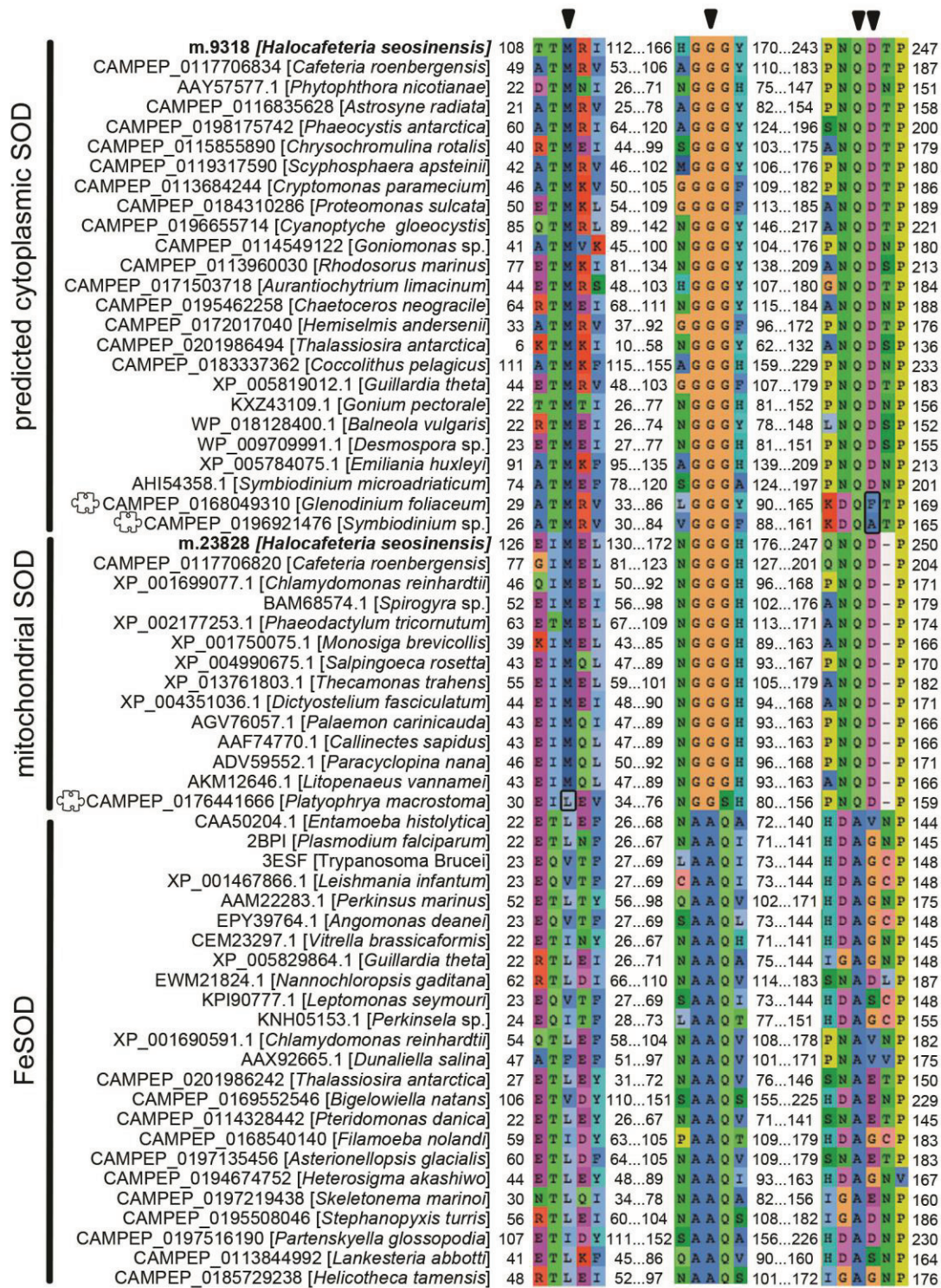
Fig. 3.24. Partial alignment of superoxide dismutase (SOD) sequences showing residues exclusively conserved in manganese-dependent SOD (triangle) as described by Wintjens *et al.* (2004). Selected taxa are listed according to the clade in which they cluster in Figure 3.23. Puzzle pieces indicate the three sequences out of 205 that deviate from the expected motif (boxed residues).

cytosolic Mn-dependent SOD contributed to protect H. seosinensis against ROS at high salt.

Subsequently, the reaction product of SOD (H_2O_2) is degraded by peroxidases and catalases. One enzyme related to peroxidase, ORF m.79082, had a striking expression pattern, with average transcript level raised 18-fold from 161 TPM at 15% salt to 3,986 TPM at 30% salt (*i.e.* the 11th most expressed transcript at high salt; Table 3.7). It was related to a family of uncharacterized peroxidase-related bacterial enzymes (TIGR01926). The next most closely related sequences belong to the alkylhydroperoxidase (TIGR00778) and carboxymuconolactone decarboxylase (PFAM02627) families. Although no information is available yet on the enzymes of the TIGR01926 family, enzymes of the alkylhydroperoxidase family were linked to increased resistance to oxidative stress (Hillas *et al.* 2000; Paterson *et al.* 2006). *Mycobacterium tuberculosis* alkylhydroperoxidase AhpD catalyzes the oxidation of peroxides via the action of residues Glu^{118}, Cys^{130}, His^{132}, Cys^{133} and His^{137} (Koshkin *et al.* 2003). In my alignments, these residues were conserved in most members of the above-mentioned families, including m.79082, with the exception of His^{132} (motif $Ex_{11}Cx_2Cx_3H$, Figure 3.25).
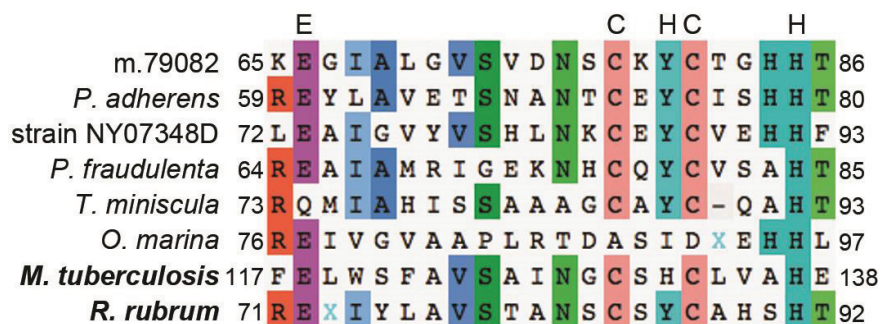
Fig. 3.25. Partial alignment of peroxidase sequences showing residues essential for catalytic activity (as displayed on top of the alignment) in *Mycobacterium tuberculosis* AhpD protein (in bold, ALB19631) as described by Koshkin *et al.* (2003). This gene, commonly found in bacteria, was also detected in *H. seosinensis* (m.79082) and in MMETSP transcriptomes generated from *Protocruzia adherens* (CAMPEP_0114993680), an undescribed Labyrinthulomycetes sp. strain NY07348D (CAMPEP_0205091550), *Pseudonitzschia fraudulenta* (CAMPEP_0201178858), *Thalassiosira miniscula* (CAMPEP_0184455648) and *Oxyrrhis marina* (CAMPEP_0205046082). *Rhodospirillum rubrum* AhpD sequence (in bold, 2OUW) for which a crystal structure is available is also shown as a reference.

Interestingly, m.79082 clustered with proteobacterial sequences to the exclusion of five homologous sequences harvested from the MMETSP dataset (Figure 3.26). These protist sequences did not cluster together, suggesting that independent LGT events might have occurred in each of these taxa (though see below). Protist sequences, including m.79082, had 100 nucleotide-long segments >70% identical to bacterial sequences (meeting the criterion that I normally used to discard sequences during my analyses). However, m.79082 was encoded in the genome assembly next to intron-containing genes. The extremely high expression of this gene also strongly argues against m.79082 being a contaminating bacterial sequence, since eukaryotic mRNA was poly-A selected (see section 2.1) and *bona fide* protein-coding transcripts from the food source *Haloferax* never had expression levels >10 TPM.

However, bacterial contamination in the cases of MMETSP sequences was harder to assess, particularly due to the high sequence conservation of these peroxidases (including at the nucleotide level) and because contamination from food bacteria was not as tightly controlled in the MMETSP compared to the *H. seosinensis* data generation reported here. Nonetheless, the low number of sequences homologous to m.79082 in this dataset indicated that this enzyme was, at best, not commonly found in protists, further suggesting that *H. seosinensis* acquired this peroxidase-coding gene laterally from bacteria. Since this gene was not detected in *C. roenbergensis* transcriptome, it is plausible that is was acquired relatively recently in *H. seosinensis*, plausibly co-incidentally with the adaptation to a halophilic lifestyle. Alternatively, this gene could have been lost in *C. roenbergensis* or not expressed under the growth condition tested during the MMETSP data generation. Given its high over-expression at high salt (18-fold increase) and its high transcript abundance (3,986 TPM), this LGT event likely contributed meaningfully to increase *H. seosinensis* resistance to high salt stress.
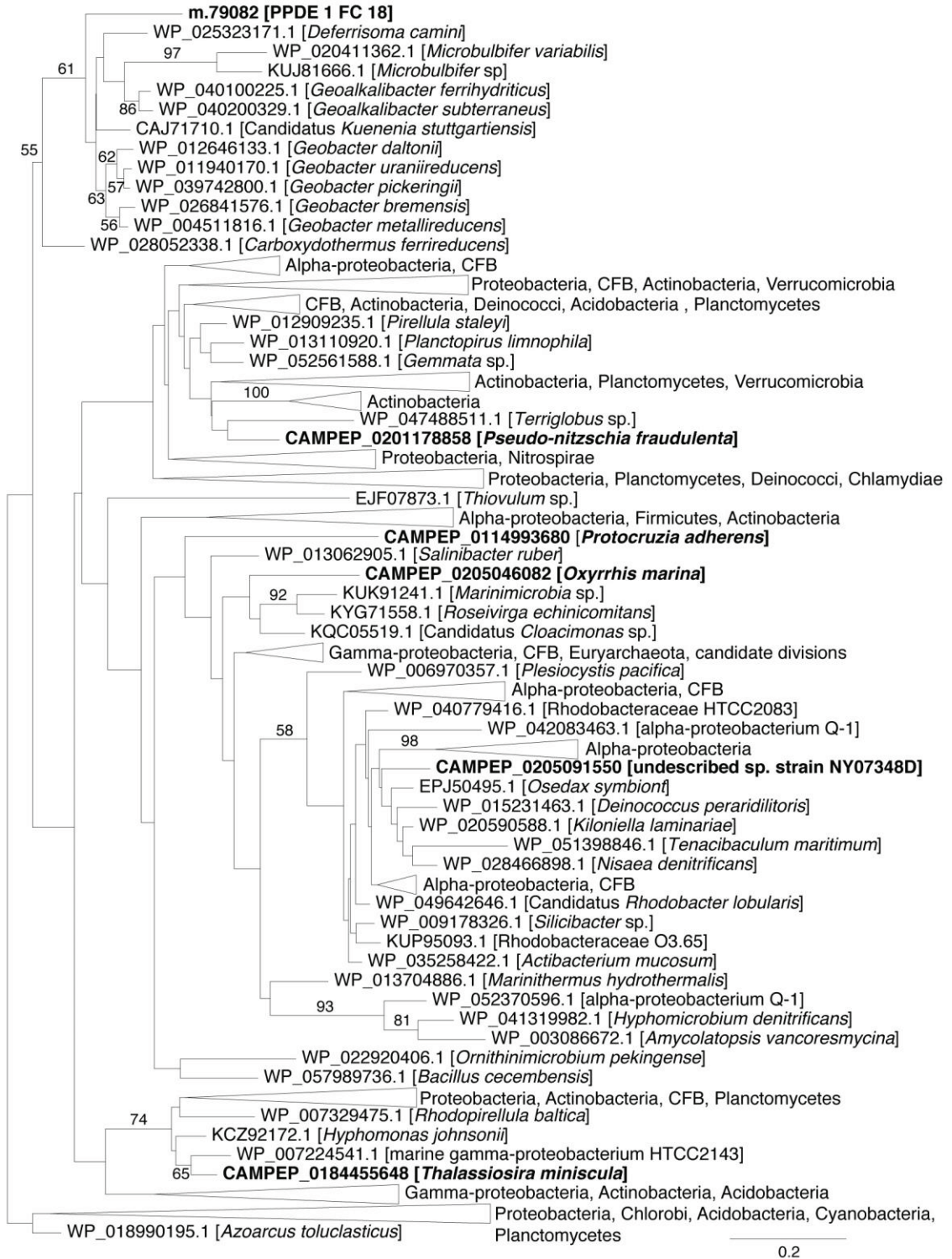
Fig. 3.26. Maximum-likelihood phylogenetic tree for peroxidase showing the relatedness of eukaryotic sequences (in bold) to prokaryotic sequences. For *H. seosinensis* sequence (m.79082), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

153

3.5.2- Glutathione-Dependent Detoxification

The glutathione system is ubiquitous in living cells. Glutathione can accumulate up to millimolar intracellular concentrations, and represents a central metabolite for removing or modifying endogenous electrophilic compounds and several xenobiotics (Meister and Anderson 1983; Deponte 2013). Glutathione is a tripeptide (Glu-Cys-Gly) that neutralizes harmful electrophilic compounds either spontaneously, or enzymatically in reactions catalyzed by glutathione transferases (GT). GTs have evolved into at least 10 classes with diverse physiological functions, activities and substrate specificities. *Halocafeteria seosinensis* expressed GTs of the mainly cytosolic classes Beta, Omega, Sigma and Theta, and of the Membrane-Associated Proteins in Eicosanoid and Glutathione metabolism (MAPEG) superfamily. High salt also up-regulated *H. seosinensis* genes coding for proteins that employ glutathione as co-factor, like glutaredoxin and peroxiredoxin, and for cinnamyl-alcohol dehydrogenase which is potentially involved in homeostasis of the NADPH that provides the reducing power for enzymes participating in ROS detoxification (*e.g.* glutathione reductase and thioredoxin reductase).

3.5.2.1- Glutathione Transferases of the Beta Class

One of the genes related to Beta class of GTs (ORF m.57692) was highly up-regulated at high salt (9.1-fold) and had high transcript abundance (average of 1280 TPM at 30% salt, rank 136). Further, m.57692 was closely related to another non-differentially expressed gene (ORF m.3188) and also to bacterial sequences, suggesting it was acquired by lateral gene transfer, followed by gene duplication and neo-functionalization. However, homologs of these proteins were also detected in the MMETSP dataset and, without a resolved phylogeny, I could not demonstrate that these genes were unambiguously acquired by *H. seosinensis* from bacteria.

In GTs of the Beta class, binding of glutathione results from a network of interactions involving conserved residues C^{10}, H^{106} and Y^{157} (*E. coli* numbering; Casalone *et al.* 1998; Allocati *et al.* 1998; Inoue *et al.* 2000). Residues S^{11}, H^{15} and E^{198} are also crucial for Beta class GTs, since substitutions of amino acid at these

positions compromise catalytic activity due to loss of structural stability (Federici *et al.* 2007; Federici *et al.* 2009). These residues are strictly conserved in the Beta class but not in the other classes (Allocati *et al.* 2009). Alignment comparison of *H. seosinensis* Beta GTs with *Escherichia coli*, *Ochrobactrum anthropi, Burkholderia xenovorans* and *Proteus mirabilis* GT sequences, for which crystal structures have been studied, demonstrated the conservation of all these crucial residues, supporting the inferred identity for these proteins in *H. seosinensis* (Figure 3.27)*.*

Beta class GTs participate in detoxification reactions against a variety of xenobiotics and metabolites resulting from oxidative and chemical stresses. In *O. anthropi*, increased GT expression is observed with exposure to membrane-active toxins, like chlorinated phenols or atrazine, and under oxidative stress derived from aromatic substrate degradation, but hydrogen peroxide does not induce expression directly (Favaloro *et al.* 2000; Tamburro *et al.* 2004). By contrast, in *P. mirabilis*, GT B1-1 mRNA and protein levels increase as a direct response to H_2O_2 exposure, and the enzyme is involved in detoxification of antimicrobial agents (Allocati *et al.* 2003). However, these GTs predominantly localize to the periplasmic space (Allocati *et al.* 1994; Tamburro *et al.* 2004). This compartment is absent in eukaryotes, and the absence of predictable target peptides in *H. seosinensis* GT sequences suggests that these enzymes most likely reside in the cytoplasm. This implies that, while these enzymes probably kept glutathione as substrate (as indicated by functional residue conservation), they might have adapted to different co-substrates, localized in different compartments.

3.5.2.2- Microsomal Glutathione Transferases

Microsomal glutathione transferases (MGT) of the MAPEG superfamily also play a critical role in oxidative stress-related detoxification. In *H. seosinensis*, one gene related to MGT3, ORF m.21576, was 2.7-fold up-regulated and had a high average expression at high salt (1,064 TPM, rank 420; Table 3.7), suggesting it had an important stress-related function. It probably originated from a recent gene duplication event, though the statistical support was only moderate (bootstrap
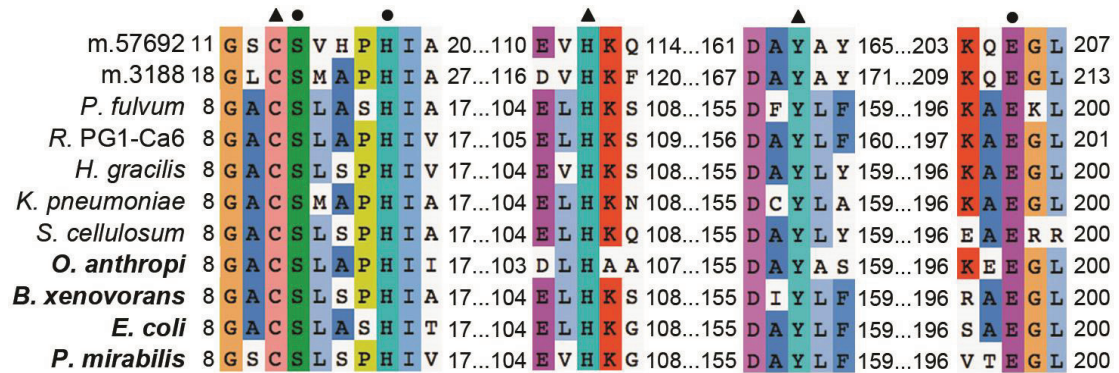
Fig. 3.27. Partial alignment of glutathione transferases of the Beta class showing conservation of residues interacting with the substrate (triangle) and residues essential for structural stability (circle) based on studies of enzymes (in bold) from *Ochrobactrum anthropi* (2NTO), *Burkholderia xenovorans* (2DSA), *Escherichia coli* (1N2A) and *Proteus mirabilis* (2PMT; Casalone *et al.* 1998; Allocati *et al.* 2000; Inoue et al. 2000; Federici *et al.* 2007; Federici *et al.* 2007). The alignment also includes *H. seosinensis* sequences (m.57692, m.3188) and their closest related sequences in the NR database from *Phaeospirillum fulvum* (WP_021133055.1), Rhodocyclaceae bacterium PG1-Ca6 (AJP48267.1), *Hylemonella gracilis* (WP_035608673.1), *Klebsiella pneumoniae* (WP_040214185.1) and *Sorangium cellulosum* (WP_044968822.1).

support 78%, Figure 3.28; the putative sister ORF is not differentially expressed). MGT3s probably have glutathione-dependent peroxidase activity toward lipid hydroperoxides, concordant with their membrane localization, and wide-specificity glutathione transferase activity toward lipophilic substrates (Jakobsson *et al.* 1997; Chen *et al.* 2011). Proteins belonging to the MAPEG superfamily typically possess four transmembrane domains (Jakobsson *et al.* 1999; Bresell *et al.* 2005). Concordantly, m.21576 encoded four regions with high hydrophobicity and a signal peptide, suggesting it might function in a membrane somewhere along the secretory path.

These sequences harbored the MAPEG superfamily-wide conserved motif (Rx_3Nx_2[E/D]) and additional glutathione-interacting residues (Figure 3.29; Martinez Mollina *et al.* 2008). Based on the conservation of these residues and on the broad membrane-related detoxification function of MAPEG enzymes, I predict that this up-regulated gene in *H. seosinensis* is involved in protecting cell membranes against toxic lipophilic compounds, probably derived from high levels of ROS.

3.5.2.3- Glutaredoxins

Cysteinyl residues are particularly vulnerable to ROS since they are among the most easily oxidized residues in proteins, and thus an early cellular response to oxidative stress is increased protein thiolation (Lii *et al.* 1994; Ravichandran *et al.* 1994). As a result, protein activity and function can be affected by inter- or intra-molecular disulfide formation, as well as creation of sulfenic, sulfonic, sulfinic acid groups and also by S-nitrosylation of these cysteinyl residues (Lillig *et al.* 2008). This can be prevented by the concerted action of glutathione and glutaredoxin, as they are the thiol transferases responsible for the reduction of protein disulfides and glutathione-protein mixed disulfides (Lillig *et al.* 2008). In *S. cerevisiae*, mutants of glutaredoxins Grx1 and Grx2 are more susceptible to oxidants like hydroperoxides, paraquat and iron chloride, and conversely, overexpression of these genes improves tolerance to oxidants (Luikenhuis *et al.* 1998; Collinson *et al.* 2002; Collinson and Grant 2003).
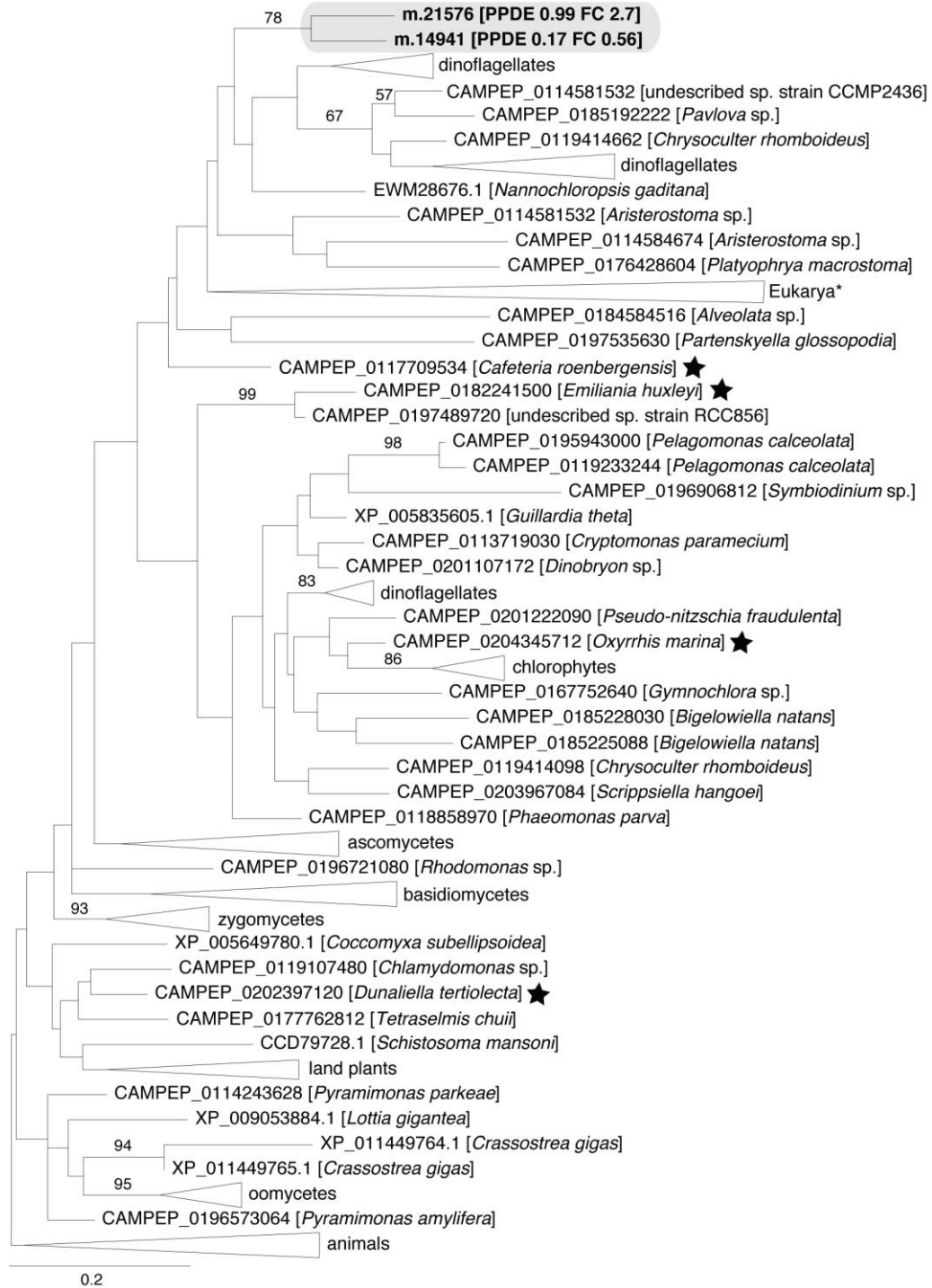
Fig. 3.28. Maximum-likelihood phylogenetic tree for gene duplication cluster 19 (grey box): microsomal glutathione transferase. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Stars indicate sequences included in Figure 3.29. The clade named Eukarya* includes sequences from chlorophytes, cryptophytes, amoebozoans, ciliates and stramenopiles. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.
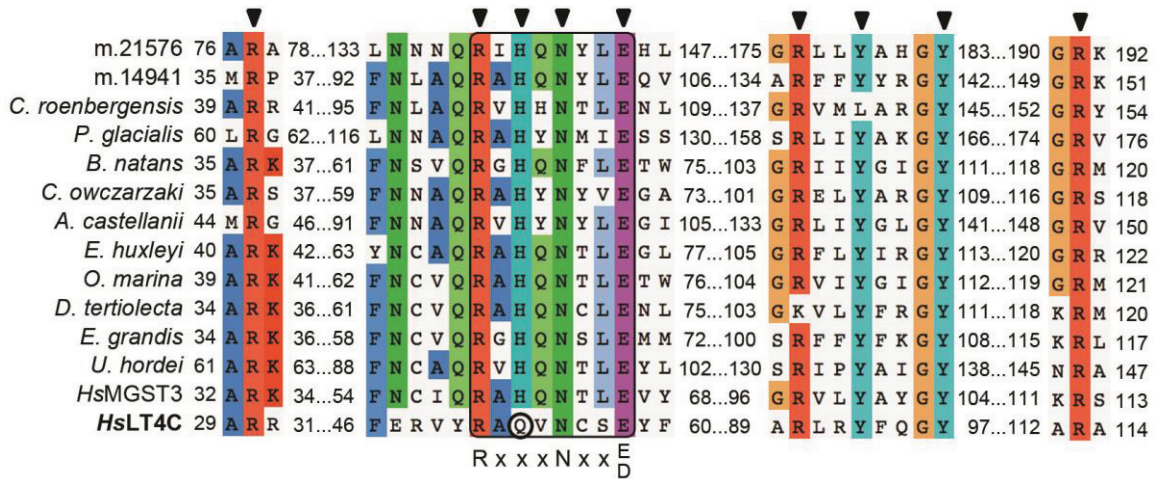
Fig. 3.29. Partial alignment of microsomal glutathione S-transferase (MGST) sequences showing the MAPEG superfamily-wide conserved motif ($Rx_3Nx_2[E/D]$, boxed) and residues that bind glutathione (triangles) in human leukotriene synthase (*Hs*LT4C in bold, 2PNO; Ago *et al.* 2007). Note that Q^{53} in *Hs*LT4C (circled) is conservatively substituted to His in MGST3 (Martinez Molina *et al.* 2008). Human MGST3 (*Hs*MGST3, AAB82609.1) is displayed as a reference in addition to *H. seosinensis* sequences (m.21576 and m.14941) and other related sequences from *Cafeteria roenbergensis* (CAMPEP_0117709534), *Polarella glacialis* (CAMPEP_0197963974), *Bigelowiella natans* (CAMPEP_0185225088), *Capsaspora owczarzaki* (XP_004364452.1), *Acanthamoeba castellanii* (XP_004333244.1), *Emiliania huxlei* (CAMPEP_0182241500), *Oxyrrhis marina* (CAMPEP_0204345712), *Dunalilella tertiolecta* (CAMPEP_0202397120), *Eucalyptus grandis* (XP_010062340.1) and *Ustilago hordei* (CCF50363.1).

In *H. seosinensis*, two genes related to glutaredoxins were up-regulated at high salt (m.72582 with 2.8-fold increase and m.39259 with 4.8-fold increase). Glutaredoxins are divided in two groups: dithiol glutaredoxins that harbor two active-site cysteines, as for *H. seosinensis* sequences (Figure 3.30), and monothiol glutaredoxins that have only one. Dithiol glutaredoxins reduce disulfide bridges and protein-glutathione conjuguates by forming a transient disulfide between the target protein and the N-terminal glutaredoxin cysteine that is subsequently reduced by the C-terminal cysteine, releasing the restored protein (Rouhier *et al.* 2008). Two more glutathione molecules are needed to regenerate glutaredoxin, and finally glutathione reductase regenerates the resulting glutathione disulfide (Lillig *et al.* 2008).

m.39259	12	C	K	H	C	15
m.72582	25	C	P	F	C	28
m.84602	62	C	P	F	C	65
m.85098	93	C	P	F	C	96
*Sc*GRX1	27	C	P	Y	C	30
*Sc*GRX2	61	C	P	Y	C	64
*Sc*GRX8	25	C	P	D	C	28
*Sc*GRX3	211	C	G	F	S	214
*Sc*GRX4	171	C	G	F	S	174
*Sc*GRX5	60	C	G	F	S	63
*Sc*GRX6	136	C	S	Y	S	139
*Sc*GRX7	108	C	P	Y	S	111

(dithiol: *Sc*GRX1, *Sc*GRX2, *Sc*GRX8; monothiol: *Sc*GRX3–*Sc*GRX7)

Fig. 3.30. Partial alignment of glutaredoxin (GRX) protein sequences showing the two active-site cysteines of dithiol glutaredoxins (*S. cerevisiae* GRX1, 2 and 8 as references and *H. seosinensis* sequences in bold) and the unique cysteine of monothiol glutaredoxins (as in *Sc*GRX3 – *Sc*GRX7).

Given that glutaredoxins belong to a highly versatile enzyme family with common functionality, the precise roles of m.72584 and m.39259 are not definable from sequence evidence alone. However, based on the over-expression of glutathione transferases, one possibility is that the glutaredoxins deglutathionylate proteins (Murphy 2012). In mammalian, yeast, plant and nematode cells, glutathione conjugates are excreted by multidrug-resistance proteins (Keppler 1999). Homologs of these ATP-dependent exporters were detected in *H. seosinensis* (ORFs m.20315 and m.77447; Table 3.7) but were not substantially differentially expressed. As secretion of glutathione conjugates results in wasting precious resources, regeneration by glutaredoxin might be an advantageous way to reclaim glutathione and the restored protein.

Alternatively, glutaredoxins can also directly reduce peroxides (Lee *et al.* 2002), can act as electron donor for a variety of enzymes like ribonucleoside-diphosphate reductase (Holmgren 1979) and may be involved in iron homeostasis and Fe/S cluster assembly (Lillig *et al.* 2008). In addition, besides the protective role of glutathione and glutaredoxin against irreversible inactivation of protein thiols, evidence is accumulating that glutathionylation can act as a regulatory mechanism by specifically increasing or reducing protein activity (Shelton *et al.* 2005; Gallogly and Mieyal 2007). Proteins probably regulated by reversible glutathionylation include actin (Wang *et al.* 2001), glyceraldehyde 3-phosphate dehydrogenase (Lind *et al.* 1998; Grant *et al.* 1999; Zaffagnini *et al.* 2007), protein tyrosine phosphatase 1B (Barrett *et al.* 1999) and NFκB subunit p50 (Pineda-Molina *et al.* 2001).

3.5.2.4- Peroxiredoxins

Glutathione is also the physiological reductant for most enzymes of the 1-Cys peroxideroxin (Prdx) 6 family that, contrary to members of the other Prdx families, can bind and reduce phospholipid hydroperoxides (Fisher 2011). In *H. seosinensis,* one ORF (m.14635) related to Prdx 6 had average transcript abundance rising from 47 TPM to 473 TPM with increased salinity (Table 3.7). Although Prdx 6 can also reduce H_2O_2, overexpression of this enzyme suggests that phospholipids were

threatened by oxidants in *H. seosinensis*. Interestingly, Prdx 6 was previously shown to be transcriptionally regulated during oxidative stress in human systems and *Plasmodium yoelii* (Kawazu *et al.* 2003; Kim *et al.* 2003; Chowdhury *et al.* 2009; O'Flaherty and de Souza 2011).

In Prdx 6, substrate oxidation is catalyzed by a single peroxidatic cysteine residue included in the conserved motif PVCTTE, also found in the *H. seosinensis* sequence (Figure 3.31; Nevalainen 2010). A putative catalytic triad His^{39}-Cys^{47}-Arg^{132}, determined in a crystal structure study of human Prdx 6 (Choi *et al.* 1998), was also conserved in *H. seosinensis.* Without a second resolving cysteine, as is the case for the other types of Prdx, Prdx 6 requires a thiol-containing electron donor like glutathione to be regenerated, and this is done, at least for some of them, by the concerted activity of GT (Manevich *et al.* 2004).
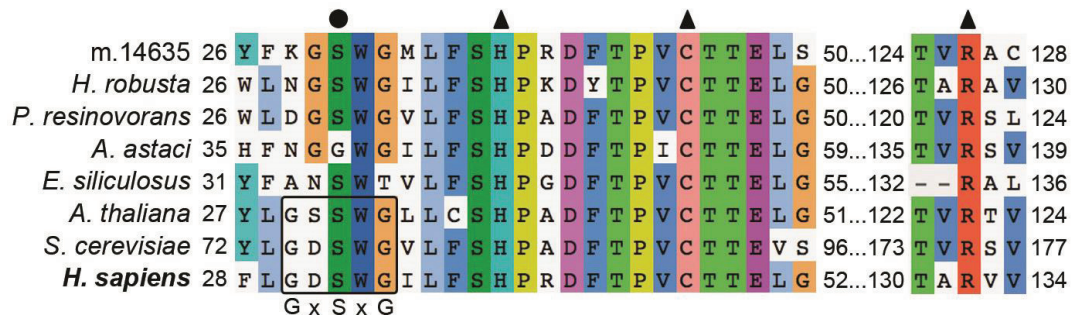
Fig. 3.31. Partial alignment of peroxiredoxin (Prdx) sequences showing the putative catalytic triad (triangles) based on study of human Prdx 6 crystal structure (in bold, NP_004896.1; Choi *et al.* 1998). Sequences from *Helobdella robusta* (XP_009018281.1) and *Pseudomonas resinovorans* (WP_028629099.1) that, like *H. seosinensis* sequence (m.14635), deviate from the lipase motif GxSxG (boxed) are shown. These sequences nonetheless include the putative serine catalytic residue (circle). For comparative purposes, sequences from *Aphanomyces astaci* (XP_009823721.1)*, Ectocarpus siliculosus* (CBN79130.1), *Arabidopsis thaliana* (NP_175247.1) and *S. cerevisiae* (AJQ03409.1) are included in the alignment.

Members of the Prdx 6 family display a second activity for phospholipid hydrolysis that is dependent on the serine residue included in the conserved lipase motif GxSxG (Kim *et al.* 1997). At these positions, m.14635 differed by having GSWG, but still included the putative catalytic serine residue. More than 15 variants of serine-containing putative lipase motifs exist (Nevalainen 2010) and comparative analysis of sequences related to m.14635 identified a few other Prdx 6-related enzymes containing this GSWG motif (from the leech *Helobdella robusta* and the bacterium *Pseudomonas resinovorans*). Whether the latter motif confers phospholipase activity remains to be answered experimentally.

Both functions are dependent on the pH and thus on the localization of the enzyme, where the phospholipase activity is stimulated by low pH occurring in lysosomes and the peroxidase activity is maximal at cytosolic pH 7 (Chen *et al.* 2000). The *H. seosinensis* enzyme is assumed here to be cytosolic based on the lack of predicted sequence signal. Notably, the catalytic sites for these functions seem to act independently since mutants for the lipase catalytic serine lose lipase activity but keep peroxidase activity towards H_2O_2 and short-chain organic hydroperoxides (Chen *et al.* 2000).

3.5.2.5- NADPH-Dependent Alcohol Dehydrogenase

NADPH provides the reducing power required to quench ROS, acting as the cofactor for enzymes such as glutathione reductase and thioredoxin reductase, for example (Murphy 2012; Mailloux *et al.* 2013). Notably, one sequence (ORF m.77193) related to cinnamyl-alcohol dehydrogenase (CD08297) had an expression level rising from an average of 2 TPM at 15% salt to 627 TPM at high salt (202-fold up-regulation; Table 3.7). Cinnamyl-alcohol dehydrogenase oxidizes a wide range of alcohols to aldehydes using NADP, and potentially contributes to maintaining NADP/NADPH balance (Larroy *et al.* 2002; Vidal *et al.* 2009). Indirect evidence of high ROS level at high salt suggested that high NADP$^+$ levels probably occurred in *H. seosinensis* and overexpression of cinnamyl-alcohol dehydrogenase could contribute to regeneration of NADPH.

Inspection of a multiple alignment including NADH- and NADPH-dependent alcohol dehydrogenases indicated conservation of Zn^{2+}-binding and NADPH-specific residues in the *H. seosinensis* sequence (Figure 3.32). Interactions between *S. cerevisiae* Adh6p and NADPH are mediated by residues Ser^{210}, R^{211} and K^{215} (Valencia *et al.* 2004). These are conserved in the *H. seosinensis* sequence. In NADH-dependent enzymes, the small, neutral amino acid at position 210 is substituted by a negatively charged residue that forms hydrogen bonds with the 2'- and 3'-hydroxyl groups of the coenzyme adenosine ribose (Fan *et al.* 1991; Bellamacina 1996). These observations suggest that this *H. seosinensis* enzyme is a NADPH-dependent zinc-binding alcohol dehydrogenase.

The cinnamyl-alcohol dehydrogenase sequence was closely related to certain bacterial sequences (Figure 3.33), however, it was encoded on a genomic contig surrounded by intron-containing genes and had high transcript abundance, showing that it was not a bacterial contamination. The phylogenetic analysis (Figure 3.33) grouped this sequence with a zinc-dependent alcohol dehydrogenase sequence from the halotolerant cyanobacterium *Halothece* sp. PCC 7418 (Garcia-Pichel *et al.* 1998) with maximum support, whereas other eukaryotic sequences, except the ones from *A. castellanii* and *G. sulphuraria*, formed a clade to the exclusion of all prokaryotic sequences (84% bootstrap support). As with the lateral acquisition of the peroxidase and the glutathione transferase, gain of this alcohol dehydrogenase gene from a prokaryote might plausibly have played a role in *H. seosinensis* adaptation to high salt. However, further experimental work is required to determine whether this gene truly contributes to NADP homeostasis.
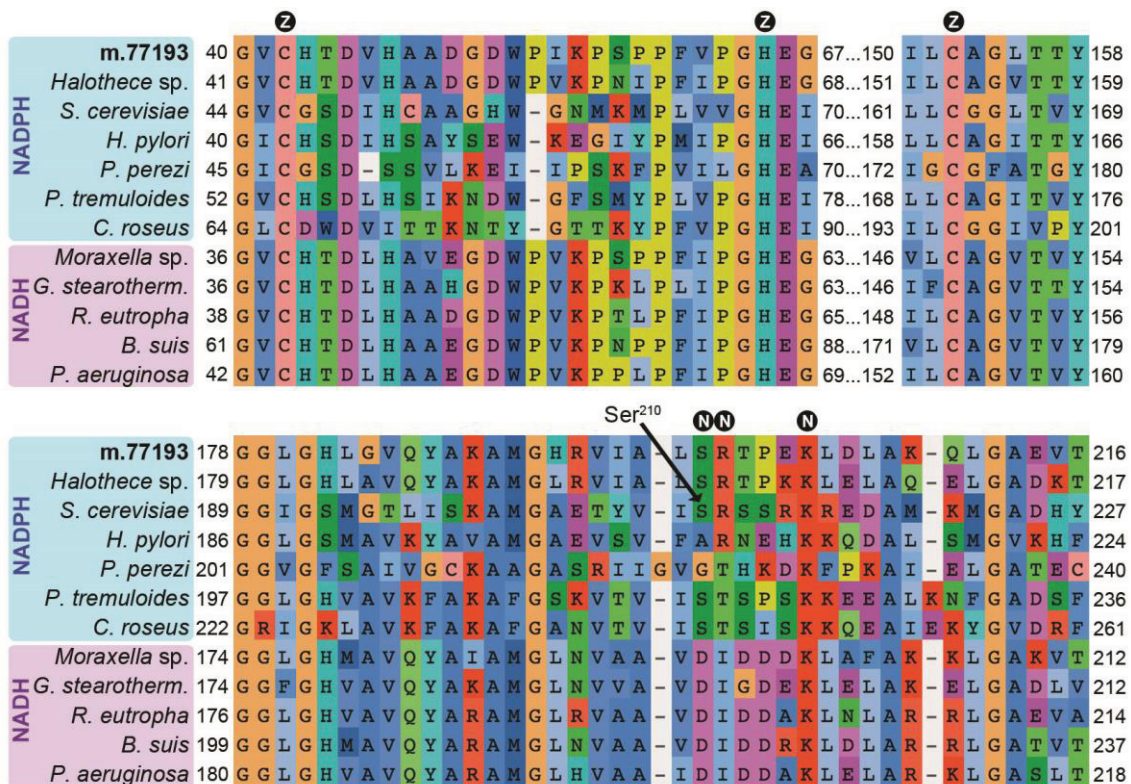
Fig. 3.32. Partial alignment of zinc-dependent alcohol dehydrogenase sequences showing conservation of residues binding the catalytic zinc ion (circled Z) and NAD(P)H (circled N). At position 210 (Ser[210], *S. cerevisiae* 1Q1N numbering), small neutral amino acids (Ala, Gly, Ser) confer specificity for NADPH while NADH-dependent enzymes harbor a negatively charged residue. In NADPH-dependent dehydrogenase, the following position contains a positively charged residue as in *S. cerevisiae* enzyme or a threonine as in the amphibian enzyme (*Pelophylax perezi* 1P0C; Rosell et al. 2003) that interacts with the NADPH terminal phosphate group. The alignment contains *H. seosinensis* sequence (in bold) and its closest sequence in the NR database (*Halothece* sp. WP_041596283.1), and sequences for characterized NADPH-dependent enzymes from *Helicobacter Pylori* (3TWO), *Populus tremuloides* (1YQD) and *Catharanthus roseus* (5H81), and for NADH-dependent enzymes from *Moraxella* sp. (4Z6K), *Geobacillus stearothermophilus* (3PII), *Ralstonia eutropha* (3S1L), *Brucella suis* (3MEQ) and *Pseudomonas aeruginosa* (1LLU).
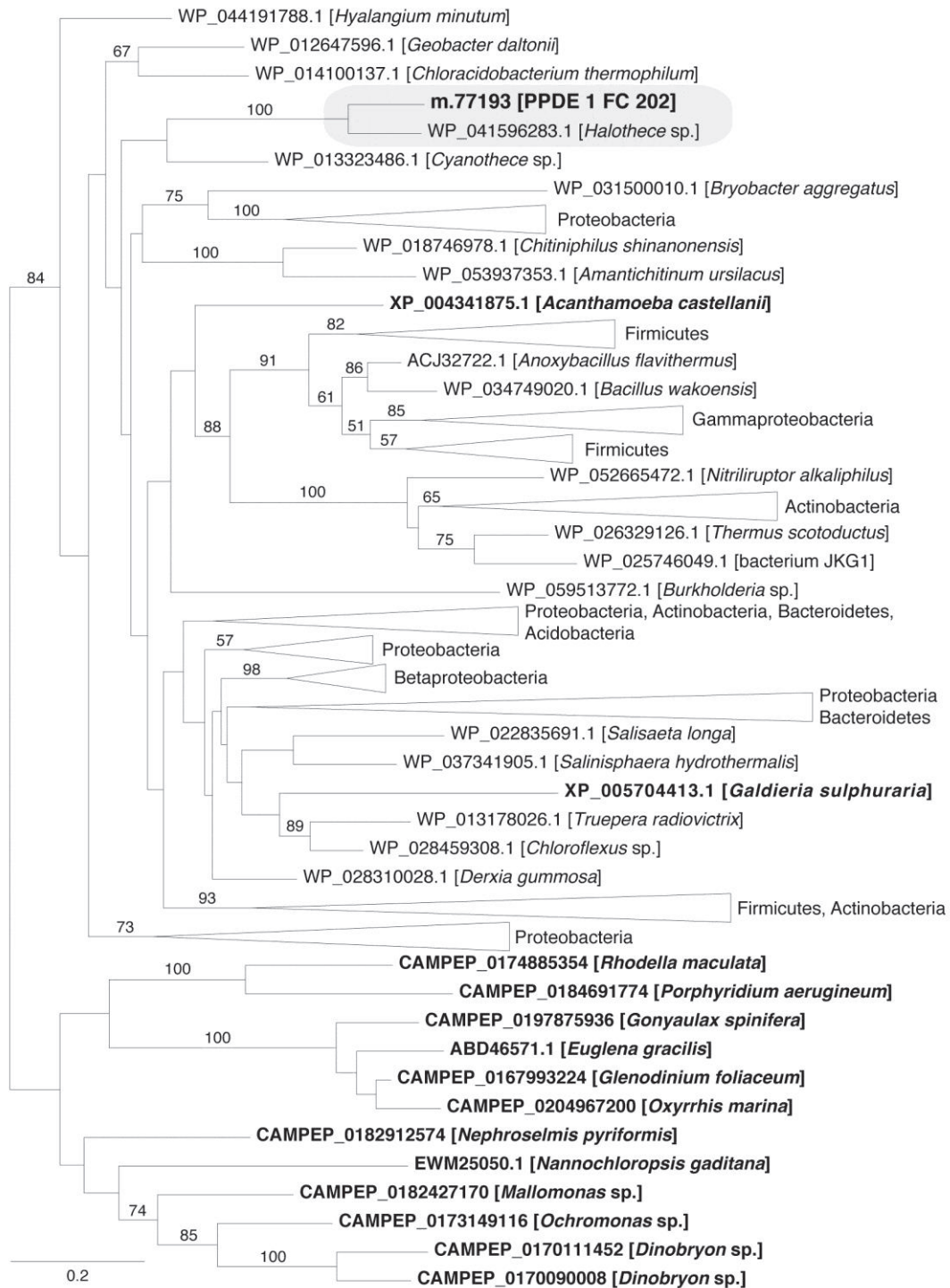
Fig. 3.33. Maximum-likelihood phylogenetic tree for zinc-dependent alcohol dehydrogenase showing robust clustering of sequences from *H. seosinensis* (m.77193; PPDE, Posterior Probability of being Differentially Expressed and FC, Fold Change 30% over 15% salt) and the cyanobacterium *Halothece* sp. (grey box) at the exclusion of all other bacterial and eukaryotic (in bold) sequences. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

166

3.5.3- Chaperones

In addition to up-regulating genes for directly neutralizing ROS or removing electrophilic groups from proteins, lipids and metabolites, *H. seosinensis* also overexpressed several chaperones that repair misfolded proteins, or at least minimize their detrimental impact.

All genes encoding the alpha-crystallin domain of the small heat shock proteins (sHSPs, PFAM00011) in *H. seosinensis* were differentially expressed, with three being 3.8- to 6.7-fold up-regulated and one (m.8880) being 10-fold repressed (Table 3.7). sHSPs binds to denatured proteins, thus preventing irreversible aggregation until these proteins are refolded by the ATP-dependent HSP70 system (Ehrnsperger *et al.* 1997; Lee and Vierling 2000) or potentially degraded (Lee *et al.* 1997). The functions of sHSPs are diverse partly due to their dynamic oligomerization involving varying subunit combinations (Sun and MacRae 2005).

Interestingly, two sHSP alpha-crystallin-domain-containing proteins with contrasting expression patterns (4.2-fold up-regulated m.65830 and 9.7-fold repressed m.8880) might have arisen through gene duplication, but this was not resolvable by phylogenetic analysis of these very short proteins (data not shown).

Concordantly, another chaperone involved in protein disaggregation was up-regulated at high salt (6.3-fold increase): the ATP-dependent chaperone ClpB (m.91451). While sHSPs prevent protein-protein aggregation by binding to misfolded proteins, ClpB dynamically mediates the disaggregation of stress-damaged proteins (Hodson *et al.* 2012). Subsequently, proper folding is achieved by the intervention of the Hsp70 chaperone system (see below). The ClpB chaperone is particularly important since, while the Hsp70 system can independently correct populations of small aggregates, resolubilisation of large aggregates requires ClpB (Goloubinoff *et al.* 1999; Diamant *et al.* 2000).

Aggregation can be the result of protein conformational changes induced by ROS oxidation (Mirzaei and Regnier 2008). However, since higher salinity correlates

with increased hydrophobic effect, it is tempting to speculate that this indirect evidence of protein aggregation at high salt (overexpression of aggregation-preventing chaperones) could partly result from higher intracellular salt content. It is still unknown whether salt concentration varies in *H.* seosinensis (direct measurements of cell ion content have not been performed), however I have good indirect indications that the intracellular salt content is probably higher that in marine protists (see chapter 2). It is reasonable to assume that salt intrusion occurs into the cytosol, at least incidentally during phagotrophy, and that this is more intense when extracellular salinity is higher, potentially leading to greater protein-protein aggregation.

As mentioned above, stress-denatured proteins re-gain their native conformation due to the action of the Hsp70 system. *Halocafeteria seosinensis* expressed eight genes encoding the Hsp70 domain (PFAM00012) with one being strongly up-regulated at high salt (m.81151; Table 3.7). Average transcript abundance for this gene rose 15-fold from 83 TPM at 15% salt to 1,657 TPM at 30% salt, where it corresponded to the 110th most expressed transcript. Based on the co-expression status of this gene, coding for a cytosolic Hsp70, and the ClpB chaperone in *H. seosinensis*, I hypothesize that these two enzymes collaborate to disaggregate cytoplasmic protein.

However, Hsp70 chaperones display a wide diversity of physiological functions, both during normal protein biogenesis and upon stress-induced unfolding, that are mediated by partner co-chaperones (Clerico *et al.* 2015). For example, a proteomic survey identified ~700 proteins interacting with DnaK (the major bacterial Hsp70) in *E. coli* (Calloni *et al.* 2012). Substrate specificity of Hsp70s is also driven by cofactors: the J proteins (Kampinga and Craig 2010). In *H. seosinensis*, 29 genes encoded a DnaJ domain (PFAM00226, present in members of the J-protein family), but none of them was up-regulated at high salt. Despite the complexity of the Hsp70 system and thus my limited ability to identify the specific cellular role of m.81151, the overexpression of this gene combined with the up-regulation of other chaperones suggested that *H. seosinensis* protein pool was threatened at high salt.

Lastly, *H. seosinensis* grown at high salt up-regulated by a factor of 2.2 another gene involved in protein repair, protein-L-isoaspartyl/D-aspartyl O-methyltransferase (PIMT; Table 3.7). This enzyme repairs L-isoaspartyl residues generated from the spontaneous deamidation of asparaginyl residues and isomerization of aspartyl residues that otherwise can induce loss of protein function (Brennan *et al.* 1994; Hayes and Setlow 1997; Smith *et al.* 2002). This type of lesion is encountered especially during environmental stresses or during long-term storage, conditions in which PIMT favors viability. For example, *E. coli* cells lacking this gene show reduced viability at stationary phase when exposed to environmental stresses, including high salt concentrations and exposure to ROS generated by paraquat (Visick *et al.* 1998). Similarly, exposure to salt and dehydration stresses increases PIMT expression and activity in wheat seedlings (Mudgett and Clarke 1994). It is noteworthy that the increased occurrence of isoaspartate observed during oxidative stress in damaged proteins (*e.g.* Ingrosso *et al.* 2002) is probably an indirect effect, since ROS are not chemically expected to directly react with asparagines and aspartates (O'Connor 2006).

3.5.4- Control of Reactive Semiquinones

Quinones are essential to aerobic life forms due to their role in the electron transport chain, and their reduced forms, hydroquinones, can also confer protection by directly reducing environmental oxidizers (Cotelle *et al.* 1991; Hino *et al.* 1998). However, one-electron reduction of quinones generates unstable semiquinones that, in the presence of molecular oxygen, can participate in the formation of highly reactive oxygen species (Wrobel *et al.* 2002). One mechanism for preventing the occurrence of this damaging situation is the divalent reduction of quinones to fully oxidized hydroquinones by NAD(P)H : quinone oxidoreductases (NQO; Brock *et al.* 1995; Sparla *et al.* 1999; Jaiswal 2000; Ross *et al.* 2000; Jensen *et al.* 2002; Wrobel *et al.* 2002; Adams and Jia 2005). In addition to preventing the formation of harmful semiquinones and through production of hydroquinones with antioxidant properties, NQO also has the ability to directly scavenge superoxide radicals (Siegel *et al.* 2004). Similar to other detoxifying enzymes, quinone reductase flavoenzymes

169

commonly have broad substrate specificity, suggesting they can protect cells against an array of potentially damaging quinones (Jensen *et al.* 2002; Wrobel *et al.* 2002), and are induced by stress exposure (Akileswaran *et al.* 1999; Pomposiello *et al.* 2001; Cohen *et al.* 2004; Ross 2004).

Halocafeteria seosinensis expresses three related genes encoding a NQO type IV domain (TIGR01755), and all were up-regulated at high salt (2.1- to 12-fold increase) suggesting that tight control of any source of ROS was required. Furthermore, one of them (m.35863) had very high transcript abundance at 30% salt (average of 1,581 TPM, rank 118). Phylogenetic analysis indicated that the genes probably arose through duplications after divergence from *Cafeteria roenbergensis* (Figure 3.34). These sequences showed a high degree of similarity (48~62% identical) with NQO type IV sequences from fungal and plant species, for which enzymatic activities were characterized (Brock *et al.* 1995; Jensen *et al.* 2002; Laskowski *et al.* 2002; Wrobel *et al.* 2002), and from *E. coli*, for which the protein crystal structure is available as well (Andrade *et al.* 2007). Inspection of an alignment of these sequences identified conserved functional residues, supporting the inferred annotation for these proteins in *H. seosinensis* (Figure 3.35).

3.5.5- Summary of section 3.5

In summary, *H. seosinensis* up-regulated several genes at high salt known to limit cellular damage induced by stress, especially ROS-derived stress. ROS would be expected to be neutralized enzymatically by superoxide dismutases, including a probable cytosolic Mn-dependent superoxide dismutase that is not commonly described in protists, and by a peroxidase that was likely acquired by LGT. The glutathione system would protect proteins and metabolites through the action of cytoplasmic and membrane glutathione transferases, peroxiredoxin 6 and glutaredoxins that would regenerate glutathionylated proteins. Among the latter enzymes, one highly expressed and up-regulated glutathione transferase of the bacterial Beta class was probably acquired by LGT, suggesting that acquisition of genes from foreign organisms contributed to increasing *H. seosinensis* resilience to
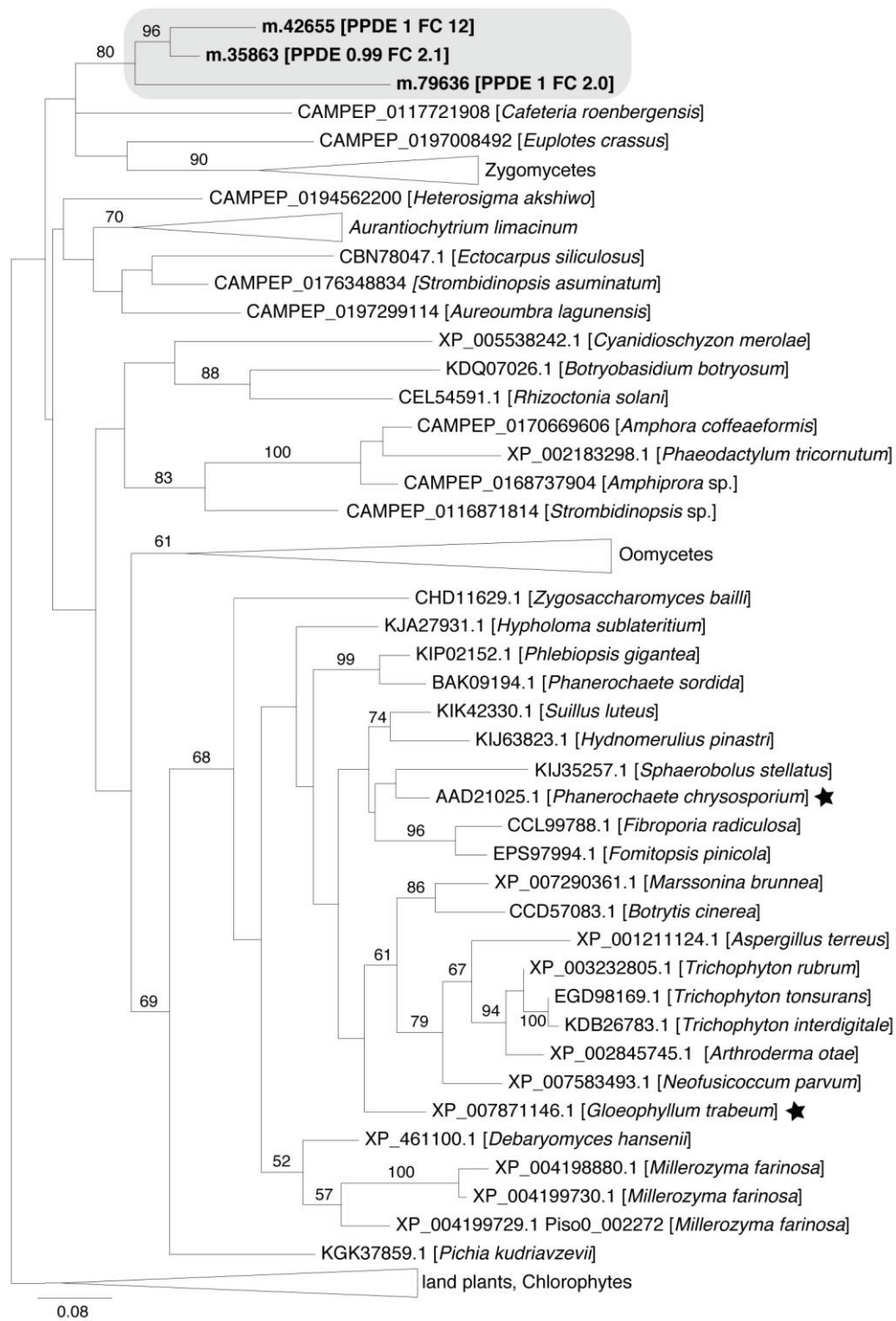
Fig. 3.34. Maximum-likelihood phylogenetic tree for gene duplication cluster 270 (grey box): NAD(P)H : quinone oxidoreductase. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Stars indicate sequences from characterized enzymes included in Figure 3.35. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.
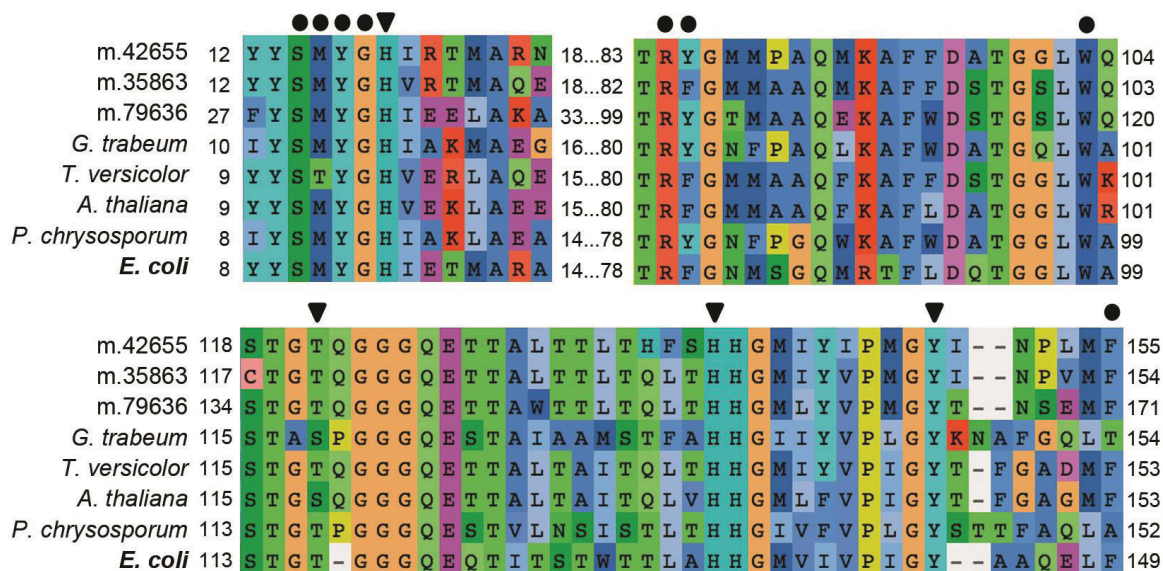
Fig. 3.35. Partial alignment of NAD(P)H:quinone oxidoreductase sequences showing structural residues that shape the active-site cavity and contribute to orient the substrate and cofactors (circles), and residues that interact through hydrogen bonds with the cofactor FMN (triangles) in *E. coli* WrbA (in bold, 3B6I) based on Andrade *et al.* (2007) and Wolfova *et al.* (2009). The alignment also includes *H. seosinensis* sequences (m.42655, m.35863 and m.79636), and sequences for which the enzymatic activity was characterized in *Gloeophyllum trabeum* (XP_007871146.1; Jensen *et al.* 2002), *Triphysaria versicolor* (AAG53945.1, Wrobel *et al.* 2002), *A. thaliana* (NP_200261.1, Laskowshi *et al.* 2002) and *Phanerochaete chrysosporium* (AAD21025.1, Brock *et al.* 1995).

high salt stress. Several chaperones, including small alpha-crystallin-domain-containing heat shock proteins, the ClpB chaperone, Hsp70 and protein-L-isoaspartyl/D-aspartyl O-methyltransferase would be expected to prevent protein aggregation, mis-folding and spontaneous deamidation of asparaginyl residues and isomerization of aspartyl residues. Finally, up-regulation of NAD(P)H : quinone oxidoreductases could have helped reduce the detriments of oxidative stress by preventing the formation of reactive semiquinones and by directly or indirectly (through synthesis of antioxidant hydroquinones) scavenging ROS. Overall, expression of these genes, many of which were abundantly transcribed at high salt, most likely contributes to increase *H. seosinensis* tolerance of hypersaline conditions.

3.6- Salt Adaptation and Membrane Lipid Remodelling

Extracellular salinity affects biological membranes by favouring transition from the bilayer (lamellar) phase to the hexagonal-II (non-bilayer) phase of certain lipids and by inducing membrane fluidity variations (Harlos and Eibl 1981; Seddon *et al.* 1983; Russell 1989; Sutton *et al.* 1990; Beney and Gervais 2001; Simonin *et al.* 2008). Consequently, as a response to variations in salinity, organisms adapt the composition of membrane lipids by adjusting the length, branching and saturation of fatty acids and the relative proportion of phospholipid head groups (Quinn 1981).

Generally, in halotolerant and halophilic bacteria, anionic lipids like phosphatidylglycerol increase at the expense of zwitterionic lipids like phosphatidylethanolamine (Vreeland *et al.* 1984; Russell 1989, 1993). Notably, while phosphatidylethanolamine adopts a hexagonal-II phase, phosphatidylglycerol maintains a lamellar phase under varying conditions and is able to suppress the transition of phosphatidylethanolamine to the hexagonal-II phase, suggesting the adjusted ratio at higher salinities contributes to avoid the formation of microdomains of hexagonal-II phase lipids, and thus prevents alterations of the membrane permeability (Russell 1989).

Eukaryotes display membrane lipid compositions that differ from these of bacteria, with the exception of the mitochondrial and the thylakoid membranes, for which the lipid composition is reminiscent of their bacterial origins (Vothknecht and Westhoff 2001; van Meer *et al.* 2008). For example, eukaryotic membranes contain sterols that are thought to decrease membrane fluidity by reducing lipid acyl chain mobility (Demel and De Kruyff 1976). Considerable information is available on the salt response in certain mesophilic organisms, primarily plants and yeasts, however these studies are not directly applicable to halophilic systems since the salt concentrations tested rarely exceed the millimolar range. The few studies on halotolerant and halophilic yeasts have indicated that species respond differently in terms of sterol and phospholipid content, and fatty acid saturation level, implying that different mechanisms are being used to regulate membrane fluidity (Tunblad-Johansson *et al.* 1987; Hosono 1992; Khaware *et al.* 1995; Andreishcheva *et al.* 1999; Turk *et al.* 2004; Smolyanyuk *et al.* 2013). One factor that seems to characterize halophilic yeasts is a low sterol:phospholipid ratio, and thus a capacity to maintain membrane fluidity along a broad range of salt concentrations (Turk *et al.* 2011).

Eukaryotes also differ from bacteria by having intracellular compartments. Since the plasma membrane is exposed to the extracellular milieu but endomembranes are not, salt exposure influences their respective lipid composition differently and can complicate lipid characterization studies. For example, when the halotolerant alga *Dunaliella salina* is exposed to increased salinity, an increase in phosphatidylethanolamine and polyunsaturated fatty acids and a decrease in diacylglycero-trimethylserines were measured in the total lipids (Al-Hasan *et al.* 1987). In contrast, investigation of an enriched plasma membrane fraction showed the opposite, *i.e.* decrease of phosphatidylethanolamine and polyunsaturated fatty acids and increase of diacylglycero-trimethylserines (Peeler *et al.* 1989). At higher salinity, *D. salina* synthesizes large amounts of glycerol, which potentially impacts the properties of endomembranes, as suggested by Azachi *et al.* (2002), who reported a higher ratio of unsaturated to saturated fatty acids in microsomes compared to the plasma membrane or thylakoid membranes.

Even though membrane adaptation to salinity is a complex phenomenon and transcriptomic analyses can only provide a very partial picture, investigation of the *H. seosinensis* transcriptional program strongly suggested that lipid metabolism was affected.

3.6.1- Sterol Synthesis, Regulation and Transport

Vertebrates, fungi and plants synthesize differing pools of sterols: cholesterol being the main sterol in vertebrates, ergosterol in fungi and phytosterols in plants. Most biosynthetic steps are shared between the three sterol types (although the order varies) and are catalyzed by homologous enzymes acting on different substrates with high structural similarity (Desmond and Gribaldo 2009). Because of this, it is challenging to infer the sterol profile of *H. seosinensis* (which is distantly related to animals, fungi and plants) based on sequence information alone. However, I identified enzymes indicating that *H. seosinensis* synthesized sterols *de novo* (Figure 3.36 and Table 3.8), using cycloartenol as precursor, as is the case for several protists investigated previously (Rees *et al.* 1969; Anding *et al.* 1971; Raederstorff and Rohmer 1987a, b; Nes *et al.* 1990). Importantly, when genes involved in sterol synthesis were identified as transcriptionally regulated in *H. seosinensis*, they were repressed at high salt. In addition, sterol insertion in the membrane was probably reduced due to esterification cycles.

Three lines of evidence suggested that sterol synthesis in *H. seosinensis* employed the phytosterol branch of the pathway. Firstly, although the sequence most closely related to oxidosqualene cyclase in *H. seosinensis* (m.78304) did not phylogenetically cluster specifically with either the vertebrate/fungi version (lanosterol synthase) or the plant version (cycloartenol synthase; Figure 3.37), it possessed all functional residues shown to confer specificity for cycloartenol

On next page: Fig. 3.36. ORFs in *H. seosinensis* coding for enzymes acting in sterol biosynthesis*. For differentially expressed ORFs (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Biosynthesis of plant-like sterols in *H. seosinensis* is supported especially by sequence analysis of enzymes indicated by a star.

squalene

squalene monooxygenase
m.88587 [PPDE 1 FC 0.2]

squalene-2,3-oxide

lanosterol synthase

cycloartenol synthase
m.78304 ★

lanosterol

cycloartenol

sterol 24-C-
methyltransferase
m.73826 [PPDE 1 FC 0.4]

sterol 14-demethylase

4,4-Dimethyl-5alpha-cholesta
-8,14,24-trien-3beta-ol

24-Methylene-
cycloartanol

delta14-sterol reductase

4-methylsterol oxidase
m.22592

14-Demethyllanosterol

cycloeucalenol

cycloeucanelol
cycloisomerase ★
m.88947 [PPDE 1 FC 0.4]

methylsterol monooxygenase

4alpha-Methylzymosterol-4-
carboxylate

obtusifoliol

sterol 14-demethylase
m.88175 ★

sterol-4alpha-carboxylate 3-dehydrogenase

4alpha-Methyl-5alpha
-ergosta-8,14,24(28)-
trien-3beta-ol

3-Keto-4-methylzymosterol

delta14-sterol reductase
m.70478

3-keto steroid reductase

4alpha-
Methylfecosterol

4alpha-Methylzymosterol

ergosterol biosyntheis

sterol delta-isomerase
m.66451

zymosterol

24-Methyl-
enelophenol

delta(24)-sterol reductase

4-alpha-methyl-delta7-sterol
-4alpha-methyl oxidase
?

5alpha-Cholest-8-
en-3beta-ol

episterol

cholesterol delta-isomerase

delta7-sterol 5-desaturase
m.31330

lathosterol

5-Dehydroepisterol

delta7-sterol 5-desaturase

7-dehydrosterol reductase
m.91444

7-Dehydrocholesterol

24-Methyl-
enecholesterol

7-dehydrocholesterol reductase

delta(24)-sterol reductase
m.68329

cholesterol

campesterol

Table 3.8. Expression of genes involved in sterol metabolism in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC	
m.88587	136.80	34.08	1.00	0.19	2.36E-17	-2.35	3.53E-05	-2.46	Squalene monooxygenase
m.78304	41.97	67.88	0.98	1.24	7.91E-02	0.32	5.12E-02	0.34	Cycloartenol synthase
m.73826	264.80	137.63	1.00	0.40	2.00E-13	-1.31	1.05E-04	-1.30	Sterol 24-C-methyltransferase
m.22592	15.38	20.86	0.06	1.04	8.69E-01	0.07	7.89E-01	0.06	4-methylsterol oxidase
m.88947	29.65	14.27	1.00	0.37	1.10E-13	-1.43	8.75E-05	-1.44	Cycloeucanelol cycloisomerase
m.88175	171.10	174.27	0.99	0.79	3.21E-02	-0.33	5.32E-02	-0.31	Sterol 14-demethylase
m.70478	67.27	48.70	1.00	0.56	2.27E-09	-0.83	5.29E-04	-0.81	Delta14-sterol reductase
m.66451	50.07	80.53	0.92	1.24	1.48E-01	0.32	5.88E-02	0.36	Sterol delta-isomerase
m.31330	126.83	190.54	0.14	1.14	4.38E-01	0.20	2.51E-01	0.22	Delta7-sterol 5-desaturase
m.91444	213.87	170.12	1.00	0.61	3.42E-05	-0.70	3.02E-03	-0.68	7-dehydrosterol reductase
m.68329	79.89	140.67	0.06	1.43	1.32E-01	0.51	4.37E-02	0.62	Delta(24)-sterol reductase
m.26733	19.07	0.10	1.00	0.00	2.25E-83	-7.53	5.92E-06	-7.78	NPC1-related proteins
m.35144	13.30	1.87	1.00	0.11	4.73E-29	-3.12	9.78E-06	-3.21	NPC1-related proteins
m.88487	33.35	8.23	1.00	0.19	6.14E-22	-2.32	2.23E-05	-2.32	NPC1-related proteins
m.41605	32.75	10.48	1.00	0.24	6.22E-18	-2.02	4.07E-05	-2.04	NPC1-related proteins
m.26739	2.65	7.31	0.08	2.49	6.65E-01	0.54	1.48E-01	3.46	NPC1-related proteins
m.51528	33.44	23.36	1.00	0.54	8.69E-02	0.59	2.17E-03	-0.85	NPC1 N-terminal domain
m.17881	5.03	38.83	1.00	5.89	2.27E-31	2.53	3.60E-06	2.61	Sterol O-acyltransferase
m.78053	20.30	42.31	1.00	1.58	1.59E-02	0.66	1.18E-02	0.72	Sterol O-acyltransferase
m.20854	2.64	2.40	0.17	0.77	5.70E-01	-0.35	5.15E-01	-0.29	Diacylglycerol O-acyltransferase
m.33861	24.96	28.87	0.74	0.89	3.68E-01	-0.16	3.28E-01	-0.14	Sterol esterase

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC	
m.85697	0.27	0.27	0.16	0.92	9.53E-01	-0.08	6.75E-01	-0.47	Sterol esterase
m.85682	0.26	1.50	0.93	4.47	2.23E-02	1.87	2.35E-02	1.98	Sterol esterase
m.85677	7.31	17.33	1.00	1.87	1.44E-03	0.89	5.41E-03	0.90	Sterol esterase

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed; Post Fold Change, posterior fold change (30% over 15% salt); log_2FC, log_2 fold change (30% over 15% salt).
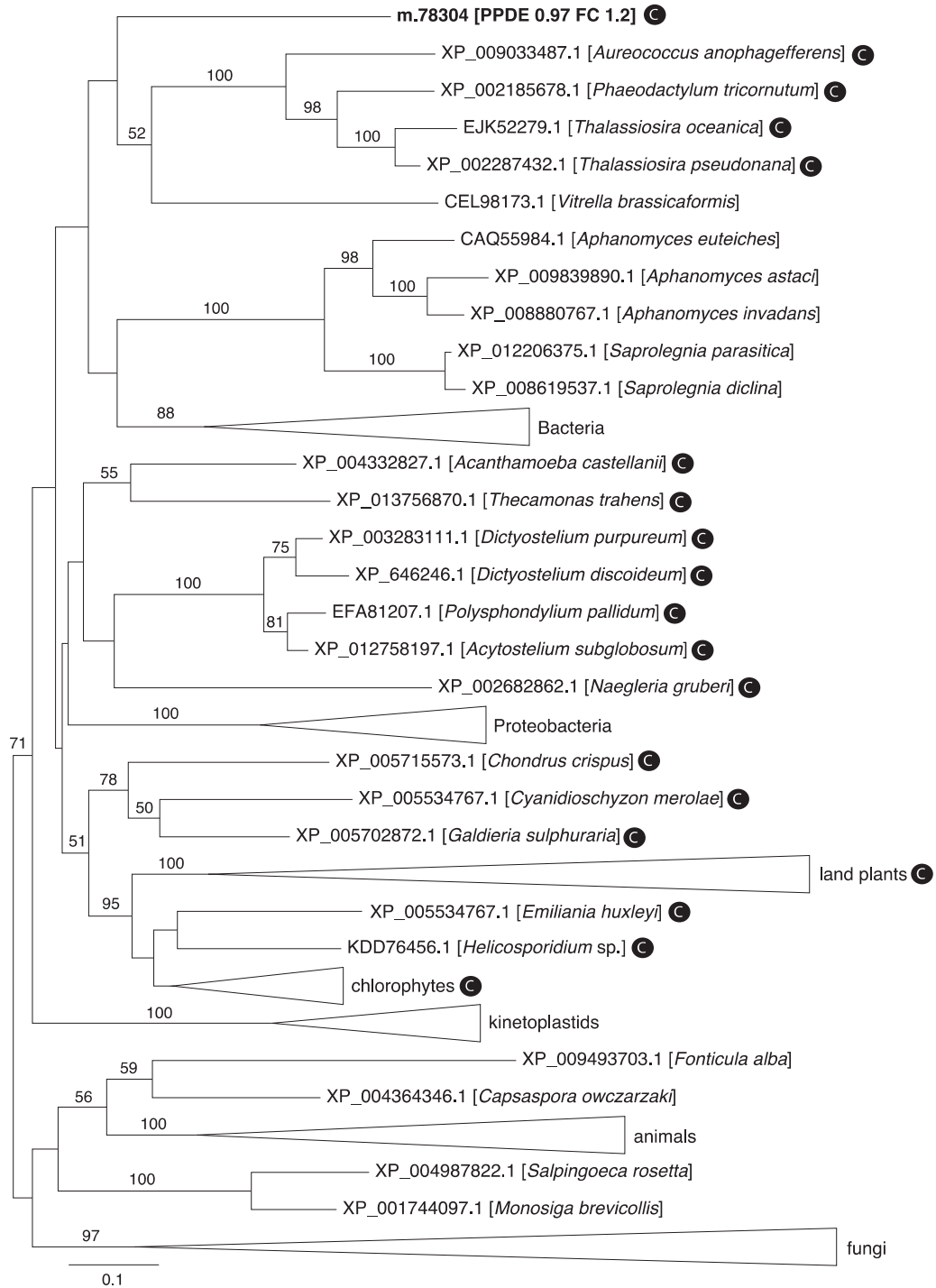
Fig. 3.37. Maximum-likelihood phylogenetic tree for oxidosqualene cyclase sequences showing the independent grouping of *H. seosinensis* sequence (m.78304 in bold, PPDE: Posterior Probability of being Differentially Expressed, FC: Fold Change 30% over 15% salt) relative to the enzymes from plants, animals and fungi. Circled "C" indicates protist sequences predicted to encode cycloartenol synthase based on residue conservation (taxa boxed in Figure 3.38). Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

production (Figure 3.38; Hart *et al.* 1999; Herrera *et al.* 2000; Meyer *et al.* 2002; Segura *et al.* 2002; Suzuki *et al.* 2006). Secondly, *H. seosinensis* expressed cycloeucalenol isomerase, which produces obtusifoliol along the plant-like branch of the pathway (Desmond and Gribaldo 2009). Thirdly, the sequence related to sterol demethylase in *H. seosinensis* contained a conserved residue specifying preference for the plant-specific substrate obtusifoliol (Figure 3.39; Gotoh 1992; Lepesheva *et al.* 2006).

Although the end products of sterol synthesis in *H. seosinensis* remain to be determined experimentally, transcriptional repression of several genes of the pathway indicated that probably less sterol accumulated in the cell at high salt. Cycloeucalenol isomerase (mentioned above) showed 2.7-fold repression, while other repressed enzymes were related to squalene monooxygenase (5.3-fold repression) and sterol 24-C-methyltransferase (2.5-fold repression). Interestingly, a decrease in membrane sterol content at increasing salinity was measured directly in the halotolerant yeasts *Yarrowia lipolitica* and *Debaryomyces hansenii* (Tunblad-Johansson *et al.* 1987; Andreishcheva *et al.* 1999)*.*

Lower sterol synthesis at high salt implies reduced sterol transport. Proteins involved in transport of sterol and sterol-modified lipids include the Patched (Ptc) domain-containing proteins (Kuwabara and Labouesse 2002). Interestingly, in *H. seosinensis*, among the five expressed proteins containing a Ptc domain (PFAM02460), four were repressed at high salt (4.2- to 220-fold repression; NPC1-related proteins in Table 3.8). This protein family includes the Niemann-Pick type C1 (NPC1) protein, to which all *H. seosinensis* proteins were evolutionarily affiliated (TIGR00917). As previously reported for NPC1 proteins of *N. gruberi* and a few stramenopiles and ciliates (Adebali *et al.* 2016), *H. seosinensis* proteins lacked the N-terminal domain commonly found in metazoa, fungi, amoebozoa and plant NPC1, but expressed this domain as a separate protein (ORF m.51528). The NPC1 N-terminal domain is potentially involved in transfer of cholesterol from NPC2 to NPC1, thus suggesting that this interaction is probably non-existent or occurs differently outside Amorphea and plants (Li *et al.* 2016).

Fig. 3.38. Partial alignment of oxidosqualene cyclase sequences showing the catalytic aspartate residue (circle) and residues that confer specificity for cycloartenol (triangle) in *A. thaliana* cycloartenol synthase (in bold, P38605.2; Hart *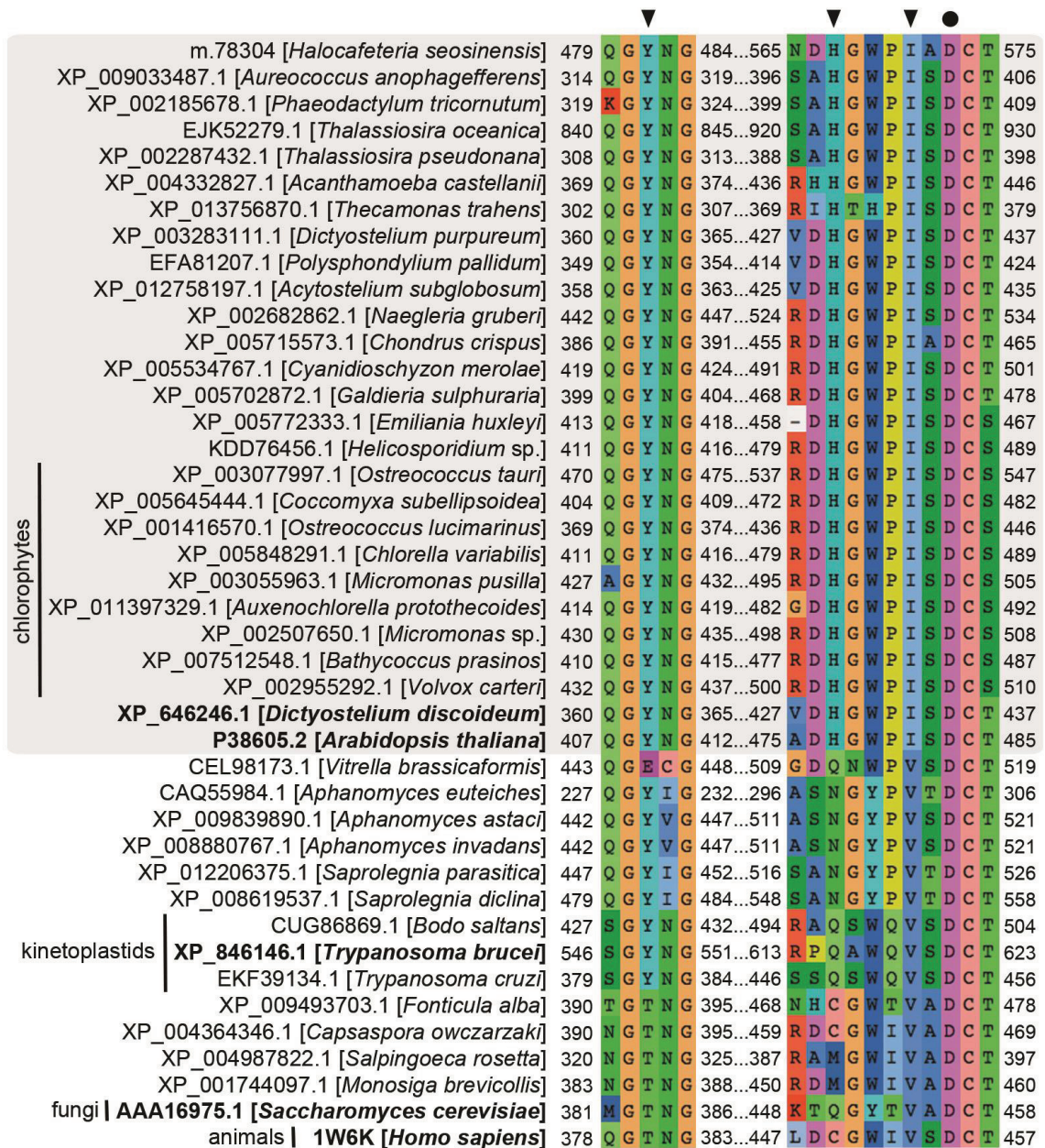et al.* 1999; Herrera *et al.* 2000; Segura *et al.* 2002). Protist sequences perfectly matching these residues are boxed. Sequences included in collapsed clades in Figure 3.37 are indicated (chlorophytes, kinetoplastids, fungi and animals). Enzymes experimentally characterized as cycloartenol synthase in *D. discoideum* (Godzina *et al.* 2000), and as lanosterol synthase in *T. brucei* (Buckner *et al.* 2000), *S. cerevisiae* (Corey *et al.* 1994) and *H. sapiens* (Thoma *et al.* 2004) are indicated in bold.

Sequence	start	motif 1	range	motif 2	end
m.78304 [*Halocafeteria seosinensis*]	479	Q G Y N G	484...565	N D H G W P I A D C T	575
XP_009033487.1 [*Aureococcus anophagefferens*]	314	Q G Y N G	319...396	S A H G W P I S D C T	406
XP_002185678.1 [*Phaeodactylum tricornutum*]	319	K G Y N G	324...399	S A H G W P I S D C T	409
EJK52279.1 [*Thalassiosira oceanica*]	840	Q G Y N G	845...920	S A H G W P I S D C T	930
XP_002287432.1 [*Thalassiosira pseudonana*]	308	Q G Y N G	313...388	S A H G W P I S D C T	398
XP_004332827.1 [*Acanthamoeba castellanii*]	369	Q G Y N G	374...436	R H H G W P I S D C T	446
XP_013756870.1 [*Thecamonas trahens*]	302	Q G Y N G	307...369	R I H T H P I S D C T	379
XP_003283111.1 [*Dictyostelium purpureum*]	360	Q G Y N G	365...427	V D H G W P I S D C T	437
EFA81207.1 [*Polysphondylium pallidum*]	349	Q G Y N G	354...414	V D H G W P I S D C T	424
XP_012758197.1 [*Acytostelium subglobosum*]	358	Q G Y N G	363...425	V D H G W P I S D C T	435
XP_002682862.1 [*Naegleria gruberi*]	442	Q G Y N G	447...524	R D H G W P I S D C T	534
XP_005715573.1 [*Chondrus crispus*]	386	Q G Y N G	391...455	R D H G W P I A D C T	465
XP_005534767.1 [*Cyanidioschyzon merolae*]	419	Q G Y N G	424...491	R D H G W P I S D C T	501
XP_005702872.1 [*Galdieria sulphuraria*]	399	Q G Y N G	404...468	R D H G W P I S D C T	478
XP_005772333.1 [*Emiliania huxleyi*]	413	Q G Y N G	418...458	- D H G W P I S D C S	467
KDD76456.1 [*Helicosporidium* sp.]	411	Q G Y N G	416...479	R D H G W P I S D C S	489
XP_003077997.1 [*Ostreococcus tauri*]	470	Q G Y N G	475...537	R D H G W P I S D C S	547
XP_005645444.1 [*Coccomyxa subellipsoidea*]	404	Q G Y N G	409...472	R D H G W P I S D C S	482
XP_001416570.1 [*Ostreococcus lucimarinus*]	369	Q G Y N G	374...436	R D H G W P I S D C S	446
XP_005848291.1 [*Chlorella variabilis*]	411	Q G Y N G	416...479	R D H G W P I S D C S	489
XP_003055963.1 [*Micromonas pusilla*]	427	A G Y N G	432...495	R D H G W P I S D C S	505
XP_011397329.1 [*Auxenochlorella protothecoides*]	414	Q G Y N G	419...482	G D H G W P I S D C S	492
XP_002507650.1 [*Micromonas* sp.]	430	Q G Y N G	435...498	R D H G W P I S D C S	508
XP_007512548.1 [*Bathycoccus prasinos*]	410	Q G Y N G	415...477	R D H G W P I S D C S	487
XP_002955292.1 [*Volvox carteri*]	432	Q G Y N G	437...500	R D H G W P I S D C S	510
XP_646246.1 [*Dictyostelium discoideum*]	360	Q G Y N G	365...427	V D H G W P I S D C T	437
P38605.2 [*Arabidopsis thaliana*]	407	Q G Y N G	412...475	A D H G W P I S D C T	485
CEL98173.1 [*Vitrella brassicaformis*]	443	Q G E C G	448...509	G D Q N W P V S D C T	519
CAQ55984.1 [*Aphanomyces euteiches*]	227	Q G Y I G	232...296	A S N G Y P V T D C T	306
XP_009839890.1 [*Aphanomyces astaci*]	442	Q G Y V G	447...511	A S N G Y P V S D C T	521
XP_008880767.1 [*Aphanomyces invadans*]	442	Q G Y V G	447...511	A S N G Y P V S D C T	521
XP_012206375.1 [*Saprolegnia parasitica*]	447	Q G Y I G	452...516	S A N G Y P V T D C T	526
XP_008619537.1 [*Saprolegnia diclina*]	479	Q G Y I G	484...548	S A N G Y P V T D C T	558
CUG86869.1 [*Bodo saltans*]	427	S G Y N G	432...494	R A Q S W Q V S D C T	504
XP_846146.1 [*Trypanosoma brucei*]	546	S G Y N G	551...613	R P Q A W Q V S D C T	623
EKF39134.1 [*Trypanosoma cruzi*]	379	S G Y N G	384...446	S S Q S W Q V S D C T	456
XP_009493703.1 [*Fonticula alba*]	390	T G T N G	395...468	N H C G W T V A D C T	478
XP_004364346.1 [*Capsaspora owczarzaki*]	390	N G T N G	395...459	R D C G W I V A D C T	469
XP_004987822.1 [*Salpingoeca rosetta*]	320	N G T N G	325...387	R A M G W I V A D C T	397
XP_001744097.1 [*Monosiga brevicollis*]	383	N G T N G	388...450	R D M G W I V A D C T	460
AAA16975.1 [*Saccharomyces cerevisiae*]	381	M G T N G	386...448	K T Q G Y T V A D C T	458
1W6K [*Homo sapiens*]	378	Q G T N G	383...447	L D C G W I V S D C T	457

Clade labels: chlorophytes (XP_003077997.1 – XP_002955292.1); kinetoplastids (CUG86869.1 – EKF39134.1); fungi (AAA16975.1); animals (1W6K).

Fig. 3.39. Partial alignment of sterol 14-alpha-demethylase sequences showing residues (boxed) specific for obtusifoliol (F) or lanosterol (I/L) based on a study of *Trypanosoma cruzi* enzyme (in bold, XP_821219.1) where a single mutation from Ile to Phe induced a shift in specificity from lanosterol to obtusifoliol (Lepesheva *et al.* 2006). Obtusifoliol specificity is also observed in *Trypanosoma brucei* (XP_828695.1), and plants like *Arabidopsis thaliana* (NP_172633.1; Kushiro et al. 2001), *Sorghum bicolor* (P93846.1; Bak et al. 1997) and *Zea mays* (NP_001168886.1; Taton and Rahier 1991), while activity specific for lanosterol is observed in fungi like *Phanerochaete chrysosporium* (ACI23621.1; Warrilow et al. 2008), *Candida albicans* (AIX03612.1; Song et al. 2004) and *Saccharomyces cerevisiae* (GAA23717.1; Aoyama et al. 1984), and animals like *Danio rerio* (NP_001001730.2; Morrison et al. 2014), *Mus musculus* (AAF73986.1; Keber et al. 2011) and *Homo sapiens* (3LD6; Strömstedt et al. 1996).

NPC1 proteins also encode a five transmembrane sterol-sensing domain (SSD, PFAM12349) that potentially monitors free sterol level in the membrane, consequently affecting sterol transport (Li *et al.* 2016). Each *H. seosinensis* NPC1-related sequence encoded at least one SSD, and in three cases, it contained two SSDs, which is uncommon. These domains are also found in a variety of animal proteins involved in cholesterol and lipid homeostasis, like the Sterol regulatory element-binding protein Cleavage-Activation Protein and 7-dehydrocholesterol reductase, in which it binds sterol embedded in the lipid bilayer (Ohgami *et al.* 2004; Feldman *et al.* 2015).

The mechanism of action of NPC1 is still ambiguous, but in mammals it is involved in intracellular cholesterol trafficking, especially in neuron cells, and in sphingolipid metabolism and transport (Sugimoto *et al.* 2001; Zervas *et al.* 2001; Zhang *et al.* 2001; Yu *et al.* 2014). Similar to mammalian cells, *npc1* mutants of the filamentous fungus *Fusarium graminearum* have impaired sterol transport (Breakspear *et al.* 2011). However, defective NPC1 not only results in impaired transport of cholesterol but also of glycosphingolipids and spingosine (Lloyd-Evans and Platt 2010). Furthermore, no sterol homeostasis disruption is observed, but sphingolipid metabolism is perturbed in *A. thaliana* when the NPC1-like protein is knocked-out or in *S. cerevisiae* when the SSD of the NPC1 homolog is mutated (Malathi *et al.* 2004; Feldman *et al.* 2015). The latter observations suggest that NPC1 might be a multi-substrate transporter, but indicate that these repressed genes in *H. seosinensis* might be involved at least in lipid transport. Therefore, although functions for these genes in *H. seosinensis* stay undetermined, I assume that sterol/sphingolipid transport for structural purposes was affected at high salt, in line with repression of sterol-related genes.

In addition to transcriptional regulation, sterol homeostasis is also achieved by cycles of esterification and hydrolysis that respectively induce sterol withdrawal from the membrane and sterol insertion in the membrane. Sterol esterification, catalyzed by sterol O-acyltransferases (SOAT), leads to accumulation of sterol esters in cytoplasmic fat droplets since these molecules are less soluble in the membrane

bilayer (Rogers *et al.* 2015). Conversely, hydrolysis by sterol esterases, results in free sterols being inserted back in the membrane. This represents the main short-term sterol regulation mechanism in mammals, yeasts and plants, and was also described in the apicomplexan parasite *Toxoplasma gondii* (Yang *et al.* 1996; Schaller 2004; Lige *et al.* 2013; Rogers *et al.* 2015). *Halocafeteria seosinensis* expressed two genes related to SOATs (ORFs m.17881 and m.78053). Consistent with repression of sterol synthesis, m.17881 was up-regulated at high salt (5.9-fold increase; Table 3.8), suggesting that membrane sterol level was potentially lower in this condition.

SOAT are enzymes with varying levels of specificity/preference. For example, although cholesterol is the best substrate of acyl-CoA cholesterol acyltransferase, the mammalian SOAT, the enzyme can also esterify other types of sterol including oxysterol, plant sterols and yeast sterol (Zhang *et al.* 2003; Rogers *et al.* 2015). The same is also observed for *A. thaliana* sterol acyltransferase 1 (*At*SAT1) that esterifies lanosterol, stigmasterol, sitosterol, with *in vitro* preference for cycloartenol (Chen *et al.* 2007). Higher levels of SOAT lead to more sterol esters as shown in *A. thaliana*, where overexpression of *At*SAT1 led to a twofold increase of sterol ester and a reduction from 59% of free sterol in wild type to 28% in transgenic plants (Chen *et al.* 2007). Thus, over-expression of these genes in *H. seosinensis* could be linked to membrane sterol homeostasis, perhaps with varying substrate specificity, where less membrane sterol would be required to ensure a fluid membrane a high salt.

Halocafeteria seosinensis sequences contain the MBOAT family motif FYxDWWN shown to be essential for SOAT catalytic activity (Figure 3.40; Guo et al. 2001). However, SOATs belong to the membrane-bound O-acyltransferase (MBOAT) family that also includes diacylglycerol O-acyltransferases (DGAT). Unfortunately, a survey of papers that identified functional residues in SOAT and DGAT based on sequence analyses and mutation studies indicated that no functional residues are universally and exclusively conserved in each group (Cao *et al.* 1996; Oelkers *et al.* 1998; Guo *et al.* 2001; Lin *et al.* 2003; Guo *et al.* 2005; An *et al.* 2006; Cao 2011). Therefore, although phylogenetic analysis indicated that *H. seosinensis* putative

Fig. 3.40. Partial alignment of membrane-bound O-acyltransferase sequences showing the family motif FYxDWWN (boxed).

SOAT sequences clustered with other SOAT-related sequences at the exclusion of DGAT sequences (with high bootstrap support, 99% and 90%; Figure 3.41), I could not determine confidently the function of these enzymes based on conserved functional residues.

Sterol is inserted back in the plasma membrane after hydrolysis of the acyl chain by sterol esterase. Interestingly, the gene closest to yeast sterol esterases in *H. seosinensis* (*i.e. S. cerevisiae* genes *TGL1*, *YEH1* and *YEH2*; Valachovic *et al.* 2002; Koeffel *et al.* 2005) was not differentially expressed. Other genes in *H. seosinensis* encoded the lipase domain alpha beta hydrolase but could not be associated to sterol esterase based on sequence conservation. This suggests that *H. seosinensis* does not transcriptionally regulate sterol esterase as a mechanism of generating more free sterol to insert in the membrane.

Twelve molecules of oxygen are required to synthesize one sterol molecule (Summons *et al.* 2006). As mentioned previously, oxygen solubility is depressed at high salt, raising the question of whether the repression of sterol synthesis at high salt was actually a result of lower oxygen availability. However, I infer that this is unlikely: although yeasts are auxotrophic for sterol when grown in complete absence of oxygen (Andreasen and Stier 1953), low oxygen availability actually stimulates transcription of sterol biosynthetic enzymes (Hughes *et al.* 2005; Todd *et al.* 2006; Chun *et al.* 2007; Synnott *et al.* 2010). Meanwhile, the transcriptome of *H. seosinensis* indicates that it is respiring at high salt, implying some oxygen availability. This argues against an oxygen-dependent repression of sterol synthesis at high salt in *H. seosinensis* and suggests that another factor, potentially a need for increased membrane fluidity, was involved.

3.6.2- *Mitochondrial Membrane Restructuring by Cardiolipin Remodeling*

Cardiolipin, a signature lipid of the mitochondrial inner membrane, represents an essential structural component that modulates the membrane properties and interacts with several membrane proteins (Luévano-Martínez and Kowaltowski 2015). It also contributes to biogenesis, assembly and stability of the
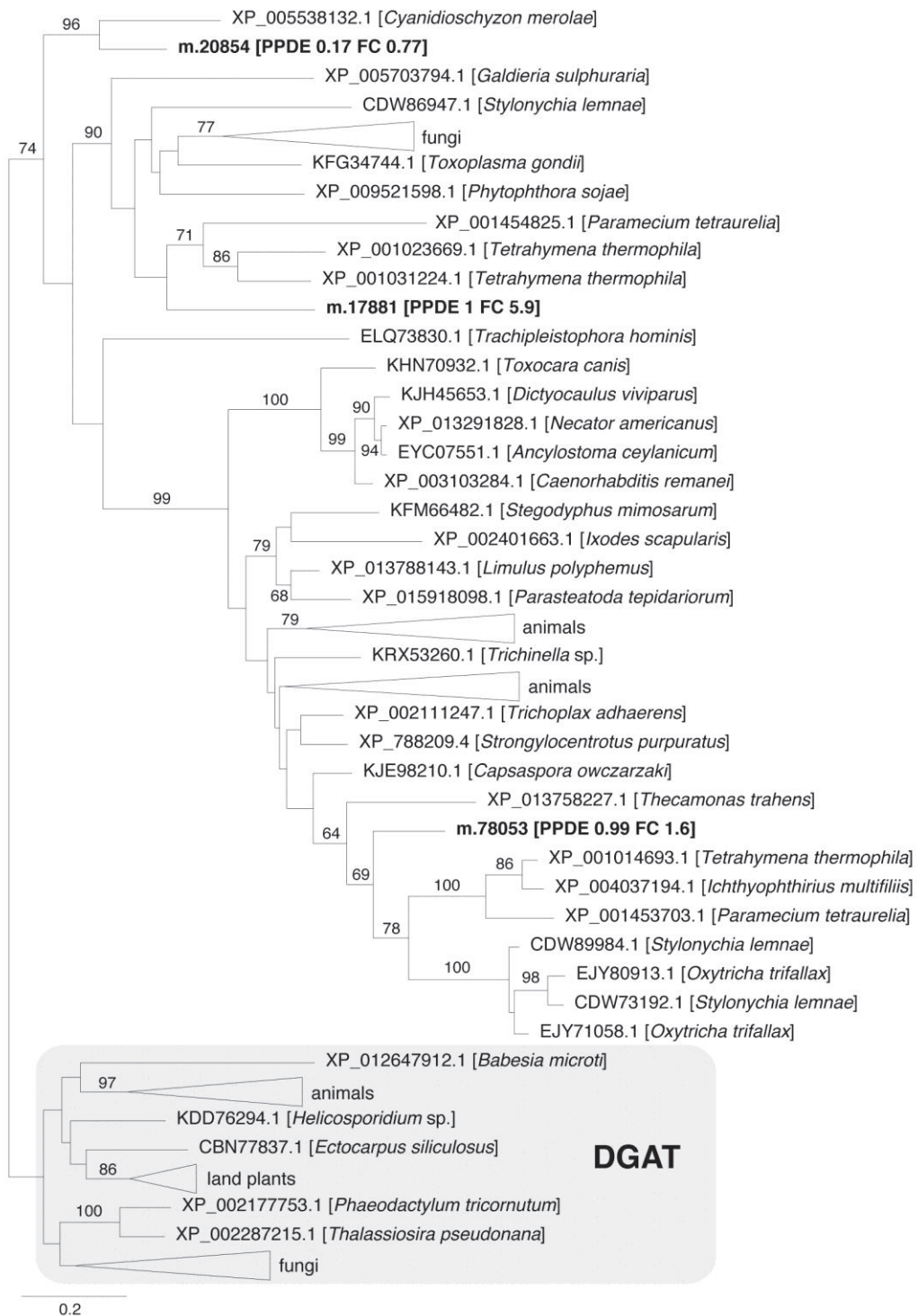
Fig. 3.41. Maximum-likelihood phylogenetic tree for membrane-bound O-acyltransferases including sequences annotated as diacylglycerol O-acyltransferase (DGAT, grey box) and sterol O-acyltransferase (outside the grey box). For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site.

respiratory chain supercomplexes (Zhang *et al.* 2005; Joshi *et al.* 2009; Acehan *et al.* 2011), to mitochondrial DNA stability and segregation (Luévano-Martínez *et al.* 2015) and it also anchors cytochrome *c* specifically and irreversibly to the membrane (Rytömaa and Kinnunen 1994).

Cardiolipin results from the condensation of cytidine diphosphodiacylglycerol with phosphatidylglycerol catalyzed by cardiolipin synthase (Schlame *et al.* 1993). This enzyme was present but not differentially expressed in *H. seosinensis* (Table 3.9). It is then remodelled to mature, more unsaturated cardiolipin by deacylation forming monolysocardiolipin followed by reacylation with another fatty acid catalyzed by tafazzin (Schlame and Rüstow 1990; Gu *et al.* 2004). In rat, the enzyme responsible for deacylation of immature cardiolipin is a mitochondrial phospholipase A2 (Zachman *et al.* 2010), while in yeast, this reaction is catalyzed by the lipase Cld1p (Beranek *et al.* 2009). Alternatively, acyl remodelling can potentially occur via a single transacylase step without the intervention of a lipase (Vreken *et al.* 2000; Xu *et al.* 2003; Schlame 2013). In *H. seosinensis*, I could not identify the enzyme performing the lipase step since several proteins were distantly related to the rat and yeast sequences. However, a gene related to tafazzin (ORF m.48339), for which the predicted protein encoded a mitochondrial transit peptide, was 2.1-fold up-regulated at high salt, suggesting that increased cardiolipin remodelling was part of *H. seosinensis* high salt transcriptional program.

As previously discussed, *H. seosinensis* was most probably exposed to high levels of ROS at high salt. Given that mature cardiolipin is vulnerable to oxidative damage, especially due to its polyunsaturated state (Kim *et al.* 2011), a plausible explanation of tafazzin up-regulation and induced remodelling would be as a way to remove ROS-damaged acyl chains in order to replace them with unaltered chains, as suggested by Baile *et al.* (2013). If not repaired, peroxidized cardiolipin leads to decreased activity of complexes I, III and IV, for instance (Paradies *et al.* 2000, 2001,

Table 3.9. Expression of genes involved in lipid metabolism in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	$\log_2$FC	Adjusted p-value	$\log_2$FC	
Cardiolipin synthesis									
m.62852	32.35	40.47	0.00	0.97	8.99E-01	-0.03	9.78E-01	0.01	Cardiolipin synthase
m.48339	14.10	37.24	1.00	2.11	2.04E-07	1.07	3.91E-04	1.10	Tafazzin
Phosphatodylglycerol (PG) / phosphatidylinositol (PI) transport									
m.67395	1959.17	156.93	1.00	0.06	5.60E-44	-3.91	1.61E-06	-3.98	PG/PI transfer potein
m.67401	1000.75	5366.43	1.00	4.03	7.31E-07	1.94	1.04E-03	2.05	PG/PI transfer potein
PI and phosphatidylethanolamine synthesis									
m.10411	144.53	60.14	1.00	0.32	2.40E-09	-1.63	4.27E-03	-1.59	PI synthase
m.47108	82.70	53.75	1.00	0.50	1.06E-05	-0.99	1.04E-02	-0.98	Phosphoethanolamine cytidylyltransferase
Fatty acid desaturases (FAD)									
m.39033	197.76	116.38	1.00	0.45	7.54E-23	-1.14	5.81E-05	-1.12	Delta12 FAD
m.34478	145.42	204.30	0.09	1.08	5.23E-01	0.11	3.71E-01	0.13	Delta12 FAD
m.67707	244.20	457.97	0.96	1.50	4.77E-03	0.58	8.13E-03	0.63	Delta9 FAD
m.88596	60.19	87.56	0.09	1.16	3.86E-01	0.21	1.83E-01	0.25	Delta6 FAD
m.75259	126.28	119.98	1.00	0.73	1.22E-03	-0.44	1.69E-02	-0.41	Delta6 FAD
m.8345	166.22	133.92	0.75	0.61	1.65E-03	-0.69	7.35E-03	-0.68	Delta6 FAD
m.68423	260.18	326.82	0.02	0.96	8.58E-01	-0.04	8.61E-01	-0.03	Delta4-sphingolipid FAD
Fatty acid elongases (FAE)									
m.26086	32.01	35.33	0.06	0.85	4.66E-01	-0.21	4.19E-01	-0.17	Long-chain FAE
m.59689	85.46	52.47	1.00	0.48	1.98E-15	-1.05	1.08E-04	-1.03	Long-chain FAE
m.60210	170.73	92.31	1.00	0.41	1.05E-14	-1.26	8.56E-05	-1.25	Long-chain FAE
m.45555	71.26	36.19	0.97	0.38	1.03E-05	-1.34	1.86E-03	-1.30	Long-chain FAE

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC	
Eukaryotic long-chain fatty acid CoA synthetase (ELFCS)									
m.10561	159.56	108.87	0.96	0.52	NA	-0.93	2.17E-03	-0.91	ELFCS
m.78263	6.92	12.05	0.96	1.35	3.99E-02	0.44	3.49E-02	0.43	ELFCS
m.30351	3.45	7.62	1.00	1.69	5.29E-04	0.76	4.63E-03	0.75	ELFCS
m.83169	13.62	17.75	0.05	1.02	9.28E-01	0.04	9.90E-01	0.00	ELFCS
m.33721	33.24	43.51	0.01	1.01	8.76E-01	0.02	7.14E-01	0.05	ELFCS
m.81834	40.82	63.54	0.06	1.25	2.37E-01	0.32	9.02E-02	0.37	ELFCS
m.85588	66.00	106.41	0.19	1.24	3.08E-01	0.31	1.86E-01	0.32	ELFCS
m.1208	83.24	119.13	0.41	1.12	3.08E-01	0.16	2.16E-01	0.18	ELFCS

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); log_2FC, log_2 fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.

2002; Musatov 2006). Concordantly, yeasts and human cells lacking tafazzin suffer from increased oxidative damage (Chen *et al.* 2008; Gonzalvez *et al.* 2013).

Alternatively, cardiolipin, as a fundamental structural component of the mitochondrial membrane, confers specific elastic and viscous properties, allowing for reversible expansion/stretching and contraction. This is especially important in non-permissive growth conditions (high temperature, salt or osmolarity: Luévano-Martínez and Kowaltowski 2015). Indeed, cardiolipin-lacking mitochondria display weakened osmotic stability and are potentially more subject to leakage due to altered swelling and shrinkage behaviours (Koshkin and Greenberg 2002). Tafazzin-mediated remodelling impacts the organelle architecture since tafazzin, for which substrate specificity depends on the membrane physical properties, can quickly reshuffle acyl groups locally, conferring flexibility of the lipid composition, as a 'membrane chaperone' that allows fusion, fission, bending and flattening of the membrane (Schlame *et al.* 2012). Therefore, up-regulation of tafazzin might be a way to increase mitochondrial morphological adaptability at high salt in *H. seosinensis*. Salt intrusion is likely to occur, at least accidentally, in a medium with salinity as extreme as 30%, and a higher level of tafazzin might contribute to maintaining homeostatic conditions and preventing mitochondrial dysfunction.

3.6.3- Lipid Transport and Synthesis

Two genes encoding lipid recognition domains were identified that had recently duplicated (*i.e.* after the divergence of *Cafeteria roenbergensis* and *H. seosinensis* from their common ancestor) but had drastically different expression patterns, and thus represented excellent ecoparalog candidates (Figure 3.42). They were encoded next to each other on the genome. One was 16-fold repressed at high salt (m.67395) while the other was 4.0-fold up-regulated (m.67401), and both of them had transcript abundance among the highest in their respective favored salinity (1,959 TPM, rank 95 at 15% salt, and 5,366 TPM, rank 7 at 30% salt respectively, Table 3.9). The homologous *Cafeteria roenbergensis* mRNA transcript (from the MMETSP) also had high relative abundance (4,317 TPM, rank 32)
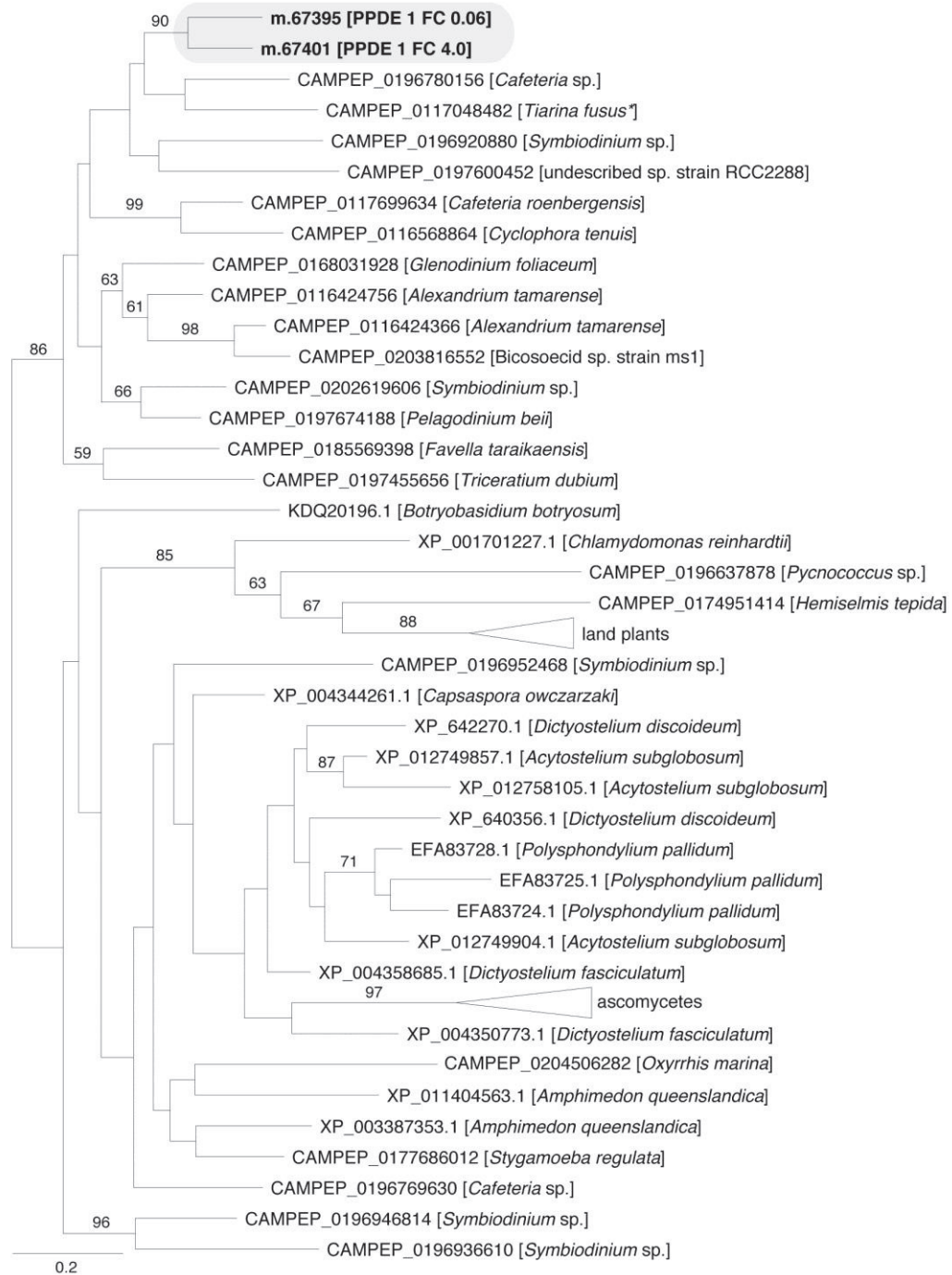
Fig. 3.42. Maximum-likelihood phylogenetic tree for gene duplication cluster 455 (grey box): phosphatidylglycerol/phosphatidylinositol transfer proteins. For *H. seosinensis* sequences (in bold), the Posterior Probability of being Differentially Expressed (PPDE) and the Fold Change (FC, 30% over 15% salt) is indicated. Bootstrap values (>50%) are indicated at branch nodes. The scale bar indicates the substitution rate/site. * Sequences generated from *Tiarina fusus* were potentially contaminated with sequences from the food source *Rhodomonas lens*.

192

suggesting these proteins have important functions. While m.67395 had no targeting sequence predicted with confidence, the probability that m.67401 (up-regulated at high salt) encoded a signal peptide was relatively high, at 0.83.

These genes encoded a MD-2-related lipid recognition domain (PFAM02221) characteristic of a family of four groups of proteins that mediate diverse biological functions involving lipid metabolism and transport, possibly through specific interactions with lipids (Inohara and Nunez 2002). The *H. seosinensis* ecoparalogs were most closely affiliated to a group of phosphatidylglycerol / phosphatidylinositol transfer proteins (PG/PI-TP) that are responsible for the intermembrane movement of phospholipids, although their physiological function is undetermined (Wirtz 1991). The closest characterized homolog available, expressed by *Aspergillus oryzae*, was shown to transfer preferentially PG and PI but also phosphatidylcholine, phosphatidylethanolamine and phosphatidylserine (Record *et al.* 1995). Similarity between *A. oryzae* PG/PI-TP sequence and m.67395 was detectable by BLASTP comparison (28% identical) while multiple alignment by eye was necessary to reveal that m.67401 was 18% identical to this protein. Transcription of the *A. oryzae* gene was stimulated by phospholipid supplementation of the medium and co-accumulation of mRNA transcripts and related proteins was observed (Record *et al.* 1999).

The substantial difference in expression pattern between these two genes in *H. seosinensis* suggested that membrane phospholipids needed to be adjusted as a response to increased salinity. Further experiments are required to determine what types of phospholipid were transferred as well as their cellular localisation, but this observation prompted me to investigate lipid synthesis in search of additional clues regarding substrate specificity of these transfer proteins.

Identifying lipid biosynthetic enzymes was complicated, since *H. seosinensis* sequences were typically distantly affiliated to proteins that had been studied and because many of them putatively have broad substrate specificity, like choline/ethanolamine phosphotransferases and kinases that can process either

choline- or ethanolamine- containing phospholipids (Lykidis 2007). In any case, most genes predicted to be involved in lipid synthesis in *H. seosinensis* were not differentially expressed, with the exception of two enzymes: phosphatidylinositol synthase (PIS) and phosphoethanolamine cytidyltransferase (PEC). These proteins were unambiguously identified by phylogenetic analyses and by specific domain arrangement.

PIS synthesizes phosphatidylinositol using CDP-diacylglycerol and myo-inositol as substrates. Only one sequence was related to this enzyme in *H. seosinensis* (ORF m.10411, 42% identical to rat PI synthase; Monaco *et al.* 1994) and it was 3.1-fold repressed at high salt (Table 3.9). The sequence predicts a fourfold-spanning membrane protein encoding a signal peptide and the ER retention signal KKxx. This suggests that it localized in the ER membrane, like the fungal, plant and mammalian enzymes (Antonsson 1997; Löfke *et al.* 2008; Bochud and Conzelmann 2015). A salt-induced decrease of phosphatidylinositol was also reported in the halotolerant yeast *D. hansenii* (Tunblad-Johansson *et al.* 1987). However, in addition to its structural role, phosphatidylinositol is the precursor for several derivatives involved in a diverse array of biological functions including glycolipid anchoring of proteins (Paulick and Bertozzi 2008), cell signalling (Divecha and Irvine 1995), vesicle trafficking (Martin 2001) and endocytosis (Sun *et al.* 2007). Therefore, with so many roles, deciphering the reason for transcriptional repression of PIS at high salt would require further investigation.

PEC participates in phosphatidylethanolamine synthesis by condensing phosphoethanolamine with CDP. In *H. seosinensis*, the most closely related gene (m.47108), 2.0-fold repressed at high salt, encoded two cytidyltransferase domains (CD02174 and CD02173) which is a common feature of this enzyme (Lykidis 2007). No signal sequence could be predicted, consistent with the cytoplasmic localization of the enzyme in mammals (Vermeulen *et al.* 1993; Bleijerveld *et al.* 2004), *T. brucei* (Gibellini *et al.* 2009) and *P. falciparium* (Maheshwari *et al.* 2013).

In contrast, PEC in plants and in the alga *Chlamydomonas reinhardtii* localizes to the outer membrane of mitochondria or the ER membrane (Wang and Moore 1991; Yang *et al.* 2004; Mizoi *et al.* 2006). Signal peptides can be predicted for these sequences only with low or moderate support (probability between 0.38-0.79). These enzymes possess an extra N-terminal transmembrane region, not detected in the cytoplasmic versions (*i.e.* from mammals, *T. brucei* and *P. falciparium*), that is also found in *H. seosinensis* predicted protein. Therefore, PEC in *H. seosinensis* is potentially more similar to the plant version in terms of regulation and physiological activity.

Nevertheless, in both mammals and plants, PEC is highly specific for phosphoethanolamine and, as the rate-limiting step of the pathway, is considered the key-regulatory enzyme of phosphatidylethanolamine synthesis (Sundler and Akesson 1975; Wang and Moore 1991; Vermeulen *et al.* 1994; Bladergroen and van Golde 1997; Tang and Moore 1997; Maheshwari *et al.* 2013). Altogether, this suggested that phosphatidylethanolamine synthesis was repressed at high salt in *H. seosinensis* as observed in the halotolerant yeast *D. hansenii* and the halophilic yeast *Phaeotheca triangularis* (Andreishcheva *et al.* 1999; Turk *et al.* 2004)*.* This result also relates to salt adaptation in bacterial membranes, where phosphatidylethanolamine is predicted to destabilize the bilayer phase at higher salinities (Russell 1989).

Regarding phosphatidylglycerol synthesis, *H. seosinensis* putative phosphatidylglycerophosphate synthase (m.79658), that catalyzes the committed step of the pathway (Greenberg and Lopes 1996; Alder 2012) was not differentially expressed. This sequence encoded the catalytic domains of eukaryotic phosphatidylglycerophosphate synthase (CD09135 and CD09137) and a mitochondrial targeting peptide. In non-photosynthetic eukaryotes, phosphatidylglycerol is found exclusively in mitochondrial membrane as a minor constituent where it is rapidly converted to cardiolipin (Daum 1985; Horvath and Daum 2013). Other than being a precursor for cardiolipin, phosphatidylglycerol was

therefore unlikely to influence the property of membranes, especially the plasma membrane, in response to variations in extracellular salinity.

In summary, I found evidence for the transcriptional regulation of genes involved in phosphatidylinositol and phosphatidylethanolamine synthesis. Two enzymes, PIS and PEC, indicated that synthesis of these lipids was potentially repressed at high salt. It is tempting to speculate that the ecoparalog related to lipid transfer proteins that was up-regulated at low salt might be involved in the transport of phosphatidylinositol or phosphatidylethanolamine. Obviously, testing this hypothesis will require additional experimental work.

3.6.4- A Note on Lipid Saturation and Elongation

The length and saturation level of fatty acyl chains impact membrane fluidity, with a greater number of double bonds and shorter chain lengths increasing fluidity (Lodish *et al.* 2000; Beney and Gervais 2001). In *H. seosinensis*, putative desaturases (with domain PFAM00487) were not differentially expressed, with the exception of a gene related to delta12 fatty acid desaturases (CD03507) that was 2.2-fold repressed at high salt (Table 3.9). The related enzyme potentially functions in the ER since it encoded the retention signal KDEL. If this enzyme acted on lipid exported to the plasma membrane, its down-regulation at high salt is unexpected since it would theoretically lead to lower membrane fluidity in a condition that favors reduced fluidity.

By contrast, a set of genes encoding long chain fatty acid elongases (PFAM01151) were repressed at high salt (2.1- to 2.6-fold), concordantly with a theoretical need for shorter acyl chains in this condition (Table 3.9). Enzymes of this family are involved in the synthesis of very long fatty acids (>20 carbons; Oh *et al.* 1997). Another set of genes related to eukaryotic long-chain fatty acid CoA synthetases (CD05927) that activate fatty acid with chain lengths between 12 and 20 were not differentially expressed.

196

3.6.5- Summary of Section 3.6

In summary, genes involved in sterol biosynthesis and transport were repressed at high salt, while genes involved in generating steryl esters were up-regulated, suggesting that levels of free (unesterified) sterol might be lower at high salt. I also detected repression of phosphatidylinositol and phosphatidylethanolamine synthesis and fatty acid elongases at high salt. In addition, high salt indirectly affected mitochondrial membranes by stimulating remodeling of cardiolipin, a signature lipid of the inner mitochondrial membrane. Finally, two recently duplicated genes with strikingly different expression patterns indicated that lipid intermembrane flow differed in both conditions supporting the idea that membranes were affected by the extracellular salinity.

Based on these transcriptomic data, a tentative model for lipid metabolism in *H. seosinensis* as a response to higher extracellular salinity includes: repression of synthesis, transport and insertion of sterol in plasma membrane, remodelling of mitochondrial cardiolipin, and decreased fatty acid elongation. Expression of several of these genes might result in membrane fluidity adjustment. Especially, intracellular content is likely to influence the properties of membrane systems. For example, *H. seosinensis* potentially synthesized ectoine and hydroxyectoine (see chapter 2), molecules that have been shown to have a fluidizing effect on bilayer systems (Harishchandra *et al.* 2010), suggesting that, in addition to their putative role as organic osmolytes, these molecules could influence membrane fluidity. Concordantly, the enzyme synthesizing hydroxyectoine was 220-fold up-regulated at high salt. Looking forward, direct measurement of lipids in varying membrane fractions is required to describe further *H. seosinensis* adaptation and to confirm the role of differentially expressed genes identified by my transcriptome analysis.

3.7- Carbohydrate Metabolism is Affected by Salinity Variation

A substantial proportion of genes related to glycoside hydrolases and sugar transporters were up-regulated at high salt. More specifically, six of them (2.1- to 9.1-fold increase) were related to hexaminidase, polygalacturonase, α-xylosidase and α-galactosidase (glycoside hydrolase families 20, 28, 31 and 36), and five genes (3.1- to 44-fold increase) were related to transporters of the Major Facilitator Superfamily (domains of transporters for sugar – PFAM00083, for nucleotide-sugar – COG5070, and for triose-phosphate – PFAM03151; Table 3.10). Enhanced expression of these genes indicated that the intracellular carbohydrate content might be higher at high salt. In line with this possibility was the up-regulation (2.4-fold) of fructosamine-3-kinase (PFAM03881). This enzyme removes unwanted fructosamine residues added to proteins during spontaneous glycation, a phenomenon that is proportional to intracellular glucose concentration (Delpierre and Van Schaftingen 2003). This enzyme in *H. seosinensis* possessed the conserved aminoglycoside kinase motif PxLXHGDLWSxN (Figure 3.43).

Fig. 3.43. Partial alignment of fructosamine-3-kinase sequences showing the aminoglycoside kinase motif (PxLxHGDLWSxN, boxed).

Table 3.10. Expression of genes involved in carbohydrate metabolism in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log_2FC	Adjusted p-value	log_2FC
Glycosyl hydrolases								
m.38126	0.23	2.67	1.00	9.09	0.00E+00	3.09	0.00E+00	3.36
m.17615	1.66	8.01	1.00	4.02	1.80E-04	1.89	1.18E-03	2.05
m.13232	1.60	7.90	1.00	3.97	4.12E-14	1.96	6.71E-05	2.01
m.67061	12.85	39.68	1.00	2.46	2.93E-10	1.29	1.38E-04	1.32
m.46314	20.07	57.53	1.00	2.30	3.16E-04	1.18	2.51E-03	1.25
m.78119	14.77	38.35	1.00	2.08	1.09E-03	1.03	4.49E-03	1.05
m.40656	11.71	25.06	0.96	1.69	4.47E-03	0.75	7.79E-03	0.78
m.49098	12.07	20.47	1.00	1.30	1.68E-02	0.39	1.97E-02	0.43
m.66740	7.58	7.88	0.46	0.79	4.44E-01	-0.34	1.61E-01	-0.44
m.86396	28.21	35.19	0.06	0.96	7.89E-01	-0.05	8.28E-01	-0.03
m.7502	97.99	128.78	0.02	1.02	8.85E-01	0.03	7.24E-01	0.06
m.33616	123.92	131.14	1.00	0.81	1.24E-02	-0.30	7.61E-02	-0.28
Sugar transporters								
m.89759	1.17	68.97	1.00	44.03	7.78E-114	5.41	1.62E-07	5.56
m.27262	4.84	61.52	1.00	9.40	1.44E-26	3.17	5.39E-06	3.41
m.51724	1.84	10.60	1.00	4.95	NA	2.09	2.90E-03	2.42
m.52982	1.69	7.36	1.00	3.47	1.10E-10	1.76	1.19E-04	1.78
m.15888	6.44	25.34	1.00	3.11	1.20E-12	1.62	8.29E-05	1.62
m.27173	1.61	4.68	0.64	2.31	4.19E-02	1.13	3.78E-02	1.10
m.1448	17.98	12.83	1.00	0.54	3.15E-05	-0.87	1.78E-03	-0.87
m.69236	13.73	20.67	0.13	1.19	2.62E-01	0.25	1.49E-01	0.26
m.93427	32.87	35.63	0.26	0.83	1.64E-01	-0.25	1.54E-01	-0.22
m.79269	34.78	45.45	0.03	1.04	8.22E-01	0.06	5.98E-01	0.09

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA	
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-value	log$_2$FC	Adjusted p-value	log$_2$FC
Sugar transporters (continued)								
m.86348	2.07	7.71	0.18	3.40	6.53E-01	0.53	7.78E-02	4.44
m.46478	32.57	28.43	1.00	0.66	4.74E-04	-0.58	5.75E-03	-0.58
m.81966	29.35	46.58	0.47	1.26	9.16E-02	0.33	4.73E-02	0.36
m.74602	25.63	20.38	1.00	0.62	2.03E-04	-0.67	3.59E-03	-0.65
m.93347	15.60	38.26	1.00	1.96	1.67E-03	0.95	5.12E-03	0.97
m.7151	187.25	154.88	0.55	0.63	NA	-0.64	1.38E-02	-0.64
m.33602	6.36	10.59	0.12	1.36	NA	0.43	2.28E-01	0.44
m.32446	144.50	179.47	0.02	0.94	7.71E-01	-0.08	7.69E-01	-0.06
m.28531	2.19	2.38	0.17	0.88	8.28E-01	-0.17	5.83E-01	-0.29
Fructosamine-3-kinase								
m.82266	61.75	194.56	1.00	2.44	5.38E-05	1.26	2.11E-03	1.28
Protein glycosylation								
m.59896	0.07	5.34	1.00	56.68	8.41E-23	5.38	4.39E-05	5.89
m.15355	8.28	22.77	1.00	2.22	4.31E-06	1.14	6.53E-04	1.16
m.27749	3.04	4.76	0.09	1.16	7.54E-01	0.20	7.40E-01	0.13
m.26405	4.44	8.70	0.98	1.47	9.41E-02	0.55	2.51E-02	0.70
m.37811	15.17	28.59	0.04	1.53	3.29E-01	0.58	2.01E-01	0.67

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt); NA, not available due to an extreme count outlier in one of the samples.

Metabolized carbohydrates are commonly used as organic osmolytes. They can also stabilize cell membranes under stress conditions by interacting with the polar heads of phospholipids (Rudolph *et al.* 1986). However, since the reactive reducing end of sugars could threaten other cellular components, for example during protein glycation (mentioned above), osmoprotective carbohydrates are typically non-reducing saccharides, like trehalose, or they are modified by addition of a small neutral molecule, like glycerol, glyceramide or glyceric acid (Roberts 2005). As discussed below I could not unambiguously identify enzymes involved in the synthesis or transport of such carbohydrates in *H. seosinensis.*

For example, *H. seosinensis* expressed two non-differentially expressed genes (ORFs m.91570 and m.88152) containing both domains required for trehalose synthesis: trehalose-6-phosphate synthase (PFAM00982) and trehalose-6-phosphate phosphatase (PFAM02358). In several eukaryotes, the trehalose-synthesizing enzyme is encoded by a fusion of these genes, but no clear evidence suggests that the fusion protein can catalyze both required reactions, *i.e.* formation of trehalose-6-phosphate from glucose-6-phosphate followed by dephosphorylation (Avonce *et al.* 2006; Sook Chung 2008). Both proteins in *H. seosinensis* encoded a partial domain non-reciprocally, *i.e.,* m.88152 encoded a partial synthase domain but a complete phosphatase domain and m.91570 encoded a partial phosphatase domain but a complete synthase domain, indicating that they might function cooperatively. However, investigation of functional residues only identified partial conservation for both proteins (Collet *et al.* 1998; Morais *et al.* 2000; Gibson *et al.* 2002; Rao *et al.* 2006). Direct measurements of intracellular content are therefore greatly needed to determine whether up-regulation of glycoside hydrolases and sugar transporters resulted in carbohydrate accumulation.

Hypothetically, increased carbohydrate-related activity could also be linked to protein glycosylation, potentially leading to increased protein solubility in conditions with lower water activity (Schülke and Schmid 1988; Tams and Welinder 1995; Tams *et al.* 1999). Concordantly, among the five genes encoding the domain of the glycosyl transferase family 41 (PFAM13844, *O*-linked β-*N*-acetylglucosamine

transferases), two were 2.2- and 57-fold up-regulated, suggesting that protein glycosylation was stimulated at high salt (Table 3.10). These enzymes transfer N-acetylglucosamine to serine and threonine residues in proteins that are normally localized in the nucleus or in the cytoplasm.

O-linked glycosylation can have myriad effects on proteins. For example, it can affect protein structure by increasing stability, can regulate enzymatic activity, and modulate proteolytic cleavage influencing protein expression and processing (Van den Steen *et al.* 1998). However, in response to exposure to a variety of stressors (salt, hydrogen peroxide, heat, UV light, heavy metals), *O*-linked glycosylation increases on a large number of proteins as a protective mechanism in metazoan cells (Zachara *et al.* 2004; Selvan *et al.* 2015). The protective effect of *O*-linked glycosylation is partly explained by modulation of HSP70 and HSP40 expression and persistence (Zachara *et al.* 2004). It is possible that the salt-dependent overexpression of *O*-linked β-*N*-acetylglucosamine transferases in *H. seosinensis* is related to a similar protective role.

In summary, overrepresentation of upregulated genes assigned to the class 'Carbohydrate metabolism and transport' indicated that high salt stimulated the expression of genes encoding for proteins that interact with sugars. These were related to glycoside hydrolases, sugar transporters and *O*-linked β-*N*-acetylglucosamine transferases. Potentially, up-regulation of these genes can lead to intracellular accumulation of carbohydrates, used as osmolytes, or to increased protein glycosylation. Consistently, overexpression of a gene related to fructosamine-3-kinase suggested that intracellular sugar concentration was higher at high salt. Investigation of the intracellular metabolites by H-NMR, HPLC or mass spectrometry and of glycosylated proteins, for example by lectin purification followed by a labeling procedure, is required to confirm the likelihood of these hypotheses.

3.8- Other Highly Up-Regulated Genes

The systems described above represented the best-case examples describing aspects of the physiological response of *H. seosinensis* exposed to very high salt concentration. This description resulted from an exhaustive examination of the most up-regulated genes and their affiliated molecular partners. In practice ~70% of genes that were highly up-regulated (> 5-fold change), highly abundant (> 100TPM), and that were assigned to informative COG classes (*i.e.* not considering categories listed in section 2.5) are discussed in the above sections or in chapter 2 (*e.g.* ectoine hydroxylase and amino acid transporters) illustrating that these discussed cases represent a large portion of the whole phenomenon.

Other sequences that fell in this category but were not discussed above included genes encoding domains for serine carboxypeptidase (PFAM00450, 70-fold increase), iron-sulfur assembly scaffold (PFAM01592, 5.3-fold increase), proline 4-hydroxylase (COG3751, 31-fold increase), an ADP/ATP transporter (PFAM00153, 68-fold increase), alpha/beta hydrolase (PFAM12695, 6.9-fold increase), aldehyde dehydrogenase (PFAM00171, 45-fold increase) and ATP-dependent exonuclease V, helicase superfamily I (COG0507, 65-fold increase; Table 3.11). Although proteins encoded by these genes were homologous to annotated sequences, the context in which they operate remained elusive.

Table 3.11. Expression of highly up-regulated genes in *Halocafeteria seosinensis*

ORF names	Abundance (TPM)		EBSeq		DESeq2		VOOM-LIMMA		Annotation
	15% salt	30% salt	PPDE	Post fold change	Adjusted p-values	log$_2$ FC	Adjusted p-values	log$_2$ FC	
m.61335	2.17	195.63	1.00	70.44	8.91E-269	6.09	3.65E-05	6.15	carboxypeptidase
m.43550	27.81	177.31	1.00	5.31	1.67E-06	2.28	1.75E-02	2.33	FeS assembly protein
m.90674	23.02	967.87	1.00	31.27	3.12E-76	4.88	1.18E-04	5.09	proline 4-hydroxylase
m.17098	31.67	2759.06	1.00	67.98	6.91E-97	5.95	8.19E-05	6.16	ADP/ATP transporter
m.19219	1.29	116.44	1.00	65.07	9.19E-43	5.78	5.93E-04	6.32	exonuclease V
m.18900	17.15	154.48	1.00	6.92	3.23E-19	2.73	7.30E-04	2.84	alpha beta hydrolase fold
m.69600	6.10	371.06	1.00	45.32	2.26E-95	5.41	8.19E-05	5.64	aldehyde dehydrogenase

Abbreviations: TPM, averaged transcripts per million; PPDE, Probability of being Differentially Expressed, Post Fold Change, posterior fold change (30% over 15% salt); log$_2$FC, log$_2$ fold change (30% over 15% salt).

4- Conclusion

The ability of protists to grow preferentially at very high salt concentrations was only recently recognized. Although molecular information about halophilic microbial eukaryotes like fungi and algae has been accumulating in the past years, this study represents the first in-depth examination of gene expression in a halophilic protozoon. The transcriptomic response of *H. seosinensis* grown at very high salt concentration included the up-regulation of genes involved in signal transduction, ion homeostasis, stress response, lipid and carbohydrate metabolism. These overexpressed genes most likely allowed *H. seosinensis* to adapt to a sustained higher level of ionic and oxidative stresses, and to acclimate its plasma membrane to enhanced hypersaline conditions. In addition, several gene duplication and LGT events potentially contributed to increase *H. seosinensis* salt adaptation over evolutionary time.

This study was only the first step in understanding *H. seosinensis* halophilicity at the molecular level. Since transcript level is not necessarily correlated with protein activity, and other control mechanisms in addition to transcriptional regulation occur in cells, further work, like quantitative proteomics and determination of enzymatic activity, is required in order to validate the results presented in this chapter. Moreover, future experiments designed to dissociate the influence on gene expression of physico-chemical parameters co-varying with salinity (*e.g.* concentration of dissolved oxygen) are greatly recommended.

Chapter 4 - Conclusion

This thesis demonstrates how deep-sequencing techniques can contribute to obtain a substantial amount of information about organisms that are difficult to study *in vitro*, in this case mainly due to the non-axenic nature of most protozoan cultures. It allowed the identification of cellular components that are pivotal to salt adaptation and that need to be further studied and quantified, for example sterols or ectoine/hydroxyectoine. As with any pioneering work, it also raised several working hypotheses.

In chapter 2, I suggested that intracellular ion content might be higher in halophilic protists relative to marine protists. This conclusion needs to be validated by direct measurements of intracellular salt content. However, the non-axenic nature of our protozoan cultures hinders any techniques that use cell pellets as starting material, since the signals coming from the organism of interest and from the prokaryotic prey need to be dissociated. Techniques that do not require isolating cells by centrifugation include the patch-clamp technique (Miedema and Assmann 1998) or X-ray microanalysis (Oren *et al.* 1997). Since the former technique is contingent on physical contact between a cell and an electrode, *P. kirbyi* would be a better candidate since it is bigger than *H. seosinensis* and does not move as frantically. Measuring optimal activity of enzymes as a function of salt could also provide insights into intracellular salt content, assuming that cytoplasmic enzymes are adapted to the cytosolic salinity.

Similarly, cytosolic accumulation of organic osmolytes must be directly measured, possibly by proton nuclear magnetic resonance (^{1}H NMR) and mass spectrometry, to validate the differential gene expression results presented in chapter 2. Effort must be invested at bringing the organism of interest to feed on a prokaryotic isolate known to not synthesize the osmolyte meant for detection, or various controls must be included in order to measure intracellular metabolites of

the prokaryotic population present in the culture and subtract the prokaryotic signal from the mixed signals. Preliminary experiments using the latter strategy indicated that this is a challenging task. Amongst other problems, analysis of the generated spectra was complicated by the high molecular diversity in the extracts and by the presence of numerous peaks unassignable to know molecules in reference databases.

Halocafeteria seosinensis expressed ectoine hydroxylase that was 200-fold up-regulated at 30% salt and that produces hydroxyectoine, a commonly used osmolyte in bacteria that is also a superior protector against desiccation. Therefore, accumulation of hydroxyectoine could, in principle, protect *H. seosinensis* in the case of complete desiccation though vitrification, as shown in certain bacteria (Tanne *et al.* 2014). This ability has ecological implications since it could favor *H. seosinensis* dispersal, similarly to the ability to form cysts (which has not yet been reported in *H. seosinensis*), and contribute to its worldwide distribution.

In chapter 3, I described the transcriptomic response of *H. seosinensis* to a long-term exposure to high salt. A complementary investigation would be to characterize its response to short-term salinity variations. One could expect this system to be more dynamic, for instance at the level of ion transport. For example, mRNA accumulation for plasma membrane H^+-ATPase in the halophytic plant *Atriplex nummularia* and in tobacco culture cells is recorded only during acute salt adaptation implying that, once the appropriate proton gradient is established, the abundance of H^+-ATPase returned to the same steady state that preceded exposure to salt (Niu *et al.* 1993; Perez-Prat *et al.* 1994).

Repression of genes involved in cellular respiration suggested that variations in oxygen level potentially affected gene expression. An experimental framework to test the impact of such a factor would be to reproduce the gene expression study in conditions where salinity is stable but oxygen varies. This approach may allow the disentanglement of the transcriptomic response related to differences in oxygen availability from the response induced by salinity variation (that includes a concomitant oxygen-dependent component).

Prey selectivity is known to affect the growth of protozoa and still little is known about feeding behavior (Verity 1991; Matz *et al.* 2002; Beardsley *et al.* 2003; Pernthaler 2005). In the context of my study, this was a major concern since prey bacteria seemed to influence the quality of RNA extracts, where bacteria non-adapted to high salt, like *Salinivibrio*, would trigger sample degradation but not extreme halophiles like *Haloferax*. Since *P. kirbyi* did not grow well on any tested Haloarchaea, I did not manage to obtain intact RNA extracts for a differential expression study of *P. kirbyi.* Therefore, further characterizations of *P. kirbyi* gene expression using transcriptomics will require an exhaustive search for the appropriate food source.

However, these feeding behaviors point at future ecological research avenues. Since halophilic protozoa can control prokaryotic abundance (Guixa-Boixareu *et al.* 1996; Pedrós-Alió *et al.* 2000; Park *et al.* 2003), how does prey selectivity affect grazing rates and thus microbial community composition in hypersaline environments? Discriminant feeding is easy to observe in *P. kirbyi*. In Petri dishes, amoebae of *P. kirbyi* tend to explore streaks of *Haloferax* without dividing while, on *Salinivibrio*, the migrating front quickly progresses, leaving a cleared zone on its way. My isolates of *Salinivibrio* and *Haloferax* divides optimally at 5% and 18% salt respectively. Therefore, in a three-organism culture with *P. kirbyi* as predator*,* we can propose that a lower salt concentration (*e.g.* 10% salt) might lead to gradual enrichment of *Haloferax* relative to *Salinivibrio* over time, while higher salinities (*e.g.* 20% salt) would disfavor *Salinivibrio* growth and subsequently *P. kirbyi*. In contrast, *H. seosinensis* feeds well on both *Salinivibrio* and *Haloferax*, probably contributing to its success in a wide range of geographical location. Hence, prey selectivity and availability is likely to influence both protist and prokaryotic diversity.

Finally, while employing a comparative approach as much as possible, the work presented in this thesis was nonetheless restricted to single strains of two species. Over the course of recent research on halophilic protists, it has become increasingly clear that (borderline) extreme halophily has evolved a large number of times within eukaryotes and that a wide range of organisms can be studied in a

laboratory setting (Park *et al.* 2006; Park *et al.* 2007; Cho *et al.* 2008; Park *et al.* 2009; Park and Simpson 2010, 2011; Foissner *et al.* 2014; Park and Simpson 2015). The growing availability of diverse halophilic protist cultures therefore unlocks the possibility of more systematic and comprehensive investigations. Acquisition of genomes and transcriptomes from other halophilic protists and related organisms should indicate whether molecular features observed in *H. seosinensis* and *P. kirbyi* represent converging commonalities in the halophile world or species-specific adaptations. Pieces of the puzzle of protist evolution are starting to fall into place, providing a better picture as discoveries will undoubtedly accumulate in the near future.

References

Acehan D, Malhotra A, Xu Y, Ren M, Stokes DL*, et al.* 2011. Cardiolipin affects the supramolecular organization of ATP synthase in mitochondria. Biophys J. 100: 2184-2192.

Adams MA and Jia ZC. 2005. Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*. J Biol Chem. 280: 8358-8363.

Adebali O, Reznik AO, Ory DS and Zhulin IB. 2016. Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. Genet Med.

Ago H, Kanaoka Y, Irikura D, Lam BK, Shimamura T*, et al.* 2007. Crystal structure of a human membrane protein involved in cysteinyl leukotriene biosynthesis. Nature. 448: 609-612.

Akileswaran L, Brock BJ, Cereghino JL and Gold MH. 1999. 1,4-benzoquinone reductase from *Phanerochaete chrysosporium*: cDNA cloning and regulation of expression. Appl Environ Microbiol. 65: 415-421.

Al-Hasan RH, Ghannoum MA, Sallal AK, Abu-Elteen KH and Radwan SS. 1987. Correlative changes of growth, pigmentation and lipid composition of *Dunaliella salina* in response to halostress. J Gen Microbiol. 133: 2607-2616.

Al-Qassab S, Lee WJ, Murray S, Simpson AGB and Patterson DJ. 2002. Flagellates from stromatolites and surrounding sediments in Shark Bay, Western Australia. Acta Protozool. 41: 91-144.

Al-Sheikh H. 2010. Two pathogenic species of *Pythium*: *P. aphanidermatum* and *P. diclinum* from a wheat field. Saudi J Biol Sci. 17: 347-52.

Alder N. 2012. Biogenesis of lipids and proteins within mitochondrial membranes. In: Yeagle PL, editor. *The Structure of Biological Membranes, Third Edition*. Boca Raton, FL: CRC Press. p. 315-378.

Alexander E, Stock A, Breiner HW, Behnke A, Bunge J*, et al.* 2009. Microbial eukaryotes in the hypersaline anoxic L'Atalante deep-sea basin. Environ Microbiol. 11: 360-381.

Aljanabi SM and Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25: 4692-4693.

Allen MA, Goh F, Burns BP and Neilan BA. 2009. Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geobiology. 7: 82-96.

Allen MD, Kropat J, Tottey S, Del Campo JA and Merchant SS. 2007. Manganese deficiency in *Chlamydomonas* results in loss of photosystem II and MnSOD function, sensitivity to peroxides, and secondary phosphorus and iron deficiency. Plant Physiol. 143: 263-277.

Allocati N, Casalone E, Masulli M, Polekhina G, Rossjohn J*, et al.* 2000. Evaluation of the role of two conserved active-site residues in Beta class glutathione S-transferases. Biochem J. 351: 341-346.

Allocati N, Cellini L, Aceto A, Iezzi T, Angelucci S*, et al.* 1994. Immunogold localization of glutathione transferase B1-1 in *Proteus mirabilis*. FEBS Lett. 354: 191-194.

Allocati N, Favaloro B, Masulli M, Alexeyev MF and Di Ilio C. 2003. *Proteus mirabilis* glutathione S-transferase B1-1 is involved in protective mechanisms against oxidative and chemical stresses. Biochem J. 373: 305-311.

Allocati N, Federici L, Masulli M and Di Ilio C. 2009. Glutathione transferases in bacteria. FEBS J. 276: 58-75.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. 1990. Basic Local Alignment Search Tool. J Mol Biol. 215: 403-410.

An SJ, Cho KH, Lee WS, Lee HO, Paik YK*, et al.* 2006. A critical role for the histidine residues in the catalytic function of acyl-CoA : cholesterol acyltransferase catalysis: evidence for catalytic difference between ACAT1 and ACAT2. FEBS Lett. 580: 2741-2749.

Anantharaman V and Aravind L. 2001. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci. 26: 579-582.

Anantharaman V, Koonin EV and Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol. 307: 1271-1292.

Andersen RA, Berges JA, Harrison PJ and Watanabe MM. 2005. Appendix A - Recipes for Freshwater and Seawater Media. In: Andersen RA, editor. *Algal Culturing Techniques*. Burlington: Elsevier Academic Press. p. 429-538.

Andersson JO. 2009. Horizontal Gene Transfer Between Microbial Eukaryotes. In: Gogarten MB, Gogarten JP and Olendzenski LC, editors. *Horizontal Gene Transfer: Genome in Flux*. Totawa, NJ: Humana Press. p. 473-487.

Andersson JO, Sjögren AM, Davis LAM, Embley TM and Roger AJ. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol. 13: 94-104.

Andersson JO, Sjögren AM, Horner DS, Murphy CA, Dyal PL, *et al.* 2007. A genomic survey of the fish parasite *Spironucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. BMC Genomics. 8: 51.

Anding C, Brandt RD and Ourisson G. 1971. Sterol biosynthesis in *Euglena gracilis* Z: sterol precursors in light-grown and dark-grown *Euglena gracilis* Z. Eur J Biochem. 24: 259-263.

Andrade SLA, Patridge EV, Ferry JG and Einsle O. 2007. Crystal structure of the NADH : Quinone oxidoreductase WrbA from *Escherichia coli*. J Bacteriol. 189: 9101-9107.

Andreasen AA and Stier TJB. 1953. Anaerobic nutrition of *Saccharomyces cerevisiae*. 1. Ergosterol requirement for growth in a defined medium. J Cell Comp Physiol. 41: 23-36.

Andreishcheva EN, Isakova EP, Sidorov NN, Abramova NB, Ushakova NA, *et al.* 1999. Adaptation to salt stress in a salt-tolerant strain of the yeast *Yarrowia lipolytica*. Biochem (Mosc). 64: 1061-1067.

Antonsson B. 1997. Phosphatidylinositol synthase from mammalian tissues. BBA-Lipid Lipid Met. 1348: 179-186.

Aoyama Y, Yoshida Y and Sato R. 1984. Yeast cytochrome P-450 catalyzing lanosterol 14-alpha-demethylation. 2. Lanosterol metabolism by purified P-45014DM and by intact microsomes. J Biol Chem. 259: 1661-1666.

Aravind L, Mazumder R, Vasudevan S and Koonin EV. 2002. Trends in protein evolution inferred from sequence and structure analysis. Curr Opin Struct Biol. 12: 392-399.

Arnold K, Bordoli L, Kopp J and Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 22: 195-201.

Arun N and Singh DP. 2013. Differential response of *Dunaliella salina* and *Dunaliella tertiolecta* isolated from brines of Sambhar Salt Lake of Rajasthan (India) to salinities: a study on growth, pigment and glycerol synthesis. J Mar Biol Assoc India. 55: 65-70.

Asojo OA, Schott EJ, Vasta GR and Silva AM. 2006. Structures of PmSOD1 and PmSOD2, two superoxide dismutases from the protozoan parasite *Perkinsus marinus*. Acta Crystallogr Sect F Struct Biol Cryst Commun. 62: 1072-1075.

Avonce N, Mendoza-Vargas A, Morett E and Iturriaga G. 2006. Insights on the evolution of trehalose biosynthesis. BMC Evol Biol. 6: 109.

Axelsen KB and Palmgren MG. 1998. Evolution of substrate specificities in the P-type ATPase superfamily. J Mol Evol. 46: 84-101.

Ayadi H, Toumi N, Abid O, Medhioub K, Hammami M, *et al.* 2002. Étude qualitative et quantitative des peuplements phyto- et zooplanctoniques dans les bassins de la saline de Sfax, Tunisie. Rev Sci Eau/J Water Sci. 15: 123-135.

Azachi M, Sadka A, Fisher M, Goldshlag P, Gokhman I, *et al.* 2002. Salt induction of fatty acid elongase and membrane lipid modifications in the extreme halotolerant alga *Dunaliella salina*. Plant Physiol. 129: 1320-1329.

Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, *et al.* 1983. The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser. 10: 257-263.

Baile MG, Whited K and Claypool SM. 2013. Deacylation on the matrix side of the mitochondrial inner membrane regulates cardiolipin remodeling. Mol Biol Cell. 24: 2008-2020.

Bak S, Kahn RA, Olsen CE and Halkier BA. 1997. Cloning and expression in *Escherichia coli* of the obtusifoliol 14 alpha-demethylase of *Sorghum bicolor* (L) Moench, a cytochrome P450 orthologous to the sterol 14 alpha-demethylases (CYP51) from fungi and mammals. Plant J. 11: 191-201.

Baker DA and Kelly JM. 2004. Structure, function and evolution of microbial adenylyl and guanylyl cyclases. Mol Microbiol. 52: 1229-1242.

Baldoni E, Genga A and Cominelli E. 2015. Plant MYB transcription factors: their role in drought response mechanisms. Int J Mol Sci. 16: 15811-15851.

Barrero-Gil J, Garciadeblás B and Benito B. 2005. Sodium, potassium-ATPases in algae and oomycetes. J Bioenerg Biomembr. 37: 269-278.

Barrett WC, DeGnore JP, Konig S, Fales HM, Keng YF, *et al.* 1999. Regulation of PTP1B via glutathionylation of the active site cysteine 215. Biochemistry. 38: 6699-6705.

Battino R, Rettich TR and Tominaga T. 1983. The solubility of oxygen and ozone in liquids. J Phys Chem Ref Data. 12: 163-178.

Battistuzzi FU and Brown A. 2015. Rates of evolution under extreme and mesophilic conditions. In: Bakermans C, editor. *Microbial Evolution under Extreme Conditions*. Berlin: Walter de Gruyter. p. 247-268.

Baumgartner M, Eberhardt S, De Jonckheere JF and Stetter KO. 2009. *Tetramitus thermacidophilus* n. sp., an amoeboflagellate from acidic hot springs. J Eukaryot Microbiol. 56: 201-206.

Bayer T, Aranda M, Sunagawa S, Yum LK, DeSalvo MK, *et al.* 2012. *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. PLoS One. 7: e35269e35269.

Beardsley C, Pernthaler J, Wosniok W and Amann R. 2003. Are readily culturable bacteria in coastal North Sea waters suppressed by selective grazing mortality? Appl Environ Microbiol. 69: 2624-2630.

Bécuwe P, Gratepanche S, Fourmaux MN, VanBeeumen J, Samyn B, *et al.* 1996. Characterization of iron-dependent endogenous superoxide dismutase of *Plasmodium falciparum*. Mol Biochem Parasitol. 76: 125-134.

Behnke MS, Wootton JC, Lehmann MM, Radke JB, Lucas O, *et al.* 2010. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of *Toxoplasma gondii*. Plos One. 5: 20e12354.

Belbahri L, Calmin G, Mauch F and Andersson JO. 2008. Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. Gene. 408: 1-8.

Bellamacina CR. 1996. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. FASEB J. 10: 1257-1269.

Beney L and Gervais P. 2001. Influence of the fluidity of the membrane on the response of microorganisms to environmental stresses. Appl Microbiol Biotechnol. 57: 34-42.

Beranek A, Rechberger G, Knauer H, Wolinski H, Kohlwein SD, *et al.* 2009. Identification of a cardiolipin-specific phospholipase encoded by the gene *CLD1* (YGR110W) in yeast. J Biol Chem. 284: 11572-11578.

Bernard T, Jebbar M, Rassouli Y, Himdikabbab S, Hamelin J, *et al.* 1993. Ectoine accumulation and osmotic regulation in *Brevibacterium linens*. J Gen Microbiol. 139: 129-136.

Bernhard JM, Kormas K, Pachiadaki MG, Rocke E, Beaudoin DJ, *et al.* 2014. Benthic protists and fungi of Mediterranean deep hypsersaline anoxic basin redoxcline sediments. Front Microbiol. 5.

Betsche T, Schaller D and Melkonian M. 1992. Identification and characterization of glycolate oxidase and related enzymes from the endocyanotic alga *Cyanophora paradoxa* and from pea leaves. Plant Physiol. 98: 887-893.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, *et al.* 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42: W252-W258.

Blackman LM, Mitchell HJ and Hardham AR. 2005. Characterisation of manganese superoxide dismutase from *Phytophthora nicotianae*. Mycol Res. 109: 1171-1183.

Bladergroen BA and van Golde LMG. 1997. CTP: phosphoethanolamine cytidylyltransferase. BBA-Lipid Lipid Met. 1348: 91-99.

Bleijerveld OB, Klein W, Vaandrager AB, Helms JB and Houweling M. 2004. Control of the CDPethanolamine pathway in mammalian cells: effect of CTP : phosphoethanolamine cytidylyltransferase overexpression and the amount of intracellular diacylglycerol. Biochem J. 379: 711-719.

Blumwald E, Aharon GS and Apse MP. 2000. Sodium transport in plant cells. BBA-Biomembranes. 1465: 140-151.

Bochud A and Conzelmann A. 2015. The active site of yeast phosphatidylinositol synthase Pis1 is facing the cytosol. BBA-Mol Cell Biol L. 1851: 629-640.

Bolen DW and Baskakov IV. 2001. The osmophobic effect: natural selection of a thermodynamic force in protein folding. J Mol Biol. 310: 955-963.

Bolger AM, Lohse M and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30: 2114-2120.

Bonete MJ, Perez-Pomares F, Diaz S, Ferrer J and Oren A. 2003. Occurrence of two different glutamate dehydrogenase activities in the halophilic bacterium *Salinibacter ruber*. FEMS Microbiol Lett. 226: 181-186.

Borin S, Crotti E, Mapelli F, Tamagnini I, Corselli C*, et al.* 2008. DNA is preserved and maintains transforming potential after contact with brines of the deep anoxic hypersaline lakes of the Eastern Mediterranean Sea. Saline Syst. 4: 10.

Borowitzka LJ and Brown AD. 1974. Salt relations of marine and halophilic species of unicellular green alga *Dunaliella*: role of glycerol as a compatible solute. Arch Microbiol. 96: 37-52.

Breakspear A, Pasquali M, Broz K, Dong Y and Kistler HC. 2011. *Npc1* is involved in sterol trafficking in the filamentous fungus *Fusarium graminearum*. Fungal Genet Biol. 48: 725-730.

Brennan TV, Anderson JW, Jia ZC, Waygood EB and Clarke S. 1994. Repair of spontaneously deamidated HPr phosphocarrier protein catalyzed by the L-isoaspartate-(D-aspartate) O-methyltransferase. J Biol Chem. 269: 24586-24595.

Bresciani G, da Cruz IBM and Gonzalez-Gallego J. 2015. Manganese superoxide dismutase and oxidative stress modulation. Adv Clin Chem. 68: 87-130.

Bresell A, Weinander R, Lundqvist G, Raza H, Shimoji M, *et al.* 2005. Bioinformatic and enzymatic characterization of the MAPEG superfamily. FEBS J. 272: 1688-1703.

Brock BJ, Rieble S and Gold MH. 1995. Purification and characterization of a 1,4-benzoquinone reductase from the basidiomycete *Phanerochaete chrysosporium*. Appl Environ Microbiol. 61: 3076-3081.

Brown MW, Kolisko M, Silberman JD and Roger AJ. 2012. Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. Curr Biol. 22: 1123-1127.

Brown MW, Spiegel FW and Silberman JD. 2009. Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. Mol Biol Evol. 26: 2699-2709.

Buckner FS, Nguyen LN, Joubert BM and Matsuda SPT. 2000. Cloning and heterologous expression of the *Trypanosoma brucei* lanosterol synthase gene. Mol Biochem Parasitol. 110: 399-403.

Burki F, Okamoto N, Pombert JF and Keeling PJ. 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc R Soc Lond [Biol]. 279: 2246-2254.

Burnstock G and Knight GE. 2004. Cellular distribution and functions of P2 receptor subtypes in different systems. In: Jeon KW, editor. *International Review of Cytology - A Survey of Cell Biology*. p. 31-304.

Bursy J, Kuhlmann AU, Pittelkow M, Hartmann H, Jebbar M, *et al.* 2008. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. Appl Environ Microbiol. 74: 7286-7296.

Bursy J, Pierik AJ, Pica N and Bremer E. 2007. Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. J Biol Chem. 282: 31147-31155.

Buskey EJ, Liu HB, Collumb C and Bersano JGF. 2001. The decline and recovery of a persistent Texas brown tide algal bloom in the Laguna Madre (Texas, USA). Estuaries. 24: 337-346.

Buskey EJ, Wysor B and Hyatt C. 1998. The role of hypersalinity in the persistence of the Texas 'brown tide' in the Laguna Madre. J Plankton Res. 20: 1553-1565.

Butinar L, Santos S, Spencer-Martins I, Oren A and Gunde-Cimerman N. 2005. Yeast diversity in hypersaline habitats. FEMS Microbiol Lett. 244: 229-234.

Butinar L, Sonjak S, Zalar P, Plemenitaš A and Gunde-Cimerman N. 2005. Melanized halophilic fungi are eukaryotic members of microbial communities in hypersaline waters of solar salterns. Bot Mar. 48: 73-79.

Butler G. 2013. Hypoxia and gene expression in eukaryotic microbes. Annu Rev Microbiol. 67: 291-312.

Butschinsky P. 1897. Die Protozooen-Fauna der Salzsee-Limane bei Odessa. Zool Anz. 20: 194-197.

Cai X. 2012. P2X receptor homologs in basal fungi. Purinergic Signal. 8: 11-13.

Calderón MI, Vargas C, Rojo F, Iglesias-Guerra F, Csonka LN, *et al.* 2004. Complex regulation of the synthesis of the compatible solute ectoine in the halophilic bacterium *Chromohalobacter salexigens* DSM 3043. Microbiol-SGM. 150: 3051-3063.

Calloni G, Chen T, Schermann SM, Chang H-c, Genevaux P, *et al.* 2012. DnaK functions as a central hub in the *E. coli* chaperone network. Cell Rep. 1: 251-264.

Cánovas D, Borges N, Vargas C, Ventosa A, Nieto JJ, *et al.* 1999. Role of *N* gamma-acetyldiaminobutyrate as an enzyme stabilizer and an intermediate in the biosynthesis of hydroxyectoine. Appl Environ Microbiol. 65: 3774-3779.

Cantrell SA, Casillas-Martínez L and Molina M. 2006. Characterization of fungi from hypersaline environments of solar salterns using morphological and molecular techniques. Mycol Res. 110: 962-970.

Cao GQ, Goldstein JL and Brown MS. 1996. Complementation of mutation in acyl-CoA:cholesterol acyltransferase (ACAT) fails to restore sterol regulation in ACAT-defective sterol-resistant hamster cells. J Biol Chem. 271: 14642-14648.

Cao H. 2011. Structure-function analysis of diacylglycerol acyltransferase sequences from 70 organisms. BMC research notes. 4: 249-249.

Casalone E, Allocati N, Ceccarelli I, Masulli M, Rossjohn J, *et al.* 1998. Site-directed mutagenesis of the *Proteus mirabilis* glutathione transferase B1-1 G-site. FEBS Lett. 423: 122-124.

Casamayor EO, Massana R, Benlloch S, Øvreås L, Díez B, *et al.* 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ Microbiol. 4: 338-348.

Casillas-Martinez L, Gonzalez ML, Fuentes-Figueroa Z, Castro CM, Nieves-Mendez D, *et al.* 2005. Community structure, geochemical characteristics and mineralogy of a hypersaline microbial mat, Cabo Rojo, PR. Geomicrobiol J. 22: 269-281.

Chaudhari VR, Vyawahare A, Bhattacharjee SK and Rao BJ. 2015. Enhanced excision repair and lack of PSII activity contribute to higher UV survival of *Chlamydomonas reinhardtii* cells in dark. Plant Physiol Biochem. 88: 60-69.

Chen H and Jiang J-G. 2009a. Osmotic responses of *Dunaliella* to the changes of salinity. J Cell Physiol. 219: 251-258.

Chen H and Jiang JG. 2009b. Osmotic responses of *Dunaliella* to the changes of salinity. J Cell Physiol. 219: 251-258.

Chen HL and Zhou HX. 2005. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. Nucleic Acids Res. 33: 3193-3199.

Chen J, Xiao S, Deng Y, Du X and Ye Z. 2011. Cloning of a novel glutathione S-transferase 3 (GST3) gene and expression analysis in pearl oyster, *Pinctada martensii*. Fish Shellfish Immunol. 31: 823-830.

Chen JW, Dodia C, Feinstein SI, Jain MK and Fisher AB. 2000. 1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities. J Biol Chem. 275: 28421-28427.

Chen Q, Steinhauer L, Hammerlindl J, Keller W and Zou J. 2007. Biosynthesis of phytosterol esters: identification of a sterol O-acyltransferase in *Arabidopsis*. Plant Physiol. 145: 974-984.

Chen S, He Q and Greenberg ML. 2008. Loss of tafazzin in yeast leads to increased oxidative stress during respiratory growth. Mol Microbiol. 68: 1061-1072.

Chen S, Seiler J, Santiago-Reichelt M, Felbe K, Grummt I, *et al.* 2013. Repression of RNA polymerase I upon stress is caused by inhibition of RNA-dependent deacetylation of PAF53 by SIRT7. Mol Cell. 52: 303-313.

Chen YH, Yang XY, He K, Liu MH, Li JG, *et al.* 2006. The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol. 60: 107-124.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, *et al.* 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14: 1147-1159.

Cho BC, Park JS, Xu KD and Choi JK. 2008. Morphology and molecular phylogeny of *Trimyema koreanum* n. sp., a ciliate from the hypersaline water of a solar saltern. J Eukaryot Microbiol. 55: 417-426.

Choi HJ, Kang SW, Yang CH, Rhee SG and Ryu SE. 1998. Crystal structure of a novel human peroxidase enzyme at 2.0 angstrom resolution. Nat Struct Biol. 5: 400-406.

Chowdhury I, Mo Y, Gao L, Kazi A, Fisher AB, *et al.* 2009. Oxidant stress stimulates expression of the human peroxiredoxin 6 gene by a transcriptional mechanism involving an antioxidant response element. Free Radical Biol Med. 46: 146-153.

Chun CD, Liu OW and Madhani HD. 2007. A link between virulence and homeostatic responses to hypoxia during infection by the human fungal pathogen *Cryptococcus neoformans*. PLoS Path. 3: 225-238e22.

Cita MB. 2006. Exhumation of Messinian evaporites in the deep-sea and creation of deep anoxic brine-filled collapsed basins. Sediment Geol. 188: 357-378.

Claros MG and Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem. 241: 779-786.

Clavero E, Hernández-Mariné M, Grimalt JO and Garcia-Pichel F. 2000. Salinity tolerance of diatoms from thalassic hypersaline environments. J Phycol. 36: 1021-1034.

Clerico EM, Tilitsky JM, Meng W and Gierasch LM. 2015. How Hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. J Mol Biol. 427: 1575-1588.

Cock JM, Sterck L, Rouze P, Scornet D, Allen AE, *et al.* 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature. 465: 617-621.

Cohen R, Suzuki MR and Hammel KE. 2004. Differential stress-induced regulation of two quinone reductases in the brown rot basidiomycete *Gloeophyllum trabeum*. Appl Environ Microbiol. 70: 324-331.

Cohen Y, Krumbein WE and Shilo M. 1977. Solar Lake (Sinai). 3. Bacterial distribution and production. Limnol Oceanogr. 22: 621-634.

Collet JF, Stroobant V, Pirard M, Delpierre G and Van Schaftingen E. 1998. A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. J Biol Chem. 273: 14107-14112.

Collinson EJ and Grant CM. 2003. Role of yeast glutaredoxins as glutathione S-transferases. J Biol Chem. 278: 22492-22497.

Collinson EJ, Wheeler GL, Garrido EO, Avery AM, Avery SV, *et al.* 2002. The yeast glutaredoxins are active as glutathione peroxidases. J Biol Chem. 277: 16712-16717.

Conibear E and Stevens TH. 1998. Multiple sorting pathways between the late Golgi and the vacuole in yeast. Biochim Biophys Acta. 140: 211-230.

Conti M and Beavo J. 2007. Biochemistry and physiology of cyclic nucleotide phosphocliesterases: essential components in cyclic nucleotide signaling. Annu Rev Biochem. 76: 481-511.

Cooper DMF, Schell MJ, Thorn P and Irvine RF. 1998. Regulation of adenylyl cyclase by membrane potential. J Biol Chem. 273: 27703-27707.

Corey EJ, Matsuda SPT and Bartel B. 1994. Molecular cloning, characterization, and overexpression of *ERG7*, the *Saccharomyces cerevisiae* gene encoding lanosterol synthase. Proc Natl Acad Sci USA. 91: 2211-2215.

Coronado MJ, Vargas C, Mellado E, Tegos G, Drainas C, *et al.* 2000. The alpha-amylase gene *amyH* of the moderate halophile *Halomonas meridiana*: cloning and molecular characterization. Microbiology. 146: 861-868.

Cotelle N, Moreau S, Bernier JL, Catteau JP and Henichart JP. 1991. Antioxidant properties of natural hydroquinones from the marine colonial tunicate *Aplidium californicum*. Free Radical Biol Med. 11: 63-68.

Cowan AK, Rose PD and Horne LG. 1992. *Dunaliella salina*: a model system for studying the response of plant cells to stress. J Exp Bot. 43: 1535-1547.

Criscuolo A and Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10: 210.

Daffonchio D, Borin S, Brusa T, Brusetti L, van der Wielen P, *et al.* 2006. Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. Nature. 440: 203-207.

Danovaro R, Corinaldesi C, Dell'Anno A, Fabiano M and Corselli C. 2005. Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea. Environ Microbiol. 7: 586-592.

Daum G. 1985. Lipids of mitochondria. Biochim Biophys Acta. 822: 1-42.

Davis JS. 2000. Structure, function, and management of the biological system for seasonal solar saltworks. Global Nest: the Int J. 2: 217-226.

De Jonckheere JF, Baumgartner M, Eberhardt S, Opperdoes FR and Stetter KO. 2011. *Oramoeba fumarolia* gen. nov., sp nov., a new marine heterolobosean amoeboflagellate growing at 54 °C. Eur J Protistol. 47: 16-23.

De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, *et al.* 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. Proc Natl Acad Sci USA. 105: 8393-8398.

Debenay JP, Geslin E, Eichler BB, Duleba W, Sylvestre F, *et al.* 2001. Foraminiferal assemblages in a hypersaline lagoon, Araruama (RJ) Brazil. J Foraminifer Res. 31: 133-151.

del Río LA, Corpas FJ, Sandalio LM, Palma JM, Gómez M, *et al.* 2002. Reactive oxygen species, antioxidant systems and nitric oxide in peroxisomes. J Exp Bot. 53: 1255-1272.

Delpierre G and Van Schaftingen E. 2003. Fructosamine 3-kinase, an enzyme involved in protein deglycation. Biochem Soc Trans. 31: 1354-1357.

Demel RA and De Kruyff B. 1976. The function of sterols in membranes. Biochim Biophys Acta. 457: 109-132.

Deole R, Challacombe J, Raiford DW and Hoff WD. 2013. An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. J Biol Chem. 288: 581-588.

Deponte M. 2013. Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. BBA-Gen Subjects. 1830: 3217-3266.

Desmond E and Gribaldo S. 2009. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. Genome Biol Evol. 1: 364-381.

Dey S and Vlot AC. 2015. Ethylene responsive factors in the orchestration of stress responses in monocotyledonous plants. Front Plant Sci. 6: 640.

Diamant S, Ben-Zvi AP, Bukau B and Goloubinoff P. 2000. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. J Biol Chem. 275: 21107-21113.

Diray-Arce J, Clement M, Gul B, Khan MA and Nielsen BL. 2015. Transcriptome assembly, profiling and differential gene expression analysis of the halophyte *Suaeda fruticosa* provides insights into salt tolerance. BMC Genomics. 16: 353.

Divecha N and Irvine RF. 1995. Phospholipid signaling. Cell. 80: 269-278.

Dolapsakis NP, Tafas T, Abatzopoulos TJ, Ziller S and Economou-Amilli A. 2005. Abundance and growth response of microalgae at Megalon Embolon solar saltworks in northern Greece: an aquaculture prospect. J Appl Phycol. 17: 39-49.

Eddy SR. 1998. Profile hidden Markov models. Bioinformatics. 14: 755-763.

Edgcomb V, Orsi W, Leslin C, Epstein SS, Bunge J, *et al.* 2009. Protistan community patterns within the brine and halocline of deep hypersaline anoxic basins in the eastern Mediterranean Sea. Extremophiles. 13: 151-167.

Edgcomb VP and Bernhard JM. 2013. Heterotrophic protists in hypersaline microbial mats and deep hypersaline basin water columns. Life. 3: 346-362.

Edgcomb VP, Orsi W, Breiner HW, Stock A, Filker S, *et al.* 2011. Novel active kinetoplastids associated with hypersaline anoxic basins in the Eastern Mediterranean deep-sea. Deep-Sea Res Pt I. 58: 1040-1048.

Ehrnsperger M, Gräber S, Gaestel M and Buchner J. 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. EMBO J. 16: 221-229.

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, *et al.* 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature. 435: 43-57.

Elcock AH and McCammon JA. 1998. Electrostatic contributions to the stability of halophilic proteins. J Mol Biol. 280: 731-748.

Elevi Bardavid R and Oren A. 2012. The amino acid composition of proteins from anaerobic halophilic bacteria of the order Halanaerobiales. Extremophiles. 16: 567-572.

Elloumi J, Carrias JF, Ayadi H, Sime-Ngando T and Bouaïn A. 2009. Communities structure of the planktonic halophiles in the solar saltern of Sfax, Tunisia. Estuar Coast Shelf Sci. 81: 19-26.

Elloumi J, Carrias JF, Ayadi H, Sime-Ngando T, Boukhris M, *et al.* 2006. Composition and distribution of planktonic ciliates from ponds of different salinity in the solar saltwork of Sfax, Tunisia. Estuar Coast Shelf Sci. 67: 21-29.

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: 1005-1016.

Entz GS. 1904. Die Fauna der kontinentalen kochsalzwässer. Math Naturwiss Ber Ungarn. 19: 89-124.

Enz G. 1879. Über einige Infusorien des Salzteiches zu Szamosfalva. Termes. Füzetek. 3: 1-40.

EPSAG, List of Media and Recipes, available from http://www.uni-goettingen.de/en/list-of-media-and-recipes/186449.html, last accessed in 2015.

Esteve I, Martínez-Alonso M, Mir J and Guerrero R. 1992. Distribution, typology and structure of microbial mat communities in Spain: a preliminary study. Limnetica. 8: 185-195.

Estrada M, Peter H, Gasol MJ, Casamayor OE and Pedrós-Alió C. 2004. Diversity of planktonic photoautotrophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. FEMS Microbiol Ecol. 49: 281–293.

Fan F, Lorenzen JA and Plapp BV. 1991. An aspartate residue in yeast alcohol dehydrogenase I determines the specificity for coenzyme. Biochemistry. 30: 6397-6401.

Farías ME, Contreras M, Rasuk MC, Kurth D, Flores MR, *et al.* 2014. Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. Extremophiles. 18: 311-329.

Favaloro B, Tamburro A, Trofino MA, Bologna L, Rotilio D, *et al.* 2000. Modulation of the glutathione S-transferase in *Ochrobactrum anthropi*: function of xenobiotic substrates and other forms of stress. Biochem J. 346: 553-559.

Feazel LM, Spear JR, Berger AB, Harris JK, Frank DN, *et al.* 2008. Eucaryotic diversity in a hypersaline microbial mat. Appl Environ Microbiol. 74: 329-332.

Federici L, Masulli M, Bonivento D, Di Matteo A, Gianni S, *et al.* 2007. Role of Ser[11] in the stabilization of the structure of *Ochrobactrum anthropi* glutathione transferase. Biochem J. 403: 267-274.

Federici L, Masulli M, Gianni S, Di Ilio C and Allocati N. 2009. A conserved hydrogen-bond network stabilizes the structure of Beta class glutathione S-transferases. Biochem Biophys Res Commun. 382: 525-529.

Feige JN and Auwerx J. 2007. Transcriptional coregulators in the control of energy homeostasis. Trends Cell Biol. 17: 292-301.

Feige JN and Auwerx J. 2008. Transcriptional targets of sirtuins in the coordination of mammalian physiology. Curr Opin Cell Biol. 20: 303-309.

Feldman MJ, Poirier BC and Lange BM. 2015. Misexpression of the Niemann-Pick disease type C1 (NPC1)-like protein in *Arabidopsis* causes sphingolipid accumulation and reproductive defects. Planta. 242: 921-933.

Ferrer M, Werner J, Chernikova TN, Bargiela R, Fernández L, *et al.* 2012. Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study. Environ Microbiol. 14: 268-281.

Filker S, Gimmler A, Dunthorn M, Mahé F and Stoeck T. 2015. Deep sequencing uncovers protistan plankton diversity in the Portuguese Ria Formosa solar saltern ponds. Extremophiles. 19: 283-295.

Filker S, Stock A, Breiner HW, Edgcomb V, Orsi W, *et al.* 2013. Environmental selection of protistan plankton communities in hypersaline anoxic deep-sea basins, Eastern Mediterranean Sea. Microbiologyopen. 2: 54-63.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, *et al.* 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44: D279-85.

Fiore-Donno AM, Meyer M, Baldauf SL and Pawlowski J. 2008. Evolution of dark-spored Myxomycetes (slime-molds): molecules versus morphology. Mol Phylogen Evol. 46: 878-889.

Fisher AB. 2011. Peroxiredoxin 6: a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities. Antioxid Redox Signal. 15: 831-844.

Florentin R. 1899. Études sur la faune des mares salées de Lorraine. Ann Sci Nat (Zool). 10: 209-350.

Foissner W, Jung J-H, Filker S, Rudolph J and Stoeck T. 2014. Morphology, ontogenesis and molecular phylogeny of *Platynematum salinarum* nov. spec., a new scuticociliate (Ciliophora, Scuticociliatia) from a solar saltern. Eur J Protistol. 50: 174-184.

Ford E, Voit R, Liszt G, Magin C, GrumMt I, *et al.* 2006. Mammalian Sir2 homolog SIRT7 is an activator of RNA polymerase I transcription. Genes Dev. 20: 1075-1080.

Fountain SJ, Cao L, Young MT and North RA. 2008. Permeation properties of a P2X receptor in the green algae *Ostreococcus tauri*. J Biol Chem. 283: 15122-15126.

Fountain SJ, Parkinson K, Young MT, Cao L, Thompson CRL, *et al.* 2007. An intracellular P2X receptor required for osmoregulation in *Dictyostelium discoideum*. Nature. 448: 200-203.

Fréalle E, Noël C, Viscogliosi E, Camus D, Dei-Cas E, *et al.* 2005. Manganese superoxide dismutase in pathogenic fungi: an issue with pathophysiological and phylogenetic involvements. FEMS Immunol Med Microbiol. 45: 411-422.

Fridovich I. 1995. Superoxide radical and superoxide dismutases. Annu Rev Biochem. 64: 97-112.

Frolow F, Harel M, Sussman JL, Mevarech M and Shoham M. 1996. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. Nat Struct Biol. 3: 452-458.

Frye RA. 2000. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. Biochem Biophys Res Commun. 273: 793-798.

Fukada K, Inoue T and Shiraishi H. 2006. A posttranslationally regulated protease, VheA, is involved in the liberation of juveniles from parental spheroids in *Volvox carteri*. Plant Cell. 18: 2554-2566.

Galinski EA. 1995. Osmoadaptation in bacteria. Adv Microb Physiol. 37: 272-328.

Gallogly MM and Mieyal JJ. 2007. Mechanisms of reversible protein glutathionylation in redox signaling and oxidative stress. Curr Opin Pharm. 7: 381-391.

Galotti A, Finlay BJ, Jiménez-Gómez F, Guerrero F and Esteban GF. 2014. Most ciliated protozoa in extreme environments are cryptic in the 'seed bank'. Aquat Microb Ecol. 72: 187-193.

Gao R and Stock AM. 2009. Biological insights from structures of two-component proteins. Annu Rev Microbiol. 63: 133-154.

García CM and Niell FX. 1993. Seasonal change in a saline temporary lake (Fuente-de-Piedra, southern Spain). Hydrobiologia. 267: 211-223.

García-Estepa R, Argandoña M, Reina-Bueno M, Capote N, Iglesias-Guerra F, *et al.* 2006. The *ectD* gene, which is involved in the synthesis of the compatible solute hydroxyectoine, is essential for thermoprotection of the halophilic bacterium *Chromohalobacter salexigens*. J Bacteriol. 188: 3774-3784.

Garcia-Perez A and Burg MB. 1991. Role of organic osmolytes in adaptation of renal cells to high osmolarity. J Membr Biol. 119: 1-13.

Garcia-Pichel F, Nübel U and Muyzer G. 1998. The phylogeny of unicellular, extremely halotolerant cyanobacteria. Arch Microbiol. 169: 469-482.

Gaynor EC, Heesen ST, Graham TR, Aebi M and Emr SD. 1994. Signal-mediated retrieval of a membrane protein from the Golgi to the ER in yeast. J Cell Biol. 127: 653-665.

Geiger SR, Lorenzen K, Schreieck A, Hanecker P, Kostrewa D, *et al.* 2010. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Mol Cell. 39: 583-594.

Gibellini F, Hunter WN and Smith TK. 2009. The ethanolamine branch of the Kennedy pathway is essential in the bloodstream form of *Trypanosoma brucei*. Mol Microbiol. 73: 826-843.

Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R and Davies GJ. 2002. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem Biol. 9: 1337-1346.

Giese AC. 1973. Ecology and Culture. In. *Blepharisma: The Biology of a Light-Sensitive Protozoan*. Stanford: Stanford University Press. p. 94-122.

Ginger ML, Fritz-Laylin LK, Fulton C, Cande WZ and Dawson SC. 2010. Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes. Protist. 161: 642-671.

Ginzburg M, Weizinger G, Cohen M and Ginzburg BZ. 1990. The adaptation of *Dunaliella* to widely differing salt concentrations. J Exp Bot. 41: 685-692.

Godzina SM, Lovato MA, Meyer MM, Foster KA, Wilson WK, *et al.* 2000. Cloning and characterization of the *Dictyostelium discoideum* cycloartenol synthase cDNA. Lipids. 35: 249-255.

Goloubinoff P, Mogk A, Ben Zvi AP, Tomoyasu T and Bukau B. 1999. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc Natl Acad Sci USA. 96: 13732-13737.

Gonzalvez F, D'Aurelio M, Boutant M, Moustapha A, Puech J-P, *et al.* 2013. Barth syndrome: Cellular compensation of mitochondrial dysfunction and apoptosis inhibition due to changes in cardiolipin remodeling linked to tafazzin (*TAZ*) gene mutation. BBA-Mol Basis Dis. 1832: 1194-1206.

Gorjan A and Plemenitaš A. 2006. Identification and characterization of ENA ATPases *HwENA1* and *HwENA2* from the halophilic black yeast *Hortaea werneckii*. FEMS Microbiol Lett. 265: 41-50.

Gossett DR, Banks SW, Millhollon EP and Lucas MC. 1996. Antioxidant response to NaCl stress in a control and an NaCl-tolerant cotton cell line grown in the presence of paraquat, buthionine sulfoximine, and exogenous glutathione. Plant Physiol. 112: 803-809.

Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem. 267: 83-90.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29: 644-52.

Grant CM, Quinn KA and Dawes IW. 1999. Differential protein S-thiolation of glyceraldehyde-3-phosphate dehydrogenase isoenzymes influences sensitivity to oxidative stress. Mol Cell Biol. 19: 2650-2656.

Greenberg ML and Lopes JM. 1996. Genetic regulation of phospholipid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev. 60: 1-&.

Gu ZM, Valianpour F, Chen SL, Vaz FM, Hakkaart GA, *et al.* 2004. Aberrant cardiolipin metabolism in the yeast *taz1* mutant: a model for Barth syndrome. Mol Microbiol. 51: 149-158.

Gueta-Dahan Y, Yaniv Z, Zilinskas BA and Ben-Hayyim G. 1997. Salt and oxidative stress: similar and specific responses and their relation to salt tolerance in *Citrus*. Planta. 203: 460-469.

Guixa-Boixareu N, Calderón-Paz JI, Heldal M, Bratbak G and Pedrós-Alió C. 1996. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat Microb Ecol. 11: 215-227.

Gunde-Cimerman N, Ramos J and Plemenitaš A. 2009. Halotolerant and halophilic fungi. Mycol Res. 113: 1231-1241.

Guo ZM, Cromley D, Billheimer JT and Sturley SL. 2001. Identification of potential substrate-binding sites in yeast and human acyl-CoA sterol acyltransferases by mutagenesis of conserved sequences. J Lipid Res. 42: 1282-1291.

Guo ZY, Lin S, Heinen JA, Chang CCY and Chang TY. 2005. The active site His-460 of human acyl-coenzyme A : cholesterol acyltransferase 1 resides in a hitherto undisclosed transmembrane domain. J Biol Chem. 280: 37814-37826.

Haft DH, Selengut JD and White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res. 31: 371-373.

Haigis MC and Sinclair DA. 2010. Mammalian sirtuins: biological insights and disease relevance. Annu Rev Pathol-Mech. 5: 253-295.

Harding T, Brown MW, Plotnikov A, Selivanova E, Park JS, *et al.* 2013. Amoeba stages in the deepest branching heteroloboseans, including *Pharyngomonas*: evolutionary and systematic implications. Protist. 164: 272-286.

Harding T, Brown MW, Simpson AGB and Roger AJ. 2016. Osmoadaptive strategy and its molecular signature in obligately halophilic heterotrophic protists. Genome Biol Evol. 8: 2241-2258.

Harishchandra RK, Wulff S, Lentzen G, Neuhaus T and Galla H-J. 2010. The effect of compatible solute ectoines on the structural organization of lipid monolayer and bilayer membranes. Biophys Chem. 150: 37-46.

Harlos K and Eibl H. 1981. Hexagonal phases in phospholipids with saturated chains: phosphatidylethanolamines and phosphatidic acids. Biochemistry. 20: 2888-2892.

Hart EA, Hua L, Darr LB, Wilson WK, Pang JH, *et al.* 1999. Directed evolution to investigate steric control of enzymatic oxidosqualene cyclization: an isoleucine-to-valine mutation in cycloartenol synthase allows lanosterol and parkeol biosynthesis. J Am Chem Soc. 121: 9887-9888.

Hauer G and Rogerson A. 2005. Heterotrophic protozoa from hypersaline environments. In: Gunde-Cimerman N, Oren A and Plemenitaš A, editors. *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*. Dordrecht: Springer. p. 519-539.

Hayes CS and Setlow P. 1997. Analysis of deamidation of small, acid-soluble spore proteins from *Bacillus subtilis* in vitro and in vivo. J Bacteriol. 179: 6020-6027.

Heidelberg KB, Nelson WC, Holm JB, Eisenkolb N, Andrade K, *et al.* 2013. Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia. Front Microbiol. 4: 115.

Hellmer J, Teubner A and Zeilinger C. 2003. Conserved arginine and aspartate residues are critical for function of MjNhaP1, a Na+/H+ antiporter of *M. jannaschii*. FEBS Lett. 542: 32-36.

Hernández JA, Jimenez A, Mullineaux P and Sevilla F. 2000. Tolerance of pea (*Pisum sativum* L.) to long-term salt stress is associated with induction of antioxidant defences. Plant Cell Environ. 23: 853-862.

Hernández JA, Olmos E, Corpas FJ, Sevilla F and Del Río LA. 1995. Salt-induced oxidative stress in chloroplasts of pea plants. Plant Sci. 105: 151-167.

Herrera JBR, Wilson WK and Matsuda SPT. 2000. A tyrosine-to-threonine mutation converts cycloartenol synthase to an oxidosqualene cyclase that forms lanosterol as its major product. J Am Chem Soc. 122: 6765-6766.

Hillas PJ, del Alba FS, Oyarzabal J, Wilks A and de Montellano PRO. 2000. The AhpC and AhpD antioxidant defense system of *Mycobacterium tuberculosis*. J Biol Chem. 275: 18801-18809.

Hino T, Kawanishi S, Yasui H, Oka S and Sakurai H. 1998. HTHQ (1-O-hexyl-2,3,5-trimethylhydroquinone), an anti-lipid-peroxidative compound: its chemical and biochemical characterizations. BBA-Gen Subjects. 1425: 47-60.

Hodson S, Marshall JJT and Burston SG. 2012. Mapping the road to recovery: The ClpB/Hsp104 molecular chaperone. J Struct Biol. 179: 161-171.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M and Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32: 767-769.

Hohl HR and Raper KB. 1963. Nutrition of cellular slime molds III. specific growth requirements of *Polyspondium pallidum*. J Bacteriol. 86: 1314-&.

Hohmann S. 2002. Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev. 66: 300-372.

Holley AK, Dhar SK, Xu Y and St Clair DK. 2012. Manganese superoxide dismutase: beyond life and death. Amino Acids. 42: 139-158.

Holmgren A. 1979. Glutathione-dependent synthesis of deoxyribonucleotides: purification and characterization of glutaredoxin from *Escherichia coli*. J Biol Chem. 254: 3664-3671.

Höppner A, Widderich N, Lenders M, Bremer E and Smits SHJ. 2014. Crystal structure of the ectoine hydroxylase, a snapshot of the active site. J Biol Chem. 289: 29570-29583.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, *et al.* 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35: W585-W587.

Horvath SE and Daum G. 2013. Lipids of mitochondria. Prog Lipid Res. 52: 590-614.

Hosono K. 1992. Effect of salt stress on lipid composition and membrane fluidity of the salt-tolerant yeast *Zygosaccharomyces rouxii*. J Gen Microbiol. 138: 91-96.

Hughes AL, Todd BL and Espenshade PJ. 2005. SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. Cell. 120: 831-842.

Ingrosso D, Cimmino A, D'Angelo S, Alfinito F, Zappia V, *et al.* 2002. Protein methylation as a marker of aspartate damage in glucose-6-phosphate dehydrogenase-deficient erythrocytes: role of oxidative stress. Eur J Biochem. 269: 2032-2039.

Inohara N and Nunez G. 2002. ML - a conserved domain involved in innate immunity and lipid metabolism. Trends Biochem Sci. 27: 219-221.

Inoue H, Nishida M and Takahashi K. 2000. Effects of Cys10 mutation to Ala in glutathione transferase from *Escherichia coli*. J Organomet Chem. 611: 593-595.

Jaiswal AK. 2000. Regulation of genes encoding NAD(P)H : quinone oxidoreductases. Free Radical Biol Med. 29: 254-262.

Jakobsson PJ, Mancini JA, Riendeau D and FordHutchinson AW. 1997. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. J Biol Chem. 272: 22934-22939.

Jakobsson PJ, Morgenstern R, Mancini J, Ford-Hutchinson A and Persson B. 1999. Common structural features of MAPEG - A widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. Protein Sci. 8: 689-692.

Javor B. 1989. *Hypersaline Environments: Microbiology and Biogeochemistry*. Berlin: Springer.

Javor BJ. 2002. Industrial microbiology of solar salt production. J Ind Microbiol Biotechnol. 28: 42-47.

Jensen KA, Ryan ZC, Wymelenberg AV, Cullen D and Hammel KE. 2002. An NADH : quinone oxidoreductase active during biodegradation by the brown-rot basidiomycete *Gloeophyllum trabeum*. Appl Environ Microbiol. 68: 2699-2703.

Ji Y, Song Y, Kim N, Youn H, Kang M, *et al.* 2014. Cleavage of cytoplasm within the oligonucleate zoosporangia of *Allomyces macrogynus*. Mycologia. 106: 369-378.

Joint I, Henriksen P, Garde K and Riemann B. 2002. Primary production, nutrient assimilation and microzooplankton grazing along a hypersaline gradient. FEMS Microbiol Ecol. 39: 245-257.

Jones E, Oliphant T and Peterson P, SciPy: Open Source Scientific Tools for Python, available from http://www.scipy.org/,

Joshi AS, Zhou J, Gohil VM, Chen S and Greenberg ML. 2009. Cellular functions of cardiolipin in yeast. BBA-Mol Cell Res. 1793: 212-218.

Käll L, Krogh A and Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 338: 1027-1036.

Kampinga HH and Craig EA. 2010. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol. 11: 579-592.

Kanai R, Ogawa H, Vilsen B, Cornelius F and Toyoshima C. 2013. Crystal structure of a Na^+-bound Na^+,K^+-ATPase preceding the E1P state. Nature. 502: 201-206.

Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44: D457-62.

Kanematsu S and Asada K. 1979. Ferric and manganic superoxide dismutases in *Euglena gracilis*. Arch Biochem Biophys. 195: 535-545.

Kapatai G, Large A, Benesch JLP, Robinson CV, Carrascosa JL, *et al.* 2006. All three chaperonin genes in the archaeon *Haloferax volcanii* are individually dispensable. Mol Microbiol. 61: 1583-1597.

Katoh K, Misawa K, Kuma K and Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059-3066.

Katz A, Pick U and Avron M. 1992. Modulation of Na^+/H^+ antiporter activity by extreme pH and salt in the halotolerant alga *Dunaliella salina*. Plant Physiol. 100: 1224-1229.

Kawazu S, Nozaki T, Tsuboi T, Nakano Y, Komaki-Yasuda K, *et al.* 2003. Expression profiles of peroxiredoxin proteins of the rodent malaria parasite *Plasmodium yoelii*. Int J Parasitol. 33: 1455-1461.

Ke H, Wang H and Ye M. 2011. Structural insight into the substrate specificity of phosphodiesterases. In: Francis SH, Conti M and Houslay MD, editors. *Handbook of Experimental Pharmacology*. Heidelberg: Springer. p. 121-134.

Keber R, Motaln H, Wagner KD, Debeljak N, Rassoulzadegan M, *et al.* 2011. Mouse knockout of the cholesterogenic cytochrome P450 lanosterol 14 alpha-Demethylase (Cyp51) resembles Antley-Bixler syndrome. J Biol Chem. 286: 29086-29097.

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, *et al.* 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12: e1001889.

Keeling PJ and Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9: 605-618.

Keppler D. 1999. Export pumps for glutathione S-conjugates. Free Radical Biol Med. 27: 985-991.

Khakh BS. 2001. Molecular physiology of P2X receptors and ATP signalling at synapses. Nature Rev Neurosci. 2: 165-174.

Khan N, Shen JG, Chang TY, Chang CC, Fung PCW, *et al.* 2003. Plasma membrane cholesterol: a possible barrier to intracellular oxygen in normal and mutant CHO cells defective in cholesterol metabolism. Biochemistry. 42: 23-29.

Khaware RK, Koul A and Prasad R. 1995. High membrane fluidity is related to NaCl stress in *Candida membranefaciens*. Biochem Mol Biol Int. 35: 875-880.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, *et al.* 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14: R36.

Kim HS, Manevich Y, Feinstein SI, Pak JH, Ho YS, *et al.* 2003. Induction of 1-Cys peroxiredoxin expression by oxidative stress in lung epithelial cells. Am J Physiol Lung Cell Mol Physiol. 285: L363-L369.

Kim J, Minkler PE, Salomon RG, Anderson VE and Hoppel CL. 2011. Cardiolipin: characterization of distinct oxidized molecular species. J Lipid Res. 52: 125-135.

Kim M, Park S, Polle JE and Jin E. 2010. Gene expression profiling of *Dunaliella* sp. acclimated to different salinities. Phycol Res. 58: 17-28.

Kim TS, Sundaresh CS, Feinstein SI, Dodia C, Skach WR*, et al.* 1997. Identification of a human cDNA clone for lysosomal type Ca^{2+}-independent phospholipase A2 and properties of the expressed protein. J Biol Chem. 272: 2542-2550.

Kirby H. 1932. Two protozoa from brine. T Am Microscop Soc. 51: 8-15.

Kirby WA, Tikhonenkov DV, Mylnikov AP, Janouškovec J, Lax G*, et al.* 2015. Characterization of *Tulamoeba bucina* n. sp., an extremely halotolerant amoeboflagellate heterolobosean belonging to the *Tulamoeba-Pleurostomum* clade (Tulamoebidae n. fam.). J Eukaryot Microbiol. 62: 227-238.

Klages K, Boldingh H and Smith GS. 1999. Accumulation of myo-inositol in *Actinidia* seedlings subjected to salt stress. Ann Bot. 84: 521-527.

Koeffel R, Tiwari R, Falquet L and Schneiter R. 2005. The *Saccharomyces cerevisiae YLL012/YEH1*, *YLR020/YEH2*, and *TGL1* genes encode a novel family of membrane-anchored lipases that are required for steryl ester hydrolysis. Mol Cell Biol. 25: 1655-1668.

Kogej T, Gorbushina AA and Gunde-Cimerman N. 2006. Hypersaline conditions induce changes in cell-wall melanization and colony structure in a halophilic and a xerophilic black yeast species of the genus *Trimmatostroma*. Mycol Res. 110: 713-724.

Kogej T, Ramos J, Plemenitaš A and Gunde-Cimerman N. 2005. The halophilic fungus *Hortaea werneckii* and the halotolerant fungus *Aureobasidium pullulans* maintain low intracellular cation concentrations in hypersaline environments. Appl Environ Microbiol. 71: 6600-6605.

Koshkin A, Nunn CM, Djordjevic S and de Montellano PRO. 2003. The mechanism of *Mycobacterium tuberculosis* alkylhydroperoxidase AhpD as defined by mutagenesis, crystallography, and kinetics. J Biol Chem. 278: 29502-29508.

Koshkin V and Greenberg ML. 2002. Cardiolipin prevents rate-dependent uncoupling and provides osmotic stability in yeast mitochondria. Biochem J. 364: 317-322.

Krogh A, Larsson B, von Heijne G and Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305: 567-580.

Krueger T, Fisher PL, Becker S, Pontasch S, Dove S*, et al.* 2015. Transcriptomic characterization of the enzymatic antioxidants FeSOD, MnSOD, APX and KatG in the dinoflagellate genus *Symbiodinium*. BMC Evol Biol. 15: 48.

Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ*, et al.* 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature. 524: 427-432.

Kunin V, Raes J, Harris JK, Spear JR, Walker JJ, *et al.* 2008. Millimeter‐scale genetic gradients and community‐level molecular convergence in a hypersaline microbial mat. Mol Syst Biol. 4: 198.

Kushiro M, Nakano T, Sato K, Yamagishi K, Asami T, *et al.* 2001. Obtusifoliol 14 alpha-demethylase (CYP51) antisense Arabidopsis shows slow growth and long life. Biochem Biophys Res Commun. 285: 98-104.

Kushner DJ. 1978. Life in high salt and solute concentrations. In: Kushner DJ, editor. *Microbial Life in Extreme Environments*. London: Academic Press. p. 317-368.

Kuwabara PE and Labouesse M. 2002. The sterol-sensing domain: multiple families, a unique role? Trends Genet. 18: 193-201.

Kyte J and Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157: 105-132.

Langmead B, Trapnell C, Pop M and Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25R25.

Lanyi JK. 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriol Rev. 38: 272-290.

Larroy C, Parés X and Biosca JA. 2002. Characterization of a *Saccharomyces cerevisiae* NADP(H)-dependent alcohol dehydrogenase (ADHVII), a member of the cinnamyl alcohol dehydrogenase family. Eur J Biochem. 269: 5738-5745.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 30: 3276-3278.

Laskowski MJ, Dreher KA, Gehring MA, Abel S, Gensler AL*, et al.* 2002. FQR1, a novel primary auxin-response gene, encodes a flavin mononucleotide-binding quinone reductase. Plant Physiol. 128: 578-590.

Law CW, Chen Y, Shi W and Smyth GK. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15: R29.

Laybourn-Parry J, Quayle W and Henshaw T. 2002. The biology and evolution of Antarctic saline lakes in relation to salinity and trophy. Polar Biol. 25: 542-552.

Lee GJ, Roseman AM, Saibil HR and Vierling E. 1997. A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. EMBO J. 16: 659-671.

Lee GJ and Vierling E. 2000. A small heat shock protein cooperates with Heat Shock Protein 70 systems to reactivate a heat-denatured protein. Plant Physiol. 122: 189-197.

Lee KO, Lee JR, Yoo JY, Jang HH, Moon JC*, et al.* 2002. GSH-dependent peroxidase activity of the rice (*Oryza sativa*) glutaredoxin, a thioltransferase. Biochem Biophys Res Commun. 296: 1152-1156.

Lei YL, Xu KD, Choi JK, Hong HP and Wickham SA. 2009. Community structure and seasonal dynamics of planktonic ciliates along salinity gradients. Eur J Protistol. 45: 305-319.

Lenassi M, Gostinčar C, Jackman S, Turk M, Sadowski I*, et al.* 2013. Whole genome duplication and enrichment of metal cation transporters revealed by *de novo* genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. Plos One. 8: e71328.

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI*, et al.* 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 29: 1035-1043.

Lepesheva GI, Zaitseva NG, Nes WD, Zhou WX, Arase M*, et al.* 2006. Cyp51 from *Trypanosoma cruzi* - A phyla-specific residue in the B' helix defines substrate preferences of sterol 14 alpha-demethylase. J Biol Chem. 281: 3577-3585.

Li B and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12: 323.

Li SJ, Hua ZS, Huang LN, Li J, Shi SH*, et al.* 2014. Microbial communities evolve faster in extreme environments. Sci Rep. 4: 6205.

Li X, Wang J, Coutavas E, Shi H, Hao Q*, et al.* 2016. Structure of human Niemann–Pick C1 protein. Proc Natl Acad Sci USA.

Licausi F, Ohme-Takagi M and Perata P. 2013. APETALA/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol. 199: 639-649.

Lige B, Sampels V and Coppens I. 2013. Characterization of a second sterol-esterifying enzyme in *Toxoplasma* highlights the importance of cholesterol storage pathways for the parasite. Mol Microbiol. 87: 951-967.

Lii CK, Chai YC, Zhao W, Thomas JA and Hendrich S. 1994. S-thiolation and irreversible oxidation of sulfhydryls on carbonic andhydrase III during oxidative stess: a method for studying protein modification in intact cells and tissues. Arch Biochem Biophys. 308: 231-239.

Lillig CH, Berndt C and Holmgren A. 2008. Glutaredoxin systems. BBA-Gen Subjects. 1780: 1304-1317.

Lin S, Lu XH, Chang CCY and Chang TY. 2003. Human acyl-coenzyme A : cholesterol acyltransferase expressed in Chinese hamster ovary cells: membrane topology and active site location. Mol Biol Cell. 14: 2447-2460.

Lind C, Gerdes R, Schuppe-Koistinen I and Cotgreave IA. 1998. Studies on the mechanism of oxidative modification of human glyceraldehyde-3-phosphate dehydrogenase by glutathione: catalysis by glutaredoxin. Biochem Biophys Res Commun. 247: 481-486.

Linder JU. 2005. Substrate selection by class III adenylyl cyclases and guanylyl cyclases. IUBMB Life. 57: 797-803.

Lippert K and Galinski EA. 1992. Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. Appl Microbiol Biotechnol. 37: 61-65.

Liu J-P and Chen R. 2015. Stressed SIRT7: facing a crossroad of senescence and immortality. Clin Exp Pharmacol Physiol. 42: 567-569.

Liu JL, Zhang DX and Hong L. 2015. Isolation, characterization and functional annotation of the salt tolerance genes through screening the high-quality cDNA library of the halophytic green alga *Dunaliella salina* (Chlorophyta). Ann Microbiol. 65: 1293-1302.

Lloyd-Evans E and Platt FM. 2010. Lipids on trial: the search for the offending metabolite in Niemann-Pick type C disease. Traffic. 11: 419-428.

Lo C-C, Bonner CA, Xie G, D'Souza M and Jensen RA. 2009. Cohesion group approach for evolutionary analysis of aspartokinase, an enzyme that feeds a branched network of many biochemical pathways. Microbiol Mol Biol Rev. 73: 594-651.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, *et al.* 2000. *Molecular Cell Biology, Fourth Edition*. New York: W. H. Freeman.

Löfke C, Ischebeck T, König S, Freitag S and Heiliviann I. 2008. Alternative metabolic fates of phosphatidylinositol produced by phosphatidylinositol synthase isoforms in *Arabidopsis thaliana*. Biochem J. 413: 115-124.

Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, *et al.* 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature. 433: 865-868.

Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, *et al.* 2009. Infrequent marine-freshwater transitions in the microbial world. Trends Microbiol. 17: 414-422.

Long M, Betrán E, Thornton K and Wang W. 2003. The origin of new genes: glimpses from the young and old. Nat Rev Genet. 4: 865-875.

Love MI, Huber W and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Ludlow MJ, Durai L and Ennion SJ. 2009. Functional characterization of intracellular *Dictyostelium discoideum* P2X receptors. J Biol Chem. 284: 35227-35239.

Luévano-Martínez LA, Forni MF, dos Santos VT, Souza-Pinto NC and Kowaltowski AJ. 2015. Cardiolipin is a key determinant for mtDNA stability and segregation during mitochondrial stress. BBA-Bioenergetics. 1847: 587-598.

Luévano-Martínez LA and Kowaltowski AJ. 2015. Phosphatidylglycerol-derived phospholipids have a universal, domain-crossing role in stress responses. Arch Biochem Biophys. 585: 90-97.

Luikenhuis S, Perrone G, Dawes IW and Grant CM. 1998. The yeast *Saccharomyces cerevisiae* contains two glutaredoxin genes that are required for protection against reactive oxygen species. Mol Biol Cell. 9: 1081-1091.

Lunter G and Goodson M. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21: 936-939.

Lushchak VI. 2011. Environmentally induced oxidative stress in aquatic animals. Aquat Toxicol. 101: 13-30.

Lykidis A. 2007. Comparative genomics and evolution of eukaryotic phospholipid biosynthesis. Prog Lipid Res. 46: 171-199.

Magnani E, Sjolander K and Hake S. 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. Plant Cell. 16: 2265-2277.

Maheshwari S, Lavigne M, Contet A, Alberge B, Pihan E, *et al.* 2013. Biochemical characterization of *Plasmodium falciparum* CTP: phosphoethanolamine cytidylyltransferase shows that only one of the two cytidylyltransferase domains is active. Biochem J. 450: 159-167.

Mailloux RJ, McBride SL and Harper M-E. 2013. Unearthing the secrets of mitochondrial ROS and glutathione in bioenergetics. Trends Biochem Sci. 38: 592-602.

Majee M, Maitra S, Dastidar KG, Pattnaik S, Chatterjee A, *et al.* 2004. A novel salt-tolerant L-myo-inositol-1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka, a halophytic wild rice - Molecular cloning, bacterial overexpression, characterization, and functional introgression into tobacco-conferring salt tolerance phenotype. J Biol Chem. 279: 28539-28552.

Malathi K, Higaki K, Tinkelenberg AH, Balderes DA, Almanzar-Paramio D, *et al.* 2004. Mutagenesis of the putative sterol-sensing domain of yeast Niemann Pick C-related protein reveals a primordial role in subcellular sphingolipid distribution. J Cell Biol. 164: 547-556.

Manevich Y, Feinstein SI and Fisher AB. 2004. Activation of the antioxidant enzyme 1-Cys peroxiredoxin requires glutathionylation mediated by heterodimerization with pi GST. Proc Natl Acad Sci USA. 101: 3780-3785.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, *et al.* 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43: D222-D226.

Marguet E and Forterre P. 1994. DNA stability at temperature typical for hyperthermophiles. Nucleic Acids Res. 22: 1681-1686.

Marguet E and Forterre P. 1998. Protection of DNA by salts against thermodegradation at temperatures typical for hyperthermophiles. Extremophiles. 2: 115-122.

Martin TFJ. 2001. PI(4,5)P-2 regulation of surface membrane traffic. Curr Opin Cell Biol. 13: 493-499.

Martinez Molina D, Eshaghi S and Nordlund P. 2008. Catalysis within the lipid bilayer-structure and mechanism of the MAPEG family of integral membrane proteins. Curr Opin Struct Biol. 18: 442-449.

Máthé I, Borsodi AK, Tóth EM, Felföldi T, Jurecska L, *et al.* 2014. Vertical physico-chemical gradients with distinct microbial communities in the hypersaline and heliothermal Lake Ursu (Sovata, Romania). Extremophiles. 18: 501-514.

Maturrano L, Santos F, Rosselló-Mora R and Antón J. 2006. Microbial diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. Appl Environ Microbiol. 72: 3887-3895.

Matz C, Boenigk J, Arndt H and Jürgens K. 2002. Role of bacterial phenotypic traits in selective feeding of the heterotrophic nanoflagellate *Spumella* sp. Aquat Microb Ecol. 27: 137-148.

Mavromatis K, Ivanova N, Anderson I, Lykidis A, Hooper SD, *et al.* 2009. Genome analysis of the anaerobic thermohalophilic bacterium *Halothermothrix orenii*. Plos One. 4: e4192.

Meister A and Anderson ME. 1983. Glutathione. Annu Rev Biochem. 52: 711-760.

Meyer MM, Xu R and Matsuda SPT. 2002. Directed evolution to generate cycloartenol synthase mutants that produce lanosterol. Org Lett. 4: 1395-1398.

Miedema H and Assmann SM. 1998. The calculation of intracellular ion concentrations and membrane potential from cell-attached and excised patch measurements. Cytosolic K$^+$ concentration and membrane potential in *Vicia faba* guard cells. J Membr Biol. 166: 101-110.

Miller A-F. 2012. Superoxide dismutases: ancient enzymes and new insights. FEBS Lett. 586: 585-595.

Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L*, et al.* 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 14: 193-202.

Min XJ. 2010. Evaluation of computational methods for secreted protein prediction in different eukaryotes. J Proteomics Bioinform. 3: 143-147.

Minz D, Fishbain S, Green SJ, Muyzer G, Cohen Y*, et al.* 1999. Unexpected population distribution in a microbial mat community: sulfate-reducing bacteria localized to the highly oxic chemocline in contrast to a eukaryotic preference for anoxia. Appl Environ Microbiol. 65: 4659-4665.

Miriyala S, Spasojevic I, Tovmasyan A, Salvemini D, Vujaskovic Z*, et al.* 2012. Manganese superoxide dismutase, MnSOD and its mimics. BBA-Mol Basis Dis. 1822: 794-814.

Mirzaei H and Regnier F. 2008. Protein : protein aggregation induced by protein oxidation. J Chromatogr B Analyt Technol Biomed Life Sci. 873: 8-14.

Misner I, Blouin N, Leonard G, Richards TA and Lane CE. 2015. The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. Genome Biol Evol. 7: 120-135.

Mizoi J, Nakamura M and Nishida I. 2006. Defects in CTP: phosphorylethanolamine cytidyltransferase affect embryonic and postembryonic development in *Arabidopsis*. Plant Cell. 18: 3370-3385.

MMETSP, Marine Microbial Eukaryote Transcriptome Sequencing Project Data, available from http://data.imicrobe.us/project/view/104, last accessed in 2015.

Moestrup O and Sengco M. 2001. Ultrastructural studies on *Bigelowiella natans*, gen. et sp. nov., a chlorarachniophyte flagellate. J Phycol. 37: 624-646.

Monaco ME, Feldman M and Kleinberg DL. 1994. Identification of rat liver phosphatidylinositol synthase as a 21 kDa protein. Biochem J. 304: 301-305.

Mongodin EF, Nelson KE, Daugherty S, DeBoy RT, Wister J*, et al.* 2005. The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc Natl Acad Sci USA. 102: 18147-18152.

Morais MC, Zhang WH, Baker AS, Zhang GF, Dunaway-Mariano D, *et al.* 2000. The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. Biochemistry. 39: 10385-10396.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC and Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182-W185.

Morrison AMS, Goldstone JV, Lamb DC, Kubota A, Lemaire B, *et al.* 2014. Identification, modeling and ligand affinity of early deuterostome CYP51s, and functional characterization of recombinant zebrafish sterol 14 alpha-demethylase. BBA-Gen Subjects. 1840: 1825-1836.

Morth JP, Pedersen BP, Buch-Pedersen MJ, Andersen JP, Vilsen B, *et al.* 2011. A structural overview of the plasma membrane Na^+, K^+-ATPase and H^+-ATPase ion pumps. Nat Rev Mol Cell Biol. 12: 60-70.

Mougel C and Zhulin IB. 2001. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci. 26: 582-584.

Mudgett MB and Clarke S. 1994. Hormonal and environmental responsiveness of a developmentally-regulated protein repair L-isoaspartyl methyltransferase in wheat. J Biol Chem. 269: 25605-25612.

Murphy MP. 2012. Mitochondrial thiols in antioxidant protection and redox signaling: distinct roles for glutathionylation and other thiol modifications. Antioxid Redox Signal. 16: 476-495.

Musatov A. 2006. Contribution of peroxidized cardiolipin to inactivation of bovine heart cytochrome c oxidase. Free Radical Biol Med. 41: 238-246.

Mustakhimov II, Reshetnikov AS, Khmelenina VN and Trotsenko YA. 2010. Regulatory aspects of ectoine biosynthesis in halophilic bacteria. Microbiology. 79: 583-592.

Nagano N, Matsui S, Kuramura T, Taoka Y, Honda D, *et al.* 2011. The distribution of extracellular cellulase activity in marine eukaryotes, thraustochytrids. Mar Biotechnol. 13: 133-136.

Nagata S, Adachi K and Sano H. 1996. NMR analyses of compatible solutes in a halotolerant *Brevibacterium* sp. Microbiology. 142: 3355-3362.

Nagata S, Wang Y, Oshima A, Zhang L, Miyake H, *et al.* 2008. Efficient cyclic system to yield ectoine using *Brevibacterium* sp JCM 6894 subjected to osmotic downshock. Biotechnol Bioeng. 99: 941-948.

Namyslowski B. 1913. Über unbekannte halophile Mikroorganismen aus dem Innern des Salzbergwerkes Wieliczka. Bull Int Acad Sci Krakow B. 3/4: 88-104.

Nass R, Cunningham KW and Rao R. 1997. Intracellular sequestration of sodium by a novel Na^+/H^+ exchanger in yeast is enhanced by mutations in the plasma membrane H^+-ATPase: insights into mechanisms of sodium tolerance. J Biol Chem. 272: 26145-26152.

NCMA, Algal Media Recipes, available from https://ncma.bigelow.org/algal-recipes, last accessed in 2014.

Nelson DL and Cox MM. 2005. *Lehninger Principles of Biochemistry (4th ed.)*. New York: W. H. Freeman.

Nes WD, Norton RA, Crumley FG, Madigan SJ and Katz ER. 1990. Sterol phylogenesis and algal evolution. Proc Natl Acad Sci USA. 87: 7565-7569.

Neudegger T, Verghese J, Hayer-Hartl M, Hartl FU and Bracher A. 2016. Structure of human heat-shock transcription factor 1 in complex with DNA. Nat Struct Mol Biol. 23: 140-146.

Nevalainen TJ. 2010. 1-Cysteine peroxiredoxin: a dual-function enzyme with peroxidase and acidic Ca2+-independent phospholipase A(2) activities. Biochimie. 92: 638-644.

Nielsen H, Engelbrecht J, Brunak S and von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10: 1-6.

Niu X, Bressan RA, Hasegawa PM and Pardo JM. 1995. Ion homeostasis in NaCl stress environments. Plant Physiol. 109: 735-742.

Niu XM, Zhu JK, Narasimhan ML, Bressan RA and Hasegawa PM. 1993. Plasma membrane H+-ATPase gene expression is regulated by NaCl in cells of the halophyte *Atriplex nummularia* L. Planta. 190: 433-438.

Nongpiur R, Soni P, Karan R, Singla-Pareek SL and Pareek A. 2012. Histidine kinases in plants: cross talk between hormone and stress responses. Plant Signal Behav. 7: 1230-7.

North RA. 2002. Molecular physiology of P2X receptors. Physiol Rev. 82: 1013-1067.

Nübel U, Garcia-Pichel F, Clavero E and Muyzer G. 2000. Matching molecular diversity and ecophysiology of benthic cyanobacteria and diatoms in communities along a salinity gradient. Environ Microbiol. 2: 217-226.

Nyblom M, Poulsen H, Gourdon P, Reinhard L, Andersson M, *et al.* 2013. Crystal Structure of Na^+, K^+-ATPase in the Na^+-Bound State. Science. 342: 123-127.

Nývltová E, Stairs CW, Hrdý I, Rídl J, Mach J, *et al.* 2015. Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. Mol Biol Evol. 32: 1039-1055.

O'Connor CM. 2006. Protein L-isoaspartyl, D-aspartyl O-methyltransferases: catalysts for protein repair. Enzymes. 24: 385-433.

O'Flaherty C and de Souza AR. 2011. Hydrogen peroxide modifies human sperm peroxiredoxins in a dose-dependent manner. Biol Reprod. 84: 238-247.

Oberstaller J, Pumpalova Y, Schieler A, Llinas M and Kissinger JC. 2014. The *Cryptosporidium parvum* ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems. Nucleic Acids Res. 42: 8271-8284.

Oelkers P, Behari A, Cromley D, Billheimer JT and Sturley SL. 1998. Characterization of two human genes encoding acyl coenzyme A: Cholesterol acyltransferase-related enzymes. J Biol Chem. 273: 26765-26771.

Ogawa H and Toyoshima C. 2002. Homology modeling of the cation binding sites of Na^+K^+-ATPase. Proc Natl Acad Sci USA. 99: 15977-15982.

Oh CS, Toke DA, Mandala S and Martin CE. 1997. *ELO2* and *ELO3*, homologues of the *Saccharomyces cerevisiae ELO1* gene, function in fatty acid elongation and are required for sphingolipid formation. J Biol Chem. 272: 17376-17384.

Oh D-H, Leidi E, Zhang Q, Hwang S-M, Li Y, *et al.* 2009. Loss of halophytism by interference with SOS1 expression. Plant Physiol. 151: 210-222.

Ohgami N, Ko DC, Thomas M, Scott MP, Chang CCY, *et al.* 2004. Binding between the Niemann-Pick C1 protein and a photoactivatable cholesterol analog requires a functional sterol-sensing domain. Proc Natl Acad Sci USA. 101: 12473-12478.

Oliveira L, Paiva ACM and Vriend G. 1999. A low resolution model for the interaction of G proteins with G protein-coupled receptors. Protein Eng. 12: 1087-1095.

Oren A. 1999. Bioenergetic aspects of halophilism. Microbiol Mol Biol Rev. 63: 334-348.

Oren A. 2014. The ecology of *Dunaliella* in high-salt environments. J Biol Res Thessalon. 21: 23.

Oren A. 1994. The ecology of the extremely halophilic archaea. FEMS Microbiol Rev. 13: 415-439.

Oren A. 2002a. *Halophilic Microorganisms and their Environments*. Dordrecht: Kluwer Academic Publishers.

Oren A. 2002b. Intracellular salt concentration and ion metabolism in halophilic microorganisms. In: Seckbach J, editor. *Halophilic Microorganisms and their Environments*. Dordrecht: Kluwyer Academic Publishers. p. 207-231.

Oren A. 2013. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. Front Microbiol. 4: 315.

Oren A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Syst. 4: 2.

Oren A, Heldal M and Norland S. 1997. X-ray microanalysis of intracellular ions in the anaerobic halophilic eubacterium *Haloanaerobium praevalens*. Can J Microbiol. 43: 588-592.

Oren A, Heldal M, Norland S and Galinski EA. 2002. Intracellular ion and organic solute concentrations of the extremely halophilic bacterium *Salinibacter ruber*. Extremophiles. 6: 491-498.

Oren A, Larimer F, Richardson P, Lapidus A and Csonka LN. 2005. How to be moderately halophilic with broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. Extremophiles. 9: 275-279.

Pachiadaki MG, Yakimov MM, LaCono V, Leadbetter E and Edgcomb V. 2014. Unveiling microbial activities along the halocline of Thetis, a deep-sea hypersaline anoxic basin. ISME J. 8: 2478-2489.

Panda AK and Das AB. 2005. Salt tolerance and salinity effects on plants: a review. Ecotoxicol Environ Saf. 60: 324-349.

Panne D, Maniatis T and Harrison SC. 2004. Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. EMBO J. 23: 4384-4393.

Paradies G, Petrosillo G, Pistolese M and Ruggiero FM. 2000. The effect of reactive oxygen species generated from the mitochondrial electron transport chain on the cytochrome c oxidase activity and on the cardiolipin content in bovine heart submitochondrial particles. FEBS Lett. 466: 323-326.

Paradies G, Petrosillo G, Pistolese M and Ruggiero FM. 2002. Reactive oxygen species affect mitochondrial electron transport complex I activity through oxidative cardiolipin damage. Gene. 286: 135-141.

Paradies G, Petrosillo G, Pistolese M and Ruggiero FM. 2001. Reactive oxygen species generated by the mitochondrial respiratory chain affect the complex III activity via cardiolipin peroxidation in beef-heart submitochondrial particles. Mitochondrion. 1: 151-159.

Park JS. 2012. Effects of different ion compositions on growth of obligately halophilic protozoan *Halocafeteria seosinensis*. Extremophiles. 16: 161-164.

Park JS, Cho BC and Simpson AGB. 2006. *Halocafeteria seosinensis* gen. et sp. nov. (Bicosoecida), a halophilic bacterivorous nanoflagellate isolated from a solar saltern. Extremophiles. 10: 493-504.

Park JS, Kim HJ, Choi DH and Cho BC. 2003. Active flagellates grazing on prokaryotes in high salinity waters of a solar saltern. Aquat Microb Ecol. 33: 173-179.

Park JS, Simpson AG, Brown S and Cho BC. 2009. Ultrastructure and molecular phylogeny of two heterolobosean amoebae, *Euplaesiobystra hypersalinica* gen. et sp. nov. and *Tulamoeba peronaphora* gen. et sp. nov., isolated from an extremely hypersaline habitat. Protist. 160: 265-83.

Park JS and Simpson AGB. 2016. Characterization of a deep-branching heterolobosean, *Pharyngomonas turkanaensis* n. sp., isolated from a non-hypersaline habitat, and ultrastructural comparison of cysts and amoebae among *Pharyngomonas* strains. J Eukaryot Microbiol. 63: 100-111.

Park JS and Simpson AGB. 2010. Characterization of halotolerant Bicosoecida and Placididea (Stramenopila) that are distinct from marine forms, and the phylogenetic pattern of salinity preference in heterotrophic stramenopiles. Environ Microbiol. 12: 1173-1184.

Park JS and Simpson AGB. 2011. Characterization of *Pharyngomonas kirbyi* (="Macropharyngomonas halophila" nomen nudum), a very deep-branching, obligately halophilic Heterolobosean. Protist. 162: 691-709.

Park JS and Simpson AGB. 2015. Diversity of heterotrophic protists from extremely hypersaline habitats. Protist. 166: 422-437.

Park JS, Simpson AGB, Lee WJ and Cho BC. 2007. Ultrastructure and phylogenetic placement within Heterolobosea of the previously unclassified, extremely halophilic heterotrophic flagellate *Pleurostomum flabellatum* (Ruinen 1938). Protist. 158: 397-413.

Parks DH, Tyson GW, Hugenholtz P and Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 30: 3123-3124.

Paterson GK, Blue CE and Mitchell TJ. 2006. An operon in *Streptococcus pneumoniae* containing a putative alkylhydroperoxidase D homologue contributes to virulence and the response to oxidative stress. Microb Pathog. 40: 152-160.

Patterson DJ and Simpson AGB. 1996. Heterotrophic flagellates from coastal marine and hypersaline sediments in Western Australia. Eur J Protistol. 32: 423-448.

Paul S, Bag SK, Das S, Harvill ET and Dutta C. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Biol. 9: R70.

Paulick MG and Bertozzi CR. 2008. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry. 47: 6991-7000.

Pavlidis P and Noble WS. 2003. Matrix2png: a utility for visualizing matrix data. Bioinformatics. 19: 295-296.

Pedrós-Alió C, Calderón-Paz JI, MacLean MH, Medina G, Marrasé C, *et al.* 2000. The microbial food web along salinity gradients. FEMS Microbiol Ecol. 32: 143-155.

Peeler TC, Stephenson MB, Einspahr KJ and Thompson GA. 1989. Lipid characterization of an enriched plasma membrane fraction of *Dunaliella salina* grown in media of varying salinity. Plant Physiol. 89: 970-976.

Pelham HRB and Bienz M. 1982. A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. EMBO J. 1: 1473-1477.

Perez-Prat E, Narasimhan ML, Niu X, Botella MA, Bressan RA, *et al.* 1994. Growth cycle stage-independent NaCl induction of plasma membrane H+-ATPase mRNA accumulation in de-adapted tobacco cells. Plant Cell Environ. 17: 327-333.

Pernthaler J. 2005. Predation on prokaryotes in the water column and its ecological implications. Nature Rev Microbiol. 3: 537-546.

Perriss SJ and Laybourn-Parry J. 1997. Microbial communities in saline lakes of the Vestfold hills (eastern Antarctica). Polar Biol. 18: 135-144.

Petersen TN, Brunak S, von Heijne G and Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8: 785-786.

Petrovič U. 2006. Role of oxidative stress in the extremely salt-tolerant yeast *Hortaea werneckii*. FEMS Yeast Res. 6: 816-822.

Petrovič U, Gunde-Cimerman N and Plemenitaš A. 2002. Cellular responses to environmental salinity in the halophilic black yeast *Hortaea werneckii*. Mol Microbiol. 45: 665-672.

Phillips JE and Gomer RH. 2015. Partial genetic suppression of a loss-of-function mutant of the neuronal ceroid lipofuscinosis-associated protease TPP1 in *Dictyostelium discoideum*. Dis Model Mech. 8: 147-156.

Pineda-Molina E, Klatt P, Vazquez J, Marina A, de Lacoba MG, *et al.* 2001. Glutathionylation of the p50 subunit of NF-kappa B: a mechanism for redox-induced inhibition of DNA binding. Biochemistry. 40: 14134-14142.

Pireyre M and Burow M. 2015. Regulation of MYB and bHLH transcription factors: a glance at the protein level. Mol Plant. 8: 378-388.

Pitcher GC, Cembella AD, Krock B, Macey BM, Mansfield L, *et al.* 2014. Identification of the marine diatom *Pseudonitzschia multiseries* (Bacillariophyceae) as a source of the toxin domoic acid in Algoa Bay, South Africa. Afr J Mar Sci. 36: 523-528.

Plemenitaš A, Lenassi M, Konte T, Kejžar A, Zajc J, *et al.* 2014. Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective. Front Microbiol. 5: 199.

Plemenitaš A, Vaupotič T, Lenassi M, Kogej T and Gunde-Cimerman N. 2008. Adaptation of extremely halotolerant black yeast *Hortaea werneckii* to increased osmolarity: a molecular perspective at a glance. Stud Mycol. 61: 67-75.

Plotnikov AO, Mylnikov AP and Selivanova EA. 2015. Morphology and life cycle of amoeboflagellate *Pharyngomonas* sp. (Heterolobosea, Excavata) from hypersaline inland Razval Lake. Biol Bull. 42: 759-769.

Pomposiello PJ, Bennik MHJ and Demple B. 2001. Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. J Bacteriol. 183: 3890-3902.

Por FD. 1980. A classification of hypersaline waters, based on trophic criteria. Mar Ecol. 1: 121-131.

Post FJ. 1977. The microbial ecology of the Great Salt Lake. Microb Ecol. 3: 143-165.

Post FJ, Borowitzka LJ, Borowitzka MA, Mackay B and Moulton T. 1983. The protozoa of a western Australian hypersaline lagoon. Hydrobiologia. 105: 95-113.

Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, *et al.* 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res. 42: D231-D239.

Prescott DM. 1994. The DNA of ciliated protozoa. Microbiol Rev. 58: 233-267.

Price MN, Dehal PS and Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 26: 1641-1650.

Prista C, Almagro A, Loureiro-Dias MC and Ramos J. 1997. Physiological basis for the high salt tolerance of *Debaryomyces hansenii*. Appl Environ Microbiol. 63: 4005-4009.

Prista C, Loureiro-Dias MC, Montiel V, Garcia R and Ramos J. 2005. Mechanisms underlying the halotolerant way of *Debaryomyces hansenii*. FEMS Yeast Res. 5: 693-701.

Proft M and Serrano R. 1999. Repressors and upstream repressing sequences of the stress-regulated *ENA1* gene in *Saccharomyces cerevisiae*: bZIP protein Sko1p confers HOG-dependent osmotic regulation. Mol Cell Biol. 19: 537-546.

Quinn PJ. 1981. The fluidity of cell membranes and its regulation. Prog Biophys Mol Biol. 38: 1-104.

Quintero FJ, Martinez-Atienza J, Villalta I, Jiang X, Kim W-Y, *et al.* 2011. Activation of the plasma membrane Na/H antiporter Salt-Overly-Sensitive 1 (SOS1) by phosphorylation of an auto-inhibitory C-terminal domain. Proc Natl Acad Sci USA. 108: 2611-2616.

Raederstorff D and Rohmer M. 1987a. The action of the systemic fungicides tridemorph and fenpropimorph on sterol biosynthesis by the soil amoeba *Acanthamoeba polyphaga*. Eur J Biochem. 164: 421-426.

Raederstorff D and Rohmer M. 1987b. Sterol biosynthesis via cycloartenol and other biochemical features related to photosynthetic phyla in the amoebae *Naegleria lovaniensis* and *Naegleria gruberi*. Eur J Biochem. 164: 427-434.

Rao KN, Kumaran D, Seetharaman J, Bonanno JB, Burley SK, *et al.* 2006. Crystal structure of trehalose-6-phosphate phosphatase-related protein: biochemical and biological implications. Protein Sci. 15: 1735-1744.

Rasuk MC, Kurth D, Regina Flores M, Contreras M, Novoa F, *et al.* 2014. Microbial characterization of microbial ecosystems associated to evaporites domes of gypsum in Salar de Llamara in Atacama Desert. Microb Ecol. 68: 483-494.

Ravichandran V, Seres T, Moriguchi T, Thomas JA and Johnston RB. 1994. S-thiolation of glyceraldehyde-3-phosphate dehydrogenase induced by the phagocytosis-associated respiratory burst in blood monocytes. J Biol Chem. 269: 25010-25015.

Raymond JA and Kim HJ. 2012. Possible role of horizontal gene transfer in the colonization of sea ice by algae. Plos One. 7: e35968.

Record E, Asther M, Marion D and Asther M. 1995. Purification and characterization of a novel specific phosphatidylglycerol-phosphatidylinositol transfer protein with high activity from *Aspergillus oryzae*. BBA-Lipid Lipid Met. 1256: 18-24.

Record E, Moukha S and Asther M. 1999. Characterization and expression of the cDNA encoding a new kind of phospholipid transfer protein, the phosphatidylglycerol/phosphatidylinositol transfer protein from *Aspergillus oryzae*: evidence of a putative membrane targeted phospholipid transfer protein in fungi. BBA-Gene Struct Expr. 1444: 276-282.

Reddy R, Smith D, Wayman G, Wu ZL, Villacres EC, *et al.* 1995. Voltage-sensitive adenylyl-cyclase activity in cultured neurons - a calcium-independent phenomenon. J Biol Chem. 270: 14340-14346.

Rees HH, Goad LJ and Goodwin TW. 1969. 2,3-oxidosqualene cycloartenol cyclase from *Ochromonas malhamensis*. Biochim Biophys Acta. 176: 892-&.

Rengpipat S, Lowe SE and Zeikus JG. 1988. Effect of extreme salt concentrations on the physiology and biochemistry of *Halobacteroides acetoethylicus*. J Bacteriol. 170: 3065-3071.

Reshetnikov AS, Mustakhimov, II, Khmelenina VN and Trotsenko YA. 2005. Cloning, purification, and characterization of diaminobutyrate acetyltransferase from the halotolerant methanotroph *Methylomicrobium alcaliphilum* 20Z. Biochem (Mosc). 70: 878-883.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3: 217-223.

Revell LJ and Harrison AS. 2008. PCCA: a program for phylogenetic canonical correlation analysis. Bioinformatics. 24: 1018-1020.

Rhodes ME, Spear JR, Oren A and House CH. 2011. Differences in lateral gene transfer in hypersaline versus thermal environments. BMC Evol Biol. 11: 199.

Ricard G, McEwan NR, Dutilh BE, Jouany JP, Macheboeuf D, *et al.* 2006. Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics. 7: 22.

Richard SB, Madern D, Garcin E and Zaccai G. 2000. Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. Biochemistry. 39: 992-1000.

Richards TA, Hirt RP, Williams BAP and Embley TM. 2003. Horizontal gene transfer and the evolution of parasitic protozoa. Protist. 154: 17-32.

Rio DC, Ares M, Hannon GJ and Nilsen TW. 2010. Purification of RNA using TRIzol (TRI Reagent). Cold Spring Harb Protoc: doi:10.1101/pdb.prot5439.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, *et al.* 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47.

Roberts MF. 2005. Organic compatible solutes of halotolerant and halophilic microorganisms. Saline Syst. 1.

Robison GA, Butcher RW and Sutherland EW. 1971. *Cyclic AMP*. New York: Academic Press.

Rodríguez-Navarro A. 2000. Potassium transport in fungi and plants. BBA-Rev Biomembranes. 1469: 1-30.

Rogers MA, Liu J, Song B-L, Li B-L, Chang CCY, *et al.* 2015. Acyl-CoA:cholesterol acyltransferases (ACATs/SOATs): enzymes with multiple sterols as substrates and as activators. J Steroid Biochem Mol Biol. 151: 102-107.

Rogerson A and Hauer G. 2002. Naked amoebae (Protozoa) of the Salton Sea, California. Hydrobiologia. 473: 161-177.

Rokitta SD, Von Dassow P, Rost B and John U. 2014. *Emiliania huxleyi* endures N-limitation with an efficient metabolic budgeting and effective ATP synthesis. BMC Genomics. 15: 10511051.

Rosell A, Valencia E, Parés X, Fita I, Farrés J, *et al.* 2003. Crystal structure of the vertebrate NADP(H)-dependent alcohol dehydrogenase (ADH8). J Mol Biol. 330: 75-85.

Ross D. 2004. Quinone reductases multitasking in the metabolic world. Drug Metab Rev. 36: 639-654.

Ross D, Kepa JK, Winski SL, Beall HD, Anwar A, *et al.* 2000. NAD(P)H : quinone oxidoreductase 1 (NQO1): chemoprotection, bioactivation, gene regulation and genetic polymorphisms. Chem-Biol Interact. 129: 77-97.

Rouhier N, Lemaire SD and Jacquot J-P. 2008. The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. Annu Rev Plant Biol. 59: 143-166.

Roy S. 2016. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. Plant Signal Behav. 11: e1117723.

Rudolph AS, Crowe JH and Crowe LM. 1986. Effects of 3 stabilizing agents proline, betaine, and trehalose on membrane phospholipids. Arch Biochem Biophys. 245: 134-143.

Ruinen J. 1938. Notizen über Salzflagellaten. II. Über die Verbreitung der Salzflagellaten. Arch Protistenkd. 90: 210-258.

Russell NJ. 1989. Adaptive modification in membranes of halotolerant and halophilic microorganisms. J Bioenerg Biomembr. 21: 93-113.

Russell NJ. 1993. Lipids of halophilic and halotolerant microorganisms. In: Vreeland RH and Hochstein LI, editors. *The Biology of Halophilic Bacteria*. Boca Raton: CRC Press. p. 163-210.

Rytömaa M and Kinnunen PKJ. 1994. Evidence for two distinct acidic phospholipid-binding sites in cytochrome *c*. J Biol Chem. 269: 1770-1774.

Salonen K and Jokinen S. 1988. Flagellate grazing on bacteria in a small dystrophic lake. Hydrobiologia. 161: 203-209.

Salway JG. 1999. *Metabolism at a Glance* Oxford: Blackwell Science.

Sanchez-Perez G, Mira A, Nyirő G, Pašić L and Rodriguez-Valera F. 2008. Adapting to environmental changes using specialized paralogs. Trends Genet. 24: 154-158.

Sarsour EH, Kalen AL and Goswami PC. 2014. Manganese superoxide dismutase regulates a redox cycle within the cell cycle. Antioxid Redox Signal. 20: 1618-1627.

Schaller H. 2004. New aspects of sterol biosynthesis in growth and development of higher plants. Plant Physiol Biochem. 42: 465-476.

Schiraldi C, Maresca C, Catapano A, Galinski EA and De Rosa M. 2006. High-yield cultivation of *Marinococcus* M52 for production and recovery of hydroxyectoine. Res Microbiol. 157: 693-699.

Schlame M. 2013. Cardiolipin remodeling and the function of tafazzin. BBA-Mol Cell Biol L. 1831: 582-588.

Schlame M, Acehan D, Berno B, Xu Y, Valvo S*, et al.* 2012. The physical state of lipid substrates provides transacylation specificity for tafazzin. Nat Chem Biol. 8: 862-869.

Schlame M, Brody S and Hostetler KY. 1993. Mitochondrial cardiolipin in diverse eukaryotes - comparison of biosynthetic reactions and molecular acyl species. Eur J Biochem. 212: 727-735.

Schlame M and Rüstow B. 1990. Lysocardiolipin formation and reacylation in isolated rat liver mitochondria. Biochem J. 272: 589-595.

Schofield CJ and Zhang ZH. 1999. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. Curr Opin Struct Biol. 9: 722-731.

Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, *et al.* 2013. Gene transfer from Bacteria and Archaea facilitated evolution of an extremophilic eukaryote. Science. 339: 1207-1210.

Schönknecht G, Weber APM and Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. BioEssays. 36: 9-20.

Schülke N and Schmid FX. 1988. Effect of glycosylation on the mechanism of renaturation of invertase from yeast. J Biol Chem. 263: 8832-8837.

Seddon JM, Cevc G and Marsh D. 1983. Calorimetric studies of the gel-fluid (Lβ-Lα) and lamellar-inverted hexagonal (Lα-H$_{II}$) phase transitions in dialkyl and dicaylphosphatidylethanolamines. Biochemistry. 22: 1280-1289.

Segura MJR, Lodeiro S, Meyer MM, Patel AJ and Matsuda SPT. 2002. Directed evolution experiments reveal mutations at cycloartenol synthase residue His477 that dramatically alter catalysis. Org Lett. 4: 4459-4462.

Seip B, Galinski EA and Kurz M. 2011. Natural and engineered hydroxyectoine production based on the *Pseudomonas stutzeri ectABCD-ask* gene cluster. Appl Environ Microbiol. 77: 1368-1374.

Selvan N, Mariappa D, van den Toorn HWP, Heck AJR, Ferenbach AT*, et al.* 2015. The early Metazoan *Trichoplax adhaerens* possesses a functional *O-*Glc*N*Ac system. J Biol Chem. 290: 11969-11982.

Severin J, Wohlfarth A and Galinski EA. 1992. The predominant role of recently discovered tetrahydropyrimidine for the osmoadaptation of halophilic eubacteria. J Gen Microbiol. 138: 1629-1638.

Sharma P, Jha AB, Dubey RS and Pessarakli M. 2012. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. J Bot. 2012: 217037.

Sheffer M, Fried A, Gottlieb HE, Tietz A and Avron M. 1986. Lipid composition of the plasma membrane of the halotolerant alga, *Dunaliella salina*. Biochim Biophys Acta. 857: 165-172.

Shelton MD, Chock PB and Mieyal JJ. 2005. Glutaredoxin: role in reversible protein S-glutathionylation and regulation of redox signal transduction and protein translocation. Antioxid Redox Signal. 7: 348-366.

Sherr BF, Sherr EB and Pedrós-Alió C. 1989. Simultaneous mesurement of bacterioplankton production and protozoan bacterivory in estuarine water. Mar Ecol Prog Ser. 54: 209-219.

Sherwood JE, Stagnitti F, Kokkinn MJ and Williams WD. 1991. Dissolved oxygen concentrations in hypersaline waters. Limnol Oceanogr. 36: 235-250.

Shinoda T, Ogawa H, Cornelius F and Toyoshima C. 2009. Crystal structure of the sodium-potassium pump at 2.4 angstrom resolution. Nature. 459: 446-U167.

Shono M, Hara Y, Wada M and Fujii T. 1996. A sodium pump in the plasma membrane of the marine alga *Heterosigma akashiwo*. Plant Cell Physiol. 37: 385-388.

Shono M, Wada M, Hara Y and Fujii T. 2001. Molecular cloning of Na+-ATPase cDNA from a marine alga, *Heterosigma akashiwo*. BBA-Biomembranes. 1511: 193-199.

Siegel D, Gustafson DL, Dehn DL, Han JY, Boonchoong P, *et al.* 2004. NAD(P)H : quinone oxidoreductase 1: role as a superoxide scavenger. Mol Pharmacol. 65: 1238-1247.

Silva-Graça M, Neves L and Lucas C. 2003. Outlines for the definition of halotolerance/halophily in yeasts: *Candida versatilis* (*halophila*) CBS4019 as the archetype? FEMS Yeast Res. 3: 347-362.

Simonin H, Beney L and Gervais P. 2008. Controlling the membrane fluidity of yeasts during coupled thermal and osmotic treatments. Biotechnol Bioeng. 100: 325-333.

Simpson AGB and Patterson DJ. 1996. Ultrastructure and identification of the predatory flagellate *Colpodella pugnax* Cienkowski (Apicomplexa) with a description of *Colpodella turpis* n. sp. and a review of the genus. Syst Parasitol. 33: 187-198.

Sivaramakrishnan V and Fountain SJ. 2015. Evidence for extracellular ATP as a stress signal in a single-celled organism. Eukaryot Cell. 14: 775-782.

Sivaramakrishnan V and Fountain SJ. 2013. Intracellular P2X receptors as novel calcium release channels and modulators of osmoregulation in *Dictyostelium*. Channels. 7: 43-46.

Smith CD, Carson M, Friedman AM, Skinner MM, Delucas L, *et al.* 2002. Crystal structure of human L-isoaspartyl-O-methyltransferase with S-adenosyl homocysteine at 1.6-angstrom resolution and modeling of an isoaspartyl-containing peptide at the active site. Protein Sci. 11: 625-635.

Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, *et al.* 2010. The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. BMC Plant Biol. 10: 83.

Smolyanyuk EV, Bilanenko EN, Tereshina VM, Kachalkin AV and Kamzolkina OV. 2013. Effect of sodium chloride concentration in the medium on the composition of the membrane lipids and carbohydrates in the cytosol of the fungus *Fusarium* sp. Microbiology. 82: 600-608.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Molec Biol. 3: Article 3.

Song JL, Harry JB, Eastman RT, Oliver BG and White TC. 2004. The *Candida albicans* lanosterol 14-alpha-demethylase (*ERG11*) gene promoter is maximally induced after prolonged growth with antifungal drugs. Antimicrob Agents Chemother. 48: 1136-1144.

Sook Chung J. 2008. A trehalose 6-phosphate synthase gene of the hemocytes of the blue crab, *Callinectes sapidus*: cloning, the expression, its enzyme activity and relationship to hemolymph trehalose levels. Saline Syst. 4: 18.

Soucy SM, Huang JL and Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nat Rev Genet. 16: 472-482.

Sparla F, Tedeschi G, Pupillo P and Trost P. 1999. Cloning and heterologous expression of NAD(P)H : quinone reductase of *Arabidopsis thaliana*, a functional homologue of animal DT-diaphorase. FEBS Lett. 463: 382-386.

Spear JR, Ley RE, Berger AB and Pace NR. 2003. Complexity in natural microbial ecosystems: the Guerrero Negro experience. Biol Bull. 204: 168-173.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E*, et al.* 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. Curr Biol. 24: 1176-1186.

Stamatakis A, Ludwig T and Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 21: 456-463.

Steinman H. 1982. Superoxide dismutases: protein chemistry and structure–function relationships. In: Oberley L, editor. *Superoxide dismutase, Vol. 1*. Boca Ratón, FL.: CRC Press. p. 11-68.

Stivaletta N, Barbieri R, Cevenini F and López-García P. 2011. Physicochemical conditions and microbial diversity associated with the evaporite deposits in the Laguna de la Piedra (Salar de Atacama, Chile). Geomicrobiol J. 28: 83-95.

Stock A, Breiner HW, Pachiadaki M, Edgcomb V, Filker S*, et al.* 2012. Microbial eukaryote life in the new hypersaline deep-sea basin Thetis. Extremophiles. 16: 21-34.

Stock A, Edgcomb V, Orsi W, Filker S, Breiner HW*, et al.* 2013. Evidence for isolated evolution of deep-sea ciliate communities through geological separation and environmental selection. BMC Microbiol. 13: 150.

Stock AM, Robinson VL and Goudreau PN. 2000. Two-component signal transduction. Annu Rev Biochem. 69: 183-215.

Stoeck T, Filker S, Edgcomb V, Orsi W, Yakimov M*, et al.* 2014. Living at the limits: evidence for microbial eukaryotes thriving under pressure in deep anoxic, hypersaline habitats. Adv Ecol. 2014: 532687.

Stoecker DK, Gustafson DE, Merrell JR, Black MMD and Baier CT. 1997. Excystment and growth of chrysophytes and dinoflagellates at low temperatures and high salinities in Antarctic sea-ice. J Phycol. 33: 585-595.

Stojilkovic SS, Leiva-Salcedo E, Rokic MB and Coddou C. 2014. Regulation of ATP-Gated P2X channels: from redox signaling to interactions with other proteins. Antioxid Redox Signal. 21: 953-970.

Stöveken N, Pittelkow M, Sinner T, Jensen RA, Heider J, *et al.* 2011. A specialized aspartokinase enhances the biosynthesis of the osmoprotectants ectoine and hydroxyectoine in *Pseudomonas stutzeri* A1501. J Bacteriol. 193: 4456-4468.

Strömstedt M, Rozman D and Waterman MR. 1996. The ubiquitously expressed human CYP51 encodes lanosterol 14 alpha-demethylase, a cytochrome P450 whose expression is regulated by oxysterols. Arch Biochem Biophys. 329: 73-81.

Sugimoto Y, Ninomiya H, Ohsaki Y, Higaki K, Davies JP, *et al.* 2001. Accumulation of cholera toxin and GM1 ganglioside in the early endosome of Niemann-Pick C1-deficient cells. Proc Natl Acad Sci USA. 98: 12391-12396.

Summons RE, Bradley AS, Jahnke LL and Waldbauer JR. 2006. Steroids, triterpenoids and molecular oxygen. Phil Trans R Soc B Biol Sci. 361: 951-968.

Sun Y and MacRae TH. 2005. Small heat shock proteins: molecular structure and chaperone function. Cell Mol Life Sci. 62: 2460-2476.

Sun YD, Carroll S, Kaksonen M, Toshima JY and Drubin DG. 2007. Ptdlns(4,5)P-2 turnover is required for multiple stages during clathrin- and actin-dependent endocytic internalization. J Cell Biol. 177: 355-367.

Sundler R and Akesson B. 1975. Regulation of phospholipid biosynthesis in isolated rat hepatocytes: effect of different substrates. J Biol Chem. 250: 3359-3367.

Susko E, Field C, Blouin C and Roger AJ. 2003. Estimation of rates-across-sites distributions in phylogenetic substitution models. Syst Biol. 52: 594-603.

Sutton GC, Russell NJ and Quinn PJ. 1990. The effect of salinity on the phase behavior of purified phosphatidylethanolamine and phosphatidylglycerol isolated from a moderately halophilic eubacterium. Chem Phys Lipids. 56: 135-147.

Suzuki M, Xiang T, Ohyama K, Seki H, Saito K, *et al.* 2006. Lanosterol synthase in dicotyledonous plants. Plant Cell Physiol. 47: 565-571.

Synnott JM, Guida A, Mulhern-Haughey S, Higgins DG and Butler G. 2010. Regulation of the hypoxic response in *Candida albicans*. Eukaryot Cell. 9: 1734-1746.

Taher AG, Abd El Wahab S, Philip G, Krumbein WE and Wali AM. 1995. Evaporitic sedimentation and microbial mats in a salina system (Port Fouad, Egypt). Int J Salt Lake Res. 4: 95-116.

Takishita K, Chikaraishi Y, Leger MM, Kim E, Yabuki A, *et al.* 2012. Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. Biol Direct. 7: 5.

Tamburro A, Robuffo I, Heipieper HJ, Allocati N, Rotilio D, *et al.* 2004. Expression of glutathione S-transferase and peptide methionine sulphoxide reductase in *Ochrobactrum anthropi* is correlated to the production of reactive oxygen species caused by aromatic substrates. FEMS Microbiol Lett. 241: 151-156.

Tammam AA, Fakhry EM and El-Sheekh M. 2011. Effect of salt stress on antioxidant system and the metabolism of the reactive oxygen species in *Dunaliella salina* and *Dunaliella tertiolecta*. Afr J Biotechnol. 10: 3795-3808.

Tams JW, Vind J and Welinder KG. 1999. Adapting protein solubility by glycosylation: N-glycosylation mutants of *Coprinus cinereus* peroxidase in salt and organic solutions. BBA-Protein Struct M. 1432: 214-221.

Tams JW and Welinder KG. 1995. Mild chemical deglycosylation of horseradish peroxidase yields a fully active, homogeneous enzyme. Anal Biochem. 228: 48-55.

Tan ZJ and Chen SJ. 2006. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. Biophys J. 90: 1175-1190.

Tang FQ and Moore TS. 1997. Enzymes of the primary phosphatidylethanolamine biosynthetic pathway in postgermination Castor bean endosperm: developmental profiles and partial purification of the mitochondrial CTP: ethanolaminephosphate cytidylyltransferase. Plant Physiol. 115: 1589-1597.

Tanne C, Golovina EA, Hoekstra FA, Meffert A and Galinski EA. 2014. Glass-forming property of hydroxyectoine is the cause of its superior function as a desiccation protectant. Front Microbiol. 5: 150.

Taton M and Rahier A. 1991. Properties and structural requirements for substrate specificity of cytochrome-P-450-dependent obtusifoliol 14-demethylase from maize (*Sea mays*) seedlings. Biochem J. 277: 483-492.

Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, *et al.* 2004. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature. 432: 118-122.

Todd BL, Stewart EV, Burg JS, Hughes AL and Espenshade PJ. 2006. Sterol regulatory element binding protein is a principal regulator of anaerobic gene expression in fission yeast. Mol Cell Biol. 26: 2817-2831.

Tomanek L. 2015. Proteomic responses to environmentally induced oxidative stress. J Exp Biol. 218: 1867-1879.

Torres GE, Gainetdinov RR and Caron MG. 2003. Plasma membrane monoamine transporters: structure, regulation and function. Nature Rev Neurosci. 4: 13-25.

Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, *et al.* 2007. Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in *Arabidopsis*. Proc Natl Acad Sci USA. 104: 20623-20628.

Triadó-Margarit X and Casamayor EO. 2013. High genetic diversity and novelty in planktonic protists inhabiting inland and coastal high salinity water bodies. FEMS Microbiol Ecol. 85: 27-36.

Tunblad-Johansson I, Andre L and Adler L. 1987. The sterol and phospholipid composition of the salt-tolerant yeast *Debaryomyces hansenii* grown at various concentrations of NaCl. Biochim Biophys Acta. 921: 116-123.

Turk M, Abramović Z, Plemenitaš A and Gunde-Cimerman N. 2007. Salt stress and plasma-membrane fluidity in selected extremophilic yeasts and yeast-like fungi. FEMS Yeast Res. 7: 550-557.

Turk M, Méjanelle L, Šentjurc M, Grimalt JO, Gunde-Cimerman N, *et al.* 2004. Salt-induced changes in lipid composition and membrane fluidity of halophilic yeast-like melanized fungi. Extremophiles. 8: 53-61.

Turk M, Plemenitaš A and Gunde-Cimerman N. 2011. Extremophilic yeasts: plasma-membrane fluidity as determinant of stress tolerance. Fungal Biol. 115: 950-958.

Tusnady GE and Simon I. 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics. 17: 849-850.

Valachovic M, Klobucnikova V, Griac P and Hapala I. 2002. Heme-regulated expression of two yeast acyl-CoA : sterol acyltransferases is involved in the specific response of sterol esterification to anaerobiosis. FEMS Microbiol Lett. 206: 121-125.

Valencia E, Larroy C, Ochoa WF, Parés X, Fita I, *et al.* 2004. Apo and Holo structures of an NADP(H)-dependent cinnamyl alcohol dehydrogenase from *Saccharomyces cerevisiae*. J Mol Biol. 341: 1049-1062.

van Dam H, Wilhelm D, Herr I, Steffen A, Herrlich P, *et al.* 1995. ATF-2 is preferentially activated by stress-activated protein kinases to mediate *c-jun* induction in response to genotoxic agents. EMBO J. 14: 1798-1811.

Van den Steen P, Rudd PM, Dwek RA and Opdenakker G. 1998. Concepts and principles of O-linked glycosylation. Crit Rev Biochem Mol Biol. 33: 151-208.

van der Wielen PWJJ, Bolhuis H, Borin S, Daffonchio D, Corselli C*, et al.* 2005. The enigma of prokaryotic life in deep hypersaline anoxic basins. Science. 307: 121-123.

van Es S, Virdy KJ, Pitt GS, Meima M, Sands TW*, et al.* 1996. Adenylyl cyclase G, an osmosensor controlling germination of *Dictyostelium* spores. J Biol Chem. 271: 23623-23625.

van Meer G, Voelker DR and Feigenson GW. 2008. Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol. 9: 112-124.

van Ooijen G, Knox K, Kis K, Bouget F-Y and Millar AJ. 2012. Genomic Transformation of the Picoeukaryote *Ostreococcus tauri*. J Vis Exp: UNSP e4074UNSP e4074.

Vaqué D, Gasol JM and Marrasé C. 1994. Grazing rates on bacteria: the significance of methodology and ecological factors. Mar Ecol Prog Ser. 109: 263-274.

Vaupotic T and Plemenitaš A. 2007. Differential gene expression and HogI interaction with osmoresponsive genes in the extremely halotolerant black yeast *Hortaea werneckii*. BMC Genomics. 8: 280.

Verity PG. 1991. Feeding in planktonic protozoans: evidence for non-random acquisition of prey. J Protozool. 38: 69-76.

Verkhratsky A and Burnstock G. 2014. Biology of purinergic signalling: its ancient evolutionary roots, its omnipresence and its multiple functional significance. BioEssays. 36: 697-705.

Vermeulen PS, Geelen MJH and Vangolde LMG. 1994. Substrate specificity of CTP: phosphoethanolamine cytidyltransferase purified from rat liver. BBA-Lipid Lipid Met. 1211: 343-349.

Vermeulen PS, Tijburg LBM, Geelen MJH and Vangolde LMG. 1993. Immunological characterization, lipid dependence, and subcellular localization of CTP: phosphoethanolamine cytidyltransferase purified from rat liver: comparison with CTP-phosphocholine cytidyltransferase. J Biol Chem. 268: 7458-7464.

Vidal R, López-Maury L, Guerrero MG and Florencio FJ. 2009. Characterization of an alcohol dehydrogenase from the cyanobacterium *Synechocystis* sp. strain PCC 6803 that responds to environmental stress conditions via the Hik34-Rre1 two-component system. J Bacteriol. 191: 4383-4391.

Visick JE, Cai H and Clarke S. 1998. The L-isoaspartyl protein repair methyltransferase enhances survival of aging *Escherichia coli* subjected to secondary environmental stresses. J Bacteriol. 180: 2623-2629.

Volcani B. 1944. The microorganisms of the Dead Sea. In. *Papers Collected to Commemorate the 70th Anniversary of Dr. Chaim Weizmann*. Rehovoth: Daniel Sieff Research Institute. p. 71-85.

Vonheijne G and Abrahmsen L. 1989. Species-specific variation in signal peptide design - implications for protein secretion in foreign hosts. FEBS Lett. 244: 439-446.

Vothknecht UC and Westhoff P. 2001. Biogenesis and origin of thylakoid membranes. BBA-Mol Cell Res. 1541: 91-101.

Vreeland RH, Anderson R and Murray RGE. 1984. Cell wall and phospholipid composition and their contribution to the salt tolerance of *Halomonas elongata*. J Bacteriol. 160: 879-883.

Vreken P, Valianpour F, Nijtmans LG, Grivell LA, Plecko B, *et al.* 2000. Defective remodeling of cardiolipin and phosphatidylglycerol in Barth syndrome. Biochem Biophys Res Commun. 279: 378-382.

Walker R, Gissot M, Huot L, Alayi TD, Hot D, *et al.* 2013. *Toxoplasma* transcription factor TgAP2XI-5 regulates the expression of genes involved in parasite virulence and host invasion. J Biol Chem. 288: 31127-31138.

Wang J, Boja ES, Tan WH, Tekle E, Fales HM, *et al.* 2001. Reversible glutathionylation regulates actin polymerization in A431 cells. J Biol Chem. 276: 47763-47766.

Wang M, Wang L, Yi Q, Gai Y and Song L. 2015. Molecular cloning and characterization of a cytoplasmic manganese superoxide dismutase and a mitochondrial manganese superoxide dismutase from Chinese mitten crab *Eriocheir sinensis*. Fish Shellfish Immunol. 47: 407-417.

Wang XM and Moore TS. 1991. Phosphatidylethanolamine synthesis by castor bean endosperm: intracellular distribution and characteristics of CTP:ethanolaminephosphate cytidyltransferase. J Biol Chem. 266: 19981-19987.

Wapinski I, Pfeffer A, Friedman N and Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature. 449: 54-U36.

Warrilow A, Ugochukwu C, Lamb D, Kelly D and Kelly S. 2008. Expression and Characterization of CYP51, the Ancient Sterol 14-demethylase Activity for Cytochromes P450 (CYP), in the White-Rot Fungus *Phanerochaete chrysosporium*. Lipids. 43: 1143-1153.

Webb R, Culture Methods for BD, available from http://arwh.org/sites/default/files/files-uploads/Attachment 2 Culture Methods for Bd.pdf, last accessed in 2014.

Wen H and Evans RJ. 2009. Regions of the amino terminus of the P2X(1) receptor required for modification by phorbol ester and mGluR1 alpha receptors. J Neurochem. 108: 331-340.

Widderich N, Höppner A, Pittelkow M, Heider J, Smits SHJ, *et al.* 2014. Biochemical properties of ectoine hydroxylases from extremophiles and their wider taxonomic distribution among microorganisms. Plos One. 9: e93809e93809.

Wilbert N. 1995. Benthic ciliates of salt lakes. Acta Protozool. 34: 271-288.

Wilbert N and Kahan D. 1981. Ciliates of Solar Lake on the Red Sea shore. Arch Protistenkd. 124: 70-95.

Wilkinson SR, Prathalingam SR, Taylor MC, Ahmed A, Horn D, *et al.* 2006. Functional characterisation of the iron superoxide dismutase gene repertoire in *Trypanosoma brucei*. Free Radical Biol Med. 40: 198-209.

Willoughby D and Cooper DMF. 2007. Organization and Ca^{2+} regulation of adenylyl cyclases in cAMP microdomains. Physiol Rev. 87: 965-1010.

Wintjens R, Noel C, May ACW, Gerbod D, Dufernez F, *et al.* 2004. Specificity and phenetic relationships of iron- and manganese-containing superoxide dismutases on the basis of structure and sequence comparisons. J Biol Chem. 279: 9248-9254.

Wirtz KWA. 1991. Phospholipid transfer proteins. Annu Rev Biochem. 60: 73-99.

Wolfova J, Smatanova IK, Brynda J, Mesters JR, Lapkouski M, *et al.* 2009. Structural organization of WrbA in apo- and holoprotein crystals. Biochim Biophys Acta, Proteins Proteomics. 1794: 1288-1298.

Wrobel RL, Matvienko M and Yoder JI. 2002. Heterologous expression and biochemical characterization of an NAD(P)H : quinone oxidoreductase from the hemiparasitic plant *Triphysaria versicolor*. Plant Physiol Biochem. 40: 265-272.

Wu QLL, Chatzinotas A, Wang JJ and Boenigk J. 2009. Genetic diversity of eukaryotic plankton assemblages in Eastern Tibetan lakes differing by their salinity and altitude. Microb Ecol. 58: 569-581.

Xu F, Jerlström-Hultqvist J, Kolisko M, Simpson AGB, Roger AJ, *et al.* 2016. On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. BMC Biol. 14: 62.

Xu Y, Kelley RI, Blanck TJJ and Schlame M. 2003. Remodeling of cardiolipin by phospholipid transacylation. J Biol Chem. 278: 51380-51385.

Yadav DK and Tuteja N. 2011. Rice G-protein coupled receptor (GPCR): in silico analysis and transcription regulation under abiotic stress. Plant Signal Behav. 6: 1079-86.

Yamashita A, Singh SK, Kawate T, Jin Y and Gouaux E. 2005. Crystal structure of a bacterial homologue of Na$^+$/Cl$^-$-dependent neurotransmitter transporters. Nature. 437: 215-223.

Yan SZ, Huang ZH, Shaw RS and Tang WJ. 1997. The conserved asparagine and arginine are essential for catalysis of mammalian adenylyl cyclase. J Biol Chem. 272: 12342-12349.

Yang HY, Bard M, Bruner DA, Gleeson A, Deckelbaum RJ, *et al.* 1996. Sterol esterification in yeast: a two-gene process. Science. 272: 1353-1356.

Yang WY, Mason CB, Pollock SV, Lavezzi T, Moroney JV, *et al.* 2004. Membrane lipid biosynthesis in *Chlamydomonas reinhardtii*: expression and characterization of CTP: phosphoethanolamine cytidylyltransferase. Biochem J. 382: 51-57.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24: 1586-1591.

Yau S, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, *et al.* 2013. Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. ISME J. 7: 1944-1961.

Youssef NH, Savage-Ashlock KN, McCully AL, Luedtke B, Shaw EI, *et al.* 2014. Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the Halobacteriales. ISME J. 8: 636-649.

Yu BP. 1994. Cellular defenses against damage from reactive oxygen species. Physiol Rev. 74: 139-162.

Yu X-H, Jiang N, Yao P-B, Zheng X-L, Cayabyab FS, *et al.* 2014. NPC1, intracellular cholesterol trafficking and atherosclerosis. Clinica Chimica Acta. 429: 69-75.

Zachara NE, O'Donnell N, Cheung WD, Mercer JJ, Marth JD, *et al.* 2004. Dynamic *O*-Glc*N*Ac modification of nucleocytoplasmic proteins in response to stress - a survival response of mammalian cells. J Biol Chem. 279: 30133-30142.

Zachman DK, Chicco AJ, McCune SA, Murphy RC, Moore RL, *et al.* 2010. The role of calcium-independent phospholipase A$_2$ in cardiolipin remodeling in the spontaneously hypertensive heart failure rat heart. J Lipid Res. 51: 525-534.

Zaffagnini M, Michelet L, Marchand C, Sparla F, Decottignies P, *et al.* 2007. The thioredoxin-independent isoform of chloroplastic glyceraldehyde-3-phosphate dehydrogenase is selectively regulated by glutathionylation. FEBS J. 274: 212-226.

Zajc J, Liu Y, Dai W, Yang Z, Hu J, *et al.* 2013. Genome and transcriptome sequencing of the halophilic fungus *Wallemia ichthyophaga*: haloadaptations present and absent. BMC Genomics. 14: 617.

Zalar P, de Hoog GS, Schroers HJ, Frank JM and Gunde-Cimerman N. 2005. Taxonomy and phylogeny of the xerophilic genus *Wallemia* (Wallemiomycetes and Wallemiales, cl. et ord. nov.). Anton Leeuw Int J G. 87: 311-328.

Zervas M, Dobrenis K and Walkley SU. 2001. Neurons in Niemann-Pick disease type C accumulate gangliosides as well as unesterified cholesterol and undergo dendritic and axonal alterations. Journal of Neuropathology and Experimental Neurology. 60: 49-64.

Zhang M, Dwyer NK, Neufeld EB, Love DC, Cooney A, *et al.* 2001. Sterol-modulated glycolipid sorting occurs in Niemann-Pick C1 late endosomes. J Biol Chem. 276: 3417-3425.

Zhang M, Mileykovskaya E and Dowhan W. 2005. Cardiolipin is essential for organization of complexes III and IV into a supercomplex in intact yeast mitochondria. J Biol Chem. 280: 29403-29408.

Zhang Y, Yu CJ, Liu J, Spencer TA, Chang CCY, *et al.* 2003. Cholesterol is superior to 7-ketocholesterol or 7 alpha-hydroxycholesterol as an allosteric activator for acyl-coenzyme A : cholesterol acyltransferase 1. J Biol Chem. 278: 11642-11647.

Zhao R, Cao Y, Xu H, Lv LF, Qiao DR, *et al.* 2011. Analysis of expressed sequence tags from the green alga *Dunaliella salina* (Chlorophyta). J Phycol. 47: 1454-1460.