UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

The 'original sin' hypothesis and the evolution of programmed cell death en route to eukaryogenesis.

by

SoRi La (791752)

ORCID: 0000-0002-7052-3902

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science

Supervisor: Prof. Pierre M. Durand

Co-supervisor: Dr Andrew Ndhlovu

27 April 2022 in Johannesburg

DECLARATION

I declare that this dissertation, “The ‘original sin’ hypothesis and the evolution of programmed cell death en route to eukaryogenesis”, is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of Witwatersrand, Johannesburg. It has not been submitted before for any degree of examination at any other University.

(Signature of Candidate)

___27___ day of ___April___ 2022___

at___the University of Witwatersrand_____.

ABSTRACT

Programmed Cell Death (PCD) genetic toolkits and mechanisms of death in prokaryotes, particularly archaea, are poorly understood and whether PCD mediated conflict in eukaryogenesis remains controversial. This study aimed to investigate C14 peptidases, the key effectors of PCD, in archaea and their evolutionary histories prior to eukaryogenesis. Their wide taxonomic distribution and phylogenetic reconstruction revealed that C14 peptidases were present before the divergence of prokaryotes. Furthermore, phyletic pattern and death domain architecture analyses revealed that they were likely associated with pro-survival functions before their co-option in PCD, supporting Ameisen's 'original sin' hypothesis for PCD origins. *In vivo* analysis to explore potential caspase- and metacaspase-like activity in *Halobacterium salinarum* under oxidative stress showed negative results, in accordance with PCD being a black queen trait in the microbial loop between *H. salinarum* and *Dunaliella salina*. This study provides a detailed evolutionary history of C14 peptidases prior to eukaryogenesis.

In dedication to Almighty God

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my supervisor, Prof. Pierre M. Durand, whose expertise was invaluable in formulating research questions and writing discussions, as well as insightful feedbacks which pushed me to sharpen my thinking and brought my work to a higher level.

I would also like to express my gratitude and appreciation to my co-supervisor, Dr Andrew Ndhlovu, for his guidance, expertise, sincerity, and professional ethics in formulating the methodology and producing results. Your valuable guidance throughout my MSc has allowed me to complete my dissertation successfully.

I would like to thank my colleague, Karen van Niekerk, who has advised, inspired, and assisted me with helpful conversations. In addition, I would like to thank my parents for their continuous guidance and support.

I would like to acknowledge Prof. Nitin S. Baliga and Dr Monica V. Orellana, Institute for Systems Biology, Seattle, for providing *Halobacterium salinarum* NRC-1 cultures.

Lastly, I gratefully acknowledge the financial support from the Paleontological Scientific Trust (PAST) Africa, Genus: DSI-NRF Centre of Excellence in Palaeosciences and the University of Witwatersrand.

PUBLICATIONS PRESENTED IN THIS DISSERTATION

La, S., Ndhlovu, A., Durand, P. M. (2021). The ancient origins of death domains support the 'original sin' hypothesis for the evolution of programmed cell death. Accepted by Journal of Molecular Evolution.

CONFERENCE PRESENTATIONS

La, S. (2021), 'The role of death in the rise of eukaryotes', *Life in the Universe Conference 2021*, online, 22 - 24 November.

La, S. (2021), 'The role of death in the rise of eukaryotes', *NeLLi 2021 Symposium: From New Lineages of Life to New Functions*, online, 16 September.

La, S. (2021), 'The role of death in the rise of eukaryotes: the case for archaeal caspase homologs', *ISCB-Africa ASBCB 2021*, online, 10 June.

La, S. (2020), 'Evolutionary history of caspase/haemoglobinase fold proteases en route to eukaryogenesis', *Palaeosciences lecture series 2020*, online, 29 October.

TABLE OF CONTENTS

LIST OF FIGURES

*Figures in the manuscript are not included

LIST OF TABLES

*Tables in the manuscript are not included

ABBREVIATIONS AND NOMENCLATURE

Abbreviation	Description
Ala	Alanine
AMC	Amido-4-Methylcoumarin
ANAPC3	Anaphase-Promoting Complex, Cyclosome, Subunit 3
ANAPC4_WD40	Anaphase-Promoting Complex Subunit 4 WD40 Domain
ANF_receptor	Receptor Family Ligand-Binding Region
AP-ATPase	Apoptotic ATP Synthase
Arg	Arginine
Asp	Aspartic acid
BAX	Bcl-2-associated X
BCL-2	B Cell Lymphoma-2
BiBP_C	Penicillin-Binding Protein C-terminus Family
BIR	Baculovirus IAP Repeat
BLAST	Basic Local Alignment Search Tool
BLASTN	Standard Nucleotide Basic Local Alignment Search Tool
BLASTP	Standard Protein Basic Local Alignment Search Tool
BQ	Black Queen
BQH	Black Queen Hypothesis
BSA	Bovine Serum Albumin
CA	Caspase
Ca^{2+}	Calcium
CARD	CA Activation And Recruitment Domain
CART	Cocaine and Amphetamine Regulated Transcript protein
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHAT	Caspase HetF Associated with Tprs
CHF	Caspase-Hemoglobinase Fold
CM	Conditioned Medium
Cys	Cysteine
DD	Death Domain
DED	Death-Effector Domain
DMSO	Dimethylsulfoxide
DPANN	Group including superphyla Aenigmarchaeota, Diapherotrites, Nanoarchaeota, Nanohaloarchaeota. Parvarchaeota and Woesearchaeota
DTT	Dithiothreitol
DUF	Domain of Unknown Function
DZR	Double Zinc Ribbon

E3	Entangle-Engulf-Endogenize
ER	Endoplasmic reticulum
FECA	First Eukaryotic Common Ancestor
FGE	Formylglycine-Generating
FoL	Forest of Life
Fn3	Fibronectin type III domain
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GO	Gene Ontology
H_2O_2	Hydrogen Peroxide
HC	Histidine-Cysteine
HEPES	Hydroxyethyl Piperazine Ethane Sulfonic Acid
HGT	Horizontal Gene Transfer
His	Histidine
HMM	Hidden Markov Model
HMMER	Biosequence analysis using profile hidden Markov models
Ig	Immunoglobulin
iTmm	Transmembrane helices predicted to have N-terminal domains situated intracellularly and C-terminal domains extracellularly
I-TASSER	Iterative Threading Assembly Refinement
katG	Catalase-peroxidase
LACA	Last Archaeal Common Ancestor
LECA	Last Eukaryotic Common Ancestor
Leu	Leucine
LRR	Leucine-rich repeat
LUCAS	Last Universal Common Ancestor State
Lys	Lysine
MAFFT	Multiple Alignment using Fast Fourier Transform
MATH	Meprin And TRAF Homology
MCA	Metacaspase
MSA	Multiple Sequence Alignment
OCA	Orthocaspase
OD	Optical Density
oTMM	Transmembrane helices predicted to have the reverse topology
PBS	Phosphate-Buffered Saline
PCA	Paracaspase
PCD	Programmed Cell Death

Peripla_BP_5	Periplasmic-Binding Protein Domain
Phe	Phenylalanine
PKD	Polycystic Kidney Disease
PMC	Paracaspase-Metacaspase-Caspase
PPC	Bacterial Pre-Peptidase C-Terminal Domain
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool
ROS	Reactive Oxygen Species
Ser	Serine
SD	Standard Deviation
TACK	Group including superphyla Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota
TaxId	Taxonomy ID
TBLASTN	Translated Basic Local Alignment Search Tool
TbMCA-Ib	Metacaspase gene of *Trypanasoma Brucei*
TRAF	Tumour Necrosis Factor Receptor-Associated Factor
TIR	Toll–Interleukin-Receptor Domain
TNF	Tumour Necrosis Factor
TPR	Tetratricopeptide Repeat
WD40	WD or Beta-Transducin Repeat
YC	Tyrosine-Cysteine

CHAPTER 1

INTRODUCTION

1.1 Literature review

1.1.1 Introduction

Programmed cell death (PCD), a genetically controlled cellular suicide, played an important role in the major transitions in evolution (Maynard Smith and Szathmáry, 1995; Blackstone and Green, 1999; Michod, 1999; Michod and Nedelcu, 2003; Durand, Barreto Filho and Michod, 2019; Durand, 2021). A well-known example is eukaryogenesis, the evolutionary process leading to the emergence of first eukaryotic cells (Chela-flores, 1998), in which PCD acted as a conflict mediator between symbionts (Blackstone and Green, 1999; Michod and Nedelcu, 2003; Durand, 2021).

While PCD played an important role in eukaryogenesis, PCD genetic toolkits in prokaryotes, especially archaea, and how it developed in pre-eukaryotes are poorly understood and sources of controversy (Ameisen, 2002; Durand, 2021). This is because PCD was initially discovered in Metazoa (Glücksmann, 1951; Lockshin and Williams, 1964) and subsequently interpreted as a hallmark of metazoan ontogenesis. Previous arguments for the origin of PCD were philosophical in nature with three propositions being the "addiction", 'original sin' (Ameisen, 2002) and "co-evolution" (Durand, 2021) hypotheses.

Caspases (CAs), one of the main effectors of PCD, and its homologues are frequently used to study the evolutionary history of PCD. By tracing their evolutionary history in prokaryotic lineages, the origin of PCD in prokaryotes can be inferred. In addition, CAs and CA homologues require additional domains of death to partake in the PCD molecular pathway (Aravind, Dixit and Koonin, 1999a). By examining different domains of death present in prokaryotic CAs, their putative ancestral functions can be inferred (Gaasterland and Ragan, 1998; Aravind, Dixit and Koonin, 1999a; Galperin and Koonin, 2000) and this may provide insight into their role in eukaryogenesis.

A key question that still remains in elucidating the origin of PCD in prokaryotes lies in archaea. Although CA homologues were identified in archaea (Asplund-Samuelsson, Bergman and Larsson, 2012; Klemenčič *et al.*, 2019), a comparative study to investigate the origin of PCD and putative roles of archaeal PCD genetic toolkits in eukaryogenesis has not been addressed. An *in vivo* study to investigate CA- and MCA-like activity in an archaeal model organism is available (Bidle *et al.*, 2010). However, the study did not address CA- and MCA-like activity in response to oxidative stress, one of the causes of conflicts during eukaryogenesis (Blackstone and Green, 1999; Radzvilavicius and Blackstone, 2015).

The purpose of this literature review will be to discuss PCD, some of the current hypotheses regarding its origin prior to eukaryogenesis and how cell death contributed to the rise of eukaryotes. In addition, the literature review will discuss how tracing the evolutionary histories of CAs, and CA homologues will provide insight into the origin of PCD in ancestral microbes. Furthermore, a comparison of CAs, CA homologues, and associated death domains will be provided, which can be used to assess their putative functions. I will also discuss tools available to search distantly related CA homologs in prokaryotes as well as the currently available data on archaeal CA homologues. Lastly, a model organism of archaea, *Halobacterium salinarum*, is presented for use in this dissertation.

1.1.2 Programmed cell death

Cell death was initially observed in multicellular embryonic tissues with its programmed nature unknown (Collin, 1906; Ernst, 1926; Kallius, 1931; Hamburger and Levi-Montalcini, 1949). Its programmed nature and its role in metazoan ontogeny was suggested later (Glücksmann, 1951; Lockshin and Williams, 1964) and was further investigated while studying the Bcl-2 protein, the product of a putative oncogene (Vaux, Cory and Adams, 1988). Key criteria to define PCD have been challenging to formulate due to different types of PCD that exist which utilize different molecular mechanisms (Reece *et al.*, 2011). The most widely used classification system for PCD was provided in 2018, based on morphological and biochemical features, relationship with exogenous and endogenous factors associated with PCD and nature of immunogenicity (Galluzzi and Vitale, 2018). PCD is classically divided into three forms based on their mechanism: apoptosis (type 1), autophagy (type 2) and necrosis (type III) (Table 1), and additional forms of PCD have been suggested (Galluzzi *et al.*, 2007; Galluzzi and Vitale, 2018). However, this review will not focus on the different forms of PCD, but rather it will focus on why PCD was essential for evolutionary transitions (Table 1) in the ancestral microbial world (Durand, Barreto and Michod, 2019).

It's important to clarify certain terminologies of PCD as they can be confusing. There are numerous PCD definitions that are evolutionary and mechanistic in nature (Durand and Ramsey, 2019) (Table 1). A widely used mechanistic definition of PCD was provided by Berman-Frank et al. (Table 1). The evolutionary definition of PCD draws a distinction between adaptive PCD that is selected for and non-adaptive PCD (ersatz PCD) that is observed in pleiotropy (Durand and Ramsey, 2019) (Table 1). The definition of PCD that I will use in my dissertation is the mechanistic definition provided by Berman-Frank *et al.*, 2004 (Table 1).

Table 1 Definitions and nomenclature.

Term	**Definition**
Altruism	"A behaviour that is costly to the actor (i.e., the focal individual who performs the behaviour) and beneficial to the recipient; costs and benefits

	are defined in terms of the lifetime direct fitness consequences of the behaviour" (West, Griffin and Gardner, 2007).
Apoptosis	"A PCD morphotype characterized by specific morphological and biochemical features, including mitochondrial depolarization, reduction of cellular volume, chromatin condensation, nuclear fragmentation, little or no ultrastructural modifications of cytoplasmic organelles, and plasma membrane blebbing (but maintenance of its integrity until the final stages of the process)" (Kroemer *et al.*, 2009)
Autophagy	"A PCD morphotype characterized by massive cytoplasm vacuolization, accumulation of autophagic vacuoles, and lack of chromatin condensation" (Kroemer *et al.*, 2009)
Evolutionary transition	"The major leaps in the complexity of life and the rise of new types of individuals from its predecessors" (Maynard Smith and Szathmáry, 1995).
Level of selection	"A hierarchical level at which Darwinian principles (heritable variation in fitness) apply" (Nedelcu *et al.*, 2011)
Fitness	"Rate of survival and reproduction" (Lewontin, 1970)
Multi-level Selection	"The concept that selection can occur at multiple levels of biological organization, for example genes, cells, individuals, groups, populations etc" (Okasha, 2005)
Necrosis	"A form of cell death characterized by cytoplasmic and organelle swelling, rupture of the plasma membrane, and the absence of features of apoptosis or autophagy" (Kroemer *et al.*, 2009)
PCD (mechanistic)	"An active, genetically controlled, cellular self-destruction driven by a series of complex biochemical events and specialized cellular machinery" (Berman-Frank *et al.*, 2004)
PCD (evolutionary)	"An adaptation for producing cell death" (Durand and Ramsey, 2019)
PCD (Ersatz)	"PCD that is intrinsic to the cell but the trait itself has not been selected for death." (Durand and Ramsey, 2019)
Selfish behaviour	"A behaviour that incurs no cost, while continuing to benefit from the acts of altruists" (Hamilton, 1970)
Units of selection	"Entities possessing phenotypic variation, differential fitness and heritability" (Lewontin, 1970)

1.1.3 PCD in prokaryotes

Since the term PCD was developed, the concept of PCD in unicellular microbes was questioned (Ratel *et al.*, 2001; Nedelcu *et al.*, 2011; Proto, Coombs and Mottram, 2013; Ramisetty, Natarajan and Santhosh, 2015). In contrast to multicellular eukaryotes that still survives despite the death of a cell (Andryukov, Somova and Timchenko, 2018), the death of a unicellular organism is the ultimate fitness cost for itself and it was assumed to provide no advantages (Nedelcu *et al.*, 2011). A morphological and a biochemical model for PCD in prokaryotes has not yet been developed, and its biological role and regulatory mechanisms are poorly understood (Engelberg-Kulka *et al.*, 2006; Erental, Sharon and Engelberg-Kulka, 2012).

1.1.4 The origin of PCD

The 'bacterial connection' to the origin and evolution of PCD in eukaryotes has been established (Koonin and Aravind, 2002; Klim *et al.*, 2018). However, the evolutionary origins of PCD and its genetic toolkits in prokaryotes have been largely unknown (Uren *et al.*, 2000; Ameisen, 2002; Koonin and Aravind, 2002; Klim *et al.*, 2018; Hofmann, 2019) and current hypotheses are based on philosophical and conceptual works (Ameisen, 2002; Durand, 2021). This is because the study of PCD in prokaryotes has been primarily biased towards the domain bacteria. Archaea have been largely overlooked from the PCD discussion because of a previous dearth of genomic data (Bidle *et al.*, 2010).

Ameisen's hypotheses on the origin of PCD

Ameisen suggested two philosophical hypotheses for the origin of PCD: the 'addiction' hypothesis and the 'original sin' hypothesis (Ameisen, 2002).

The 'addiction' hypothesis postulates that ancestral PCD genetic toolkits were initially involved in 'murder' and PCD arose as a result of an evolutionary arms race between cells (Ameisen, 2002). An example provided is the toxin-antitoxin Maz-EF mediated cell death system of *Escherichia coli* (Aizenman, Engelberg-Kulka & Glaser, 1996). In adverse environmental conditions (e.g., nutrient shortage), *E. coli* uses the Maz-EF mediated cell death system to have an evolutionary advantage over other bacterial species. The selected 'selfish' killer genes (e.g., MazF) and antagonists (e.g., MazE) provide selective advantages to the bacterial colony by allowing a part of the cell population to survive at the expense of others (Aizenman, Engelberg-Kulka and Glaser, 1996; Engelberg-Kulka, Hazan and Amitai, 2005; Ramisetty, Natarajan and Santhosh, 2015).

The 'original sin' hypothesis postulates that ancestors of PCD genetic modules were present since the origin of life and were rooted in primordial non-death functions (e.g., cell signalling, cell cycle, and cell differentiation) which were later co-opted into the PCD molecular pathway (Ameisen, 2002). The intrinsic inability of a cell to avoid random genetic mutations and alterations, and subsequently, its inability to avoid self-destruction as a consequence of essential life-sustaining activities, is the 'original sin' of the cell. Ameisen argues that the 'original sin' hypothesis is supported by one of the main effectors of PCD, CAs, which are also responsible for cell survival

mechanisms (e.g. cell cycle control and cell differentiation) in eukaryotes (Lamkanfi *et al.*, 2007; Shalini *et al.*, 2015).

Although Ameisen suggested two different hypotheses for the origin of PCD, he argues that the two may be synergistic, of which genetic modules allowing both cell survival and self-destruction became selected for the 'addictive' advantage (Ameisen, 2002). For example, the death-inducing toxin MazF and the cell survival-related antidote MazE genetic modules were selected for the advantage of certain microbes over others (Ameisen, 2002). The complementary and synergistic nature of the two hypotheses supports the argument for life-death coevolution in the last universal common ancestor state (LUCAS) as well as PCD being one of the driving forces in the evolution of life and complexity (Durand, 2021).

Investigation for the origin of PCD

For a long time, PCD was thought to be a unique characteristic of multicellular life (Saunders, 1966; Kerr, 1971; Kerr, Wyllie and Currie, 1972) and studies on the origin of PCD were focused on eukaryotes (Blackstone and Green, 1999; Koonin and Aravind, 2002; Klim *et al.*, 2018). In turn, a strong connection between bacteria and eukaryotic PCD was established because one of the CA homologues, metacaspases (MCAs), is hypothesized to have originated from an α-proteobacterium, possibly through horizontal gene transfer (HGT) during eukaryogenesis (Koonin and Aravind, 2002). CAs are executioners of the eukaryotic PCD pathway (discussed in section 1.1.4), and they are frequently used to trace the evolutionary history of eukaryotic PCD and support hypotheses for PCD origin (Ameisen, 2002). Phylogenetic analysis showed a close relationship of eukaryotic MCAs and paracaspases (PCAs) to CA homologues of *Rhizobia sp.*, an α-proteobacterium (Koonin and Aravind, 2002). In addition, the topology of the resulting phylogenetic tree also supported the origin of eukaryotic CA homologues from a mitochondrial endosymbiont (Koonin and Aravind, 2002). Lastly, some of the 'domains of death', additional protein domains that are required within CA homologues to act as receptors and adaptors of the cell signalling pathway (discussed in section 1.1.4), were identified in bacteria (Aravind, Dixit and Koonin, 1999a).

In contrast, the discovery of PCD in bacteria and increasing evidence from other unicellular taxa (reviewed by Bidle, 2016) show that PCD is universal for unicellular and multicellular life (Engelberg-Kulka *et al.*, 2006; Allocati *et al.*, 2015; Peeters and de Jonge, 2018). These studies also suggest that PCD may have evolved earlier than previously thought. However, the origin of PCD prior to eukaryogenesis is currently unknown. This is because PCD studies in prokaryotes have been biased towards bacteria and limited archaeal genomic data were available (Bidle *et al.*, 2010). Fortunately, a number of putative CA homologues, orthocaspases (OCAs) and MCAs,) have been reported from certain archaeal phyla (Euryarchaeota, Thermoplasmatota, Thaumarchaeota, Bathyarchaeota and Heimdallarchaeota) (Asplund-Samuelsson, Bergman and Larsson, 2012; Klemenčič *et al.*, 2019). Identification of homologues of CAs in both bacteria and archaea suggests PCD may have been present prior to eukaryogenesis. As such, we can deduce the

origin of prokaryotic PCD from the comparative analysis of its genetic toolkits, similarly to the method followed for investigating eukaryotic PCD's origin. (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Aravind, Dixit and Koonin, 1999a; Hofmann, 2019).

1.1.5 The role of PCD in the rise of eukaryotes

PCD appears to have played a much more significant part in the evolution of ancestral microbial life than previously imagined (Blackstone and Green, 1999; Ameisen, 2002; Koonin and Aravind, 2002; Nedelcu, 2009; Nedelcu *et al.*, 2011; Kaczanowski, 2016; Koonin and Zhang, 2017; Durand, Barreto Filho and Michod, 2019; Durand, 2021). PCD increases ecological fitness (Orellana et al. 2013) and acts as an altruistic mechanism in complex microbial communities, such as biofilms (Lewis, 2000). Most importantly, PCD is suggested to be an important adaptive mechanism for prokaryotes in early life and it played an important role during eukaryogenesis as a conflict mediator (Blackstone and Green, 1999; Michod and Nedelcu, 2003).

Eukaryogenesis

Eukaryogenesis is the multi-step, evolutionary process leading to the origin of eukaryotic cells and one of the major transitions in the evolution of life (Margulis, 1970; Corliss, 1987; Maynard Smith and Szathmáry, 1995; Chela-flores, 1998). The most widely accepted hypothesis of eukaryogenesis is that it occurred as a result of an endosymbiotic event which resulted in the formation of mitochondria and chloroplasts from α-proteobacteria and cyanobacteria respectively, with an archaeal host (Sagan, 1967; Margulis, 1970). This hypothesis has been established over 40 years ago (Sagan, 1967; Margulis, 1970). Recent advances in phylogenetics, cell biology and the discovery of new microbial diversity are providing new insights and hypotheses into eukaryogenesis, allowing a more accurate examination of prokaryotic contributors to eukaryotic biological features (López-García and Moreira, 2019) (Figure 1). For example, recent phylogenetic analyses suggested that the archaeal host belonged to the superphylum Asgard group (Spang *et al.*, 2015, 2018; Imachi *et al.*, 2020).

It is challenging to deduce the time point of eukaryogenesis as it involved two long transitional periods (Figure 1). The first transition was from the Last Archaeal Common Ancestor (LACA) to the First Eukaryotic Common Ancestor (FECA) of which the mitochondrion was acquired, and the second transition involved FECA obtaining the full complement of traits common to the Last Eukaryotic Common Ancestor (LECA) (Figure 1). The oldest prokaryotic fossils identified are 3.48 billion years old (giga-annum, Ga) (Shen, Buick and Canfield, 2001; Ueno *et al.*, 2008; Shen *et al.*, 2009) and prokaryotes are suggested to have existed since then (Figure 1). The first eukaryotic cell is estimated to have existed around 1.7 Ga (Knoll *et al.*, 2006) (Figure 1).

Multiple questions remain in unravelling each step of the multi-stage transition from prokaryotes to eukaryotes (López-garcía and Moreira, 2015). For example, the order of the acquisition and the

emergence of complex internal cellular structures (e.g., mitochondria, the nucleus, and the primordial endomembrane organelle) are unclear (López-garcía and Moreira, 2015) (Figure 1).

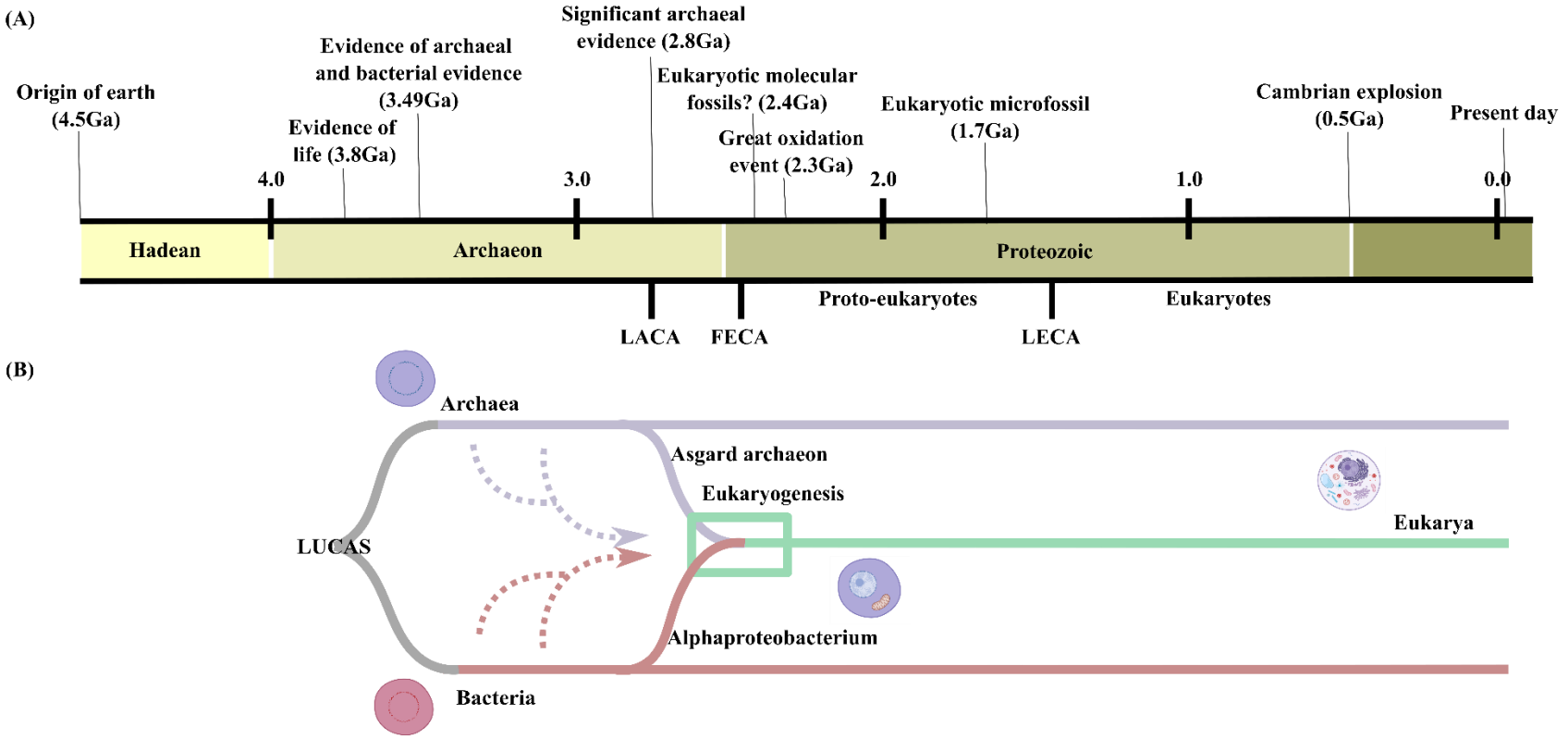

Figure 1 (A) Timeline of major events from the origin of life to extant eukaryotes on earth. The first transitional period occurred from LACA to FECA, during which the endosymbiotic event occurred. The second transitional period occurred from FECA to LECA with the acquisition of additional cellular features to develop into a eukaryotic cell. Further development of LECA resulted in current, complex eukaryotes. Bars are divided into geological epochs from the origin of Earth to the present. Scale is in billions of years (Ga). **(B)** Eukaryogenesis and the transition from LUCAS to archaea (purple) and bacteria (red), to early eukaryotes (green) possessing primordial structures, and lastly, to extant eukaryotes (green). HGT is indicated in dotted lines (adapted from Dacks, Field, Buick, *et al.*, 2016).

This is because intermediate forms showing prokaryote-eukaryote transitions are incomplete in the fossil record (López-garcía and Moreira, 2015). Different models of eukaryogenesis other than the symbiogenetic scenario have also been proposed, such as the predator-prey and the pathogen-host hypothesis (Speijer, 2020). However, the endosymbiotic model remains the most commonly accepted hypothesis.

In this dissertation, the Entangle-Engulf-Endogenize (E3) model (Imachi *et al.*, 2020) endosymbiotic model is adopted. Under the E3 model, a syntrophic/fermentative host Asgard archaeon (*Candidatus Prometheoarchaeum syntrophicum* strain MK-D1) interacted with an H_2-scavenging sulphate-reducing bacterium and a facultatively aerobic organotrophic O_2-scavenging Alphaproteobacterium (future mitochondrion). Under this model, only the aerobic Alphaproteobacterium was engulfed by the archaeal host.

PCD as a conflict mediator during eukaryogenesis

It is expected that in a major transition like eukaryogenesis, the initially independent units at a lower level of complexity (Asgard archaeon, Alphaproteobacterium and H_2-scavenging sulphate-reducing bacterium) passed through iterative cycles of cooperation, conflict and conflict mediation (Maynard Smith and Szathmáry, 1995). Cooperation is defined as "the formation and maintenance of groups due to cooperative traits and the benefit they provide to others" (Michod and Nedelcu, 2003). In the E3 model, the initial cooperation would have been the facultative symbiotic relationship where the three symbionts fed on each other's waste. However, conflict arose because the initially independent units had strong individual selection (Michod and Nedelcu, 2003). For example, the rise in the oxygen level during the Great Oxidation Event (2.3 Ga) promoted the acquisition of aerobic metabolism by the symbiotic group to adapt to the new environment. The resulting change in the metabolic complementation would have promoted a disagreement on the efficiency of production waste as the aerobic alphaproteobacterial metabolism would have yielded higher energy than the remaining symbionts (Michod and Nedelcu, 2003; Imachi *et al.*, 2020). This encouraged selfish replication and bioenergetic conflicts through reactive oxygen species (ROS) between the lower independent symbionts (Blackstone and Green, 1999; Radzvilavicius and Blackstone, 2015). This competition and selfish behaviour would have prevented the emergence of a higher level of individuality, and this conflict would have to be mediated for the symbiotic relationship to remain in a stable union as well as contribute to the emergence of individuality at a higher level (Michod and Nedelcu, 2003).

A conflict mediator refers to "a trait that reduces fitness variation in lower individuals while maintaining cooperation within the population and selecting for the higher-level units – cell-group or organism" (Michod, 2003). In the E3 model, multiple conflict mediator mechanisms have been suggested, such as HGT, endogenization of the aerobic partner by the host archaeon and the

establishment of a symbiont-to-host ATP channel (Michod and Nedelcu, 2003; Imachi *et al.*, 2020). However, in this review, I focus on a specific conflict mediator – PCD. PCD is hypothesized to have played an essential role as a conflict mediator by aligning the different evolutionary interests of individuals interacting in the primordial cell and enforcing cooperation between the initially independent units (Blackstone and Green, 1999; Blackstone, 2013, 2016). The initial hypothesis of PCD as a conflict mediator was established because of the dynamics surrounding ROS, the key role of mitochondria in the initiation of the PCD molecular pathway and its ability to manipulate the organism's fate (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998).

Blackstone suggested that primitive PCD effectors, such as CA homologues, cytochrome C and BCL-2 (B-cell lymphoma-2), were used as conflict mediators, and the resulting conflict mediation would have been through the death of the group (Blackstone and Green, 1999). The symbiotic group would have passed through iterative cycles of cooperation-conflict-conflict mediation that led to its death until a group with aligned evolutionary interests arose giving rise to FECA (Blackstone and Green, 1999; Durand, Barreto Filho and Michod, 2019; Durand, 2021) (Figure 2). Blackstone and others further suggested that the PCD molecular pathway arose as a vestige of these evolutionary conflicts (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Blackstone and Green, 1999). The constant crosstalk between the archaeon and the bacterium using PCD genetic toolkits represents a primitive PCD molecular pathway, and they suggested this pathway may have been refined by selection and converted into the eukaryotic PCD molecular pathway over time (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Blackstone and Green, 1999; Ameisen, 2002).

This hypothetical model provided by Blackstone and others is further supported by Klim and colleagues. They suggested that proto-eukaryotes possessed several apoptotic factors, such as apoptotic DNases, CA homologues, and AIFs (Apoptosis-inducing factor) (Klim *et al.*, 2018). However, they suggested that PCD arose due to an evolutionary arms race between the proto-eukaryotic host ("predators") and the alphaproteobacterium ("prey"), of which the alphaproteobacterium produced toxins as a defence mechanism against the host which later evolved into apoptotic factors (Klim *et al.*, 2018; Kaczanowski, 2020).

One of the limitations of the hypothetical model provided by Blackstone and others is that it only considered cooperation and conflict between two symbionts - the Alphaproteobacterium and the Asgard archaeon (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Blackstone and Green, 1999; Ameisen, 2002). However, the current E3 model involved an interaction between three symbionts (Deltaproteobacterium, Alphaproteobacterium and Asgard archaeon) for the rise of FECA (Imachi *et al.*, 2020). The metabolic complementation, ROS dynamics, and conflict mediation in a three-symbiont E3 model were likely different to a two-symbiont model (Michod and

Nedelcu, 2003), and it is unclear whether PCD would have arisen under similar conflict and conflict mediator mechanisms. Further analysis of the E3 model and resulting conflict is required for a more refined model of the role of PCD as a conflict mediator during eukaryogenesis.

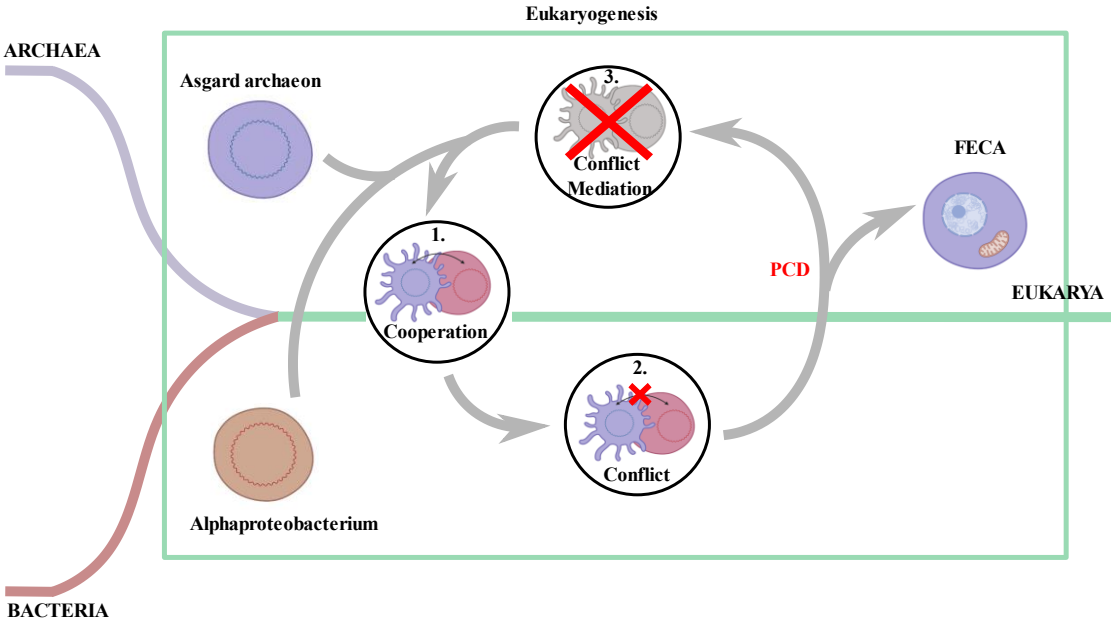

Figure 2 Iterative cycles of cooperation, conflict, and conflict mediation during eukaryogenesis. The three domains of life: archaea (purple), bacteria (red) and eukarya (green), are shown in lines to show the transition from prokaryotes to eukaryotes. The two symbionts, Asgard archaeon (purple) and Alphaproteobacteria (red), underwent cycles of cooperation-conflict-conflict mediation with PCD as the conflict mediator until FECA arose.

1.1.6 Classification of CAs and CA homologues: key effectors of PCD

PCD requires multiple essential and accessory components in its molecular pathway (reviewed by Galluzzi and Vitale, 2018). However, this study focuses on one of the main effectors in the PCD molecular pathway – CAs and CA homologues. CA- and MCA-independent PCD exists as well (Cheng *et al.*, 2001; Bloss, Witze and Rothman, 2003; Tait and Green, 2008; Carmona-Gutierrez *et al.*, 2011) but they will not be discussed in this dissertation as their mechanisms are not well understood in prokaryotes.

CAs are classified as a part of the caspase-hemoglobinase fold (CHF) family of proteases (Aravind and Koonin, 2002). This is based on the conserved structural core of the unique α/β-fold (Walker *et al.*, 1994; Wilson *et al.*, 1994) and the His-Cys (HC) dyad, which is a requirement for their proteolytic activity (Aravind and Koonin, 2002). Iterative sequence searches revealed the diversity of CA-related

proteases were underappreciated which led to the inclusion of diverse thiol proteases, such as gingipains, clostripains (Chen *et al.*, 1998) and separins (Uhlmann *et al.*, 2000), as well as CA homologues, PCAs and MCAs (Uren *et al.*, 2000) (Figure 3). The classification system is shown in Figure 3, based on the similarity-based clustering, symmetry of recovery in PSI-BLAST searches, phylogenetic analysis and cladistic analysis based on shared derived characters (Aravind and Koonin, 2002).

Within the CHF family of proteases, CAs (subfamily C14A) and CA homologues (subfamily C14B) are subclassified into the C14 peptidase family on the MEROPS database (Rawlings & Barrett, 1993; Rawlings, Waller, Barrett, *et al.*, 2014) (Figure 3). The C14 peptidase family consists of cytosolic endopeptidases that possess the C14 peptidase domain (Pfam ID: PF00656). The catalytic HC dyad is present within the C14 peptidase domain.

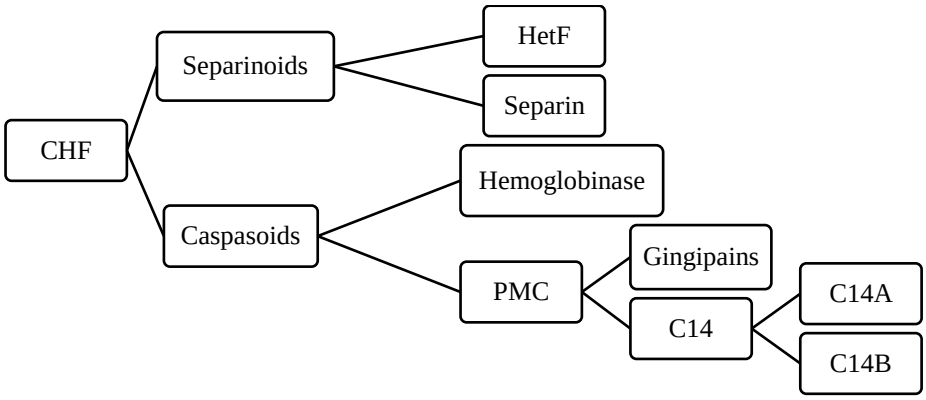

Figure 3 Classification of the CHF family by Aravind and colleagues, and the C14 peptidase family by MEROPS. PMC refers to the Paracaspase–Metacaspase–Caspase subclass (adapted from Aravind and Koonin, 2002).

1.1.7 The C14 peptidase family

The C14 peptidase family, a large protease family of cytosolic endopeptidases (Chen *et al.*, 1998), consists of CAs and CA homologues - MCAs, PCAs and OCAs (Uren *et al.*, 2000; Klemenčič, Novinec and Dolinar, 2015). MCAs and PCAs were first identified by Uren and colleagues, based on the predicted structural homologies with the catalytic domain in CAs (Uren et al., 2000). MCA-like proteases, primitive CA homologues with a simple protein domain architecture, were identified by Choi and Berges (2013). MCA-like proteases were identified later in bacteria by Klemenčič and colleagues and were named OCAs (Klemenčič, Novinec and Dolinar, 2015).

Different terminologies have been used to describe the properties of different members of the C14 peptidase family which can be misleading. To have a unified set of terminologies, the nomenclature

suggested by Minina and colleagues will be adopted in this dissertation (Minina *et al.*, 2020) (Figure 4A).

(A)

C14 peptidase	Protein domain organization	Specificty in P1	Requirement for activation	Self-inhibition
True OCA	H C p20	Arg	Unknown	Unknown
Pseudo OCA	Y C p20	Arg	Unknown	Unknown
CA	N-terminal pro-domain — H C p20 — p10	Asp	Apoptosome or initiator CAs	IAP proeins
Type I MCA	N-terminal pro-domain — H C p20 — p10	Arg and Lys	Ca2+ and/or proteolysis	N-terminal region
Type II MCA	H C p20 — : — p10	Arg and Lys	Ca2+ and/or proteolysis	N-terminal region
Type III MCA	p10 — H C p20	Arg and Lys	Ca2+ and/or proteolysis	N-terminal region
Type I PCA	N-terminal pro-domain — H C p20 — p10	Arg	Dimerization	Substrate-binding loop (L5)
Type II PCA	H C p20 — p10	Arg	Dimerization	Substrate-binding loop (L5)
MCA-like protease	H C p20	Arg	Unknown	Unknown

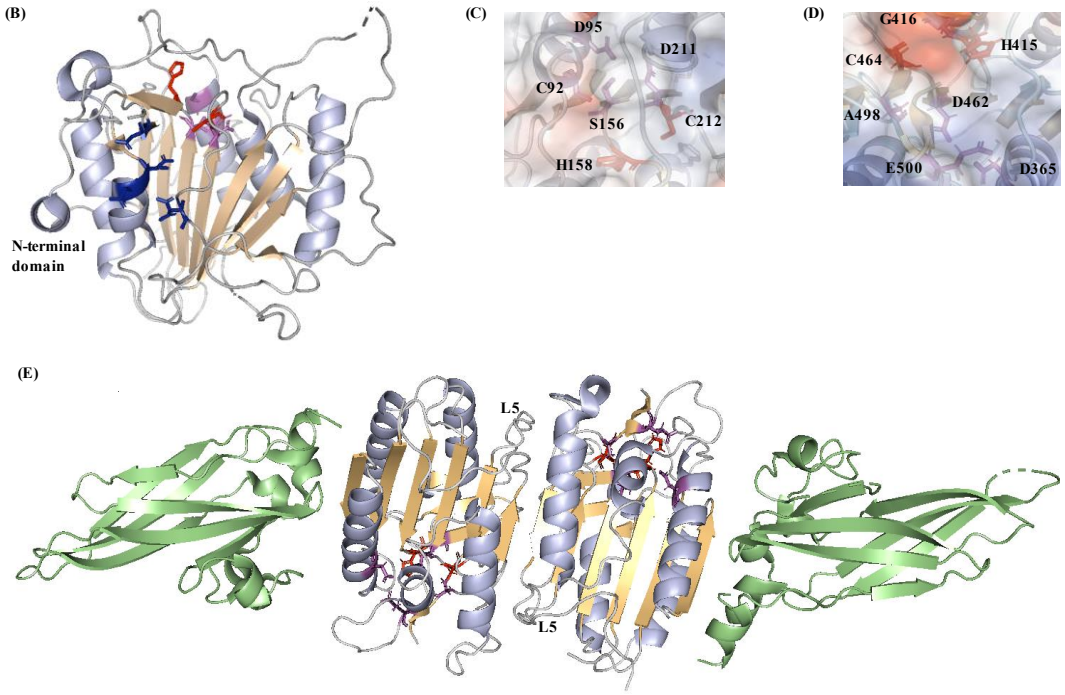

Figure 4 Domain organization, structure analysis and substrate specificities of C14 peptidases **(A)** Classification and domain organization of C14 peptidases according to the unified nomenclature of the C14 peptidase family (adapted from Minina *et al.*, 2020). Members of the family possess the catalytic p20-like region (maroon), a regulatory p10-like region (blue) and/or the linker region (thick grey line). The positions of catalytic His (H) and Cys (C) residues are shown above the p20-like region. The most common Tyrosine-Cysteine (YC) dyad is shown for the pseudo-OCA. N-terminal pro-domains are

indicated in green. The figure is not drawn to scale. **(B)** Automated I-TASSER protein structure prediction of an inactive (Cys213Gly/Ala) mutant of MCA obtained from *Trypanosoma brucei* (TbMCA-Ib) shown in ribbon representation (β-sheets in beige and α-helices in grey) visualized using PyMOL (DeLano Scientific; http://www.pymol.org). The catalytic HC dyad (red), key substrates (purple) and Ca^{2+} binding sites (blue) are shown. **(C)** Detailed view of the S1 pocket of the model structure of TbMCA-Ib. Key residues are indicated with the amino acid positions as well as the electrostatic surface potential, of which blue and red denote positively and negatively charged surface areas. **(D)** Detailed view of the S1 pocket of the model structure of MALT1 of *H. sapiens*. Key residues are indicated with the amino acid positions as well as the electrostatic surface potential, of which blue and red denote positively and negatively charged surface areas. **(E)** Automated I-TASSER protein structure prediction of the human mucosa-associated lymphoid tissue (MALT-C) homodimer of *H. sapiens* shown in ribbon representation (Immunoglobin (Ig) domain in green, β-sheets in beige and α-helices in grey) visualized using PyMOL. The catalytic HC dyad (red) and key substrates (purple) are shown.

Caspases

CAs are major effectors and regulators of apoptotic cell death pathways, proliferation and inflammation in Metazoa (Fuentes-Prior and Salvesen, 2004; MacKenzie and Clark, 2012). They have specificity for protein substrates that contain an aspartic acid (Asp) in the P1 position (Wei *et al.*, 2000). They possess the p20-like region, the p10-like region and N-terminal prodomains (Figure 4A), which includes the Asp site for proteolysis (Figure 4A). In addition to the Asp, Arg87, Gln184, Arg233 (named according to CASP7) are responsible for the specificity of the S1-binding site (Wei *et al.*, 2000; Chai *et al.*, 2001).

Metacaspases

MCAs are ancestral CA homologues (Choi and Berges, 2013) that are involved in a wide range of functions including PCD (Bozhkov *et al.*, 2005; González *et al.*, 2007; Bidle and Bender, 2008; Coll *et al.*, 2010), cell-cycle progression (Lee *et al.*, 2008), cell proliferation (Castanys-Muñoz *et al.*, 2012), endoplasmic reticulum (ER) stress response (Richie *et al.*, 2007), clearance of insoluble aggregates (Lee *et al.*, 2010) and virulence (Proto *et al.*, 2011). It is unclear if MCAs are involved in PCD in prokaryotes because the link between PCD and MCAs has only been established for selected plant and unicellular eukaryote taxa (Bozhkov *et al.*, 2005; González *et al.*, 2007; Bidle and Bender, 2008; Coll *et al.*, 2010).

MCAs have a strict preference for substrates containing basic arginine (Arg) and/or lysine (Lys) residues (Vercammen *et al.*, 2004; Watanabe and Lam, 2005; González *et al.*, 2007; Moss *et al.*, 2007). The domain architecture of an MCA consists of a p20-like region, a linker region, and a p10-like

region (Figure 4A). They can be further classified according to the domain architecture: MCA type I, MCA type II and MCA type III (Figure 4A). Type I MCA have a prodomain, a p20-like region, a short linker region (28.6±4.7) and a p10-like region. Type II MCAs don't have the prodomains and have a long linker region (161.3±32.9), and type III MCAs possess the p10-like region prior to the p20-like region (Choi and Berges, 2013).

The structural basis for many of the functional differences between the MCAs and CAs was revealed by the first MCA structure, an inactive (Cys213Gly/Ala) mutant of MCA obtained from *Trypanosoma brucei* (TbMCA-Ib) (McLuskey *et al.*, 2012). Several residues were shown to be involved in substrate binding and/or enzyme activity (McLuskey *et al.*, 2012): Cys92, Asp95, Ser156 and Asp211 (Figure 4B and 4C). In addition, MCA activity is calcium-dependent (Bozhkov *et al.*, 2005; Moss *et al.*, 2007; Meslin *et al.*, 2011; Watanabe and Lam, 2011). High-affinity Ca^{2+} binding sites were identified (Figure 4B), which are highly conserved in a primary sequence alignment of both type I and type II MCAs (McLuskey *et al.*, 2012). Additional low-affinity Ca^{2+} binding sites were identified in type III MCA as well (Asp107 or Asp113) (Klemenčič and Funk, 2018b) (Figure 4B). Furthermore, N-terminal prodomains are important for the function and activity regulation of MCAs. For example, the N-terminal region acts as a gatekeeper in TbMCA-Ib, controlling substrate access to the active site (McLuskey *et al.*, 2012) (Figure 4B).

Paracaspases

PCAs are calcium-independent active dimers (Uren *et al.*, 2000; Hachmann *et al.*, 2012) with different activation and associated upstream mechanisms to the above-mentioned C14 peptidases (Minina et al., 2020). PCAs are involved in several pro-inflammatory pathways in innate- and adaptive- immunity in Metazoa (Jaworski and Thome, 2016) as well as regulation of osmotic stress tolerance by vacuolar expansion (Saheb *et al.*, 2013).

Two different types of PCAs exist. Type I PCAs all have a human mucosa-associated lymphoid tissue translocation protein 1 (MALT1) (Yu *et al.*, 2011; Wiesmann *et al.*, 2012) like domain composition (Hulpiau *et al.*, 2016). Their full-length protein comprises an N-terminal death domain (DD), followed by two Ig-like domains (Ig1 and Ig2) and the p20-like region (Figure 4A and 4E). They are Arg-specific in the P1 position (Uren *et al.*, 2000). Type II PCAs have a simple domain architecture and only possess the C14 peptidase domain (Hulpiau *et al.*, 2016).

An available PCA structure belongs to a type I PCA - human mucosa-associated lymphoid tissue translocation protein 1 (MALT1) (Yu *et al.*, 2011; Wiesmann *et al.*, 2012) (Figure 4E). Sequence and structural analyses revealed four residues are required for the P1 Arg substrate specificity: Asp365, Gly416, Ala498 and Glu500 (Figure 4D).

Orthocaspases

OCAs are the most evolutionarily primitive CA homologues in the C14 peptidase family that only possess the conserved catalytic p20-like region (Klemenčič, Novinec and Dolinar, 2015) (Figure 4A). They were initially named MCA-like proteases in eukaryotes (Choi and Berges, 2013) until they were identified in prokaryotes (Klemenčič, Novinec and Dolinar, 2015). OCAs of cyanobacterium *Microcystis aeruginosa* PCC 7806 (MaOC1) were experimentally studied and suggested to be an endopeptidase with a preference for Arg in P1 position (Klemenčič, Novinec and Dolinar, 2015). To my knowledge, no crystallography structure is available for OCAs and their association with PCD is yet to be confirmed.

Comparison of different members of the C14 peptidase family

Sequence and structure comparison of the identified active site regions of CAs, PCAs, MCAs and OCAs revealed that they all use structurally homologous regions for inhibitor/substrate recognition and that the catalytic HC dyad resides in similar positions (Choi and Berges, 2013; McLuskey and Mottram, 2015) (Figure 4A). However, different members exhibit a variety of substrate specificities, activation and inhibition mechanisms, and possess different N-terminal domains.

CAs and the rest of the C14 peptidase members exhibit different substrate specificities. CAs are Asp-specific whereas PCAs and OCAs are Arg-specific and MCAs are Arg- or Lys- specific (Vercammen *et al.*, 2004; Watanabe and Lam, 2005; Klemenčič, Novinec and Dolinar, 2015) (Figure 4A). It is important to distinguish between the two different modes of enzyme activity owing to different substrate specificities. The presence of CA-like enzyme activity doesn't provide positive evidence for the presence of OCAs, MCAs or PCAs, or vice versa. On the other hand, serine proteases are known to exhibit CA-like activity in plants (Egger *et al.*, 2003; Coffeen and Wolpert, 2004). It is important to note that CA-like activity is observed in enzymes other than C14 peptidases whereas MCA-like activity is highly specific for C14 peptidases. Further studies to elucidate enzymes responsible for CA- and MCA-like enzyme activity are required.

1.1.8 Domains of death

Domains of death are Pfam domains identified within a C14 peptidase sequence, adjacent to the C14 peptidase domain (Aravind, Dixit and Koonin, 1999a). They are required for a C14 peptidase to be involved in PCD (Aravind, Dixit and Koonin, 1999a) (Figure 4A). Aravind and colleagues suggested a list of the most important domains of death that act as receptors and adaptors of the PCD signalling pathway. These include the Death Domain (DD), Death Effector Domain (DED), and Caspase Activation and Recruitment Domain (CARD) (Aravind, Dixit and Koonin, 1999a) (Table 2). The DD–DED–CARD triad of domains is typically found in Metazoa whereas other domains, such as Meprin-

And-TRAF-Homology (MATH) domain and Toll/interleukin-1 receptor (TIR) homology domain, are widely distributed in diverse lineages (Aravind, Dixit and Koonin, 1999a). TIR is usually found in cohorts with Leucine-rich (Leu-rich) repeat (LRR) domain and apoptotic ATP synthase (AP-ATPase) domains in plants and is involved in resistance to pathogens and stress response (Aravind, Dixit and Koonin, 1999a; DeYoung and Innes, 2006; Saucet, Esmenjaud and Van Ghelder, 2020). However, C14 peptidases possess additional domains in addition to those related to death and their functions are not fully understood. (Asplund-Samuelsson, Bergman and Larsson, 2012).

Table 2 Taxonomic distribution of domains of death in eukarya and bacteria**.** Abbreviations used: Bcl-2-associated X (BAX), B cell lymphoma-2 (BCL-2), Baculovirus IAP Repeat (BIR), CA activation and recruitment domain (CARD), Cocaine and Amphetamine Regulated Transcript protein (CART), death domain (DD), death-effector domain (DED), Leu-rich repeat (LRR), Meprin and TRAF (tumour necrosis factor receptor-associated factor) homology (MATH), PKD (Polycystic Kidney Disease), Toll–interleukin-receptor (TIR), tumour necrosis factor (TNF), and WD or beta-transducin repeat (WD40) (adapted from Aravind, Dixit and Koonin, 1999b).

Ligands, intracellular domains, adaptor domains and membrane proteins	**Eukarya**	**Bacteria**
TNF family	+	+
TNFR, extracellular domain	+	+
LRR	+	+
BCL-2/BAX	+	-
DD	+	-
DED	+	-
CARD	+	-
TIR	+	+
BIR	+	-
MATH	+	-
CART	+	-
WD40	+	+
PKD	+	+

A study by Asplund-Samuelsson and colleagues investigated the presence of these domains of death in sequenced prokaryotic genomes (Asplund-Samuelsson, Bergman and Larsson, 2012). They observed numerous, small protein-protein interaction death domains in many prokaryotic MCAs, such as the WD40, TPR_1 (Tetratricopeptide repeat_1), and TPR_2 domains (Table 2). They suggested

identification of domains of death in prokaryotic MCAs is indicative of a connection with PCD (Asplund-Samuelsson, Bergman and Larsson, 2012). In addition, many MCA-associated domains seem to have an extracellular localization, as indicated by the presence of N-terminal transmembrane helices. They suggested these MCAs perform their role in PCD on the surface of the cell or engage with neighbouring cells (Asplund-Samuelsson, Bergman and Larsson, 2012).

1.1.9 Taxonomic distribution of C14 peptidases

C14 peptidases are widely distributed across the three domains of life. However, certain members are lineage-specific. CAs and PCAs are limited to domain eukarya of which CAs are only present in Metazoa and PCAs are present in Metazoa and slime moulds (Uren *et al.*, 2000). MCAs are widely distributed in all lineages of eukarya and bacteria, except for Metazoa (Uren *et al.*, 2000; Asplund-Samuelsson, Bergman and Larsson, 2012; Choi and Berges, 2013). OCAs or MCA-like proteases are present in both unicellular eukarya, bacteria and archaea (Asplund-Samuelsson, Bergman and Larsson, 2012; Choi and Berges, 2013; Klemenčič, Novinec and Dolinar, 2015; Klemenčič *et al.*, 2019). Although C14 peptidases were identified in archaea (Asplund-Samuelsson, Bergman and Larsson, 2012; Klemenčič *et al.*, 2019), refined taxonomic distribution analysis of C14 peptidases in different archaeal lineages is absent.

The patchy distribution of different C14 peptidases across the three domains of life has been previously explained by the Black Queen hypothesis (BQH) (Ndhlovu, Durand and Ramsey, 2021). A Black Queen hypothesis aims to explain reductive genome evolutionary events in some microbial lineages and why certain essential functions are rare in certain communities (Morris, Lenski and Zinser, 2012). According to Morris, certain microbial lineages have undergone adaptive gene loss events, where the advantage was to conserve the organism's limiting resources where the gene's function is crucial to the community, enabling division of labour in microbial communities (Morris, Lenski and Zinser, 2012). An example is the production of catalase-peroxidase (katG), an enzyme that is expensive to produce but essential for defence against external hydrogen peroxide (Morris, Papoulis and Lenski, 2014). As the enzyme is ‘leaky’, some of the microbial lineages underwent adaptive gene loss events but the trait was still maintained within the population. The patchy distribution of MCAs was suggested as support for PCD as a BQ trait (Ndhlovu, Durand and Ramsey, 2021). However, further analyses are required to assess if gene loss events of C14 peptidases in certain lineages is suggestive of PCD as a BQ trait in microbial communities (Ndhlovu, Durand and Ramsey, 2021).

1.1.10 Tools for identifying distantly related CA homologues in prokaryotes

Numerous tools can be used to detect CA homologues in distantly related organisms as well as trace their evolutionary history. TBLASTN, BLASTP and PSI-BLAST (Altschul *et al.*, 1990, 1997) were previously used to search protein sequence databases to identify distantly related CA homologues and

domains of death (Aravind, Dixit and Koonin, 1999a; Uren *et al.*, 2000; Choi and Berges, 2013). However, HMMER is the currently most widely used homology detection of CAs in distantly related organisms (Asplund-Samuelsson, Bergman and Larsson, 2012; Klim *et al.*, 2018; Hofmann, 2019; Klemenčič *et al.*, 2019). It also allows the detection of distantly related domains of death that are present in a CA homologue protein sequence (Asplund-Samuelsson, Bergman and Larsson, 2012; Hofmann, 2019). HMMER is a useful tool to perform sequence analysis using profile hidden Markov models (HMMs). HMMs are probabilistic models of protein and DNA sequence domain families (Eddy, 1998). They are created by converting a multiple sequence alignment (MSA) into a position-specific scoring matrix. HMM profiles can capture the amino acid bias in certain positions of the MSA. Comparison of the HMM profile to a database of sequences allows the detection of remotely homologous proteins as sensitively as possible (Eddy, 1998) and it is often used in cohort with the Pfam database (http://pfam.xfam.org/). Identified CA homologue protein sequences and domains residing in them can be further analysed using protein domain architecture analyses and investigate their association with PCD (Aravind, Dixit and Koonin, 1999a; Asplund-Samuelsson, Bergman and Larsson, 2012).

1.1.11 C14 peptidases in archaea

Archaea, one of the major domains of life, has been largely overlooked from the PCD discussion due to the previous lack of CA or CA homologue sequences detected in these organisms (Bidle *et al.*, 2010). However, C14 peptidase sequences belonging to archaea have been the focus of recent studies (Asplund-Samuelsson, Bergman and Larsson, 2012; Klemenčič *et al.*, 2019), which has identified 93 putative C14 peptidases sequences belonging tovarious phyla of archaea. The discovery of new archaeal lineages, superphyla TACK (Elkins *et al.*, 2008; Nunoura *et al.*, 2011; Spang *et al.*, 2015) and Asgard group (Spang *et al.*, 2015; Eme *et al.*, 2017; Zaremba-Niedzwiedzka *et al.*, 2017), as well as declining costs and improved sequencing methods in the last decade have resulted in increased sequencing of prokaryotes, which has provided additional genomic information of archaea (Land *et al.*, 2015; Srivastav and Suneja, 2019; Zhang *et al.*, 2020).

The dilemma of CA-like activity in archaeal organisms with no C14 peptidases

Salt and temperature stress-induced CA-like activity have been detected in a model organism of archaea, *Haloferax volcanii* (Bidle *et al.*, 2010). Inhibition of CA-like activity by using the broad-spectrum CA inhibitor (Z-VAD-FMK) severely impaired cell growth under low and high salt stress, demonstrating a critical role in the cellular stress response. High CA-like activity was also detected in organisms *Halorubrum* sp. SC3, *Haloarcula* sp. SC8 of phylum Euryarchaeota and *Sulfolobus solfataricus* of superphylum TACK (Bidle *et al.*, 2010). Further investigation of catalytic specificity of CA-like activity in *H. volcanii* revealed they exhibit highly specific CA-like activity, closely

resembling CA-4 (Seth-Pasricha, Bidle and Bidle, 2013). Bidle and colleagues suggested that the detected archaeal CA-like activity was responsible for normal cell function and these enzymes exhibiting CA-like activity appeared to have co-evolved with other metabolic pathways, broadening their biological roles beyond apoptosis and cell death (Bidle *et al.*, 2010). However, *in silico* analyses revealed no C14 peptidase sequences in the studied organisms. On the other hand, archaea possessing C14 peptidases, *Methanosarcina* sp. and *Pyrococcus furiosus* exhibited no CA-like activity. Therefore, it is likely that archaeal C14 peptidases were not responsible for CA-like activity detected in the above-mentioned archaeal organism, and it is uncertain if archaeal C14 peptidases are associated with functions other than cell death. The MCA-like activity would likely have been detected in archaeal organisms possessing C14 peptidases.

1.1.12 Model organism: *Halobacterium salinarum* NRC-1

H. salinarum is a rod-shaped, motile organism that lives in highly saline environments (4M salt and higher) (Lanyi, 1974; Magrum, Luehrsen and Woese, 1978) (Figure 5). It belongs to the phylum Euryarchaeota, class Halobacteria, order Halobacteriales and family Halobacteriaceae (Elazari-Volcani, 1940). They can be recognized by their typical reddish colour, which originates from bacterioruberins (Shahmohammadi *et al.*, 1998). The genome of *H. salinarum* NRC-1 is 2571010 bp in total consisting of three circular replicons (a large chromosome and two novel megaplasmids) (Ng *et al.*, 2000; Pfeiffer *et al.*, 2020). Annotation of its genome found 2611 protein-coding genes (Pfeiffer *et al.*, 2020). Lateral gene transfer between halobacteria and haloarchaea are common due to the highly saline, dynamic environment (Cline and Doolittle, 1987; Becker *et al.*, 2014).

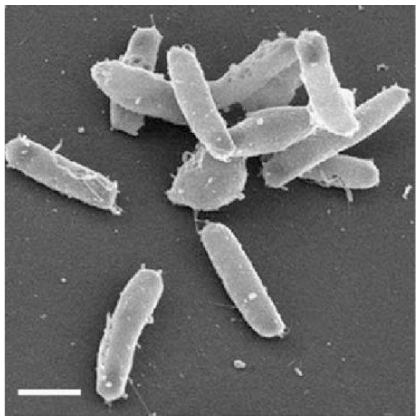

Figure 5 Scanning electron microscopy of *H. salinarum* strain NRC-1. The scale bar (white line) represents l µm (Gruber *et al.*, 2004).

H. salinarum is a good model organism for archaea that represents the halophilic branch of the domain (Peck, DasSarma and Krebs, 2000). This is because *H. salinarum* is easy to culture as it is both aerobic

and mesophilic (Ng *et al.*, 2000). In addition, previously performed large scale PCR and DNA arrays, as well as developed gene replacement and systematic knockouts, make the organism an ideal candidate for the study of archaeal genetics and functional genomics (Peck, DasSarma and Krebs, 2000). However, it is unknown if *H. salinarum* undergoes PCD or if it possesses C14 peptidases. To my knowledge, no studies to detect CA- or MCA-like activity has been performed in this organism.

1.2 Research aims and objectives

Although PCD is suggested as one of the conflict mediators during eukaryogenesis, it is unclear if PCD was a trait possessed by ancestral prokaryotes involved in this major transition. This is because there has been a previous dearth of studies of PCD in prokaryotes, especially archaea. This research focuses on investigating the presence of members of the C14 peptidase family in archaea and determining if PCD has a more ancestral origin prior to eukaryogenesis. Elucidating the evolutionary history of C14 peptidases in archaea and eventually, the three domains of life will suggest when PCD genetic toolkits originated and how they contributed as a conflict mediator during eukaryogenesis.

The broad aim of this work was to examine C14 peptidases in archaea and analyse their evolutionary history, as well as to investigate CA-like and MCA-like protease activity in a model archaeal organism, *H. salinarum* strain NRC-1. To achieve this, the following specific objectives were set:

- Construct a robust evolutionary history of archaeal C14 peptidases and re-evaluate their origin through sequence and phylogenetic analyses.
- Investigate the ancestral role of archaeal C14 peptidases through phyletic pattern analysis and investigation of death domains to provide a refined perspective on the role of C14 peptidases and PCD as a conflict mediator during eukaryogenesis.
- Elucidate if C14 peptidases in archaea are associated with PCD through *in vivo* fluorometric analysis and measure their proteolytic enzyme activity.

CHAPTER 2

THE ANCIENT ORIGINS OF DEATH DOMAINS SUPPORT THE 'ORIGINAL SIN' HYPOTHESIS FOR THE EVOLUTION OF PROGRAMMED CELL DEATH.

2.1 Introduction to the publication

The major role played by mitochondria (originating from Alphaproteobacterium) in PCD and its ability to control the organism's fate as well as the dynamics involving ROS and CAs suggested that PCD was a key mediator in the rise of eukaryotes. However, it is unknown if ancestral prokaryotes prior to eukaryogenesis possessed CA homologues. This is due to a lack of study of the origin of PCD genes in prokaryotes, with archaea being largely overlooked in the PCD discussion because of a previous dearth of genomic data. According to my knowledge, there is no comparative analysis of archaeal CA homologues and other death domains. Recent advances in genome sequencing and culturing methods have provided additional genomic archaeal data, which allowed me to perform a comparative analysis of CA homologues in archaea and determine if PCD genetic toolkits have a more ancestral origin prior to eukaryogenesis. This publication has been accepted by the Journal of Molecular Evolution. Figures and tables are numbered according to the guideline of the Journal of Molecular Evolution.

2.2 Contribution to the evolutionary history of C14 peptidases in archaea and pre-eukaryotes

The publication provides the first *in silico* analysis of archaeal CA homologues which revealed a more ancestral origin of one of the essential PCD genetic toolkits. Based on the data, CA homologues may have emerged very early on, prior to the diversification of archaea and bacteria. However, additional analyses including sequences from metagenomics and single-celled genomics are required to provide a more robust evolutionary history of C14 peptidases and death domains in the three domains of life. Single-celled genomics is a powerful tool to sequence genomes of individual unicellular organisms without any cultures of organisms (Blainey, 2013) which is an extreme advantage for organisms belonging to phyla TACK or the Asgard group as they are challenging to culture (Imachi *et al.*, 2020). It is important to note that a domain and/or sequence homology search of archaeal genomes to identify MCAs and OCAs, rather than an annotation search to identify the initial list of organisms possessing these proteins, may provide a less biased list of organisms possessing these proteins.

Phyletic pattern analysis revealed that these archaeal CA homologues were associated with cell survival than cell death, supporting the 'original sin' hypothesis of the origin of PCD. However, C14 peptidases possess highly complex domain architectures and widespread taxonomic distributions,

making it difficult to investigate the origin and evolution of PCD from a single protein family. A comparative analysis of diverse domains of death in archaea, in addition to C14 peptidases would provide a more nuanced evolutionary history of PCD prior to eukaryogenesis. The potential role of CA homologues in Asgard archaea during eukaryogenesis, according to one of the most updated Entangle-Engulf-Endogenize (E3) model of eukaryogenesis, has been provided.

The ancient origins of death domains support the 'original sin' hypothesis for the evolution of programmed cell death.

So Ri La[1], Andrew Ndhlovu[2], Pierre M. Durand[3]

[1]Evolutionary Studies Institute, University of Witwatersrand, Braamfontein, Johannesburg, South Africa. (ORCID: 0000-0002-7052-3902)
[2]Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Matieland, Stellenbosch, South Africa. (ORCID: 0000-0003-0984-240X)
[3]Evolutionary Studies Institute, University of Witwatersrand, Braamfontein, Johannesburg, South Africa. (ORCID: 0000-0003-2925-2132)

Corresponding Author: So Ri La, E-mail: lsori415@gmail.com

Acknowledgements. This work was supported by Paleontological Scientific Trust (PAST) Africa (https://past.org.za/) and Genus: DSI-NRF Centre of Excellence in Palaeosciences (https://genus.africa/).

ABSTRACT

The discovery of caspase homologs in bacteria highlighted the relationship between programmed cell death (PCD) evolution and eukaryogenesis. However, the origin of PCD genes in prokaryotes themselves (bacteria and archaea) is poorly understood and a source of controversy. Whether archaea also contain C14 peptidase enzymes and other death domains is largely unknown because of a historical dearth of genomic data. Archaeal genomic databases have grown significantly in the last decade, which allowed us to perform a detailed comparative study of the evolutionary histories of PCD-related death domains in major archaeal phyla, including the deepest branching phyla of *Candidatus* Aenigmarchaeota, *Candidatus* Woesearchaeota, and Euryarchaeota. We identified death domains associated with executioners of PCD, like the caspase homologs of the C14 peptidase family, in 321 archaea sequences. Of these, 15.58% were metacaspase type I orthologues and 84.42% were orthocaspases. Maximum likelihood phylogenetic analyses revealed a scattered distribution of orthocaspases and metacaspases in deep-branching bacteria and archaea. The tree topology was incongruent with the prokaryote 16S phylogeny suggesting a common ancestry of PCD genes in prokaryotes and subsequent massive horizontal gene transfer coinciding with the divergence of archaea and bacteria. Previous arguments for the origin of PCD were philosophical in nature with two popular propositions being the "addiction" and 'original sin' hypotheses. Our data support the 'original sin' hypothesis, which argues for a pleiotropic origin of the PCD toolkit with pro-life and pro-death functions tracing back to the emergence of cellular life - the Last Universal Common Ancestor State.

Keywords: Programmed cell death, archaea, caspase homologs, eukaryogenesis, original sin hypothesis

INTRODUCTION

Over the last few decades, it has emerged that programmed cell death (PCD) played a much more significant part in the evolution of life than previously imagined (Ameisen, 2002; Michod and Nedelcu, 2003; Segovia *et al.*, 2003; Blackstone, 2013; Iranzo *et al.*, 2014; Koonin and Zhang, 2017; Durand, 2021). PCD fulfills a critical role in the eco-evolutionary dynamics of prokaryote and unicellular eukaryote communities (Vardi *et al.*, 2007; Bidle, 2016; Durand, Sym and Michod, 2016; Abada and Segev, 2018; Ndhlovu, Durand and Ramsey, 2021) and tissue homeostasis in multicellular organisms (Glücksmann, 1951; Lockshin and Williams, 1964). In addition, PCD was essential for the evolutionary transitions in individuality that characterize the history of life, such as eukaryogenesis, multicellularity, and eusociality in insects (Iranzo et al. 2014; Durand et al. 2019).

It is important to acknowledge at the outset, that the very subject of PCD in microbes has been fraught. This is largely because the term 'PCD' was initially discovered and subsequently interpreted as a hallmark of ontogenesis in Metazoa (Glücksmann, 1951; Lockshin and Williams, 1964) and the very

concept of programmed microbial death in some unicellular lineages was questioned (Ratel *et al.*, 2001; Nedelcu *et al.*, 2011; Proto, Coombs and Mottram, 2013; Ramisetty, Natarajan and Santhosh, 2015). Different researchers often imply different things when referring to the various kinds of cell death, which can lead to a conflation of questions and ideas. To circumvent this, there are now both mechanistic and evolutionary definitions. The mechanistic definition we adopt is the general Berman-Frank interpretation that PCD is "an active, genetically controlled, cellular self-destruction driven by a series of complex biochemical events and specialized cellular machinery" (Berman-Frank *et al.*, 2004). Evolutionists use a historical account of whether there was a selection for the trait or not and draw a distinction between PCD as an adaptation (true PCD) and non-adaptive PCD when the trait is a pleiotropic phenomenon (also called ersatz PCD) (Durand and Ramsey, 2019). For the purposes of this study and for investigating the hypotheses below, we use the Berman-Frank mechanistic definition. The evolutionary definitions can be ignored because the 'selection for' and 'selection of' distinction is not drawn here.

There have been several hypotheses for the origin of PCD, based largely on philosophical and conceptual works, which can now be examined by comparative analyses of death domains in prokaryotes (bacteria and archaea) (Aravind, Dixit and Koonin, 1999b; Asplund-Samuelsson, Bergman and Larsson, 2012; Hofmann, 2019). One of the hypotheses for the emergence of PCD in the earliest cells is the `original sin' hypothesis, which postulates that ancestral effector genes of PCD were present at the origin of cellular life (Ameisen, 2002). The argument is that PCD effectors may have been rooted in primordial non-death functions, such as the cell cycle or adaptations to environmental stress (Shrestha and Megeney 2012; Shalini et al. 2015), with subsequent pro-death functions emerging. Crosstalk between molecular pathways in extant unicellular organisms is consistent with this claim (van Creveld et al. 2018). There are other hypotheses that may be complementary and synergistic such as the 'addiction' hypothesis (Ameisen, 2002) and the argument for life-death coevolution in the last universal common ancestor state (LUCAS) (Durand, 2021). The different hypotheses may not be mutually exclusive; however, the 'original sin' hypothesis is much more accessible with a comparative genomics approach and is the focus here.

Genomic data have linked eukaryogenesis with PCD in bacteria (Blackstone and Green, 1999; Koonin and Aravind, 2002; Michod and Nedelcu, 2003). Many of the 'death domains' have been identified and the "bacterial connection" between PCD and eukaryogenesis is well established (Koonin and Aravind, 2002). While it is known that PCD played a part in the rise of eukaryotes, potentially as a conflict mediator (Blackstone and Green, 1999; Michod and Nedelcu, 2003; Iranzo *et al.*, 2014), and continues to impact the social lives of microbes (Arnoult *et al.*, 2001; Durand, Rashidi and Michod, 2011; Bayles, 2014), the origin of the genetic toolkit in prokaryotes generally, is largely unknown (Koonin and Aravind, 2002; Kaczanowski, 2016; Klim *et al.*, 2018). This raises some interesting questions. Were PCD genes also present in archaea, or did they emerge in bacteria after the divergence

of the two lineages? One of the historical obstacles to addressing this question has been a dearth of sequence data from archaeal taxa. However, since 2012, much more data have become available (for example, Asplund-Samuelsson et al. 2012), presenting an opportunity to investigate the most ancient origins of the PCD genetic toolkit (Srivastav and Suneja, 2019; Zhang *et al.*, 2020). In addition, there has been some "tantalizing" empirical evidence for PCD genes in archaea (Asplund-Samuelsson et al. 2012; Klemenčič et al. 2019). Caspases (cysteine-aspartic proteases) (CAs) and their homologs in non-metazoa are one of the central mediators of PCD, activating multiple enzymes that are part of the death signaling cascade (Cohen, 1998). Caspase homologs refer to orthocaspases (OCAs), metacaspases (MCAs), and paracaspases (PCAs), as per the proposed unified nomenclature of the C14 peptidase family (Minina *et al.*, 2020) (Fig. 1). There are some reports of putative caspase homologs of the C14 peptidase family, OCAs and MCAs, in archaeal phyla (Euryarchaeota, Thermoplasmatota, Thaumarchaeota, Bathyarchaeota, and Heimdallarchaeota) (see supplementary documents in Klemenčič *et al.*, 2019).

Three homologous groups of caspases (CAs) are documented, type I, II, and III MCAs, type I and II PCAs, and OCAs (Fig. 1). All possess the p20-like region which includes the conserved Histidine-Cysteine (HC) dyad, sometimes also referred to as the Cysteine-Histidine (CH) dyad, which is situated within a characteristic caspase/haemoglobinase fold (Aravind and Koonin, 2002; McLuskey and Mottram, 2015) (Fig. 1). They are members of the CD clan, C14 peptidase family of the MEROPS peptidase database (Rawlings *et al.*, 2014). Unlike CAs that are only found in metazoa, PCAs are common in metazoa and slime molds while MCAs are universally distributed in the three domains of life, except for metazoa (Uren *et al.*, 2000). OCAs are ancestral caspase homologs present in bacteria and phytoplankton (Choi and Berges, 2013; Klemenčič, Novinec and Dolinar, 2015). OCAs in eukaryotes are commonly referred to as MCA-like proteases (Minina *et al.*, 2020). MCAs, PCAs, and OCAs have different substrate specificities and N-terminal domains compared to CAs (McLuskey and Mottram, 2015) but both sets of homologs correlate and are causally associated with cell survival and death functions (Vercammen *et al.*, 2007; Shrestha and Megeney, 2012; Minina *et al.*, 2017; van Creveld *et al.*, 2018). Cell survival functions include cell cycle regulation (Ambit *et al.*, 2008; Lee *et al.*, 2008), stress response (Richie *et al.*, 2007), and virulence mediation (Proto *et al.*, 2011; Benler and Koonin, 2020). The pleiotropic nature of caspase homologs in both death and non-death-related functions (Shrestha and Megeney, 2012; Hill and Nyström, 2015; Klemenčič, Novinec and Dolinar, 2015; Lema A *et al.*, 2021) indicate that the 'original sin' hypothesis is a potentially useful framework for investigating the evolution of the PCD toolkit.

The aim of this study was to investigate the potential presence of homologous caspase-like protein sequences in archaea and provide detailed comparative phylogenetic and sequence analyses of other possible death domains. Our analyses revealed type I MCAs and OCAs are surprisingly widespread in archaea. The phylogenetic reconstruction analysis and taxonomic distribution are suggestive of

massive horizontal gene transfer (HGT) events between bacteria and archaea, and that at least some of the effectors of PCD were likely present prior to the diversification of prokaryotes. In addition, numerous death domains were identified in archaeal OCAs and type I MCAs, and subsequent phyletic pattern analyses inferred their putative ancestral functions. Our data provide strong support for the 'original sin' hypothesis of PCD. This lays the foundation for understanding the evolution and role of PCD in prokaryotic communities and as a conflict mediator in eukaryogenesis.

RESULTS

Caspase homologs in archaea

In total, 321 caspase homologous sequences were identified spanning eleven archaeal phyla (Fig. 2 and Supplementary Table S1). Sequences identified as caspase homologs were further classified according to their structural subtypes (meta-, ortho-, and paracaspases) (Fig. 1). Fifty sequences (15.58%) possessed both the p20-like region and the p10-like region and hence classified as MCAs. The length of the linker region for the fifty sequences was 7.34 ± 27.34 amino acids (Supplementary Table S2). Following the guidelines and criteria of (Choi and Berges, 2013), the presence of a short linker region (161.3±32.9 aa) was used to classify the identified MCAs as type I MCAs (Supplementary Table S1) versus type II MCAs with longer linker regions. The remaining 271 sequences only contained the p20-like region and were therefore classified as OCAs (84.42%) (Supplementary Table S1).

Archaeal OCAs and type I MCAs identified in this study showed a wide taxonomic distribution with sequences belonging to the three major superphyla (TACK, Asgard group, and DPANN) and two phyla of archaea (Thermoplasmatota and Euryarchaeota) (Fig. 2). TACK is an acronym for the superphyla Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (Guy and Ettema, 2011) and DPANN the Aenigmarchaeota, Diapherotrites, Nanoarchaeota, Nanohaloarchaeota. Parvarchaeota and Woesearchaeota (Rinke *et al.*, 2013; Castelle *et al.*, 2015). The majority of OCAs and type I MCAs were identified in the phylum Euryarchaeota (51.09%) and phyla with the least number of homologs were *Candidatus* Aenigmarchaeota and *Candidatus* Woesearchaeota belonging to superphylum DPANN (Fig. 2). Unclassified organisms refer to sequences obtained from environmental samples that do not have taxonomic annotation information.

Sequence analysis of archaeal orthocaspases and type I metacaspases

Multiple sequence alignment (MSA) and sequence logo analyses of archaeal OCAs and type I MCAs revealed the highest sequence conservation occurred in amino acid residues surrounding the catalytic HC dyad (Supplementary Fig. S1 and S2). Additional four key amino acid residues are known to have importance for the formation of the S1 pocket and enzyme specificity in eukaryotic type I MCAs (McLuskey *et al.*, 2012). Of these, two Asp residues were conserved in archaeal type I MCAs and present in most OCAs (Supplementary Fig. S1 and S2). The remaining two amino acids, acidic Cys and Ser, showed variation in residues in both archaeal OCAs and type I MCAs (Supplementary Fig. S1

and S2). The variation in residues corresponded to the amino acids observed in CAs, PCAs, or MCAs of eukaryotes (Wei *et al.*, 2000; Yu *et al.*, 2011; McLuskey *et al.*, 2012). The acidic residues important for substrate specificity in type I MCAs (Y31, numbered according to TbMCA-Ib) and PCAs (E500, numbered according to MALT1), were absent. Proline-rich repeats, which are usually found in the N-terminal region of type I MCAs (Uren *et al.*, 2000), were identified in the linker region of archaeal type I MCAs (Supplementary Fig. S2). Archaeal type I MCAs possessed high affinity and low affinity Ca^{2+} binding sites (Supplementary Fig. S2 and S4). They were absent in archaeal OCAs.

According to the automated I-TASSER protein structure prediction, structural mimicry of archaeal type I MCAs and OCAs to TbMCA-Ib of *Trypanosoma brucei* (UniProt ID: Q585F3) was observed. Type I MCA of *Candidatus* Thorarchaeota archaeon (GenBank ID: TFG95767.1) and OCA of *Candidatus* Prometheoarchaeum syntrophicum (GenBank ID: QEE14607.1) was used as the model template for archaeal type I MCAs (Supplementary Fig. S4) and OCAs (GenBank ID: QEE14607.1) (Supplementary Fig. S3), respectively. The S1 binding pocket was conserved which possessed the HC dyad and the four key residues involved in substrate specificity and enzyme activity (Supplementary Figure S3 and S4). Some differences were observed in the secondary structure and the tertiary folding of archaeal OCAs and type I MCAs, including the number of the β-sheets and α-helices, as well as the missing N-terminal region in archaeal type I MCAs (Supplementary Fig. S3 and S4).

Phylogenetic analysis

Archaeal OCAs and type I MCAs phylogenies were reconstructed using the MSA of the p20-like region and the C14 peptidase domain, respectively (Supplementary Fig. S1 and S2). Both the OCAs and type I MCAs phylogenetic distributions were incongruent with 16S rRNA phylogeny and revealed a diffuse, scattered distribution (Supplementary Fig. S1 and S2). Tree topologies revealed the clades were clustered according to the different key residues required for the S1 pocket formation and enzyme activation, which corresponded with the residues observed in CAs, PCAs, or type I MCA of *T. Brucei*. Both trees were well supported with robust bootstrap values above 0.80 (Supplementary Fig. S1 and S2).

Phylogenetic reconstruction of the p20-like region of bacterial and archaeal OCAs, and eukaryotic CAs, PCAs and type I MCAs (Fig. 3) revealed an incongruency with the species tree provided by Hug and colleagues (Hug *et al.*, 2016). P20-like regions of PCAs and CAs of Metazoa clustered together, while eukaryotic type I MCAs and PCA of *D. discodeum* (UniProt ID: Q9GPM2) were placed in separate clades. CAs, PCAs, and eukaryotic type I MCAs all branch from bacterial OCAs, except for PCA of *D. discodeum* which branches from archaeal OCAs.

The phylogenetic relationship of type I MCAs across the three domains of life: archaea, bacteria, and eukarya were incongruent with the classic 16S rRNA phylogenetic distribution (Fig. 4). Three clades were observed that were well supported with robust bootstrap values above 0.95 (Fig. 4). For the ultrafast bootstrap used by IQ-TREE, a clade is considered reliable if its support is more than 95%

(Hoang *et al.,* 2018). Clade 1 possessed the CAs and bacterial type I MCAs, which included the alphaproteobacterium *Rhizobiales* bacterium (UniProt ID: A0A2A4P607). The second clade consisted of bacterial and archaeal type I MCAs. They possessed key residues resembling type I MCA as well as the low and high affinity Ca^{2+} binding sites and proline-rich repeats. The third clade possessed type I MCAs of eukarya, bacteria, and archaea, which included the OCA of Asgard archaeon (*Candidatus* Thorarchaeota archaeon (GenBank ID: TFG95767.1)). They also possessed the key residues observed in a type I MCA as well as the low affinity and high affinity Ca^{2+} binding sites.

C14 peptidase- accompanying domains and domain architectures of archaeal OCAs and type I MCAs

In addition to the C14 Peptidase domain, 56 different C14 peptidase-accompanying domains were identified (Supplementary Table S3). These include Peptidase_C13 (34.07%), Raptor_N (10.37%), Formylglycine-generating (FGE) sulfatase (4.44%), Polycystic Kidney Disease (PKD) (4.44%) and PKD_4 (4.44%). Thirty domains were identified only once in the sequence data. GO prediction and functional information available on the Pfam database revealed the functions of the three most abundant C14 peptidase-accompanying domains (PKD, PKD_4, and FGE_sulfatase) as cell surface proteins that protect against extreme environments (Jing *et al.,* 2002), respond to environmental stimuli, and unknown, respectively (Supplementary Table S4). Archaeal OCAs and type I MCAs that possess C14 peptidase-accompanying domains associated with cell survival and PCD were identified. Single C14 peptidase-accompanying domains predicted to have both functions were common (Fig. 5 and Supplementary Table S4). In addition, C14 peptidase-accompanying domains implicated in interactions with the environment and neighboring cells were abundant. These included domains associated with cell adhesion, cell projection, and cell surface receptors (Supplementary Table S4). A complete list of C14 peptidase-accompanying domains, their abundance, and predicted functions is available in Supplementary Table S3 and S4.

Prediction of transmembrane domains using TMHMM revealed 127 (39.56%) sequences with varying transmembrane topologies (Supplementary Table S5). One hundred fourteen protein sequences were found to be type II membrane proteins, which are single-spanning membrane proteins that possess intracellular N-terminals and extracellular C-terminals (von Heijne, 2006) and resemble signaling peptides (Krogh *et al.,* 2001).

In total, 63 different domain architectures were identified (Fig. 6 and 7). Out of 321 sequences, 183 (57.01%) sequences possessed C14 peptidase-accompanying domains of which 177 sequences were OCAs and six were type I MCAs. The most abundant domain architectures were OCAs possessing a single p20-like region (29.30%), OCAs possessing a transmembrane helix followed by a p20-like region (22.43%), and type I MCAs possessing a p20-like region followed by a p10-like region (13.40%). Forty-two domain architectures were identified only once in the sequence dataset (13.08%) and no domain architecture was unique to a specific phylum.

Transmembrane OCAs are common in archaea (39.25%). The most common domain architecture of transmembrane OCAs was an extracellular C-terminal p20-like region and domains interacting with the neighboring cells and the environment (8.41%) (Fig. 6, domain architecture 3, 5-8, 10, 14,15, 17, 18, 20-22, 25-27, 30). Compared to transmembrane OCAs, cytosolic OCAs possessed a great variety of domains, including those associated with survival (Fig. 7, domain architecture 34, 37, 38, 43, 48, 55-57, 60-62) and PCD (Fig. 7, domain architecture 43, 56, 61 and 62). The death domains (Aravind, Dixit and Koonin, 1999b; Li and Roberts, 2001) identified in this study include TPR_1, TPR_2, TPR_16, and TPR_11 (Tetratricopeptide repeat), and WD_40 (WD domain) and ANAPC4_WD40 (Anaphase-promoting complex subunit 4 WD40 domain) (Supplementary Table S4). Function assignments via Pfam database, dcGO, and Pfam2GO, associated these death domains with cell survival functions, such as regulation of cell growth, cell projection assembly, and cell cycle control (Supplementary Table S4).

Archaeal type I MCAs possessed simple domain architectures, of which most did not possess any C14 peptidase-accompanying domains (86.60%). Transmembrane type I MCAs were rare (3.38%). C14 peptidase-accompanying domains in archaeal type I MCAs were associated with cell adhesion and cell surface receptors (Fig. 6, domain architecture 12 and 29, and Fig. 7, domain architecture 58).

Taxonomic distribution analysis revealed that cytosolic and transmembrane OCAs and type I MCAs associated with cell survival had wide taxonomic distributions (Supplementary Table S6). They were identified in superphyla Asgard group and TACK, and phyla Thermoplasmatota and Euryarchaeota. Archaeal cytosolic OCAs possessing death domains were confined to superphyla Asgard group and TACK (Supplementary Table S6).

DISCUSSION

Caspase homologs were identified in archaeal superphyla TACK, the Asgard group, and DPANN, and phyla Euryarchaeota and Thermoplasmatota. In total, 321 OCAs and type I MCAs were identified (Fig. 2) by stringent hmmsearch filtering, manual inspection of the HC dyad (Supplementary Fig. S1 and S2), and structural similarity (Supplementary Fig. S3 and S4). The HC dyad is a requirement for the proteolytic activity of OCAs and MCAs (Aravind and Koonin, 2002; McLuskey and Mottram, 2015), whereas the absence of the dyad indicates catalytic inactivity (Szallies, Kubata and Duszenko, 2002). It is likely, therefore, that the OCAs and MCAs reported here are active.

The OCAs were much more abundant in archaea (84.42%) than type I MCAs (15.58%) (Fig. 2), a pattern that is similar to the findings in bacteria (Klemenčič *et al.*, 2019). It is possible that some sequences classified as OCAs may be partial sequences, which can result in erroneous classifications, and future whole-genome sequencing analyses will resolve any potential inconsistencies. There was a bias in the taxonomic distribution of archaea (Fig. 2), which we attribute to the relative dearth of data from superphyla TACK, DPANN, and Asgard group. This will likely also change as more data

become available. CAs and PCAs were not identified, which is in keeping with previous phylogenetic predictions (Uren *et al.*, 2000). Similarly, type II MCAs and type III MCAs were not identified. This was unsurprising because type II and III MCAs are reportedly specific to Viridiplantae and some algal species where they are associated with secondary endosymbiotic events (Choi and Berges, 2013). In addition to the previously listed putative archaeal caspase homologs (Klemenčič *et al.*, 2019), an additional 228 sequences were identified here, including previously unclassified sequences from superphyla Asgard group and TACK.

Comparative sequence and secondary structure prediction analyses of OCAs and archaeal type I MCAs confirmed the presence of key residues and structures typical of this group of homologs (Supplementary Fig. S1, S2, S3, and S4). This includes the HC dyad, and two Asp residues present in the S1 pocket that are required for substrate specificity and enzyme activity. There was some minor variation in two of the key residues. These are the acidic Cys (Supplementary Fig. S1 and S2), which is important for substrate binding (McLuskey *et al.*, 2012), and Ser (Supplementary Fig. S1 and S2), which is important for regulation of enzyme activity (McLuskey *et al.*, 2012). Interestingly, this variation matched those of eukaryotic type I MCAs, PCAs, or CAs (Supplementary Fig. S1 and S2) and may have some functional significance. The conservation of key residues in eukaryotic type I MCAs, PCAs. or CAs suggests that they may be derivatives of archaeal OCAs and type I MCAs. Furthermore, the structural mimicry of the automated I-TASSER predicted model of archaeal type I MCA and OCAs with the X-ray crystallography of TbMCA-Ib further suggests common ancestry and structural (and presumably functional) homology (Supplementary Fig. S4).

The presence of key residues and structures required for a catalytically active C14 peptidase suggests archaeal OCAs and type I MCAs are catalytically active. A small subset of archaeal OCAs was very similar in their key residues to an OCA found in *Microcystis aeruginosa* (Supplementary Fig. S1) that is known to have MCA-like activity (Klemenčič, Novinec and Dolinar, 2015). This subset may, therefore, also have MCA-like activity. The conservation of some of the Ca^{2+} binding sites in archaeal type I MCAs suggests that Ca^{2+} remains important as a secondary messenger in the activation of archaeal type I MCAs. As described by (Klemenčič, Novinec and Dolinar, 2015), Ca^{2+} binding sites were absent in archaeal OCAs.

The N-terminal domain requirement for MCA activity (McLuskey *et al.*, 2012), and the zinc-finger domain required for cell death function (Coll *et al.*, 2010) were absent in archaeal OCAs and type I MCAs (Supplementary Fig. S3 and S4). Tyr, which functions as a latch with Ser for regulation of enzyme activation (McLuskey *et al.*, 2012) was also absent in both archaeal OCAs and type I MCAs. The observed differences in the secondary structure and some of the key residues required for enzyme activity and regulation (supplementary Fig. S1 and S3, Supplementary Material online) suggest there are likely subtle differences in substrate specificity or regulation of activity of archaeal OCAs and type I MCAs compared to TbMCA-Ib.

Origins of prokaryotic type I MCAs and OCAs

Phylogenetic reconstruction of archaeal type I MCAs and OCAs illustrates a diffuse gene distribution pattern across different phyla and is inconsistent with the species trees (Supplementary Fig. S1 and S2 and discussed further below). Both trees were well supported with robust bootstrap values above 0.80. Only one sequence belonging to the superphylum Asgard group was available (Genbank ID TFG95767.1) for type I MCA phylogenetic analysis, while none were available from superphylum DPANN, which resulted in a biased phylogeny of type I MCAs towards the phylum Euryarchaeota.

The identification of OCAs and type I MCAs in all three domains of life - archaea, bacteria, and eukarya - and their shared origins with CAs and PCAs raises new questions: were the ancestors of these PCD effectors present in the ancestors of both archaea and bacteria? Or did they emerge after the divergence of the two prokaryote lineages with subsequent HGT between them? Furthermore, since most of the eukaryotic members of the C14 peptidase family appear to have originated from the bacterial ancestors of mitochondria (Aravind, Dixit and Koonin, 2001; Koonin and Aravind, 2002; Klim *et al.*, 2018), what has been the evolutionary trajectory of the archaeal homologs? What is clear from the data here and previous publications (Uren *et al.*, 2000) is that PCAs and CAs are limited to the so-called higher, complex organisms while MCAs are widely distributed across all lineages with the notable exception of Metazoa (Uren *et al.*, 2000; Choi and Berges, 2013). OCAs are confined to unicellular organisms (Klemenčič, Novinec and Dolinar, 2015; Klemenčič and Funk, 2018a). The conservation of key residues, the structural similarities between archaeal OCAs and the p20-like region of PCAs and type I MCAs, and their presence in the deepest branching superphylum DPANN, phylum Euryarchaeota (Fig. 3), and the bacterial phyla Aquifex, Thermotogae, and Deinococci (Klemenčič *et al.*, 2019) suggest that OCAs (the p20-like region) are almost certainly the most ancestral members of the peptidase C14 clan. This supports earlier predictions (Choi and Berges, 2013).

Three clades of type I MCAs are neatly resolved. These are the bacterial caspase homologs (clade 1), prokaryotic type I MCAs (clade 2), and eukaryotic type I MCAs (clade 3). CAs were placed in one clade with bacterial type I MCAs, which included the *Rhizobiales* bacterium (UniProt ID: A0A2A4P607) bacteria (Fig. 4). This supports the origin of CAs as bacterial caspase homologs during the endosymbiotic event (Koonin and Aravind, 2002) and is consistent with previous findings tracing the evolutionary histories of CAs and MCAs across the three domains of life (Klim *et al.*, 2018).

The origin of unicellular eukaryotic type I MCAs in clade 3 is more difficult to resolve and could have emerged in either archaea or bacteria (Fig. 4). The absence of the Ca^{2+} binding sites in bacterial type I MCAs in clade 1 indicates that the ancestral type I MCAs were not regulated by Ca^{2+}, a feature that emerged later. Proline-rich repeats, which are typically found in the N-terminal domains of eukaryotic type I MCAs, were only identified in clade 2, which contains prokaryote type I MCAs. The distribution of type I MCAs across different domains indicates HGT events between prokaryotes and eukaryotes.

The incongruency between gene and species trees is most likely due to extensive HGT between archaea and bacteria (Maddison, 1997; Ponting *et al.*, 1999; Ochman, Lawrence and Groisman, 2000; Aravind, Dixit and Koonin, 2001). Tracing the key residues in the S1 pocket and additional C14 peptidase-accompanying domains present in different members of the C14 peptidase family also suggest massive HGT. This is in line with what is known about large-scale gene flows between the proto-cellular organisms during the emergence of diverse prokaryotic lineages (Koonin *et al.*, 1997; Aravind *et al.*, 1998; Nelson *et al.*, 1999; Nelson-sathi *et al.*, 2015; Koonin, 2016; Wagner *et al.*, 2017; Husnik and McCutcheon, 2018) and limits the identification of donor lineages (Akanni *et al.*, 2015). Of course, there are other potential explanations. For example, one type of a C14 peptidase may have arisen *de novo* in an ancestral archaeon with continuation in daughter lineages and HGT and gene duplication events (Fig. 8). However, this would have to include multiple duplication events with subsequent HGT of the paralogous genes. It seems more parsimonious that the OCAs were present prior to the diversification of archaea and bacteria, suggesting that ancestral OCAs were already present in the realm of the LUCAS (Last Universal Common Ancestor State). The acronym LUCAS refers here to Koonin's "forest of life" (FoL) concept, which was a hypothetical state prior to the existence of cells as we understand them today and in which genetic material flowed between semi-cellular compartments (Koonin, 2009). Another possible explanation is that these C14 peptidases originated from eukaryotes and were transferred to prokaryotes via HGT events (Rawlings and Bateman, 2019). However, their wide taxonomic distribution in all of the major phyla of archaea and bacteria suggest they were present in prokaryotes first. The mitochondrial origin of the C14 peptidase family during eukaryogenesis (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Blackstone and Green, 1999) also supports their prokaryotic origin and is consistent with Klim and colleagues' mitochondrial hypothesis for the origin of eukaryotic apoptosis (Klim *et al.*, 2018).

C14 peptidase-accompanying domains associated with pro-survival outnumber death domains in archaeal OCAs and type I MCAs

The conservation of OCAs and their ancient origins raise questions about their ancestral functions. The functional inference was based on phyletic and functional pattern analysis of C14 peptidase-accompanying domains, with reference to previous studies (Gaasterland and Ragan, 1998; Aravind, Dixit and Koonin, 1999b; Galperin and Koonin, 2000). Of course, functional inference based on phyletic pattern analysis is predictive and awaits future empirical studies. Nevertheless, a large variety of different C14 peptidase-accompanying domains were easily identified, and these were more abundant in archaeal OCAs (97.27%) than type I MCAs (2.73%) (Supplementary Table S3,). The high abundance of the Peptidase_C13 and Raptor_N domains is spurious and explained by their sequence similarities to the C14 peptidase domain and the presence of the HC dyad (Ginalski, Zhang and Grishin, 2004; Rawlings *et al.*, 2014). Functional assignment of the other C14 peptidase-accompanying domains using multiple GO databases indicated that domains associated with cell

survival are more abundant than those associated with PCD (Fig. 5). GO functions associated with cell survival were chosen based on the Gedanken experiment by Ameisen (Ameisen, 2002), which includes pro-survival functions like cell differentiation, cell cycle regulation, and cell metabolism.

Domain architecture analyses revealed archaeal OCAs resembling signaling peptides that interact with the external environment (Fig. 6). Examples include OCAs possessing a transmembrane helix followed by C14 peptidase-accompanying domains associated with cell adhesion, cell projection, and cell surface receptors. The occurrence of these membrane-bound OCAs suggests archaeal OCAs were involved in cellular responses to the surrounding environment through cell signaling mechanisms. In contrast, only a few archaeal type I MCAs possessed C14 peptidase-accompanying domains and transmembrane helices. It is acknowledged that the transmembrane components of the archaeal OCAs and type I MCAs resembling signaling peptides were predicted using the TMHMM program, which reports a ~78% precision in topology prediction (Krogh *et al.*, 2001) and should be interpreted in this light.

Cytosolic archaeal OCAs possessing death domains were observed in superphylum TACK (Fig. 7 and Supplementary Table S6). However, it is uncertain if their presence implies archaeal OCAs are involved in PCD as death domains are associated with both death and pro-survival. According to the GO prediction, the death domains are also associated with cell survival functions (a quirk of the historical use of PCD terminology), such as cell cycle control, differentiation, and metabolism (Supplementary Table S4). The abundance of C14 peptidase-accompanying domains associated with cell survival in archaeal OCAs adds to this uncertainty.

The pleiotropic nature of MCAs and the association of these effector enzymes with both PCD and cell survival functions is established (Shrestha and Megeney, 2012; Hill and Nyström, 2015). Their pleiotropic nature and the higher abundance of C14 peptidase-accompanying domains associated with pro-survival, therefore, raises questions about the role of ancestral OCAs and type I MCAs in archaea. Did they evolve because of PCD-associated cell death functions, or were they co-opted from ancestral proteins with pro-survival functions? Intuitively, it seems more parsimonious that these cytosolic archaeal OCAs were initially involved in pro-survival activities and were subsequently co-opted into death-related functions.

Archaeal OCAs and eukaryogenesis

According to the Entangle-Engulf-Enslave (E3) model, the physical interaction between a *Candidatus* Prometheoarchaeum syntrophicum MK-D1 archaeon and an alphaproteobacterium involved cellular protrusions and phagocytosis-independent engulfment (Imachi *et al.*, 2020). This is because phagocytotic machinery is absent in Asgard group, the proposed archaeal clade involved in archaea-alphaproteobacteria symbiosis (Moreira and Lopez-Garcia, 1998). Using this hypothetical framework, which is now one of the most updated hypotheses for eukaryogenesis (Imachi *et al.*, 2020), we wished to investigate the potential role of these archaeal OCAs in the rise of FECA (First Eukaryotic Common

Ancestor). Two OCAs found in *Candidatus* Prometheoarchaeum syntrophicum strain MK-D1 archaeon possessed simple domain architectures with no C14 peptidase-accompanying domains (Fig. 6, domain architecture 1 and Fig. 7, domain architecture 32). In contrast, OCAs of other Asgard archaea (Fig. 6, domain architecture 2, 3, 7, 14, 17, 19 and 24 and Fig. 7, domain architecture 36 and 60) were all transmembrane proteins with domains associated with cell metabolism, cell adhesion, cell cycle control, cell projection, cell differentiation and cell surface proteins/receptors implicated in stress responses. The abundance of cell surface and receptor-like OCAs in the Asgard group suggests these archaeal OCAs may be candidate PCD related genes that were also important in the E3 model of eukaryogenesis (Fig. 6).

The 'original sin' hypothesis and the PCD toolkit.

The phyletic pattern analysis of archaeal OCAs revealed that they are not only the most ancestral peptidase C14 members, but they also contain an abundance of pro-survival domains (Fig. 6 and 7). This is especially true of the deep-branching phyla Euryarchaeota and Thermoplasmatota, linking these ancestral OCAs with cell survival prior to eukaryogenesis. Although OCAs of DPANN only possessed the p20-like region, it is possible that additional genomic data will identify some of the C14 peptidase-accompanying domains. Although we are unsure if ancestral OCAs and type I MCAs were associated with PCD, it is clear that the ancestral effectors of PCD were already present prior to the diversification of bacteria and archaea. Type I MCAs appeared later in Thermoplasmatota and Euryarchaeota (Fig. 9) and acquired Ca^{2+} binding sites to refine their activity (Fig. 4 and 9). In addition, OCAs and type I MCAs of superphylum TACK acquired death domains and other C14-accompanying domains via HGT (Fig. 6, 7, and 9). It seems that this pre-adaptation allowed for the emergence of pro-death processes in prokaryotes and eukaryotes (Aravind, Dixit and Koonin, 1999b; Ameisen, 2002). Klemenčič and Funk have suggested before that PCD evolved from effectors involved in physiological cellular processes (Klemenčič and Funk, 2018a). The findings presented here enhance these assertions and support the 'original sin' hypothesis which is a philosophical argument put forward by Ameisen in 2002 that argues for diverse pro-survival functions for the PCD toolkit with subsequent refinement for essential death-related functions (Ameisen 2002).

CONCLUSION

OCAs and type I MCAs are present in archaea and have a wide taxonomic distribution in the three domains of life. The phylogeny of OCAs and type I MCAs suggest that these are the ancestral peptidase C14 members and were present since LUCAS. Their strong similarity with other members in bacteria and eukarya further supports this and the incongruency with the species tree proposes massive HGT of domains between prokaryotes and proto-cellular compartments prior to their divergence. The phyletic and functional pattern analyses of C14 peptidase-accompanying domains and death domains provide strong support for the 'original sin' hypothesis that OCAs in the Asgard group were perhaps

first involved in pro-survival functions with subsequent co-option for PCD functions in prokaryotes and the earliest eukaryotes.

MATERIALS AND METHODS

Sequence retrieval and identification of caspase homologs. The NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/) (Accessed 2 June 2021) was queried using search terms: "C14", "caspase ", "metacaspase" and "orthocaspase" in organism "archaea" under default parameters. Retrieved protein sequences were identified as caspase homologs by the presence of the Peptidase_C14 domain (PF00656) (http://pfam.xfam.org/) (Mistry *et al.*, 2021) using the hmmsearch tool from the HMMER v3.0 package (http://hmmer.org/) (Eddy, 1998) under a domain independent E-value ≤ 0.0001 and score/bias ratio ≥ 10. Hmmsearch uses profile hidden Markov Models (HMM) to detect sequence similarity and homology (Eddy, 1998). The raw Peptidase_C14 HMM profile (PF00656) (206 amino acids) obtained from the Pfam database was used. The aligned MSA file of 321 sequences was aligned to Peptidase_C14 HMM profile (PF00656) using the hmmalign tool under the --trim option. The presence of the conserved histidine and cysteine (HC) dyad was identified by visually inspection under Jalview v2.0 (Waterhouse *et al.*, 2009) according to the reference position (H220 and C276) provided on the MEROPS database (Rawlings *et al.*, 2014) as well as using the reference sequence of TbMCA-Ib of *T. brucei*. Sequences not possessing the HC dyad or pseudo-dyads (Eyers and Murphy, 2016) were removed.

Classification and taxonomy assignment of caspase homologs. A p20-like and p10-like region profile HMM was constructed with HMMER v3.0 hmmbuild, according to the approach followed by Klemenčič and colleagues (Klemenčič *et al.*, 2019). The first 457 and the last 211 columns of Pfam Peptidase_C14 (PF00656) seed alignment (112 sequences with gaps) were used as the input for the p20-like and the p10-like region respectively. To identify PCAs, hmmsearch was performed using the Immunoglobulin (Ig) (PF00047), Ig_2 (PF13895), or Ig_3 (PF13927) HMM profiles (http://pfam.xfam.org/), and the unaligned 321 protein sequence file under a domain independent E-value ≤ 0.0001 and score/bias ratio ≥ 10 using the –domblout parameter to produce the domain hits table as the output. MCAs were classified according to their subtypes based on sequence data of the amino acid length of the linker region between p20 and p10-like regions (Uren *et al.*, 2000; Choi and Berges, 2013) (Fig. 1). Linker region length was determined from the alignment coordinates in the domain hits table obtained from the hmmsearch hit.

Taxonkit, an NCBI taxonomic classification toolkit (Shen and Ren, 2021) was then used to convert taxonomy names to their respective taxonomy ID (TaxId) and assign the full taxonomy to 321 sequences. If no taxonomic information was available, it was assigned as 'unclassified'.

Sequence analysis and tertiary model prediction. Sequence analysis of archaeal OCAs and type I MCAS was performed by aligning respective unaligned sequences to the constructed p20-like region

HMM profiles and the Peptidase_C14 HMM profile (PF00656) using hmmalign under --trim option. A more reliable alignment was produced by realigning the produced MSA using MAFFT v.7.470 (Katoh and Standley, 2013) under default parameters. Sequences were filtered by removing sequences with no residues at the site of four key residues required for the correct formation of the S1 pocket (Cys, Asp, Ser, and Asp) (Wei *et al.*, 2000; Yu *et al.*, 2011; McLuskey *et al.*, 2012). Three Metazoa CAs, *Homo sapiens* CASP8, *H. sapiens* CASP3 and *Caenorhabditis elegans* cell death protein 3 (UniProt IDs: Q14790, P42573 and P42574), PCAs of *H. sapiens* (MALT1), *C. elegans* (MALT1), *Amphiprion percula* (MALT PCA3) and *Dictyostelium discoideum* (PCP_DICDI) (UniProt IDs: Q9UDY8, G5EG87, A0A3P8TIJ8 and Q9GPM2), and type I MCA of *T. brucei* (TbMCA-Ib) (UniProt ID: Q585F3) were included for sequence comparison. PCAs of *Dictyostelium discoideum* (UniProt ID: Q9GPM2) and *Amphiprion percula* (UniProt ID: A0A3P8TIJ8) were only used for the sequence comparison with archaeal type I MCAs as they possessed the p10-like region. Secondary structure prediction was performed using JPred (Protein Secondary Structure Prediction Server) (Drozdetskiy *et al.*, 2015) under Jalview with default parameters. To generate sequence logos of archaeal OCAs and type I MCAs, MSAs were transformed to HMMs using hmmbuild and submitted to Skylign5 (https://www.skylign.org/) (Wheeler, Clements and Finn, 2014). The full alignment of OCAs (271 sequences) or type I MCAs (50 sequences) were used using the parameter 'information above background'. Putative protein tertiary structures of archaeal OCAs and type I MCAs were predicted using I-TASSER under default parameters (https://zhanglab.dcmb.med.umich.edu/I-TASSER/) (Yang and Zhang, 2015). I-TASSER uses a hierarchical approach to predict protein structures by multiple threading approach LOMETS (Local Meta-Threading Server, version 3). Protein structures were visualized on PYMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). The protein sequence of an archaeal OCA (GenBank ID: QEE14607.1) and type I MCA (GenBank ID: TFG95767.1) was selected as the representative OCA and type I MCA model with the TbMCA-Ib of *T. brucei* (PDB ID: 4AFV) as the template.

Phylogenetic analysis. Four MSAs (archaeal type I MCA and OCA, and type I MCA and OCA across three domains of life) were created for the reconstruction of respective phylogenies. Archaeal OCA and type I MCA MSAs created for the sequence analysis were used, excluding sequences with no taxonomic classification. For the sequence selection of type I MCAs across the three domains of life, all 33 filtered archaeal type I MCA sequences were used. UniProt IDs of an identical number of type I MCAs of eukarya belonging to unicellular phyla and bacteria were obtained from supplementary documents (Klemenčič *et al.*, 2019). Respective protein sequences were obtained from the UniProt database (The UniProt Consortium, 2021). For the sequence selection of the p20-like region, 143 filtered p20-like regions of archaeal OCA sequences were used. An identical number of UniProt IDs of bacterial OCAs were obtained from supplementary documents (Klemenčič *et al.*, 2019) and the respective sequences were obtained from the UniProt database. The p20-like region of eukaryotic CAs

(*H. sapiens* CASP8, *H. sapiens* CASP3 and *C. elegans* cell death protein 3), PCAs (*H. sapiens* (MALT1), *C. elegans* (MALT1) and *D. discoideum* (PCP_DICDI)), and eleven eukaryotic type I MCAs with known functions (Tsiatsiani *et al.*, 2011) were included. Eukaryotic OCAs were not considered as the aim was to determine the origin and investigate the evolutionary history of the p20-like region from prokaryotic OCAs to eukaryotic CAs, type I MCAs, and PCAs. An identical protocol was followed for the construction of all four MSAs. Unaligned MSA file of acquired protein sequences was aligned to the p20-like region HMM profile for OCAs and the entire Peptidase_C14 HMM profile using hmmalign under --trim option. Gaps were removed using seqmagick 0.7.0 at default parameters as gaps are treated as unknown characters in IQ-TREE (Nguyen *et al.*, 2015). The resulting trimmed MSA file was then aligned using MAFFT under default parameters. Sequences were filtered by removing sequences with no residues at the site of four key residues required for the correct formation of the S1 pocket (McLuskey *et al.*, 2012). The best substitution model for phylogenetic reconstruction was chosen using IQ-TREE to correct for multiple changes at the same site under a maximum likelihood (ML) approach, including all sites (including gaps) (Kalyaanamoorthy *et al.*, 2017). Considered substitution models included a wide range of models that are supported by IQ-TREE, including advanced partition and mixture models (Minh *et al.*, 2020). Rate heterogeneity across sites was considered as well (Minh *et al.*, 2020). The phylogenetic tree was reconstructed using IQ-TREE using an ML approach (Nguyen *et al.*, 2015). The reliability of the reconstructed tree was assessed using the bootstrap method with 1000 replicates in IQ-TREE (Hoang *et al.*, 2018). The phylogenetic tree was visualized using the Interactive Tree of Life (iTOL) v5.0 online service (https://itol.embl.de/) (Letunic and Bork, 2019).

Identification of C14 peptidase-accompanying domains. C14 peptidase-accompanying domains refer to Pfam domains identified within an OCA or type I MCA sequence, adjacent to the C14 peptidase domain. The entire database of HMM profiles was retrieved from Pfam (ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam33.0/) and the HMM profile database was prepared for usage using hmmpress in the HMMER package. The identified, full-length 321 amino acid sequences of caspase homologs were subjected to a hmmsearch against the produced HMM profile database under a threshold E-value of ≤ 0.0001, independent E-value ≤ 0.0001 and score bias ratio ≥ 10, and the output was produced using --domtblout option (Eddy, 1998).

Identification of transmembrane domains. Transmembrane domains in caspase-like protein sequences were predicted using the TMHMM Online Server v2.0 under default values (http://www.cbs.dtu.dk/services/TMHMM/) (Sonnhammer, von Heijne and Krogh, 1998) with the unaligned MSA of the 321 sequences as the input file. Transmembrane domains were classified according to their topology (von Heijne, 2006). If the C14 peptidase domain was located C-terminally to a single N-terminal transmembrane helix, it was regarded as a signal peptide, following the approach of Krogh et al. (Krogh *et al.*, 2001).

Functional prediction and classification of C14 peptidase-accompanying domains. Putative functions (biological processes) of identified C14 peptidase-accompanying domains were predicted using the gene ontology (GO) information using the Pfam2Go (https://rdrr.io/github/missuse/ragp/man/pfam2go.html) and dcGO Enrichment approach (https://supfam.org/SUPERFAMILY/dcGO/), as well as the information available on Pfam website (https://pfam.xfam.org/). The Pfam release used for the functional annotation was v34.0 (Mistry *et al.*, 2021). For dcGO GO assignment, False Discovery Rate (FDR) threshold of 0.05 was chosen as the cut-off value and the remaining setting was set as default. Multiple GO databases were used for functional prediction (Forslund and Sonnhammer, 2008). Predicted functions were further classified: cell survival, PCD, and interaction with neighboring cells and the environment. Cell survival-associated functions refer to the functions listed by Ameisen (Ameisen, 2002). The classification of the predicted functions (biological processes) used was as follows: cell differentiation, cell cycle (DNA replication, transcription, and recombination, DNA repair mechanisms, cell membrane repair mechanisms, cell division, chromatin organization, remodeling or dissolution of the nuclear membrane and chromosomal migration for cell division), and cell metabolism (regulation of ionic channels). The term 'PCD-associated' refers to the death domains that act as ligands or adaptors of the PCD molecular pathway domains (Aravind, Dixit and Koonin, 1999b; Li and Roberts, 2001). Interaction with neighboring cells and the environment refers to biological processes: cell adhesion, cell projection, cell surface protein or receptor, response to stress, and interaction with neighboring cells. The remaining functions were classified under "other". If no function was predicted by the GO databases, the function of the domain was assigned as "unknown".

Domain architecture analysis. Protein domain architectures of archaeal caspase homologs were created by using a custom R script (https://github.com/Asplund-Samuelsson/caspases.git) and an input file (Sequence ID, Domain, alignment coordinates). Domain refers to the C14_peptidase domain, C14 peptidase-accompanying domains, and transmembrane helices. Domains were placed in sequential order according to the alignment coordinates. An overlap of two domains was calculated by their predicted alignment coordinate end and start position of domain1 and domain2. If two domains on a sequence overlapped more than 25%, the domain with the lower E-value was accepted. C14 peptidase domain took precedence if other domains were found. If a transmembrane helix occupied the same region as a C14 peptidase-accompanying domain, the C14 peptidase-accompanying domain took precedence. Consecutive repeats of identical domains were identified as a single domain.

Conflicts of Interests

The authors declare no conflicts of interest.

Data Availability. The data underlying this article are available at https://github.com/laso415/C14peptidase as well as in the article and its online supplementary material. The data sets were derived from sources: NCBI Protein database, UniProt database, Pfam database and GO databases (dcGO and Pfam2Go), as indicated in the Materials and Methods section and supplementary data. The specific accession numbers are provided in the respective Materials and Methods sections and the supplementary materials.

REFERENCES

Abada A, Segev E (2018) Multicellular Features of Phytoplankton. Front Mar Sci 5:144. https://doi.org/10.3389/fmars.2018.00144

Akanni WA, Siu-ting K, Creevey CJ, et al (2015) Horizontal gene flow from Eubacteria to Archaebacteria and what it means for our understanding of eukaryogenesis. Philos Trans R Soc B Biol Sci 370:20140337. https://doi.org/10.1098/rstb.2014.0337

Ambit A, Fasel N, Coombs GH, Mottram JC (2008) An essential role for the *Leishmania major* metacaspase in cell cycle progression. Cell Death Differ 15:113–122. https://doi.org/10.1038/sj.cdd.4402232

Ameisen J-C (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. Cell Death Differ 9:367–93. https://doi.org/10.1038/sj/cdd/4400950

Aravind L, Dixit VM, Koonin E V. (2001) Apoptotic molecular machinery: Vastly increased complexity in vertebrates revealed by genome comparisons. Science (80-) 291:1279–1284. https://doi.org/10.1126/science.291.5507.1279

Aravind L, Dixit VM, Koonin E V (1999) The domains of death. Elsevier Sci Ltd 0004:47–53. https://doi.org/10.1016/s0968-0004(98)01341-3

Aravind L, Koonin E V. (2002) Classification of the caspase-hemoglobinase fold: Detection of new families and implications for the origin of the eukaryotic separins. Proteins Struct Funct Genet 46:355–367. https://doi.org/10.1002/prot.10060

Aravind L, Tatusov RL, Wolf YI, et al (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet 14:442–444. https://doi.org/10.1016/S0168-9525(98)01553-4

Arnoult D, Tatischeff I, Estaquier J, et al (2001) On the evolutionary conservation of the cell death pathway: mitochondrial release of an apoptosis-inducing factor during Dictyostelium discoideum cell death. Mol Biol Cell 12:3016–3030. https://doi.org/10.1091/mbc.12.10.3016

Asplund-Samuelsson J, Bergman B, Larsson J (2012) Prokaryotic Caspase Homologs: Phylogenetic Patterns and Functional Characteristics Reveal Considerable Diversity. PLoS One 7:1–14. https://doi.org/10.1371/journal.pone.0049888

Bayles KW (2014) Bacterial programmed cell death: Making sense of a paradox. Nat Rev Microbiol

12:63–69. https://doi.org/10.1038/nrmicro3136

Benler S, Koonin E V (2020) Phage lysis-lysogeny switches and programmed cell death: Danse macabre. BioEssays 42:2000114. https://doi.org/10.1002/bies.202000114

Berman-Frank I, Bidle KD, Haramaty L, Falkowski PG (2004) The demise of the marine cyanobacterium, *Trichodesmium spp*., via an autocatalyzed cell death pathway. Limnol Oceanogr 49:997–1005. https://doi.org/10.4319/lo.2004.49.4.0997

Bidle KD (2016) Programmed Cell Death in Unicellular Phytoplankton. Curr Biol 26:R594–R607. https://doi.org/10.1016/j.cub.2016.05.056

Blackstone NW (2013) Why did eukaryotes evolve only once? Genetic and energetic aspects of conflict and conflict mediation. Philos Trans R Soc Lond B Biol Sci 368:20120266. https://doi.org/10.1098/rstb.2012.0266

Blackstone NW, Green DR (1999) The evolution of a mechanism of cell suicide. BioEssays 21:84–88. https://doi.org/10.1002/(SICI)1521-1878(199901)21:1%3C84::AID-BIES11%3E3.0.CO

Castelle CJ, Wrighton KC, Thomas BC, et al (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25:690–701. https://doi.org/10.1016/j.cub.2015.01.014

Choi CJ, Berges JA (2013) New types of metacaspases in phytoplankton reveal diverse origins of cell death proteases. Cell Death Dis 4:e490. https://doi.org/10.1038/cddis.2013.21

Cohen GM (1998) Caspases: the executioners of apoptosis. Biochem J 95:1–16. https://doi.org/10.1016/s0378-4274(98)80028-3

Coll NS, Vercammen D, Smidler A, et al (2010) Arabidopsis type I metacaspases control cell death. Science 330:1393–1397. https://doi.org/10.1126/science.1194980

Consortium TU (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100

Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43:W389–W394. https://doi.org/10.1093/nar/gkv332

Durand PM (2021) The Evolutionary Origins of Life and Death. The University of Chicago Press, Chicago

Durand PM, Barreto Filho MM, Michod RE (2019) Cell Death in Evolutionary Transitions in Individuality. Yale J Biol Med 92:1–12

Durand PM, Ramsey G (2019) The Nature of Programmed Cell Death. Biol Theory 14:30–41. https://doi.org/10.1007/s13752-018-0311-0

Durand PM, Rashidi A, Michod RE (2011) How an Organism Dies Affects the Fitness of Its Neighbors. Am Nat 177:224–232. https://doi.org/10.1086/657686

Durand PM, Sym S, Michod RE (2016) Programmed Cell Death and Complexity in Microbial Systems. Curr Biol 26:R587–R593. https://doi.org/10.1016/j.cub.2016.05.057

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763. https://doi.org/10.1093/bioinformatics/14.9.755

Eyers PA, Murphy JM (2016) The evolving world of pseudoenzymes: proteins, prejudice and zombies. BMC Biol 14:98. https://doi.org/10.1186/s12915-016-0322-x

Forslund K, Sonnhammer ELL (2008) Predicting protein function from domain content. Bioinformatics 24:1681–1687. https://doi.org/10.1093/bioinformatics/btn312

Frade JM, Michaelidis TM (1997) Origin of eukaryotic programmed cell death: a consequence of aerobic metabolism? Bioessays 19:827–832. https://doi.org/10.1002/bies.950190913

Gaasterland T, Ragan MA (1998) Microbial genescapes: phyletic and functional patterns of ORF distribution among prokaryotes. Microb Comp Genomics 3:199–217. https://doi.org/10.1089/omi.1.1998.3.199

Galperin MY, Koonin E V (2000) Who's your neighbor? New computational approaches for functional genomics. Nat Biotechnol 18:609–613. https://doi.org/10.1038/76443

Ginalski K, Zhang H, Grishin N V (2004) Raptor protein contains a caspase-like domain. Trends Biochem Sci 29:522–524. https://doi.org/10.1016/j.tibs.2004.08.006

Glücksmann A (1951) Cell Deaths in Normal Vertebrate Ontogeny. Biol Rev 26:59–86. https://doi.org/10.1111/j.1469-185X.1951.tb00774.x

Guy L, Ettema TJG (2011) The archaeal "TACK" superphylum and the origin of eukaryotes. Trends Microbiol 19:580–587. https://doi.org/10.1016/j.tim.2011.09.002

Hill SM, Nyström T (2015) The dual role of a yeast metacaspase: What doesn't kill you makes you stronger. Bioessays 37:525–531. https://doi.org/10.1002/bies.201400208

Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

Hofmann K (2019) The Evolutionary Origins of Programmed Cell Death Signaling. Cold Spring Harb Perspect Biol 12:a036442. https://doi.org/10.1101/cshperspect.a036442

Hug LA, Baker BJ, Anantharaman K, et al (2016) A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48

Husnik F, McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol 16:67–79. https://doi.org/10.1038/nrmicro.2017.137

Imachi H, Nobu MK, Nakahara N, et al (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. Nature 577:519–525. https://doi.org/10.1038/s41586-019-1916-6

Iranzo J, Lobkovsky A, I.Wolf Y, Koonin E (2014) Virus-host arms race at the joint origin of multicellularity and programmed cell death. Cell Cycle 13:3083–3088. https://doi.org/10.4161/15384101.2014.949496

Jing H, Takagi J, Liu J, et al (2002) Archaeal Surface Layer Proteins Contain β Propeller, PKD, and β Helix Domains and Are Related to Metazoan Cell Surface Proteins. Structure 10:1453–1464.

https://doi.org/10.1016/S0969-2126(02)00840-7

Kaczanowski S (2016) Apoptosis: its origin, history, maintenance and the medical implications for cancer and aging. Phys Biol 13:31001. https://doi.org/10.1088/1478-3975/13/3/031001

Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Klemenčič M, Asplund-Samuelsson J, Dolinar M, Funk C (2019) Phylogenetic Distribution and Diversity of Bacterial Pseudo-Orthocaspases Underline Their Putative Role in Photosynthesis. Front Plant Sci 14:293. https://doi.org/10.3389/fpls.2019.00293

Klemenčič M, Funk C (2018) Structural and functional diversity of caspase homologues in non-metazoan organisms. Protoplasma 255:387–397. https://doi.org/10.1007/s00709-017-1145-5

Klemenčič M, Novinec M, Dolinar M (2015) Orthocaspases are proteolytically active prokaryotic caspase homologues : the case of *Microcystis aeruginosa*. Mol Microbiol 98:142–150. https://doi.org/10.1111/mmi.13110

Klim J, Gładki A, Kucharczyk R, et al (2018) Ancestral State Reconstruction of the Apoptosis Machinery in the Common Ancestor of Eukaryotes. G3 (Bethesda) 8:2121–2134. https://doi.org/10.1534/g3.118.200295

Koonin E, Aravind L (2002) Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ 9:394–404. https://doi.org/10.1038/sj/cdd/4400991

Koonin E V., Zhang F (2017) Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. BioEssays 39:1–9. https://doi.org/10.1002/bies.201600186

Koonin E V (2016) Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. F1000Research 5:1805. https://doi.org/10.12688/f1000research.8737.1

Koonin E V (2009) On the origin of cells and viruses: primordial virus world scenario. Ann N Y Acad Sci 1178:47–64. https://doi.org/10.1111/j.1749-6632.2009.04992.x

Koonin E V, Mushegian AR, Galperin MY, Walker DR (1997) Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol Microbiol 25:619–637. https://doi.org/10.1046/j.1365-2958.1997.4821861.x

Kroemer G (1997) Mitochondrial implication in apoptosis. Towards an endosymbiont hypothesis of apoptosis evolution. Cell Death Differ 4:443–456. https://doi.org/10.1038/sj.cdd.4400266

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315

Lee REC, Puente LG, Kaern M, Megeney LA (2008) A non-death role of the yeast metacaspase: Yca1p alters cell cycle dynamics. PLoS One 3:e2956. https://doi.org/10.1371/journal.pone.0002956

Lema A S, Klemenčič M, Völlmy F, et al (2021) The Role of Pseudo-Orthocaspase (SyOC) of *Synechocystis sp.* PCC 6803 in Attenuating the Effect of Oxidative Stress. Front Microbiol 12:82. https://doi.org/10.3389/fmicb.2021.634366

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239

Li D, Roberts R (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol Life Sci 58:2085–2097. https://doi.org/10.1007/pl00000838

Lockshin A, Williams M (1964) Programmed cell death—II. Endocrine potentiation of the breakdown of the intersegmental muscles of silkmoths. J Insect Physiol 10:643–649. https://doi.org/10.1016/0022-1910(64)90034-4

Maddison WP (1997) Gene Trees in Species Trees. Syst Biol 46:523–536. https://doi.org/10.1093/sysbio/46.3.523

McLuskey K, Mottram JC (2015) Comparative structural analysis of the caspase family with other clan CD cysteine peptidases. Biochem J 466:219–232. https://doi.org/10.1042/BJ20141324

McLuskey K, Rudolf J, Proto WR, et al (2012) Crystal structure of a *Trypanosoma brucei* metacaspase. Proc Natl Acad Sci U S A 109:7469–7474. https://doi.org/10.1073/pnas.1200885109

Michod RE, Nedelcu AM (2003) Cooperation and conflict during the unicellular – multicellular and prokaryotic – eukaryotic transitions. In: Moya A, Font E (eds) Evolution: From molecules to ecosystems, 1st edn. Oxford University Press, Oxford, pp 195–208

Mignotte B, Vayssiere J-L (1998) Mitochondria and apoptosis. Eur J Biochem 252:1–15. https://doi.org/10.1046/j.1432-1327.1998.2520001.x

Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Minina EA, Coll NS, Tuominen H, Bozhkov P V. (2017) Metacaspases versus caspases in development and cell fate regulation. Cell Death Differ 24:1314–1325. https://doi.org/10.1038/cdd.2017.18

Minina EA, Staal J, Alvarez VE, et al (2020) Classification and Nomenclature of Metacaspases and Paracaspases: No More Confusion with Caspases. Mol Cell 77:927–929. https://doi.org/10.1016/j.molcel.2019.12.020

Mistry J, Chuguransky S, Williams L, et al (2021) Pfam: The protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913

Moreira D, Lopez-Garcia P (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol 47:517–530. https://doi.org/10.1007/pl00006408

Ndhlovu A, Durand P, Ramsey G (2021) Programmed cell death as a black queen in microbial communities. Mol Ecol 30:1110– 1119. https://doi.org/10.1111/mec.15757

Nedelcu AM, Driscoll WW, Durand PM, et al (2011) On the paradigm of altruistic suicide in the unicellular world. Evolution 65:3–20. https://doi.org/10.1111/j.1558-5646.2010.01103.x

Nelson-sathi S, Sousa FL, Röttger M, et al (2015) Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature 517:77–80. https://doi.org/10.1038/nature13805.Origins

Nelson KE, Clayton RA, Gill SR, et al (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323–329. https://doi.org/10.1038/20601

Nguyen L, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304. https://doi.org/10.1038/35012500

Ponting CP, Aravind L, Schultz J, et al (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J Mol Biol 289:729–745. https://doi.org/10.1006/jmbi.1999.2827

Proto WR, Castanys-Munoz E, Black A, et al (2011) *Trypanosoma brucei* Metacaspase 4 Is a Pseudopeptidase and a Virulence Factor. J Biol Chem 286:39914–39925. https://doi.org/10.1074/jbc.M111.292334

Proto WR, Coombs GH, Mottram JC (2013) Cell death in parasitic protozoa: regulated or incidental? Nat Rev Microbiol 11:58–66. https://doi.org/10.1038/nrmicro2929

Ramisetty BCM, Natarajan B, Santhosh RS (2015) MazEF-mediated programmed cell death in bacteria: “What is this?” Crit Rev Microbiol 41:89–100. https://doi.org/10.3109/1040841X.2013.804030

Ratel D, Boisseau S, Nasser V, et al (2001) Programmed Cell Death or Cell Death Programme? That is the Question. J Theor Biol 208:385–386. https://doi.org/10.1006/jtbi.2000.2218

Rawlings N, Bateman A (2019) Origins of peptidases. Biochimie 166:4–18. https://doi.org/10.1016/j.biochi.2019.07.026

Rawlings ND, Waller M, Barrett AJ, Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 42:D503–D509. https://doi.org/10.1093/nar/gkt953

Richie DL, Miley MD, Bhabhra R, et al (2007) The *Aspergillus fumigatus* metacaspases CasA and

CasB facilitate growth under conditions of endoplasmic reticulum stress. Mol Microbiol 63:591–604. https://doi.org/10.1111/j.1365-2958.2006.05534.x

Rinke C, Schwientek P, Sczyrba A, et al (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352

Segovia M, Haramaty L, Berges JA, Falkowski PG (2003) Cell death in the unicellular chlorophyte *Dunaliella tertiolecta*. A hypothesis on the evolution of apoptosis in higher plants and metazoans. Plant Physiol 132:99–105. https://doi.org/10.1104/pp.102.017129

Shalini S, Dorstyn L, Dawar S, Kumar S (2015) Old, new and emerging functions of caspases. Cell Death Differ 22:526–539. https://doi.org/10.1038/cdd.2014.216

Shen W, Ren H (2021) TaxonKit: A practical and efficient NCBI taxonomy toolkit. J Genet Genomics. https://doi.org/10.1016/j.jgg.2021.03.006

Shrestha A, Megeney L (2012) The non-death role of metacaspase proteases. Front Oncol 2:78. https://doi.org/10.3389/fonc.2012.00078.

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings Int Conf Intell Syst Mol Biol 6:175–82

Srivastav R, Suneja G (2019) Recent Advances in Microbial Genome Sequencing. In: Tripathi V, Kumar P, Tripathi P, et al. (eds) Microbial Genomics in Sustainable Agroecosystems, 1st edn. Springer Singapore, Singapore, pp 131–144

Szallies A, Kubata BK, Duszenko M (2002) A metacaspase of *Trypanosoma brucei* causes loss of respiration competence and clonal death in the yeast Saccharomyces cerevisiae. FEBS Lett 517:144–150. https://doi.org/10.1016/s0014-5793(02)02608-x

Tsiatsiani L, Van Breusegem F, Gallois P, et al (2011) Metacaspases. Cell Death Differ 18:1279–1288. https://doi.org/10.1038/cdd.2011.66

Uren AG, O'Rourke K, Aravind L, et al (2000) Identification of Paracaspases and Metacaspases. Mol Cell 6:961–967. https://doi.org/10.1016/s1097-2765(05)00086-9

van Creveld SG, Ben-Dor S, Mizrachi A, et al (2018) A redox-regulated type III metacaspase controls cell death in a marine diatom. bioRxiv 444109. https://doi.org/10.1101/444109

Vardi A, Eisenstadt D, Murik O, et al (2007) Synchronization of cell death in a dinoflagellate population is mediated by an excreted thiol protease. Environ Microbiol 9:360–369. https://doi.org/10.1111/j.1462-2920.2006.01146.x

Vercammen D, Declercq W, Vandenabeele P, Van Breusegem F (2007) Are metacaspases caspases? J Cell Biol 179:375–380. https://doi.org/10.1083/jcb.200705193

von Heijne G (2006) Membrane-protein topology. Nat Rev Mol Cell Biol 7:909–918. https://doi.org/10.1038/nrm2063

Wagner A, Whitaker RJ, Krause DJ, et al (2017) Mechanisms of gene flow in archaea. Nat Rev Microbiol 15:492–501. https://doi.org/10.1038/nrmicro.2017.41

Waterhouse AM, Procter JB, Martin DMA, et al (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wei Y, Fox T, Chambers SP, et al (2000) The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. Chem Biol 7:423–432. https://doi.org/10.1016/S1074-5521(00)00123-X

Wheeler TJ, Clements J, Finn RD (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics 15:7. https://doi.org/10.1186/1471-2105-15-7

Yang J, Zhang Y (2015) Protein Structure and Function Prediction Using I-TASSER. Curr Protoc Bioinforma 52:5.8.1-5.8.15. https://doi.org/10.1002/0471250953.bi0508s52

Yu JW, Jeffrey PD, Ha JY, et al (2011) Crystal structure of the mucosa-associated lymphoid tissue lymphoma translocation 1 (MALT1) paracaspase region. Proc Natl Acad Sci 108:21004 LP – 21009. https://doi.org/10.1073/pnas.1111708108

Zhang Z, Wang J, Wang J, et al (2020) Estimate of the sequenced proportion of the global prokaryotic genome. Microbiome 8:134. https://doi.org/10.1186/s40168-020-00903-z

FIGURE LEGENDS

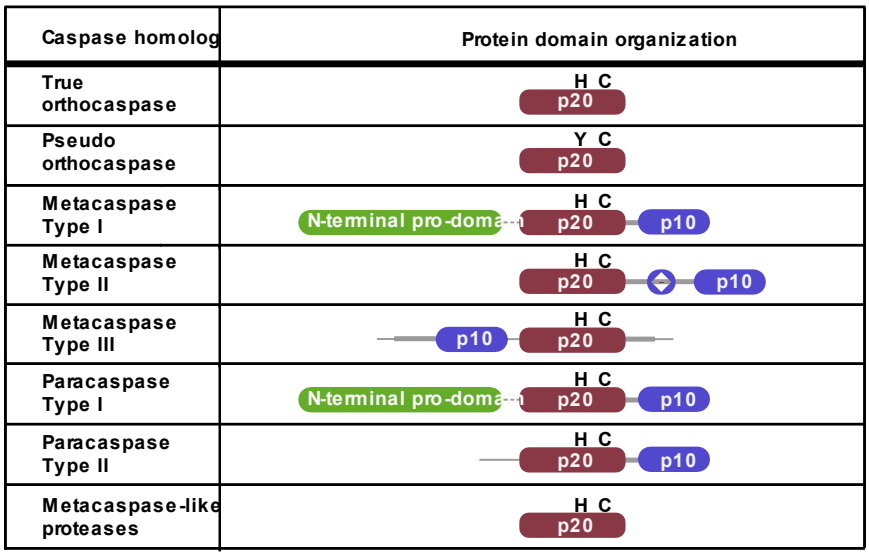

Fig. 1 Classification and domain organization of caspase homologs possessing the catalytic dyad in unicellular organisms according to the unified nomenclature of the C14 peptidase family (adapted from Minina *et al.*, 2020). Members of the family possess the catalytic p20-like region (maroon), a regulatory p10-like region (blue) and/or the linker region (thick grey line). The positions of catalytic His (H) and Cys (C) residues are shown above the p20-like region. The most common YC dyad is shown for the pseudo-orthocaspase. N-terminal pro-domains are indicated in green. Detailed classification is shown in the methodology. The figure is not drawn to scale (Color figure online)

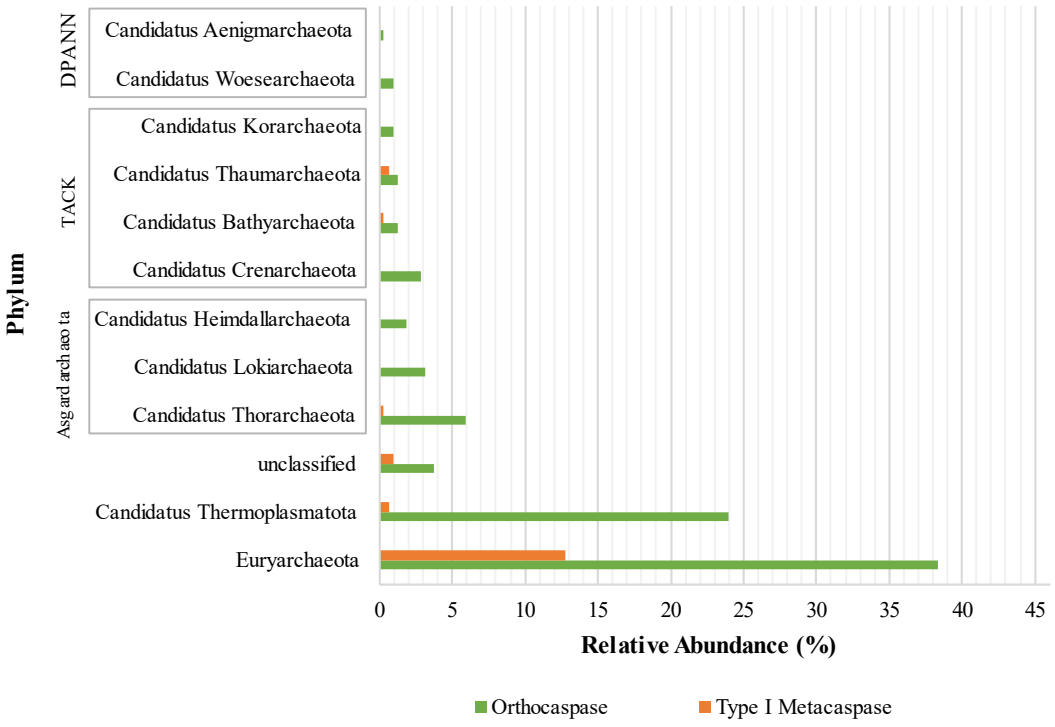

Fig. 2 Taxonomic distribution and abundance (%) of OCAs (green) and type I MCAs (orange) according to different phyla of archaea. Superphyla are indicated in grey boxes. Unclassified refers to sequences from environmental samples where taxonomic information was not available from the NCBI taxonomy database.

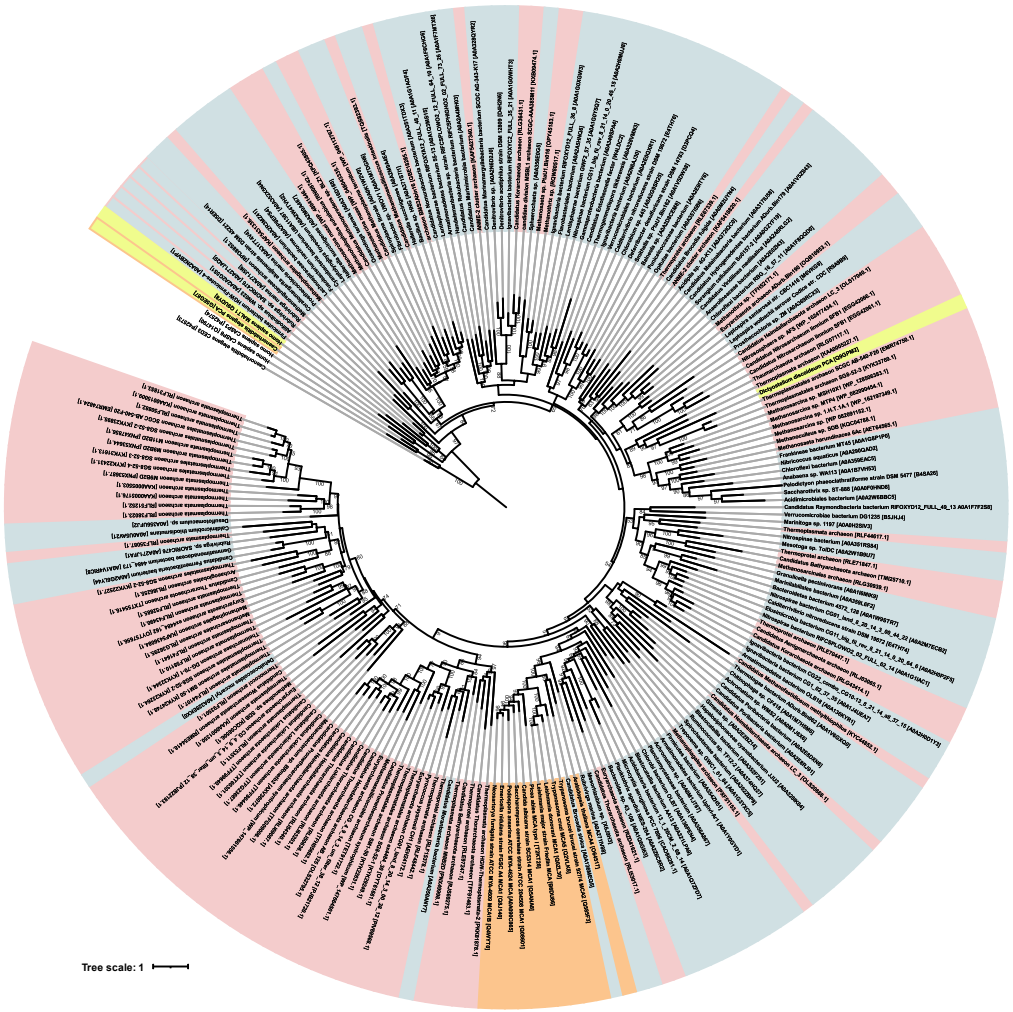

Fig. 3 Maximum likelihood phylogenetic tree of the p20-like region of 231 amino acid sequences of prokaryotic OCAs, and eukaryotic CAs, PCAs, and type I MCAs generated using the LG+R6 model under IQ-TREE. Three Metazoa CAs, *H. sapiens* CA 3 and CA 8, and *C. elegans* cell death protein 3 (UniProt IDs: Q14790, P42573, and P42574), were used as outgroup sequences. The classification is shown in different colors: eukaryotic PCAs (yellow), eukaryotic type I MCAs (orange), archaeal OCAs (red), and bacterial OCAs (blue). The tree reliability was tested using the bootstrap method with 1000 replicates, of which values below 0.50 are not shown. The tree scale bar shows substitutions per site. Corresponding Genbank IDs of archaeal OCAs, and UniProt IDs of eukaryotic CAs, PCAs, and type I MCAs, and bacterial OCAs are shown in brackets.

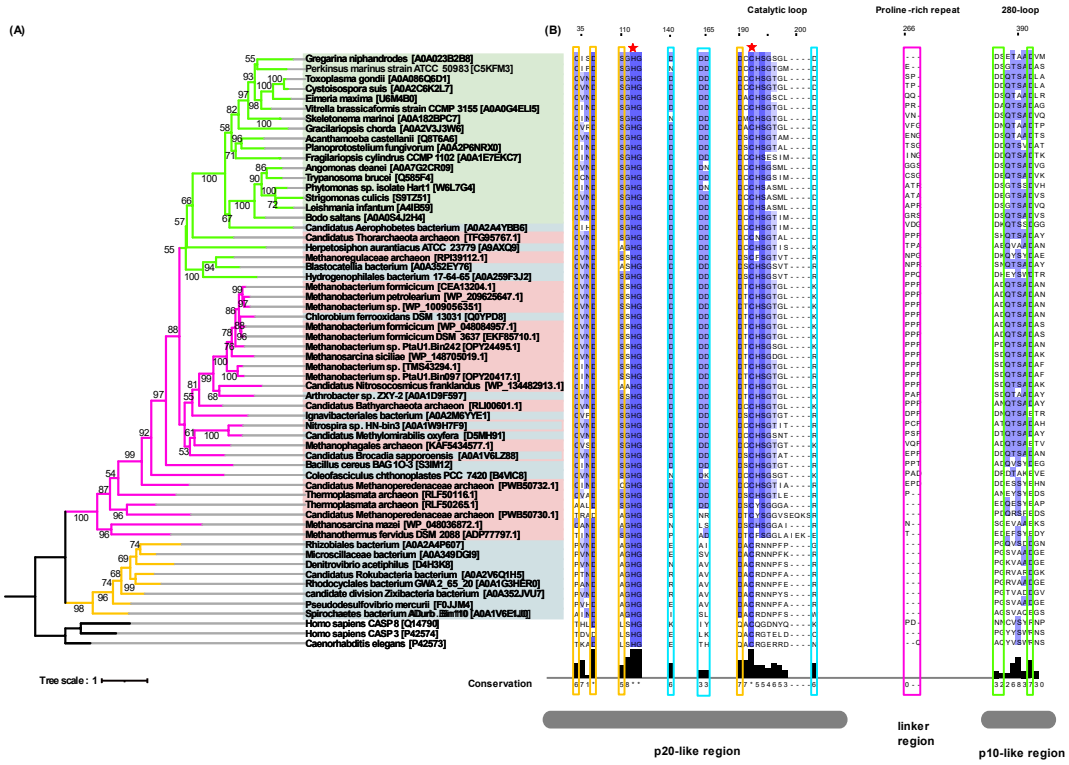

Fig. 4 (A) Maximum likelihood phylogenetic tree of MCAs in the three domains of life generated using the LG+R5 model under IQ-TREE. The tree is based on the C14_peptidase domain of 69 protein sequences. Three Metazoa CAs, *H. sapiens* CA 3 and CA 8, and *C. elegans* cell death protein 3 (UniProt IDs: Q14790, P42573 and P42574), were used as outgroup sequences. The classification according to the three domains of life is shown in different colors: eukarya (green), archaea (red), and bacteria (blue). Different clades are shown in different colors: clade 1 (orange), clade 2 (pink), and clade 3 (green). Bootstrap values based on 1000 replicates calculated by maximum likelihood. Bootstrap values below 0.50 are not shown. The tree scale bar shows substitutions per site. Corresponding Genbank IDs of archaea and UniProt IDs of eukaryotes and bacteria are shown in brackets. **(B)** Partial MSA of key residues of type I MCAs in the three domains of life under hmmalign as well as MAFFT (See Methods). The figure was generated using Jalview. Colors of the amino acid have been assigned according to the Percentage Identity color scheme on Jalview with an identity threshold above 0.50. The catalytic HC dyad is indicated with a red star. Residues involved in the formation of the S1 pocket are indicated in orange, the proline-rich repeat is shown in pink, and Ca^{2+} binding sites are shown in blue (high affinity) and green (low affinity). The p20-like, linker, and p10-like regions are shown below the MSA.

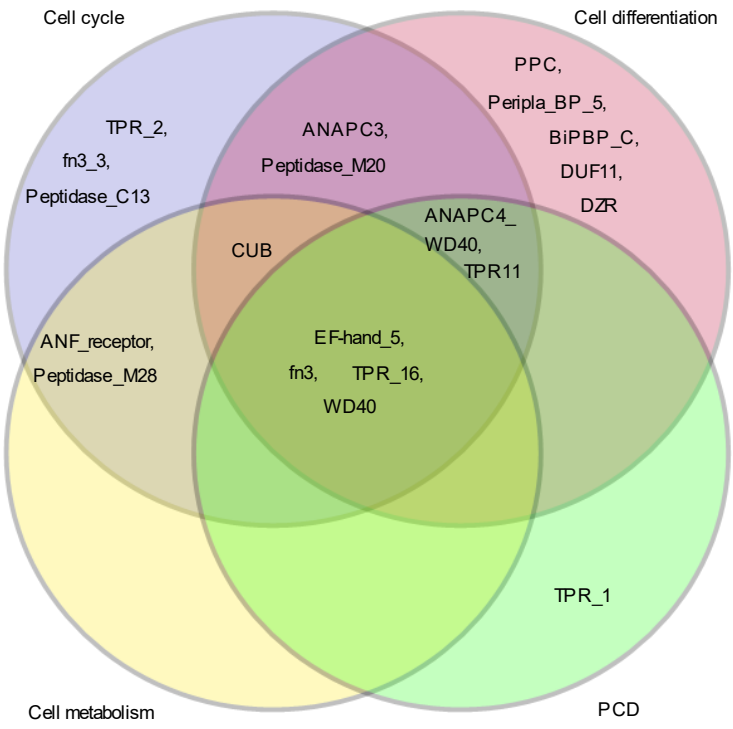

Fig. 5 Dual role of C14 peptidase-accompanying domains which were predicted to have cell survival and death-associated functions. Cell survival associated functions refer to functions mentioned by Ameisen which contributed to the rise of the PCD genetic toolkit (Ameisen, 2002). C14 peptidase-accompanying domain associated with cell survival or death has been classified according to appropriate categories: cell cycle (blue), cell metabolism (orange), cell differentiation (purple), and PCD (green). C14 peptidase-accompanying domains include ANAPC3 (Anaphase-promoting complex, cyclosome, subunit 3), ANAPC4_WD40 (Anaphase-promoting complex subunit 4 WD40 domain), ANF_receptor (Receptor family ligand binding region), BiBP_C (Penicillin-Binding Protein C-terminus Family), CUB (CUB domain), DUF11 (Domain of unknown function), DZR (Double zinc ribbon), EF-hand_5 (EF hand), fn3 (Fibronectin type III domain), fn3_3 (Polysaccharide lyase family 4, domain II), Peripla_BP_5 (Periplasmic binding protein domain), Peptidase_C13, Peptidase_M20, Peptidase_M28, PPC (Bacterial pre-peptidase C-terminal domain), TPR_1, TPR_2, TPR_11 and TPR_16 (TPR repeat), and WD_40 (WD domain, G-beta repeat). Detailed information is available in the supplementary table S4, Supplementary Materials.

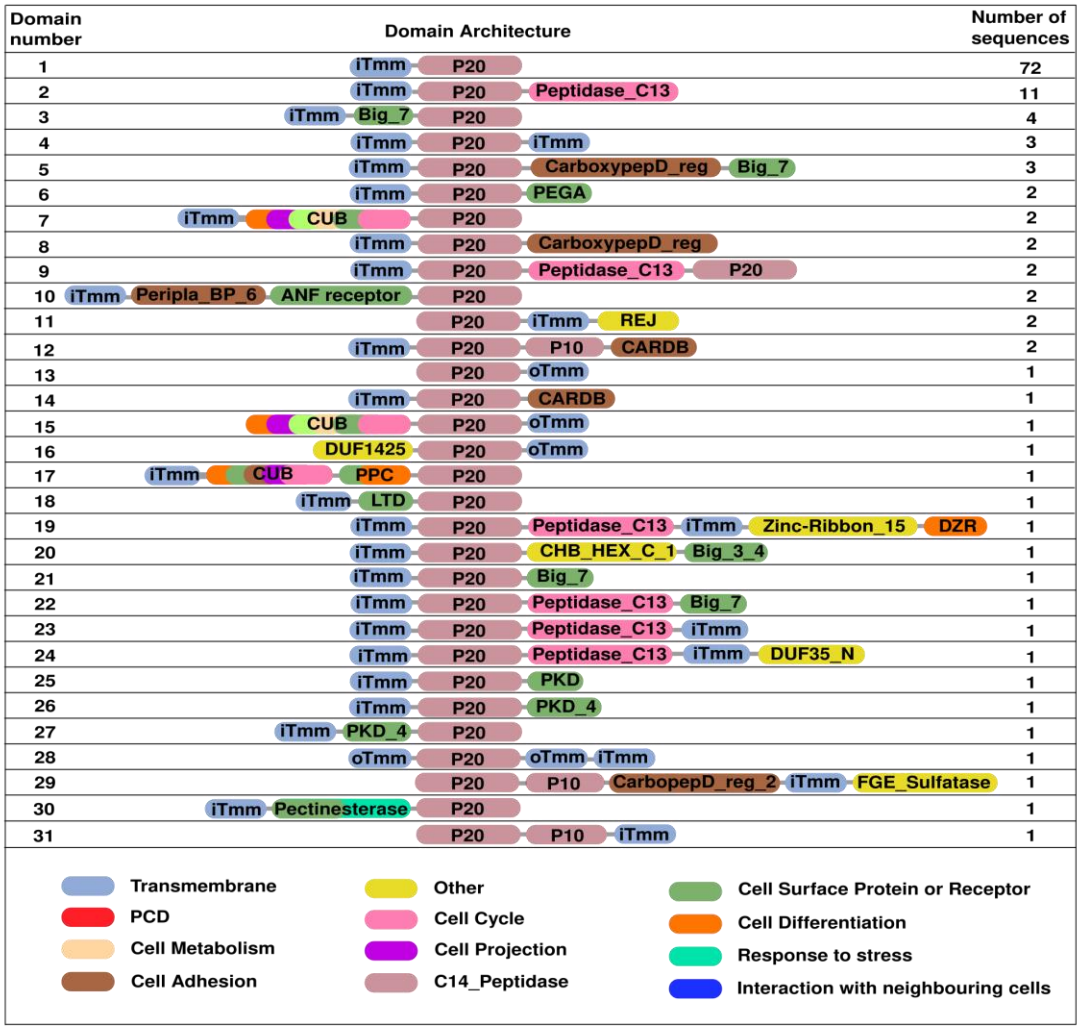

Fig. 6 Domain architectures of transmembrane archaeal OCAs and type I MCAs. Domains (oval) and distances (grey line) between domains are not drawn to scale. Domains are presented with N-terminals on the left and C-terminals on the right. iTmm refers to transmembrane helices predicted to have N-terminal domains situated intracellularly and C-terminal domains extracellularly. oTMM refers to transmembrane helices predicted to have the reverse topology. Functional categories and color keys of C14 peptidase-accompanying domains and transmembrane domains are indicated below the domain architectures. Only functions associated with cell survival, death, or interactions with neighboring cells and the environment are indicated. Functions labeled 'other' (yellow) refer to domains associated with functions other than PCD, cell survival, or cell-cell interactions. P20- and p10-like regions of the C14 peptidase domain are colored lilac. If a domain was associated with more than one function, all the potential functions were indicated with the corresponding color.

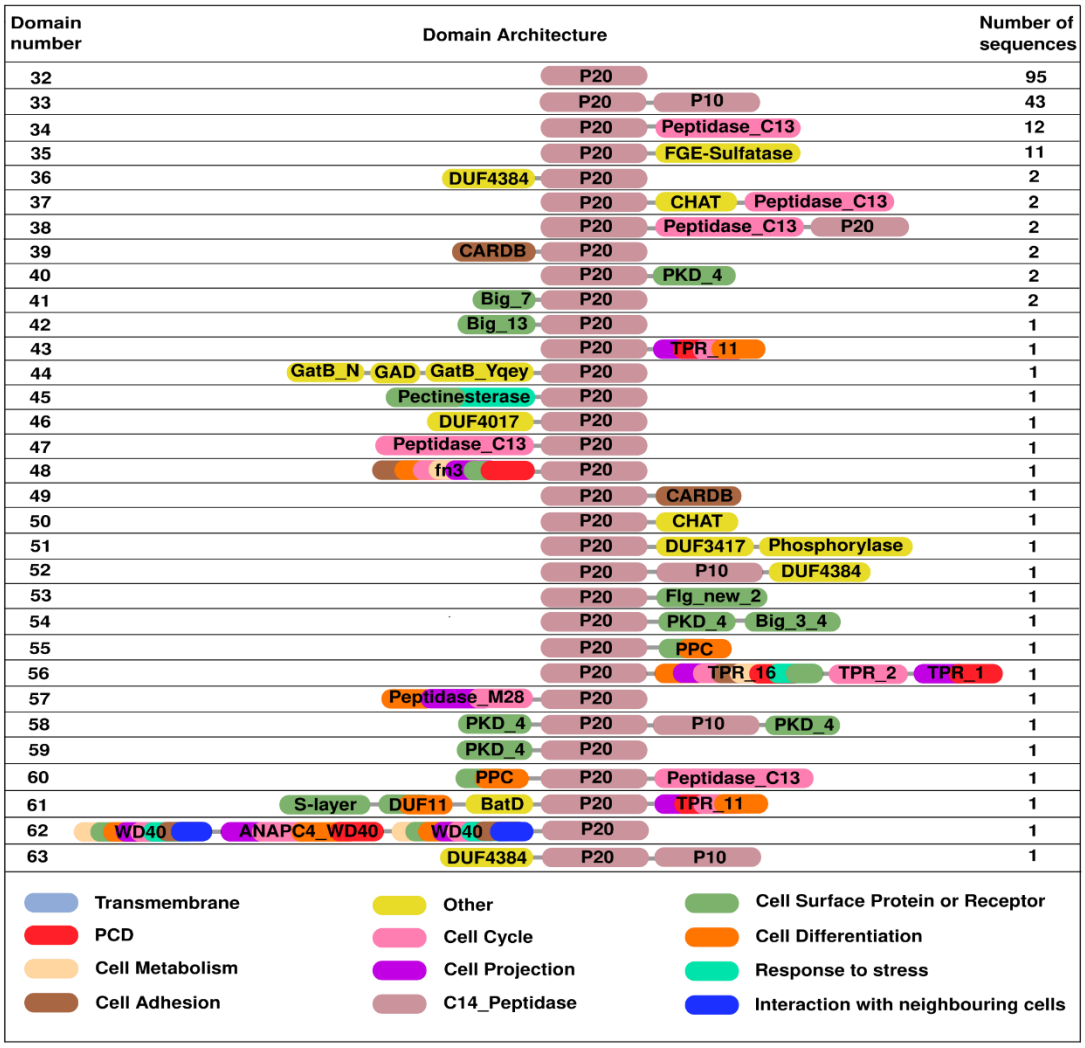

Fig. 7 Domain architectures of 195 cytosolic archaeal OCAs and type I MCAs. Domains (oval) and distances (grey line) between domains are not drawn to scale. Domains are presented with N-terminals on the left to C-terminals on the right. iTmm refers to transmembrane helices predicted to have N-terminal domains situated intracellularly and C-terminal domains extracellularly. oTMM refers to transmembrane helices predicted to have reverse topology. Functional categories associated with cell survival, death, or interaction with neighboring cells and the environment were considered, and the color key of C14 peptidase-accompanying domains and transmembrane domains is shown below the domain architectures. Functions labeled 'Other' refer to domains associated with functions other than PCD, cell survival, or interaction with neighboring cells and the environment. P20- and p10-like region of the C14 peptidase domain is colored lilac. If a domain was associated with more than one function, all the potential functions were indicated with the corresponding colors.

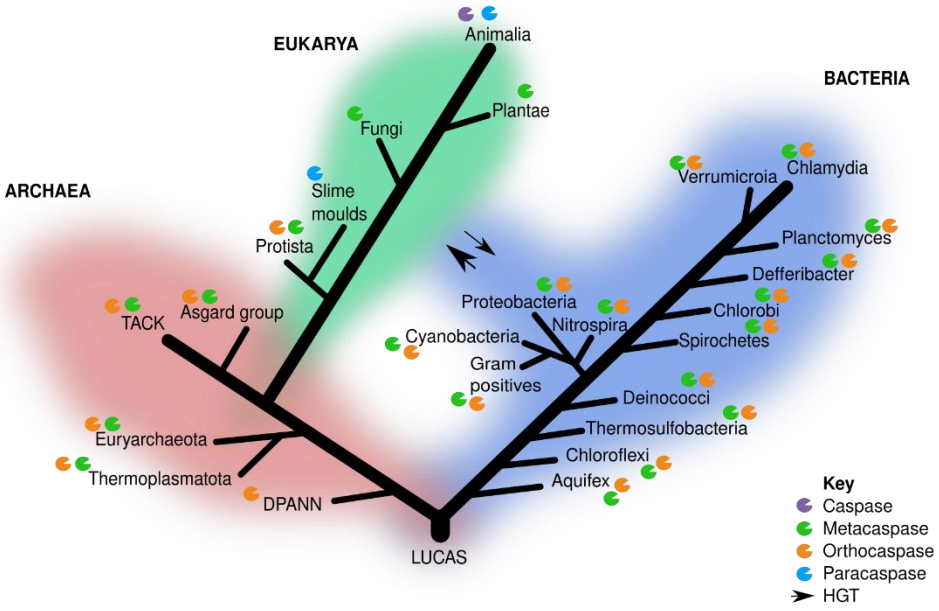

Fig. 8 Canonical 3-domain tree of life with caspase homologs. Only the groups of interest included in this study are indicated. HGT events between the domain eukarya and bacteria are indicated by the arrows and thickness indicates the frequency of the event.

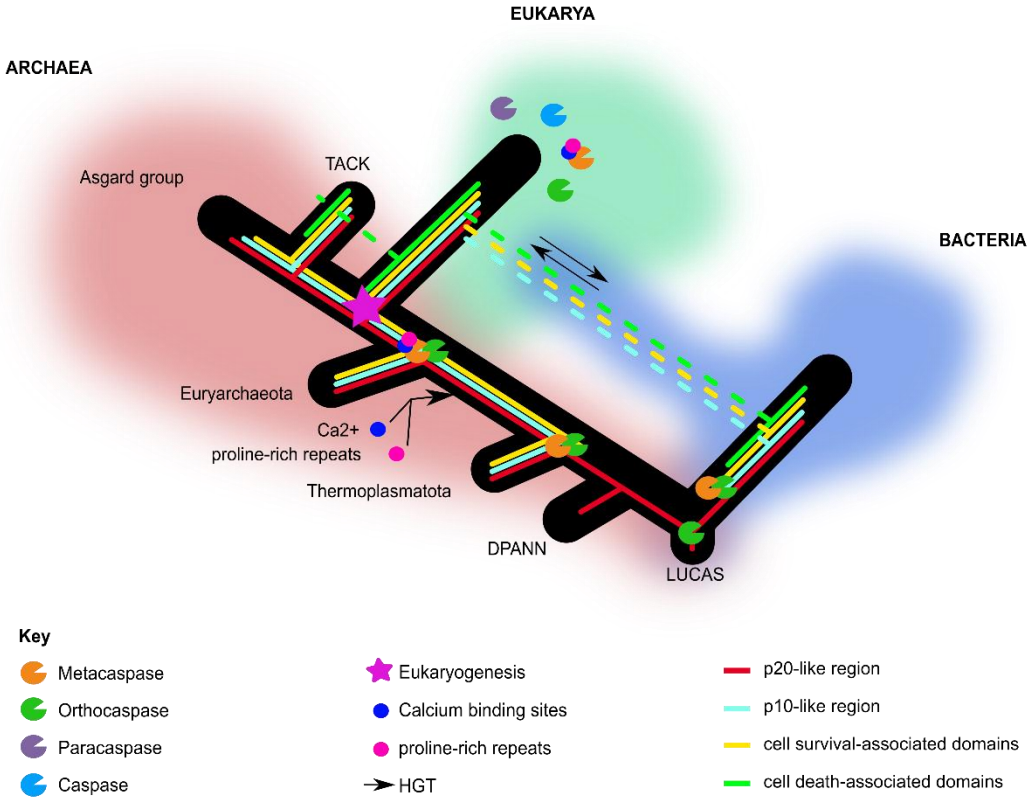

Fig. 9 Proposed acquisitions of the different C14 peptidase domains and their associated pro-survival or pro-death domains in archaea. A simple C14 peptidase distribution is shown on the tree of life: archaea (red), bacteria (blue), and eukarya (green). The acquisitions of domains are shown in lines of different colors: p20-like region (red), p10-like region (light blue), C14 peptidase-accompanying domains associated with cell survival (yellow), and death domains (green). The important regions specific to type I MCAs, the Ca^{2+} binding sites (blue circle) and proline-rich repeats (pink circle), are indicated. LUCAS refers to the Last Universal Common Ancestor State. The point of endosymbiosis leading to eukarya is shown with a star. Horizontal gene transfer events between the domain eukarya and bacteria are indicated by the arrows.

CHAPTER 3

INVESTIGATION OF PUTATIVE CASPASE- AND METACASPASE-LIKE ACTIVITY IN AN ARCHAEAL MODEL ORGANISM IN RESPONSE TO OXIDATIVE STRESS

3.1 Introduction

C14 peptidases were thought to be absent in archaea and studies of CA- and MCA-like activity in archaeal organisms have been limited (Aravind and Koonin, 2002; Bidle *et al.*, 2010). CA-like activity refers to enzyme activity that is specific for CA substrates containing Asp residues (Wei *et al.*, 2000) (Table 3) but the activity is observed in organisms that are known to be absent of CAs. These include organisms excluding metazoa. MCA-like activity refers to enzyme activity specific for MCA substrates containing basic Arg or Lys residues (Table 3) but the activity is present in organisms that do not possess MCA sequences according to the *in silico* analysis (Uren *et al.*, 2000; Choi and Berges, 2013).

Table 3 Fluorogenic peptide substrates used for the detection of activity of CA3, CA8 and MCA.

C14 peptidase	Substrate	Reference
CA3	Asp–Glu–Val–Asp-AFC (DEVD-AFC, Sigma-Aldrich)	(Nicholson *et al.*, 1995)
CA8	Ile–Glu–Trp–Asp-7-Amido-4-trifluoromethylcoumarin (IETD-AFC, Sigma-Aldrich)	(Nicholson *et al.*, 1995)
MCA	Ac-Val-Arg-Pro-Arg-AMC (Ac-VRPR-AMC, Bachem)	(Tsiatsiani *et al.*, 2011)

According to my knowledge, the only study that investigated CA-like activity in archaea was done by (Bidle *et al.*, 2010). The study revealed high CA-like activity in four archaeal organisms belonging to phyla Euryarchaeota and Crenarchaeota without recognizable C14 peptidase sequences in their genomes (Bidle *et al.*, 2010). Substrates specific for CA8 (Ile–Glu–Trp–Asp-7-Amido-4-trifluoromethylcoumarin) and CA3 (Asp–Glu–Val–Asp-AFC) were used. Hydrolytic assays using diverse protease substrates and inhibitors revealed that CA-like activity in one of the studied organisms, *Haloferax volcanii*, resemble CA4 (Seth-Pasricha, Bidle and Bidle, 2013). Subsequent analyses based on genome-enabled proteomics, structural analysis and targeted gene knockouts revealed that the CA-like activity is associated with unfolded protein responses and ER-associated degradation pathways (Seth-Pasricha *et al.*, 2019). However, the detected CA-like activity is likely caused by enzymes other than CAs. This is because CAs are exclusively found in Metazoa, whereas serine proteases are found

in plants and exhibit CA-like activity in plants (Egger *et al.*, 2003; Cedzich *et al.*, 2009). Another limitation of this study is that they did not measure MCA-like activity in archaeal organisms that possess C14 peptidases. Since MCAs are present in prokaryotes, MCA enzyme activity should be measured. For example, MCA activity should be measured in *Methanosarcina sp.* that possess C14 peptidase sequences (Bidle *et al.*, 2010) rather than CA-like activity that is not specific to C14 peptidase associated enzyme activity in prokaryotes.

In chapter two, I reported that C14 peptidases are present in archaea. The taxonomic distribution and phylogenetic analysis of C14 peptidases across three domains of life revealed C14 peptidases were possibly present since LUCAS, and their phyletic pattern analysis revealed they were more likely associated with pro-survival functions rather than PCD. This led me to examine whether there is CA- and MCA-like activity in *Halobacterium salinarum* strain NRC-1, a commonly used model organism of archaea, in response to oxidative stress, and whether the activity is associated with cellular survival or PCD. Hydrogen peroxide (H_2O_2) was chosen as the applied oxidative stress to deduce the relationship between oxidative stress and CA- and MCA-like activity. Oxidative stress caused by ROS was one of the conflicts that arose during eukaryogenesis and archaeal organisms would have had to develop a mechanism to mediate this conflict (Frade and Michaelidis, 1997; Kroemer, 1997; Mignotte and Vayssiere, 1998; Blackstone and Green, 1999).

According to Aravind and Koonin, *H. salinarum* only possesses the CHAT (Caspase HetF Associated with Tprs) domains, part of the CHF family (Aravind and Koonin, 2002). They are also classified in the C50 peptidase family according to the MEROPS database (Rawlings *et al.*, 2014). These proteins possess the HC dyad and they are suggested as the ancestors of the C14 peptidase family (Aravind and Koonin, 2002). As previous analyses by Bidle and colleagues revealed CA-like activity is detected in archaeal organisms with no C14 peptidase protein sequences (Bidle *et al.*, 2010), it was unclear if the model organism would portray enzyme activity or not.

This chapter aims to identify potential C14 peptidases or other proteases associated with CA- and MCA-like activity in response to oxidative stress. An *in silico* analysis to look for C14 peptidases in the model organism *H. salinarum* was performed as well as fluorometric enzyme assays to detect CA- and MCA-like activity. I present evidence that C14 peptidases are not present in *H. salinarum* and there is an absence of CA- and MCA-like activity in response to oxidative stress applied by H_2O_2.

3.2 Materials and Methods

3.2.1 Growth and culture conditions of *H. salinarum* strain NRC-1

An archaeal isolate, *H. salinarum* strain NRC-1, was obtained from the Institute for Systems Biology, University of Washington. It was grown aerobically with shaking (100 rpm) at 37 °C (Nuttall *et al.*, 2008), and cultured in an optimal conditioned medium (CM) 4.28 M medium which contained NaCl (250 g/l), $MgSO_4.7H_2O$ (20 g/l), Kcl (2 g/l), Na-citrate dihydrate (3 g/l), bacteriological peptone (10 g/l) and uracil without any trace metals (50 mg/l) (Bidle and Bender, 2008). The pH of the medium was adjusted to 7.25 using the STARTER 3100A pH meter (OHAUS; Parsippany, USA). Growth in liquid culture was monitored using the Biomate 5 spectrophotometer (ThermoFisher; Waltham, MA, USA).

3.2.2 Taxonomic confirmation through the 16S rRNA gene

The taxonomic identity of the isolate was confirmed using 16S rRNA sequencing. The analysis was performed at Inqaba Biotechnical Industries, South Africa. Universal primers pA/Bak-F (5'-AGAGTTTGATCCTGGCTCAG-3') and Ph/Bak-R (5'-AAGGAGGTGATCCAGCCGCA-3') were used to amplify the 16S rRNA gene target region (Weisburg *et al.*, 1991; Turner *et al.*, 1999). Results were obtained by BLASTN 2.11.0 search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.*, 1997).

3.2.3 Phylogenetic analysis

Sequences belonging to the phylum Euryarchaeota were selected to provide a phylogenetic placement of the 16S rRNA sequence of *H. salinarum* NRC-1 (Archaea Bak-R). Fourteen 16S rRNA nucleotide sequences were downloaded from https://www.ezbiocloud.net. Nucleotide sequences were aligned using MAFFT v.7.470 (Katoh & Standley, 2013) under default parameters. By using IQ-TREE and the statistical method of maximum likelihood, the best substitution model was chosen. The phylogenetic tree was constructed using IQ-TREE (Hoang *et al.*, 2018) under a maximum likelihood (ML) approach. The reliability of the constructed tree was tested using the Bootstrap method with 1000 replicates in IQ-TREE (Hoang *et al.*, 2018). The phylogenetic tree was visualized using the Interactive Tree of Life (iTOL) v5.0 online service (https://itol.embl.de) (Letunic & Bork, 2019).

3.2.4 Growth curve analysis in response to H_2O_2 stress

One hundred millilitre cultures were grown in a 125 mL Erlenmeyer flask (DWK Life Sciences GmbH, Mainz, Germany) at 37 °C. Cultures were inoculated with a log preculture to an initial optical density of 600 nm (OD_{600}) ~0.001 and shaken at 100rpm. Growth was monitored using a Biomate 5 spectrophotometer (ThermoFischer Waltham, MA, USA). For H_2O_2 exposure experiments, cultures were diluted in the CM medium to the mid-logarithmic phase cultures (OD_{600} ~0.2) and supplemented with 3% (w/v) H_2O_2 to final concentrations of 10 and 25 mM H_2O_2. Growth curve and stress exposure experiments were carried out in triplicate.

3.2.5 *In silico* analysis: searching for the C14 peptidase domain in the *H. salinarum* genome

Protein sequences from the *H. salinarum* NRC-1 genome (RefSeq ID: NZ_CP038631.1 (chromosome), NZ_CP038632.1 (pHSAL1), and NZ_CP038633.1 (pHSAL2)) were obtained from GenBank Genome Database (https://www.ncbi.nlm.nih.gov/genome/). The raw Peptidase_C14 (PF00656) and CHAT (PF12770) HMM profiles were obtained from the Pfam database (http://pfam.xfam.org/) (Mistry *et al.*, 2021). The genome was subjected to an HMMER v3.0 package hmmsearch (http://hmmer.org/) (Eddy, 1998) under a domain-independent E-value ≤ 0.0001 and score/bias ratio ≥ 10. Profile Hidden Markov Models (HMM) are used in Hmmsearch to detect sequence similarity and homology (Eddy 1998). The C14 peptidase HMM profile was aligned against the raw protein sequences of C14 peptidases of Euryarchaeota obtained from chapter two and the raw protein sequence of the CHAT domain of *H. salinarum* using hmmalign to identify regions of similarity. Subtilisin-like serine proteases are suggested to exhibit CA-like activity in plants (Hatsugai *et al.*, 2009), and the organism's genome was subjected to hmmsearch against the raw Peptidase_S8 (PF00082) and the Proteasome (PF00227) HMM profiles, specific to serine proteases, obtained from the Pfam database. Hits with E-values ≤ 0.0001 and a score/bias ratio ≥ 10 were accepted.

3.2.6 Detection of CA-like and MCA-like activity in *H. salinarum*

To test for CA- and MCA-like activity, 20 ml of *H. salinarum* culture were pelleted by centrifuging at 3000 x g and resuspended in cold phosphate-buffered saline (PBS) solution. The cells were lysed using 500 µl of lysis buffer [50 mM HEPES, 0.1% CHAPS, 10 mM DTT] (Spungin, Bidle and Berman-Frank, 2019). Cell extracts were then centrifuged (10 000 x g, 2 min) and 5 µl of supernatant was transferred into 0.5 ml tubes which contained 200 µl of NaCl-optimized assay buffer [50 mM HEPES (pH 7.3), 1.5 M NaCl, 10% sucrose, 0.1% CHAPS and 10 mM DTT] and the fluorogenic peptide substrate. Different substrates were used to detect the activity of different C14 peptidases (Table 3). Purified recombinant human CA3 (ThermoFischer; Waltham, MA, USA) and CA8 protein (Abcam; Cambridge, UK) served as a positive control (0.5 µg/ml per assay) respectively. No positive controls were used for MCAs but the AMC (amido-4-methylcoumarin) standard curve was accounted for when measuring its enzyme activity and the negative controls were carefully set out. Negative controls were a) sample blank with no substrates b) substrate blank with no cell lysates and c) DMSO control. To verify potential CA- and MCA-like activity, 2 µl of CA3 inhibitor (Ac-DEVD-CHO), broad-spectrum CA inhibitor (z-VAD-FMK) and MCA inhibitor (Antipain dihydrochloride N-(Nα-Carbony- Arg- Val-Arg-al)-Phe) were used to inhibit CA3-, CA8- and MCA-like activity, respectively. All tests were conducted in triplicate and fluorescence intensity (Ex 365 nm, Em 420 nm) was measured over one hour period with measurements taken at 10 min intervals using a Portable Fluorometer Fluo-100A (AllSheng; China). The standard AMC curve was produced by a series of seven diluted AMC solutions

in the concentration range of 50 nM to 6 µM using the NaCl-optimized assay buffer. A single factor ANOVA test was performed on Microsoft Excel (version 2112) to evaluate the significant difference between the activity of the sample exposed to inhibitors and no inhibitors. A p-value < 0.05 was considered significant.

3.2.7 Measurement of protein sample concentration

The Bradford protein assay (Bradford, 1976) was used to measure the concentration of total protein in each treatment sample. Bovine Serum Albumin (BSA) standard curve was created using a series of BSA (Sigma-Aldrich) standards in the 1.5625 µg/ml to 1000 µg/ml range. The protein concentration of each sample was calculated by adding 60 µl of unknown protein sample to 1.5 ml of Bradford reagent (Sigma-Aldrich) followed by 10 minutes incubation. The absorbance was measured at 595 nm using Biomate 5 spectrophotometer (ThermoFisher Waltham, MA, USA).

3.3 Results

3.3.1 Growth curve of *H. salinarum*

H. salinarum reached the exponential phase after 10 hours and the stationary phase after 36 hours in the CM medium. The sigmoid growth curve produced is shown in Figure 6. The culture after 36 hours of growth is shown in Figure A1, Appendix A.

To measure the effect of oxidative stress on the growth of *H. salinarum*, the organism was exposed to different levels of H_2O_2 (0 mM, 10 mM, and 25 mM) and their growth was measured at OD_{600} (Figure 7). The sigmoid growth curve of the sample exposed to 0 mM H_2O_2 (Figure 7) matched the growth curve of the organism shown in Figure 6. A decrease in cell density was observed for groups treated with 10 mM and 25 mM of H_2O_2 with a similar trend until 20 hours of exposure. There was an increase in cell density of the sample exposed to 10 mM of H_2O_2 between 3 and 5h (OD_{600} = 0.185 to OD_{600} = 0.223), and between 19h and 55h (OD_{600} = 0.133 to OD_{600} = 0.312) (Figure 7). The sample exposed to 25 mM of H_2O_2 reached the death phase after 45 hours of growth.

3.3.2 Taxonomic confirmation of *H. salinarum*

The taxonomic identity of the isolate was confirmed as *H. salinarum* strain NRC-1 The top ten BLAST results are provided in Table 4 and the rest are available in Table A1, Appendix A. The phylogenetic tree of *H. salinarum* with other organisms belonging to phylum Euryarchaeota revealed a correct phylogenetic placement within the phylum (Figure 8).

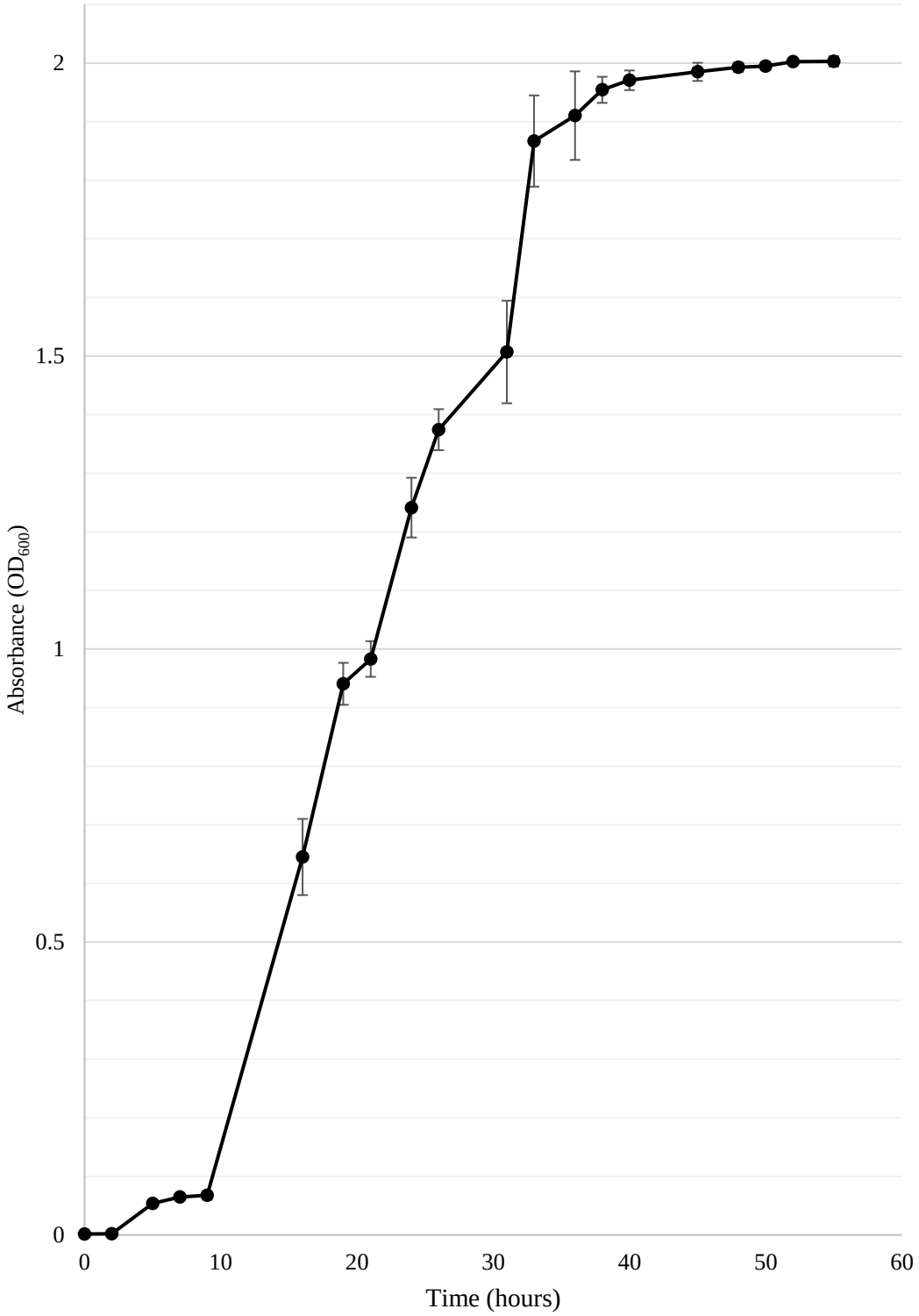

Figure 6 Growth curve of *H. salinarum* in CM media monitored by changes in optical density at OD_{600}. The mean value with ± SD (Standard Deviation) is indicated (n = 3).

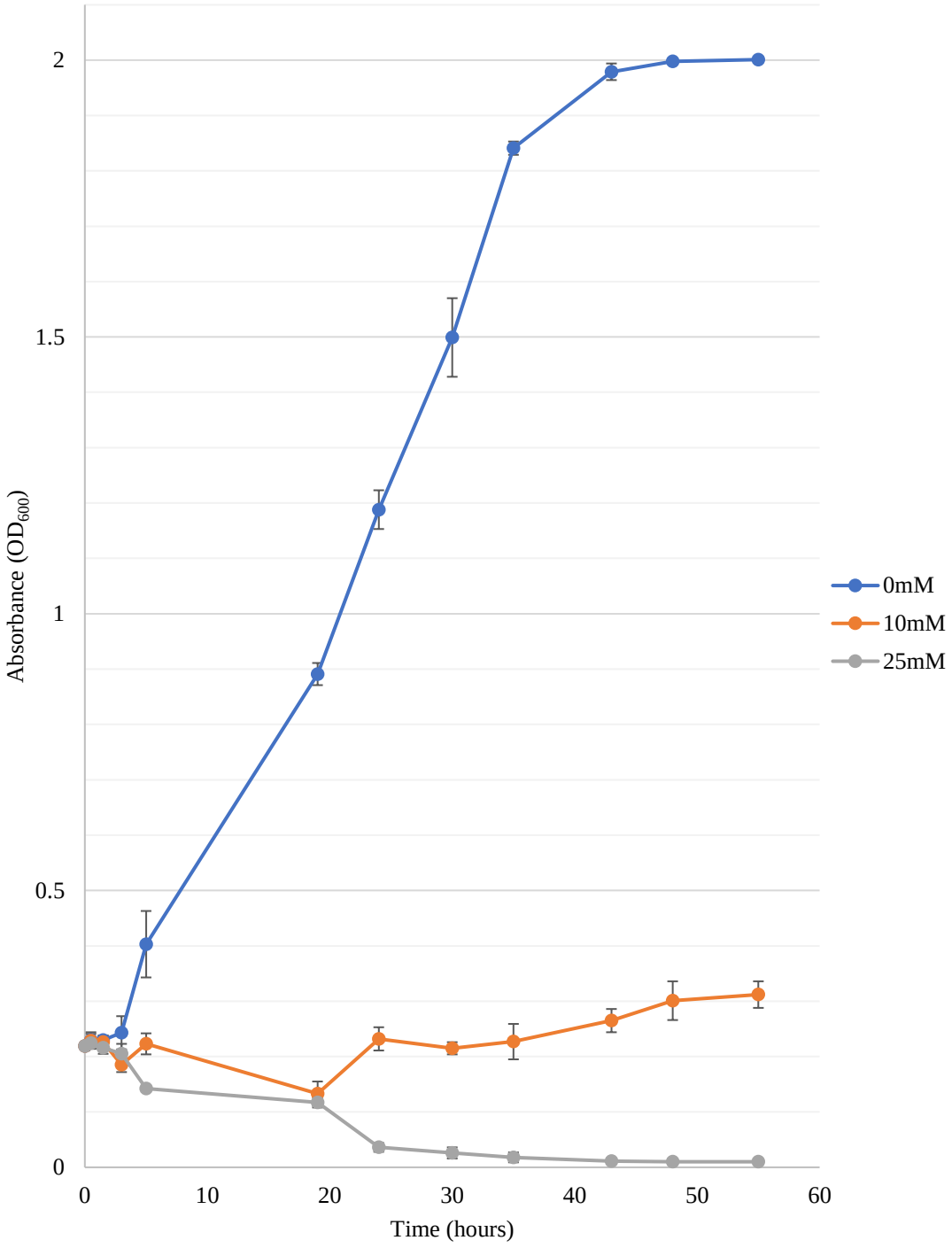

Figure 7 Growth curve of *H. salinarum* in CM media under oxidative stress. The growth of *H. salinarum* in the CM medium with 0mM (blue), 10mM (orange) and 25mM of H_2O_2 (grey) was monitored by changes in optical density at 600 nm (OD_{600}). The mean value with ± SD is indicated (n = 3).

Table 4 Top 10 BLASTn hits for taxonomic identification of the sample using the primer Archaea_BAK-R_E07_13. The full list of BLASTn results is available in Table A1, Appendix A.

Sequences producing significant alignments:	**Organism**	**Score (Bits)**	**E-value**
gi|1621075037|gb|CP038633.1|	*H. salinarum NBRC 14715*	1577	0
gi|1621075037|gb|CP038633.1|	*H. salinarum NRC-1,*	1577	0
gi|1621074903|gb|CP038632.1|	*H. salinarum strain 91-R6*	1577	0
gi|167728379|emb|AM774417.1|	*H. salinarum strain NES7*	1577	0
gi|167728226|emb|AM774416.1|	*H. salinarum strain NES5*	1577	0
gi|1700659907|tpg|BK010831.1|	*H. salinarum strain NES2*	1577	0
gi|1700659907|tpg|BK010831.1|	*H. salinarum strain NES1*	1577	0
gi|1700659701|tpg|BK010830.1|	*H. salinarum strain 67A*	1577	0
gi|1700659701|tpg|BK010830.1|	*H. salinarum strain 66A*	1577	0
gi|291369028|gb|CP001953.1|	*H. salinarum strain 64A*	1577	0

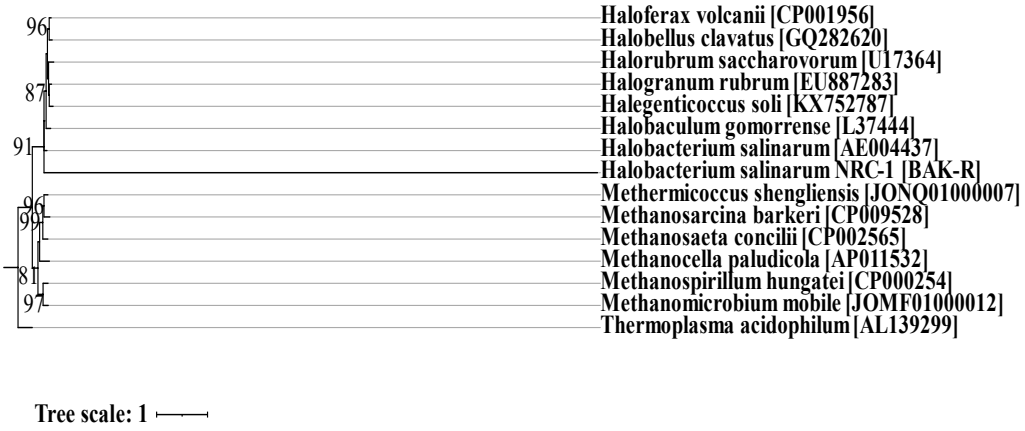

Figure 8 16S rRNA tree of taxonomic classification of the *H. salinarum* sample (Bak-R) and DNA sequences belonging to organisms of phylum Euryarchaeota reconstructed at maximum likelihood using GTR+F+R3 model. *Thermoplasmatota acidophilum* (Ezbiocloud database ID: AL139299) was

used as the outgroup. The tree scale bar shows substitutions per site. Corresponding Ezbiocloud database IDs are shown in brackets next to each organism name.

3.3.3 *In silico* analysis: searching for the C14 peptidase domain

The results from querying the *H. salinarum* genome using hmmsearch and the Peptidase_C14 HMM profile domain did not yield any statistically significant hits under the set E-value and bit score threshold. However, there was a statistically significant hit for the CHAT domain in the *H. salinarum* genome (Table 5). In addition, statistically significant hits for the Peptidase_S8 and Proteasome HMM profiles against the *H. salinarum* genome was observed (Table 5). MSA of the CHAT domain of *H. salinarum* with C14 peptidases of organisms belonging to phylum Euryarchaeota revealed only the H151 of the HC dyad to be present in the domain. The Cys of the HC dyad as well as the key residues required for substrate specificity and enzyme activity were absent (Figure 9).

Table 5 Hmmsearch hits of *H. salinarum* protein sequences against the Proteasome, CHAT and Peptidase S8 Pfam domains.

Pfam domain	Accession number	Target sequence accession number	E-value	Start position of domain	End position of domain
Proteasome	PF00227	WP_010902078.1	1.3e-59	33	217
Proteasome	PF00227	WP_012289244.1	1.6e-48	46	225
CHAT	PF12770	WP_012289324.1	2.1e-06	322	459
Peptidase_S8	PF00082	WP_136360974.1	1.9e-58	147	395
Peptidase_S8	PF00082	WP_010903433.1	2.2e-38	146	459

3.3.4 CA-like and MCA-like activity in *H. salinarum* in response to H_2O_2

CA3-, CA8- and MCA-like activity was not observed in *H. salinarum* response to H_2O_2 (Figure 10 and 11). The standard AMC curve (Figure A2, Appendix A) was used to determine CA- and MCA-like activity and the results were normalized to µg/ml protein. Negative values were reported for certain treatment groups (Table A2 and A3, Appendix A). There was no difference in activity between the sample and the inhibited sample (p-value > 0.05; Table A2 and A3, Appendix A). High CA3- and CA8- like activities (> 50-fold enzyme activity) were detected in the positive controls (Table A2 and A3, Appendix A). According to the BSA standard curve used to estimate the protein concentration of samples (Figure A3, Appendix A), the concentration of the protein extract of each treatment group exceeded the standard concentration required for the assay (Figure 12 and 13). The caveat to this approach is that proteins present in this organism were assumed as native folded proteins.

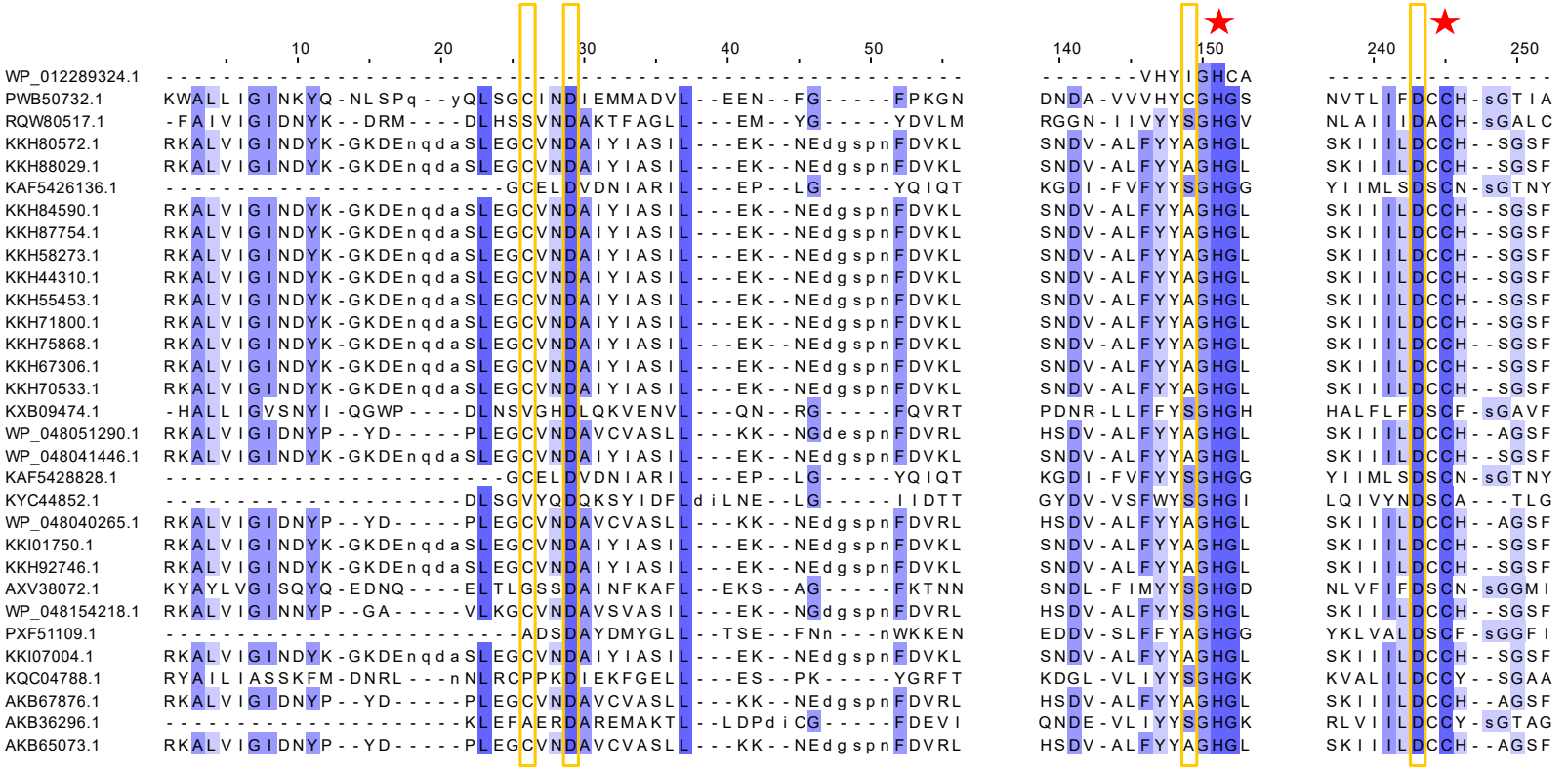

Figure 9 Partial MSA of CHAT domain of *H. salinarum* (Genbank ID: WP_012289324.1) and 30 C14 peptidase sequences belonging to phylum Euryarchaeota generated using hmmalign. The 30 C14 peptidase sequences were chosen according to the highest pairwise identity. The figure was generated using Jalview. Shades of the amino acid have been assigned according to the percentage identity colour scheme on Jalview with an identity threshold above 0.50. The catalytic HC dyad is indicated by a red star. Residues involved in the formation of the S1 pocket are indicated in orange.

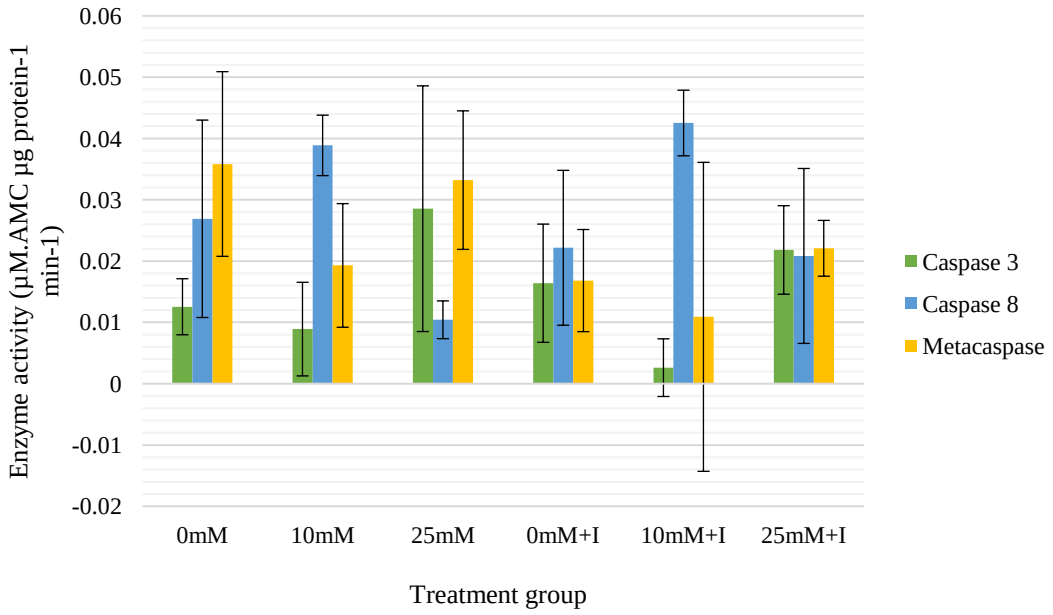

Figure 10 Measured CA3- (green), CA8- (blue) and MCA-like activity (yellow) in *H. salinarum* of each treatment group after exposure to one hour of H_2O_2. Treatment groups with added inhibitors are indicated by +I. Results are expressed in mean ± SD (n = 3). Differences in activity between samples and the inhibited samples were not significant (P-value > 0.05). Table showing the measured activity of each treatment group is available in Table A2, Appendix A.

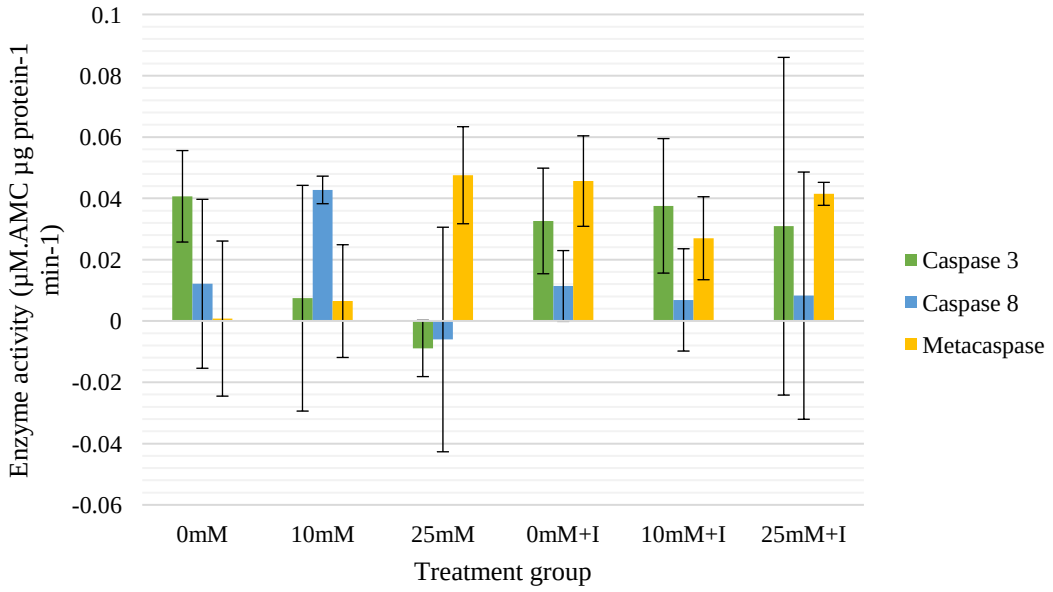

Figure 11 Measured CA3- (green), CA8- (blue) and MCA-like activity (yellow) in *H. salinarum* of each treatment group after exposure to three hours of H_2O_2. Treatment groups with added inhibitors

are indicated by +I. Results are expressed in mean ± SD (n = 3). Differences in activity between samples and inhibited samples were not significant (P-value > 0.05). Table showing the measured activity of each treatment group is available in Table A3, Appendix A.

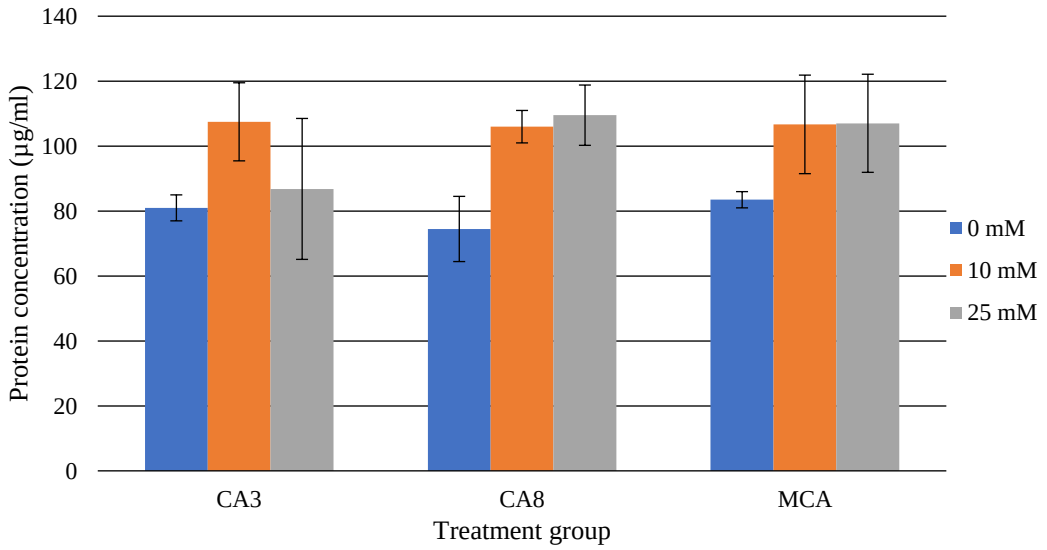

Figure 12 Concentration of extracted proteins (µg/ml) of samples of different treatment groups (CA3, CA8 and MCA) exposed to one hour of different concentrations of H_2O_2 (0 mM (blue), 10 mM (orange) and 25 mM (grey)) determined using the Bradford assay and the BSA standard curve (Figure A3, Appendix A).

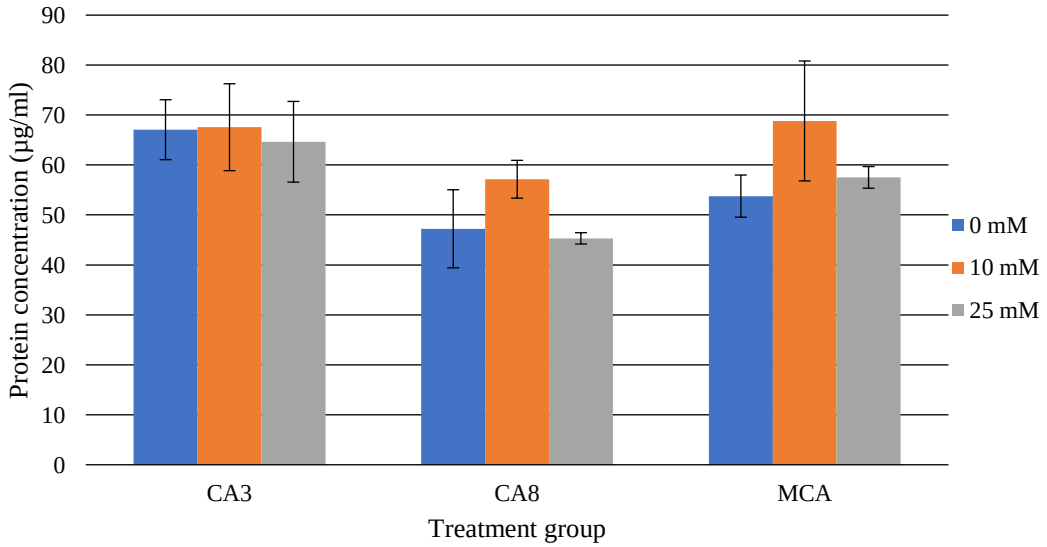

Figure 13 Concentration of extracted proteins (µg/ml) of samples of different treatment groups (CA3, CA8 and MCA) exposed to three hours of different concentrations of H_2O_2 (0 mM (blue), 10 mM

(orange) and 25 mM (grey)) determined using the Bradford assay and the BSA standard curve (Figure A3, Appendix A).

3.4 Discussion

3.4.1 Growth of *H. salinarum* in response to high levels of oxidative stress

The halophilic *H. salinarum* grew well in the CM medium with the growth curve (Figure 6) consistent with previous growth curves of this organism (Shand and Betlach, 1991; Facciotti *et al.*, 2010). The effect of oxidative stress on the growth of *H. salinarum* was analysed (Figure 7) by exposing the organism to varying levels of H_2O_2. High concentrations of H_2O_2 (10 mM and 25 mM) were used because the organism possesses catalase-peroxidase (UniProtKB ID: O73955) which reduces the oxidative stress applied by H_2O_2 (Robinson *et al.*, 2011; Sharma *et al.*, 2012). The steady rise in the growth curve of the organism after 19 hours of exposure to 10 mM of H_2O_2 suggests that the organism was recovering from oxidative stress (Figure 7). However, *H. salinarum* could not survive when exposed to a higher degree of oxidative stress (25 mM of H_2O_2) (Figure 7). A limitation of this study is that the used H_2O_2 concentrations varied by a wide margin (0 mM, 10 mM, and 25 mM) and a detailed description of how the organism grew weren't provided. A second growth curve of *H. salinarum* is required by exposing the organism to varying levels of oxidative stress within a smaller range of concentrations of H_2O_2 (i.e., 0 mM, 5 mM, 10 mM, 15 mM, 20 mM and 25 mM). This would provide a good indication of how *H. salinarum* responds to oxidative stress.

3.4.2 Absence of the C14 peptidase domain in the *H. salinarum* genome

The C14 peptidase domain was not found in the currently available protein sequences from the *H. salinarum* NRC-1 genome (RefSeq ID: NZ_CP038631.1 (chromosome), NZ_CP038632.1 (pHSAL1), and NZ_CP038633.1 (pHSAL2)), as per the previously reported absence in *Haloacterium* sp. by Aravind and Koonin (2002). In chapter two, C14 peptidases were not identified in *Halobacteria* organisms, and they seem to be absent in the entire class. In chapter two, C14 peptidases were suggested as ancestral proteins present prior to the divergence of prokaryotes because of the wide distribution of C14 peptidases in all three domains of life. C14 peptidases were found in all major phyla of archaea, including the phylum Euryarchaeota of which class Halobacteria belong. The wide distribution of C14 peptidases suggests that the absence may be explained by a gene loss event which is a dominant mechanism in the evolution of archaeal genomes (Koonin and Wolf, 2008; Csűrös and Miklós, 2009) and more frequent than horizontal gene transfer (Snel, Bork and Huynen, 2002; Kunin and Ouzounis, 2003). Therefore, the most parsimonious explanation for the absence of C14 peptidases in *Halobacterium* is a gene loss event. As archaea normally live in extreme environments, the loss of this protein may be explained by the frequent loss of proteins that is unnecessary for the function of organisms under significant energy stress (Valentine, 2007).

Aravind and Koonin previously identified the CHAT domain in *H. salinarum* and suggested the domain is related to C14 peptidases (Aravind and Koonin, 2002). In the MEROPS database, they are classified in the C50 peptidase family of the CD clan and as a sister group to the separins. The CHAT domain identified in this organism (Table 5) may be a divergent member from ancestral C14 peptidases of archaea. However, the domain does not possess the key residues required for a catalytically active C14 peptidase (Figure 9). This includes the HC dyad, two Asp residues, Cys and Ser that are present in the S1 pocket and are required for substrate specificity and enzyme activity (McLuskey *et al.*, 2012). Based on the sequence analysis, the domain will likely not exhibit any CA- and MCA-like activity in *H. salinarum.*

3.4.3 Absence vs low detection of CA- and MCA-like activity in *H. salinarum*

Low fluorescence readings were measured for CA- and MCA-like activity in *H. salinarum* compared to the activity detected in human recombinant proteins used as the positive control (Figure 10 and 11). The negative readings were observed because lower fluorescence was detected in the readings than the fluorescence of the initial reading. The most likely reason is dynamic fluorescence quenching, which refers to a biological process that decreases the fluorescence intensity of a sample, due to denaturation of a protein (Lakowicz, 2006). There was no significant difference between the level of CA- and MCA-like activity of the sample and the sample exposed to inhibitors (Table A2 and A3, Appendix A). The reason for the absence of significant differences is the absence of C14 peptidase sequences in this organism. Therefore, under the parameters set, CA- and MCA-like activity was undetected in *H. salinarum* in response to oxidative stress. Although not directly tested in this study, these results suggest that ROS stress does not trigger CA- or MCA-dependent PCD in *H. salinarum.*

Alternative studies regarded the low CA-like and MCA-like activity detected as significant (Bidle *et al.*, 2010; Spungin, Bidle and Berman-Frank, 2019). For example, CA- and MCA-like activity detected in the marine cyanobacterium *Trichodesmium* was between 0.1 nM and 5 nM for CA8 activity, and between 0 pM and 60 pM for MCA-like activity (Spungin, Bidle and Berman-Frank, 2019). The authors stated the change in the degree of detected activity in response to different stress is suggestive of significant activity (Spungin, Bidle and Berman-Frank, 2019). In addition, there was a significant difference between the level of CA- and MCA-like activity of the sample and the sample exposed to inhibitors to support the presence of the enzyme activity.

Based on the absence of measurable changes in activity levels and the absence of inhibition observed of enzyme activity, it can be concluded that CA- or MCA-like activity in the *H. salinarum* was undetected, despite the higher level of activity detected (Figure 10 and 11) than the previous study done by Bidle *et al.*, 2010. The controls set for this experiment and the experiment performed in triplicate further supports this statement. No marginal errors and significant values are supported by

the manufacturer of the assay kit (ThermoFischer; Waltham, MA, USA) to use as a standard for significant values. Although the absence of proof of CA- and MCA-like activity in *H. salinarum* is not proof of absence, there was no evidence of CA- and MCA-like activity using DEVD-AFC, IETD-AFC and Ac-VRPR-AMC substrates in this organism.

3.4.4 Absence of CA- and MCA-like activity in *H. salinarum*

The absence of detected CA- and MCA-like activity in *H. salinarum* in response to oxidative stress correlates with the *in silico* analysis of which no C14 peptidase sequences were detected in the organism and the CHAT domain did not possess the required substrates for enzyme activity (Figure 9 and Table 5). However, the absence of C14 peptidase sequences does not confirm the absence of CA-like activity. This is because CA-like activity was observed in archaeal organisms that did not possess any C14 peptidase sequences (Bidle *et al.*, 2010). In the study by Bidle and colleagues, Archaea lacking C14 peptidase sequences showed CA-like activity, whereas organisms possessing the sequences showed no activity. The absence of CA-like activity can therefore be explained by reasons other than missing sequences.

The absence of CA- and MCA-like activity can be explained by the following alternatives. Firstly, high levels of H_2O_2 (10mM and 25mM) were used for exposure of the organism to oxidative stress, and the majority of the organisms may be dead for a significant level of enzyme activity to be detected. This is observed in the decrease in cell density of samples exposed to H_2O_2 (Figure 7). No enzyme activity would be detected in the sample with 0 mM of H_2O_2 as no stress was applied. A wider range of H_2O_2 concentrations should be used to make sure the organism is exposed to an adequate degree of oxidative stress for the detection of putative enzyme activity. Another reason may be due to time points used to measure enzyme activity, as low cell densities measured in the growth curve after one and three hours of exposure to H_2O_2 (OD_{600} ~0.2) (Figure 7) suggests a majority of the organisms were dead. MCAs of *Trichodesmium sp.* exhibited a 50-fold increase in enzyme activity after 30 hours of stress induction (Spungin, Bidle and Berman-Frank, 2019). CA- and MCA-like enzyme activities should be measured when higher cell densities (i.e., when the organism reaches the log phase) have been reached after longer hours of stress exposure. Lastly, oxidative stress may not be appropriate stress for CA- and MCA-like enzyme activity to be detected in *H. salinarum.* The previous study that measured CA-like activity in *H. volcanii* was conducted under salt and temperature stress, not under the stress used in this study (Bidle *et al.*, 2010).

Serine proteases have been found to exhibit CA-like activity. This has been documented in several taxa, such as StSBTc-3 of *Solanum tuberosum* (Hatsugai *et al.*, 2009) and PBA1 of *Arabidopsis thaliana* (Cedzich *et al.*, 2009). The StSBTc-3 of *S. tuberosum* possesses the Peptidase_S8 domain (UniProt ID: PF00082.24) and PBA1 of *A. thaliana* possesses the Proteasome domain (UniProt ID: PF00227.28).

Both domains were identified in *H. salinarum* (Table 5), but significant CA-like activity was not observed. It seems serine proteases present in *H. salinarum* are not responsible for CA-like activity in *H. salinarum*. Potential reasons for the absence of activity are that the serine protease pathway is underdeveloped in *H. salinarum,* or serine proteases in this organism do not exhibit enzyme activity in response to oxidative stress.

Based on the lack of C14 peptidase sequences and associated enzyme activity, it appears the organism would not undergo CA- and MCA-dependent PCD under oxidative stress. However, it is uncertain if *H. salinarum* undergoes PCD at all in order to detect the activity of associated enzymes since no studies have been conducted for archaeal PCD.

3.4.5 Absence of C14 peptidases of *H. salinarum* is consistent with PCD as a black queen trait

The BQ hypothesis refers to 'the availability of public goods due to leakiness which provides conditions for some taxon to relinquish the cost of performing a leaky vital function' and refers to a BQ trait as an adaptive gene loss (Morris, Lenski and Zinser, 2012; Ndhlovu, Durand and Ramsey, 2021). The absence of C14 peptidase proteins as well as CA- and MCA-like activity can be explained by PCD being a Black Queen (BQ) trait. For a trait to be a BQ, it needs to satisfy four conditions: (i) leaky within the community to be used by other species; (ii) costly; (iii) vital to the community; and (iv) performed by a fraction of the community (Morris, Lenski and Zinser, 2012). PCD has been suggested as a BQ trait because it fulfils two of the four criteria with empirical data required for the other two criteria (Ndhlovu, Durand and Ramsey, 2021). Firstly, PCD is certainly costly for the organism and secondly, PCD is a leaky function resulting in public goods. In the *H. salinarum* and *Dunaliella salina* microbial loop, a small subset of *D. salina* cells undergoes PCD at the onset of darkness and release C- and N-based dissolved organic matter which is used by the *H. salinarum* for survival (Orellana *et al.*, 2013). Under continuous light conditions, *H. salinarum* can induce PCD in *D. salina* to metabolize and remineralize the dissolved organic matter available. Orellana and colleagues suggest that although the initial role of PCD in certain *D. salina* cells were for nutrient supplementation in its population, *H. salinarum* was able to exploit this process for its own needs. As seen in the loop, PCD is leaky in the community to release dissolved organics matter, it is vital to the community because *H. salinarum* is dependent on the remineralized glycerol, and it is only performed by a small subset of the *D. salina* population, which is consistent with PCD as a BQ trait in this microbial loop (Orellana *et al.*, 2013; Ndhlovu, Durand and Ramsey, 2021).

MCAs are suggested as an example to investigate PCD as a BQ trait (Ndhlovu, Durand and Ramsey, 2021). This is because of the patchy distribution of MCAs in diverse lineages (Bidle and Falkowski, 2004; Asplund-Samuelsson, Bergman and Larsson, 2012), and the gene loss events in certain taxa are suggested as adaptive gene loss events. As seen in chapter 2, C14 peptidases were identified in the

Euryarchaeota phylum but weren't identified in the Halobacteria class. The absence of the C14 peptidase gene in *H. salinarum* may be the result of an ancestral gene loss for the organism where it may have been more favourable for them to not carry the costly gene anymore. Based on the induction of PCD of *D. salina* by *H. salinarum* in the microbial loop (Orellana *et al.*, 2013), the gene loss event for the Halobacteria class is consistent with the claim that PCD is a BQ trait in the microbial loop. In addition, the adaptive gene loss of C14 peptidases in *H salinarum* and the absence of CA- and MCA-like activity in response to oxidative stress (Table A2 and A3, Appendix A) suggests *H. salinarum* does not undergo MCA-dependent PCD under oxidative stress. Although *H. salinarum*, induces PCD in *D. salina* under a continuous light regime, it is uncertain if the induction of PCD is oxidative stress or density-dependent. Further *in vivo* analysis to determine if the *H. salinarum* undergoes PCD in response to oxidative stress is required.

3.4.6 Alternative evolutionary scenarios for the absence of C14 peptidases in *H. salinarum*

Alternative evolutionary scenarios may be provided for the absence of the C14 peptidase gene and its associated enzyme activity, other than the BQ hypothesis. Firstly, it is possible that the gene was never present in the organism in the first place. However, if we assume that C14 peptidases have been present since LUCA (La, Ndhlovu, and Durand 2022), this evolutionary scenario is unlikely. Secondly, the gene may have been lost through HGT events. HGT events are common within the C14 peptidase family and this has been explored elsewhere (Wang *et al.*, 2018; La, Ndhlovu and Durand, 2022) as well as in prokaryotes (Aravind, Dixit, and Koonin 2001; Maddison 1997; Ochman, Lawrence, and Groisman 2000; Ponting et al. 1999). In addition, microbial symbionts undergo genome reduction events to be energy efficient during gene replication and only allow key genes required for survival to be retained (McCutcheon and Moran 2012; Simonsen 2022). Thirdly, the genes may have diverged into a completely novel genes over time and cannot be detected by homology search strategies. Lastly, an alternative gene other than C14 peptidases may be responsible for programmed cell death in *H. salinarum*. This may be the most likely alternative explanation as gasdermins were identified to be associated with cell death in bacteria (Johnson et al. 2022). However, further analyses are required to confirm if *H. salinarum* does indeed undergo PCD.

3.5 Conclusion

This study investigated the presence of CA- and MCA-like activity in the archaeal organism, *H. salinarum*, when exposed to H_2O_2. Based on the *in silico* analysis and the fluorometric enzyme assays, C14 peptidases are not present in its genome and CA- and MCA-like activity is absent in response to oxidative stress. Although CHAT domains that are closely related to C14 peptidases were present in the organism, no proteolytic activity was observed due to the lack of key residues. Additionally, serine

proteases that are known to exhibit CA-like activity in plants were identified in *H. salinarum*. However, obtained results suggest that they do not exhibit CA-like activity in response to oxidative stress. The results presented here provide evidence that *H. salinarum* does not possess C14 peptidases nor does it exhibit the associated enzyme activity. The lack of C14 peptidases in the Halobacteria class, including *H. salinarum*, is consistent with PCD as a BQ trait. However, further studies to assess if the loss of C14 peptidases in *H. salinarum* is an evolutionary response to PCD acting as a BQ in the microbial loop are required.

CHAPTER 4

CONCLUDING REMARKS AND FUTURE WORK

This research was aimed at examining the role of C14 peptidases in archaea and analyzing their evolutionary history prior to eukaryogenesis. Based on sequence and tertiary structure analyses, archaeal and eukaryotic C14 peptidases are homologous. The wide distribution of C14 peptidases in the three domains of life and the incongruency between the species tree and the C14 peptidase gene tree suggests the presence of C14 peptidases prior to the divergence of prokaryotes is the most parsimonious explanation. Phyletic pattern analysis revealed a higher abundance of cell survival associated C14 peptidase-accompanying domains than death domains, which supports a pro-survival related function of ancestral cell death toolkits rather than a pro-death function. Furthermore, C14 peptidases of Asgard archaea were found to possess cell surface and receptor-like OCAs. They possessed C14 peptidase-accompanying domains associated with cell adhesion and cell projection, suggesting they may be candidate PCD related genes important in the E3 model of eukaryogenesis.

Although this research illustrates the presence of C14 peptidases prior to the divergence of prokaryotes, the taxonomic distribution and phylogenetic analysis only depict a small representation of archaeal and bacterial C14 peptidases. Additional analyses including sequences from sequenced genomes and emerging approaches such as single-celled genomic may provide a more robust evolutionary history of C14 peptidases and death domains in the three domains of life. In addition, it is acknowledged that phyletic pattern and gene ontology (GO) analyses to infer putative enzyme function is limited and further *in vivo* analysis is required to determine if archaeal C14 peptidases are indeed associated with cell survival or PCD in these organisms.

Investigation of C14 peptidases in the model archaeal organism, *H. salinarum*, revealed the C14 peptidase gene is unlikely to be present. No CA- and MCA-like activity was observed in response to oxidative stress due to H_2O_2 exposure. Although a closely related CHAT domain and serine proteases were identified in the organism's genome, it seems both proteins are not associated with CA- or MCA-like activity. The absence of the C14 peptidase gene and the associated enzyme activity can be explained as an adaptive gene loss event which is consistent with PCD as a BQ trait in the microbial loop between *H. salinarum* and *D. salina.* The initial objective of this experiment was to determine the role of archaeal C14 peptidases in response to oxidative stress, as oxidative stress is suggested as one of the conflicts that lead to the rise of eukaryotic PCD. These results that *H. salinarum* do not possess the C14 peptidase gene nor a CA or MCA-like dependent response to oxidative stress. This dissertation has provided a second study in history to empirically investigate C14 peptidases in archaea.

Based on these conclusions, additional model archaeal organisms are required to investigate the association between archaeal CA- and MCA-like activity and response to oxidative stress. For example, *in vivo* analysis of CA- or MCA-like activity of an Asgard archaeon in response to oxidative stress using fluorometric enzyme assays is required. Whether Asgard archaea undergo PCD under oxidative stress or not is also an important question to ask to determine the role of PCD during eukaryogenesis.

Lastly, to further understand the implications of the noteworthy absence of C14 peptidases in certain archaeal lineages and investigate if PCD is a BQ trait in the microbial loop between *H. salinarum* and *D. salina*, future studies could include an *in vivo* PCD study in *H. salinarum* in response to oxidative stress and determine if PCD is absent in the organism.

In conclusion, this dissertation investigated the presence of C14 peptidases in archaea and analyse their evolutionary history prior to eukaryogenesis as well as the association between CA- or MCA-like protease activity and oxidative stress in *H. salinarum*. The work presented here provides a robust evidence-based evolutionary history of C14 peptidases prior to eukaryogenesis. Although no *in vivo* evidence of CA- and MCA-activity was observed in the model archaeal organism, C14 peptidases were present prior to the divergence of prokaryotes and may have played an important role as a cell surface receptor in the E3 model of eukaryogenesis. This is the first *in silico* study to investigate the putative function of C14 peptidases in archaea and their role for PCD to act as a conflict mediator. The phyletic pattern analysis of archaeal C14 peptidases supports the ‘original sin’ hypothesis of the origin of PCD and how the rise of eukaryotes and the origin of eukaryotic PCD is interconnected.

REFERENCES

Abada, A. and Segev, E. (2018) 'Multicellular Features of Phytoplankton', *Frontiers in Marine Science*, 5, p. 144.

Aizenman, E., Engelberg-Kulka, H. and Glaser, G. (1996) 'An *Escherichia coli* chromosomal "addiction module" regulated by guanosine 3',5'-bispyrophosphate: a model for programmed bacterial cell death', *Proceedings of the National Academy of Sciences*, 93(12), pp. 6059 LP – 6063.

Akanni, W.A. *et al.* (2015) 'Horizontal gene flow from Eubacteria to Archaebacteria and what it means for our understanding of eukaryogenesis', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, p. 20140337.

Allocati, N. *et al.* (2015) 'Die for the community: An overview of programmed cell death in bacteria', *Cell Death and Disease*, 6(1), pp. 1–10.

Altschul, S.F. *et al.* (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215, pp. 403–410.

Altschul, S.F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic acids research*, 25(17), pp. 3389–3402.

Ambit, A. *et al.* (2008) 'An essential role for the Leishmania major metacaspase in cell cycle progression.', *Cell death and differentiation*, 15(1), pp. 113–122.

Ameisen, J.C. (2002) 'On the origin, evolution, and nature of programmed cell death: a timeline of four billion years', *Cell death and differentiation*, 9(4), pp. 367–93.

Andryukov, B.G., Somova, L.M. and Timchenko, N.F. (2018) 'Molecular and Genetic Characteristics of Cell Death in Prokaryotes', 33(2), pp. 73–83.

Aravind, L. *et al.* (1998) 'Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles', *Trends in Genetics*, 14(11), pp. 442–444.

Aravind, L., Dixit, V.M. and Koonin, E. V. (2001) 'Apoptotic molecular machinery: Vastly increased complexity in vertebrates revealed by genome comparisons', *Science*, 291(5507), pp. 1279–1284.

Aravind, L., Dixit, V.M. and Koonin, E. V (1999a) 'The domains of death: evolution of the apoptosis machinery', *Trends in biochemical sciences*, 24(2), pp. 47–53.

Aravind, L., Dixit, V.M. and Koonin, E. V (1999b) 'The domains of death', *Elsevier Science Ltd.*,

0004(February), pp. 47–53.

Aravind, L. and Koonin, E.V. (2002) ‘Classification of the caspase-hemoglobinase fold: Detection of new families and implications for the origin of the eukaryotic separins’, *Proteins: Structure, Function and Genetics*, 46(4), pp. 355–367.

Arnoult, D. *et al.* (2001) ‘On the evolutionary conservation of the cell death pathway: mitochondrial release of an apoptosis-inducing factor during Dictyostelium discoideum cell death.’, *Molecular biology of the cell*, 12(10), pp. 3016–3030.

Asplund-Samuelsson, J., Bergman, B. and Larsson, J. (2012) ‘Prokaryotic Caspase Homologs: Phylogenetic Patterns and Functional Characteristics Reveal Considerable Diversity’, *PLoS ONE*, 7(11), pp. 1–14.

Bayles, K.W. (2014) ‘Bacterial programmed cell death: Making sense of a paradox’, *Nature Reviews Microbiology*, 12(1), pp. 63–69.

Becker, E.A. *et al.* (2014) ‘Phylogenetically Driven Sequencing of Extremely Halophilic Archaea Reveals Strategies for Static and Dynamic Osmo-response’, *PLOS Genetics*, 10(11), p. e1004784.

Benler, S. and Koonin, E. V (2020) ‘Phage lysis-lysogeny switches and programmed cell death: Danse macabre’, *BioEssays*, 42(12), p. 2000114.

Berman-Frank, I. *et al.* (2004) ‘The demise of the marine cyanobacterium, *Trichodesmium spp.*, via an autocatalyzed cell death pathway’, *Limnology and Oceanography*, 49(4), pp. 997–1005.

Bidle, K.A. *et al.* (2010) ‘Tantalizing evidence for caspase-like protein expression and activity in the cellular stress response of Archaea’, *Environmental Microbiology*, 12(5), pp. 1161–1172.

Bidle, K.D. (2016) ‘Programmed Cell Death in Unicellular Phytoplankton’, *Current Biology*, 26(13), pp. R594–R607.

Bidle, K.D. and Bender, S.J. (2008) ‘Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*.’, *Eukaryotic cell*, 7(2), pp. 223–236.

Bidle, K.D. and Falkowski, P.G. (2004) ‘Cell death in planktonic, photosynthetic microorganisms’, *Nature Reviews Microbiology*, 2(8), pp. 643–655.

Blackstone, N.W. (2013) ‘Why did eukaryotes evolve only once? Genetic and energetic aspects of conflict and conflict mediation’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1622), p. 20120266.

Blackstone, N.W. (2016) ‘An Evolutionary Framework for Understanding the Origin of Eukaryotes’, *Biology*, 5(2), p. 18.

Blackstone, N.W. and Green, D.R. (1999) ‘The evolution of a mechanism of cell suicide’, *BioEssays*, 21(1), pp. 84–88.

Blainey, P.C. (2013) ‘The future is now: single-cell genomics of bacteria and archaea’, *FEMS microbiology reviews*, 37(3), pp. 407–427.

Bloss, T.A., Witze, E.S. and Rothman, J.H. (2003) ‘Suppression of CED-3-independent apoptosis by mitochondrial βNAC in Caenorhabditis elegans’, *Nature*, 424(6952), pp. 1066–1071.

Bozhkov, P.V. *et al.* (2005) ‘Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(40), pp. 14463–14468.

Bradford, M.M. (1976) ‘A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding’, *Analytical biochemistry*, 72, pp. 248–254.

Carmona-Gutierrez, D. *et al.* (2011) ‘Ceramide triggers metacaspase-independent mitochondrial cell death in yeast.’, *Cell cycle*, 10(22), pp. 3973–3978.

Castanys-Muñoz, E. *et al.* (2012) ‘*Leishmania mexicana* metacaspase is a negative regulator of amastigote proliferation in mammalian cells’, *Cell Death & Disease*, 3(9), pp. e385–e385.

Castelle, C.J. *et al.* (2015) ‘Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling’, *Current Biology*, 25(6), pp. 690–701.

Cedzich, A. *et al.* (2009) ‘The protease-associated domain and C-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SlSBT3)’, *The Journal of biological chemistry*, 284(21), pp. 14068–14078.

Chai, J. *et al.* (2001) ‘Crystal Structure of a Procaspase-7 Zymogen: Mechanisms of Activation and Substrate Binding’, *Cell*, 107(3), pp. 399–407.

Chela-flores, J. (1998) ‘First steps in eukaryogenesis: physical phenomena in the origin and evolution of chromosome structure’, *Origins ofLife and Evolution ofthe Biosphere*, 28, pp. 215–225.

Chen, J.M. *et al.* (1998) ‘Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases’, *FEBS letters*, 441(3), pp. 361–365.

Cheng, E.H.-Y.. *et al.* (2001) ‘BCL-2, BCL-XL Sequester BH3 Domain-Only Molecules Preventing BAX- and BAK-Mediated Mitochondrial Apoptosis’, *Molecular Cell*, 8(3), pp. 705–711.

Choi, C.J. and Berges, J.A. (2013) ‘New types of metacaspases in phytoplankton reveal diverse origins of cell death proteases’, *Cell Death and Disease*, 4(2), p. e490.

Cline, S.W. and Doolittle, W.F. (1987) ‘Efficient transfection of the archaebacterium *Halobacterium halobium*’, *Journal of bacteriology*, 169(3), pp. 1341–1344.

Coffeen, W.C. and Wolpert, T.J. (2004) ‘Purification and Characterization of Serine Proteases That Exhibit Caspase-Like Activity and Are Associated with Programmed Cell Death in *Avena sativa*’, *The Plant Cell*, 16(4), pp. 857–873.

Cohen, G.M. (1998) ‘Caspases: the executioners of apoptosis’, *Biochem. J.*, 95(326), pp. 1–16.

Coll, N.S. *et al.* (2010) ‘Arabidopsis type I metacaspases control cell death’, *Science (New York, N.Y.)*, 330(6009), pp. 1393–1397.

Collin, R. (1906) ‘Recherches cytologiques sur le développment de la cellule nerveuse’, *Névrae*, 8, pp. 181–308.

Consortium, T.U. (2021) ‘UniProt: the universal protein knowledgebase in 2021’, *Nucleic Acids Research*, 49(D1), pp. D480–D489.

Corliss, J. (1987) ‘Protistan Phylogeny and Eukaryogenesis’, *International Review of Cytology*, 100, pp. 319–370.

van Creveld, S.G. *et al.* (2018) ‘A redox-regulated type III metacaspase controls cell death in a marine diatom’, *bioRxiv*, p. 444109.

Csűrös, M. and Miklós, I. (2009) ‘Streamlining and Large Ancestral Genomes in Archaea Inferred with a Phylogenetic Birth-and-Death Model’, *Molecular Biology and Evolution*, 26(9), pp. 2087–2095.

Dacks, J.B. *et al.* (2016) ‘The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together’, *Journal of Cell Science*, 129(20), pp. 3695–3703.

DeYoung, B.J. and Innes, R.W. (2006) ‘Plant NBS-LRR proteins in pathogen sensing and host defense’, *Nature immunology*, 7(12), pp. 1243–1249.

Drozdetskiy, A. *et al.* (2015) ‘JPred4: a protein secondary structure prediction server’, *Nucleic Acids Research*, 43(W1), pp. W389–W394.

Durand, P.M. (2021) *The Evolutionary Origins of Life and Death*. Chicago: The University of Chicago Press.

Durand, P.M., Barreto Filho, M.M. and Michod, R.E. (2019) ‘Cell Death in Evolutionary Transitions

in Individuality', *Yale Journal of Biology and Medicine*, 92(20), pp. 1–12.

Durand, P.M. and Ramsey, G. (2019) 'The Nature of Programmed Cell Death', *Biological Theory*, 14(1), pp. 30–41.

Durand, P.M., Rashidi, A. and Michod, R.E. (2011) 'How an Organism Dies Affects the Fitness of Its Neighbors', *The American Naturalist*, 177(2), pp. 224–232.

Durand, P.M., Sym, S. and Michod, R.E. (2016) 'Programmed Cell Death and Complexity in Microbial Systems', *Current Biology*, 26(13), pp. R587–R593.

Eddy, S.R. (1998) 'Profile hidden Markov models', *Bioinformatics*, 14(9), pp. 755–763.

Egger, L. *et al.* (2003) 'Serine proteases mediate apoptosis-like cell death and phagocytosis under caspase-inhibiting conditions', *Cell Death & Differentiation*, 10(10), pp. 1188–1203.

Elazari-Volcani, B. (1940) 'Genus XII. Halobacterium', in *R. S. Breed, E. G. D. Murray, and N. R. Smith (eds) Bergey's Manual of Determinative Bacteriology, 7th. Ed.* Baltimore: The Williams & Wilkins Company, pp. 207–212.

Elkins, J.G. *et al.* (2008) 'A korarchaeal genome reveals insights into the evolution of the Archaea', *Proceedings of the National Academy of Sciences of the United States of America*, 105(23), pp. 8102–8107.

Eme, L. *et al.* (2017) 'Archaea and the origin of eukaryotes', *Nature Reviews Microbiology*, 15(12), pp. 711–723.

Engelberg-Kulka, H. *et al.* (2006) 'Bacterial programmed cell death and multicellular behavior in bacteria', *PLoS Genetics*, 2(10), pp. 1518–1526.

Engelberg-Kulka, H., Hazan, R. and Amitai, S. (2005) 'mazEF: A chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria', *Journal of Cell Science*, 118(19), pp. 4327 LP – 4332.

Erental, A., Sharon, I. and Engelberg-Kulka, H. (2012) 'Two programmed cell death systems in *Escherichia coli*: An apoptotic-like death is inhibited by the mazEF-mediated death pathway', *PLoS Biology*, 10(3), p. e1001281.

Ernst, M. (1926) 'Über Untergang von Zellen während der normalen Entwicklung bei Wirbeltieren', *Z Anat Entwicklungsgesch*, 79, pp. 228–262.

Eyers, P.A. and Murphy, J.M. (2016) 'The evolving world of pseudoenzymes: proteins, prejudice and zombies', *BMC Biology*, 14(1), p. 98.

Facciotti, M.T. *et al.* (2010) ‘Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis’, *BMC Systems Biology*, 4(1), p. 64.

Forslund, K. and Sonnhammer, E.L.L. (2008) ‘Predicting protein function from domain content’, *Bioinformatics*, 24(15), pp. 1681–1687.

Frade, J.M. and Michaelidis, T.M. (1997) ‘Origin of eukaryotic programmed cell death: a consequence of aerobic metabolism?’, *BioEssays : news and reviews in molecular, cellular and developmental biology*, 19(9), pp. 827–832.

Fuentes-Prior, P. and Salvesen, G.S. (2004) ‘The protein structures that shape caspase activity, specificity, activation and inhibition’, *Biochemical Journal*, 384(2), pp. 201–232.

Gaasterland, T. and Ragan, M.A. (1998) ‘Microbial genescapes: phyletic and functional patterns of ORF distribution among prokaryotes’, *Microbial & comparative genomics*, 3(4), pp. 199–217.

Galluzzi, L. *et al.* (2007) ‘Cell death modalities: Classification and pathophysiological implications’, *Cell Death and Differentiation*, 14(7), pp. 1237–1243.

Galluzzi, L. and Vitale, I. (2018) ‘Molecular mechanisms of cell death : recommendations of the Nomenclature Committee on Cell Death 2018’, *Cell Death and Differentiation*, 25, pp. 486–541.

Galperin, M.Y. and Koonin, E.V. (2000) ‘Who’s your neighbor? New computational approaches for functional genomics’, *Nature biotechnology*, 18(6), pp. 609–613.

Ginalski, K., Zhang, H. and Grishin, N. V (2004) ‘Raptor protein contains a caspase-like domain.’, *Trends in biochemical sciences*, 29(10), pp. 522–524.

Glücksmann, A. (1951) ‘Cell Deaths in Normal Vertebrate Ontogeny’, *Biological Reviews*, 26(1), pp. 59–86.

González, I.J. *et al.* (2007) ‘*Leishmania major* metacaspase can replace yeast metacaspase in programmed cell death and has arginine-specific cysteine peptidase activity’, *International Journal for Parasitology*, 37(2), pp. 161–172.

Gruber, C. *et al.* (2004) ‘*Halobacterium noricense* sp. nov., an archaeal isolate from a bore core of an alpine Permian salt deposit, classification of *Halobacterium* sp. NRC-1 as a strain of *H. salinarum* and emended description of *H. salinarum*’, *Extremophiles*, 8(6), pp. 431–439.

Guy, L. and Ettema, T.J.G. (2011) ‘The archaeal “TACK” superphylum and the origin of eukaryotes’, *Trends in Microbiology*, 19(12), pp. 580–587.

Hachmann, J. *et al.* (2012) ‘Mechanism and specificity of the human paracaspase MALT1’, *The*

Biochemical journal, 443(1), pp. 287–295.

Hamburger, V. and Levi-Montalcini, R. (1949) 'Proliferation, differentiation and degeneration in the spinal ganglia of the chick embryo under normal and experimental conditions', *The Journal of experimental zoology*, 111(3), pp. 457–501.

Hamilton, W.D. (1970) 'Selfish and Spiteful Behaviour in an Evolutionary Model', *Nature*, 228(5277), pp. 1218–1220.

Hatsugai, N. *et al.* (2009) 'A novel membrane fusion-mediated plant immunity against bacterial pathogens', *Genes & development*, 23(21), pp. 2496–2506.

von Heijne, G. (2006) 'Membrane-protein topology', *Nature Reviews Molecular Cell Biology*, 7(12), pp. 909–918.

Hill, S.M. and Nyström, T. (2015) 'The dual role of a yeast metacaspase: What doesn't kill you makes you stronger.', *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(5), pp. 525–531.

Hoang, D.T. *et al.* (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35(2), pp. 518–522.

Hofmann, K. (2019) 'The Evolutionary Origins of Programmed Cell Death Signaling', *Cold Spring Harbor Perspectives in Biology*, 12(9), p. a036442.

Hug, L.A. *et al.* (2016) 'A new view of the tree of life', *Nature Microbiology*, 1(5), p. 16048.

Hulpiau, P. *et al.* (2016) 'MALT1 is not alone after all: identification of novel paracaspases', *Cellular and molecular life sciences : CMLS*, 73(5), pp. 1103–1116.

Husnik, F. and McCutcheon, J.P. (2018) 'Functional horizontal gene transfer from bacteria to eukaryotes', *Nature Reviews Microbiology*, 16(2), pp. 67–79.

Imachi, H. *et al.* (2020) 'Isolation of an archaeon at the prokaryote–eukaryote interface', *Nature*, 577(7791), pp. 519–525.

Iranzo, J. *et al.* (2014) 'Virus-host arms race at the joint origin of multicellularity and programmed cell death', *Cell cycle (Georgetown, Tex.)*, 13(19), pp. 3083–3088.

Jaworski, M. and Thome, M. (2016) 'The paracaspase MALT1: biological function and potential for therapeutic inhibition', *Cellular and molecular life sciences : CMLS*, 73(3), pp. 459–473.

Jing, H. *et al.* (2002) 'Archaeal Surface Layer Proteins Contain β Propeller, PKD, and β Helix

Domains and Are Related to Metazoan Cell Surface Proteins', *Structure*, 10(10), pp. 1453–1464.

Kaczanowski, S. (2016) 'Apoptosis: its origin, history, maintenance and the medical implications for cancer and aging', *Physical biology*, 13(3), p. 31001.

Kaczanowski, S. (2020) 'Symbiotic Origin of Apoptosis', in *Results and problems in cell differentiation*. Germany, pp. 253–280.

Kallius, E. (1931) 'Der Zelluntergang als Mechanismus bei der Histio-and Morphogenese', *Verh Anat Ges Suppl Anat Anz*, 72, pp. 10–22.

Kalyaanamoorthy, S. *et al.* (2017) 'ModelFinder: fast model selection for accurate phylogenetic estimates', *Nature Methods*, 14(6), pp. 587–589.

Katoh, K. and Standley, D.M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track', *Molecular Biology and Evolution*, 30(4), pp. 772–780.

Kerr, J.F.R. (1971) 'Shrinkage necrosis: A distinct mode of cellular death', *The Journal of Pathology*, 105(1), pp. 13–20.

Kerr, J.F.R., Wyllie, A.H. and Currie, A.R. (1972) 'Apoptosis: A basic biological phenomenon with wide-ranging implications in tissue kinetics', *Journal of Internal Medicine*, 239(26), pp. 239–257.

Klemenčič, M. *et al.* (2019) 'Phylogenetic Distribution and Diversity of Bacterial Pseudo-Orthocaspases Underline Their Putative Role in Photosynthesis', *Frontiers in plant science*, 14(10), p. 293.

Klemenčič, M. and Funk, C. (2018a) 'Structural and functional diversity of caspase homologues in non-metazoan organisms', *Protoplasma*. 2017/07/25, 255(1), pp. 387–397.

Klemenčič, M. and Funk, C. (2018b) 'Type III metacaspases: calcium-dependent activity proposes new function for the p10 domain', *The New phytologist*, 218(3), pp. 1179–1191.

Klemenčič, M., Novinec, M. and Dolinar, M. (2015) 'Orthocaspases are proteolytically active prokaryotic caspase homologues : the case of *Microcystis aeruginosa*', *Molecular Microbiology*, 98(1), pp. 142–150.

Klim, J. *et al.* (2018) 'Ancestral State Reconstruction of the Apoptosis Machinery in the Common Ancestor of Eukaryotes', *G3 (Bethesda)*, 8(6), pp. 2121–2134.

Knoll, A.H. *et al.* (2006) 'Eukaryotic organisms in Proterozoic oceans', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(May), pp. 1023–1038.

Koonin, E. and Aravind, L. (2002) ‘Origin and evolution of eukaryotic apoptosis: the bacterial connection’, *Cell death and differentiation*, 9(4), pp. 394–404.

Koonin, E. V. and Zhang, F. (2017) ‘Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices’, *BioEssays*, 39(1), pp. 1–9.

Koonin, E. V *et al.* (1997) ‘Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea’, *Molecular microbiology*, 25(4), pp. 619–637.

Koonin, E. V (2009) ‘On the origin of cells and viruses: primordial virus world scenario.’, *Annals of the New York Academy of Sciences*, 1178(1), pp. 47–64.

Koonin, E. V (2016) ‘Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions.’, *F1000Research*, 5, p. 1805.

Koonin, E. V and Wolf, Y.I. (2008) ‘Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world’, *Nucleic Acids Research*, 36(21), pp. 6688–6719.

Kroemer, G. (1997) ‘Mitochondrial implication in apoptosis. Towards an endosymbiont hypothesis of apoptosis evolution’, *Cell Death & Differentiation*, 4(6), pp. 443–456.

Kroemer, G. *et al.* (2009) ‘Classification of cell death : recommendations of the Nomenclature Committee on Cell Death 2009’, *Cell Death and Differentiation*, 16(1), pp. 3–11.

Krogh, A. *et al.* (2001) ‘Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes’, *Journal of Molecular Biology*, 305(3), pp. 567–580.

Kunin, V. and Ouzounis, C.A. (2003) ‘The Balance of Driving Forces During Genome Evolution in Prokaryotes’, *Genome Research* , 13(7), pp. 1589–1594.

La, S.R., Ndhlovu, A. and Durand, P.M. (2022) ‘The Ancient Origins of Death Domains Support the “Original Sin” Hypothesis for the Evolution of Programmed Cell Death.’, *Journal of molecular evolution* [Preprint].

Lakowicz, J.R. (2006) ‘Quenching of Fluorescence’, in Lakowicz, J.R. (ed.) *Principles of Fluorescence Spectroscopy*. Boston, MA: Springer US, pp. 277–330.

Lamkanfi, M. *et al.* (2007) ‘Caspases in cell survival, proliferation and differentiation’, *Cell Death & Differentiation*, 14(1), pp. 44–55.

Land, M. *et al.* (2015) ‘Insights from 20 years of bacterial genome sequencing’, *Functional & integrative genomics*. 2015/02/27, 15(2), pp. 141–161.

Lanyi, J.K. (1974) ‘Salt-dependent properties of proteins from extremely halophilic bacteria’, *Bacteriological reviews*, 38(3), pp. 272–290.

Lee, R.E.C. *et al.* (2008) ‘A non-death role of the yeast metacaspase: Yca1p alters cell cycle dynamics’, *PloS one*, 3(8), p. e2956.

Lee, R.E.C. *et al.* (2010) ‘Metacaspase Yca1 is required for clearance of insoluble protein aggregates’, *Proceedings of the National Academy of Sciences*, 107(30), pp. 13348–13353.

Lema A, S. *et al.* (2021) ‘The Role of Pseudo-Orthocaspase (SyOC) of Synechocystis sp. PCC 6803 in Attenuating the Effect of Oxidative Stress’, *Frontiers in Microbiology*, 12, p. 82.

Letunic, I. and Bork, P. (2019) ‘Interactive Tree Of Life (iTOL) v4: recent updates and new developments’, *Nucleic Acids Research*, 47(W1), pp. W256–W259.

Lewis, K. (2000) ‘Programmed death in bacteria’, *Microbiology and molecular biology reviews : MMBR*, 64(3), pp. 503–514.

Lewontin, R.C. (1970) ‘The Units of Selection’, *Annual Review of Ecology and Systematics*, 1, pp. 1–18.

Li, D. and Roberts, R. (2001) ‘WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases.’, *Cellular and molecular life sciences : CMLS*, 58(14), pp. 2085–2097.

Lockshin, A. and Williams, M. (1964) ‘Programmed cell death—II. Endocrine potentiation of the breakdown of the intersegmental muscles of silkmoths’, *Journal of Insect Physiology*, 10(4), pp. 643–649.

López-garcía, P. and Moreira, D. (2015) ‘Open questions on the origin of eukaryotes’, *Trends Ecol Evol.*, 30(11), pp. 697–708.

MacKenzie, S.H. and Clark, A.C. (2012) ‘Death by Caspase Dimerization’, in Matthews, J.M. (ed.) *Protein Dimerization and Oligomerization in Biology*. New York, NY: Springer New York, pp. 55–73.

Maddison, W.P. (1997) ‘Gene Trees in Species Trees’, *Systematic Biology*, 46(3), pp. 523–536.

Magrum, L.J., Luehrsen, K.R. and Woese, C.R. (1978) ‘Are extreme halophiles actually “bacteria”?’, *Journal of molecular evolution*, 11(1), pp. 1–8.

Margulis, L. (1970) *Origin of eukaryotic cells: evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth*. New Haven: Yale University Press.

Maynard Smith, J. and Szathmáry, E. (1995) *The major Transitions in Evolution*. Oxford: Oxford University Press.

McLuskey, K. *et al.* (2012) 'Crystal structure of a Trypanosoma brucei metacaspase', *Proceedings of the National Academy of Sciences of the United States of America*, 109(19), pp. 7469–7474.

McLuskey, K. and Mottram, J.C. (2015) 'Comparative structural analysis of the caspase family with other clan CD cysteine peptidases', *The Biochemical journal*, 466(2), pp. 219–232.

Meslin, B. *et al.* (2011) '*Plasmodium falciparum* Metacaspase PfMCA-1 Triggers a z-VAD-fmk Inhibitable Protease to Promote Cell Death', *PLOS ONE*, 6(8), p. e23867.

Michod, R. (1997) 'Evolution of the Individual', *The American Naturalist*, 150(S1), pp. S5–S21.

Michod, R.E. (1999) 'Darwinian dynamics : evolutionary transitions in fitness and individuality'. Princeton, N.J.: Princeton University Press, p. 280.

Michod, R.E. (2003) *Cooperation and Conflict in the Evolution of Complexiy*. Tucson.

Michod, R.E. and Nedelcu, A.M. (2003) 'Cooperation and conflict during the unicellular – multicellular and prokaryotic – eukaryotic transitions', in Moya, A. and Font, E. (eds) *Evolution: From molecules to ecosystems*. 1st edn. Oxford: Oxford University Press, pp. 195–208.

Mignotte, B. and Vayssiere, J.-L. (1998) 'Mitochondria and apoptosis', *European Journal of Biochemistry*, 252(1), pp. 1–15.

Minh, B.Q. *et al.* (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37(5), pp. 1530–1534.

Minina, E.A. *et al.* (2017) 'Metacaspases versus caspases in development and cell fate regulation', *Cell Death and Differentiation*, 24(8), pp. 1314–1325.

Minina, E.A. *et al.* (2020) 'Classification and Nomenclature of Metacaspases and Paracaspases: No More Confusion with Caspases', *Molecular Cell*, 77(5), pp. 927–929.

Mistry, J. *et al.* (2021) 'Pfam: The protein families database in 2021', *Nucleic Acids Research*, 49(D1), pp. D412–D419.

Moreira, D. and Lopez-Garcia, P. (1998) 'Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis.', *Journal of molecular evolution*, 47(5), pp. 517–530.

Morris, J., Lenski, R. and Zinser, E. (2012) 'The Black Queen Hypothesis: evolution of dependencies

through adaptive gene loss', *American Society for Microbiology*, 3(2), pp. e00036-12.

Morris, J.J., Papoulis, S.E. and Lenski, R.E. (2014) 'Coexistence of evolving bacteria stabilized by a shared black queen function', *Evolution*, 68(10), pp. 2960–2971.

Moss, C.X. *et al.* (2007) 'Metacaspase 2 of *Trypanosoma brucei* is a calcium-dependent cysteine peptidase active without processing', *FEBS Letters*, 581(29), pp. 5635–5639.

Ndhlovu, A., Durand, P. and Ramsey, G. (2021) 'Programmed cell death as a black queen in microbial communities', *Molecular Ecology*, 30, pp. 1110– 1119.

Nedelcu, A.M. (2009) 'Comparative genomics of phylogenetically diverse unicellular eukaryotes provide new insights into the genetic basis for the evolution of the programmed cell death machinery', *Journal of molecular evolution*, 68(3), pp. 256–268.

Nedelcu, A.M. *et al.* (2011) 'On the paradigm of altruistic suicide in the unicellular world', *Evolution; international journal of organic evolution*, 65(1), pp. 3–20.

Nelson-sathi, S. *et al.* (2015) 'Origins of major archaeal clades correspond to gene acquisitions from bacteria', *Nature*, 517(7532), pp. 77–80.

Nelson, K.E. *et al.* (1999) 'Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima.', *Nature*, 399(6734), pp. 323–329.

Ng, W.V. *et al.* (2000) 'Genome sequence of *Halobacterium* species NRC-1', *Proceedings of the National Academy of Sciences*, 97(22), pp. 12176 LP – 12181.

Nguyen, L. *et al.* (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32(1), pp. 268–274.

Nicholson, D.W. *et al.* (1995) 'Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis', *Nature*, 376(6535), pp. 37–43.

Nunoura, T. *et al.* (2011) 'Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group', *Nucleic Acids Research*, 39(8), pp. 3204–3223.

Nuttall, S. *et al.* (2008) *The Halohandbook, Online*.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) 'Lateral gene transfer and the nature of bacterial innovation', *Nature*, 405(6784), pp. 299–304.

Okasha, S. (2005) 'Multilevel Selection and the Major Transitions in Evolution', *Philosophy of*

Science, 72(5), pp. 1013–1025.

Orellana, M.V. *et al.* (2013) 'A Role for Programmed Cell Death in the Microbial Loop', *PLoS ONE*, 8(5), p. e62595.

Peck, R.F., DasSarma, S. and Krebs, M.P. (2000) 'Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker', *Molecular Microbiology*, 35(3), pp. 667–676.

Peeters, S.H. and de Jonge, M.I. (2018) 'For the greater good: Programmed cell death in bacterial communities', *Microbiological research*, 207, pp. 161–169.

Pfeiffer, F. *et al.* (2020) 'Whole-genome comparison between the type strain of Halobacterium salinarum (DSM 3754T) and the laboratory strains R1 and NRC-1', *MicrobiologyOpen*, 9(2), p. e974.

Ponting, C.P. *et al.* (1999) 'Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer', *Journal of Molecular Biology*, 289(4), pp. 729–745.

Proto, W.R. *et al.* (2011) '*Trypanosoma brucei* Metacaspase 4 Is a Pseudopeptidase and a Virulence Factor', *Journal of Biological Chemistry*, 286(46), pp. 39914–39925.

Proto, W.R., Coombs, G.H. and Mottram, J.C. (2013) 'Cell death in parasitic protozoa: regulated or incidental?', *Nature reviews. Microbiology*, 11(1), pp. 58–66.

Radzvilavicius, A.L. and Blackstone, N.W. (2015) 'Conflict and cooperation in eukaryogenesis: Implications for the timing of endosymbiosis and the evolution of sex', *Journal of the Royal Society Interface*, 12(111), p. 20150584.

Ramisetty, B.C.M., Natarajan, B. and Santhosh, R.S. (2015) 'MazEF-mediated programmed cell death in bacteria: "What is this?"', *Critical Reviews in Microbiology*, 41(1), pp. 89–100.

Ratel, D. *et al.* (2001) 'Programmed Cell Death or Cell Death Programme? That is the Question', *Journal of theoretical biology*, 208, pp. 385–386.

Rawlings, N. and Bateman, A. (2019) 'Origins of peptidases', *Biochimie*, 166, pp. 4–18.

Rawlings, N.D. *et al.* (2014) 'MEROPS: the database of proteolytic enzymes, their substrates and inhibitors', *Nucleic Acids Research*, 42(D1), pp. D503–D509.

Rawlings, N.D. and Barrett, A.J. (1993) 'Evolutionary families of peptidases', *The Biochemical journal*, 290(1), pp. 205–218.

Reece, S.E. *et al.* (2011) 'The Meaning of Death : Evolution and Ecology of Apoptosis in Protozoan

Parasites', *PLoS Pathogens*, 7(12), pp. 1–9.

Richie, D.L. *et al.* (2007) 'The Aspergillus fumigatus metacaspases CasA and CasB facilitate growth under conditions of endoplasmic reticulum stress', *Molecular microbiology*, 63(2), pp. 591–604.

Rinke, C. *et al.* (2013) 'Insights into the phylogeny and coding potential of microbial dark matter', *Nature*, 499(7459), pp. 431–437.

Robinson, C.K. *et al.* (2011) 'A major role for nonenzymatic antioxidant processes in the radioresistance of *Halobacterium salinarum*', *Journal of bacteriology*. 2011/01/28, 193(7), pp. 1653–1662.

Sagan, L. (1967) 'On the Origin of Mitosing Cells', *Journal of Theoretical Biology*, 14, pp. 225–274.

Saheb, E. *et al.* (2013) 'A functional connection of *Dictyostelium* paracaspase with the contractile vacuole and a possible partner of the vacuolar proton ATPase', *Journal of biosciences*, 38(3), pp. 509–521.

Saucet, S.B., Esmenjaud, D. and Van Ghelder, C. (2020) 'Integrity of the Post-LRR Domain Is Required for TIR-NB-LRR Function', *Molecular Plant-Microbe Interactions®*, 34(3), pp. 286–296.

Saunders, J.J. (1966) 'Death in Embryonic Systems', *Science*, 154(3749), pp. 604–612.

Segovia, M. *et al.* (2003) 'Cell death in the unicellular chlorophyte Dunaliella tertiolecta. A hypothesis on the evolution of apoptosis in higher plants and metazoans', *Plant physiology*, 132(1), pp. 99–105.

Seth-Pasricha, M. *et al.* (2019) 'Catalytic linkage between caspase activity and proteostasis in Archaea', *Environmental Microbiology*, 21(1), pp. 286–298.

Seth-Pasricha, M., Bidle, K.A. and Bidle, K.D. (2013) 'Specificity of archaeal caspase activity in the extreme halophile *Haloferax volcanii*', *Environmental Microbiology Reports*, 5(2), pp. 263–271.

Shahmohammadi, H.R. *et al.* (1998) 'Protective Roles of Bacterioruberin and Intracellular KCl in the Resistance of *Halobacterium salinarium* against DNA-damaging Agents', *Journal of Radiation Research*, 39(4), pp. 251–262.

Shalini, S. *et al.* (2015) 'Old, new and emerging functions of caspases', *Cell death and differentiation*, 22(4), pp. 526–539.

Shand, R.F. and Betlach, M.C. (1991) 'Expression of the bop gene cluster of *Halobacterium halobium* is induced by low oxygen tension and by light', *Journal of bacteriology*, 173(15), pp. 4692–4699.

Sharma, K. *et al.* (2012) 'The RosR transcription factor is required for gene expression dynamics in

response to extreme oxidative stress in a hypersaline-adapted archaeon', *BMC genomics*, 13(1471–264), p. 351.

Shen, W. and Ren, H. (2021) 'TaxonKit: A practical and efficient NCBI taxonomy toolkit', *Journal of Genetics and Genomics* [Preprint].

Shen, Y. *et al.* (2009) 'Evaluating the role of microbial sulfate reduction in the early Archean using quadruple isotope systematics', *Earth and Planetary Science Letters*, 279(3), pp. 383–391.

Shen, Y., Buick, R. and Canfield, D.E. (2001) 'Isotopic evidence for microbial sulphate reduction in the early Archaean era', *Nature*, 410(6824), pp. 77–81.

Shrestha, A. and Megeney, L. (2012) 'The non-death role of metacaspase proteases', *Frontiers in Oncology*, 2, p. 78.

Snel, B., Bork, P. and Huynen, M.A. (2002) 'Genomes in Flux: The Evolution of Archaeal and Proteobacterial Gene Content', *Genome Research* , 12(1), pp. 17–25.

Sonnhammer, E.L., von Heijne, G. and Krogh, A. (1998) 'A hidden Markov model for predicting transmembrane helices in protein sequences.', *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 6, pp. 175–82.

Spang, A. *et al.* (2015) 'Complex archaea that bridge the gap between prokaryotes and eukaryotes', *Nature*, 521(7551), pp. 173–179.

Spang, A. *et al.* (2018) 'Asgard archaea are the closest prokaryotic relatives of eukaryotes', *PLOS Genetics*, 14(3), p. e1007080.

Speijer, D. (2020) 'Debating Eukaryogenesis—Part 1: Does Eukaryogenesis Presuppose Symbiosis Before Uptake?', *BioEssays*, 42(4), p. 1900157.

Spungin, D., Bidle, K.D. and Berman-Frank, I. (2019) 'Metacaspase involvement in programmed cell death of the marine cyanobacterium Trichodesmium', *Environmental Microbiology*, 21(2), pp. 667–681.

Srivastav, R. and Suneja, G. (2019) 'Recent Advances in Microbial Genome Sequencing', in Tripathi, V. et al. (eds) *Microbial Genomics in Sustainable Agroecosystems*. 1st edn. Singapore: Springer Singapore, pp. 131–144.

Szallies, A., Kubata, B.K. and Duszenko, M. (2002) 'A metacaspase of Trypanosoma brucei causes loss of respiration competence and clonal death in the yeast Saccharomyces cerevisiae', *FEBS letters*, 517(1–3), pp. 144–150.

Tait, S.W.G. and Green, D.R. (2008) ‘Caspase-independent cell death: leaving the set without the final cut’, *Oncogene*, 27(50), pp. 6452–6461.

Tsiatsiani, L. *et al.* (2011) ‘Metacaspases’, *Cell Death and Differentiation*, 18(8), pp. 1279–1288.

Turner, S. *et al.* (1999) ‘Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis’, *The Journal of eukaryotic microbiology*, 46(4), pp. 327–338.

Ueno, Y. *et al.* (2008) ‘Quadruple sulfur isotope analysis of ca. 3.5Ga Dresser Formation: New evidence for microbial sulfate reduction in the early Archean’, *Geochimica et Cosmochimica Acta*, 72(23), pp. 5675–5691.

Uhlmann, F. *et al.* (2000) ‘Cleavage of Cohesin by the CD Clan Protease Separin Triggers Anaphase in Yeast’, *Cell*, 103(3), pp. 375–386.

Uren, A.G. *et al.* (2000) ‘Identification of Paracaspases and Metacaspases’, *Molecular Cell*, 6(4), pp. 961–967.

Valentine, D.L. (2007) ‘Adaptations to energy stress dictate the ecology and evolution of the Archaea’, *Nature Reviews Microbiology*, 5(4), pp. 316–323.

Vardi, A. *et al.* (2007) ‘Synchronization of cell death in a dinoflagellate population is mediated by an excreted thiol protease’, *Environmental Microbiology*, 9(2), pp. 360–369.

Vaux, D.L., Cory, S. and Adams, J.M. (1988) ‘Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells’, *Nature*, 335(6189), pp. 440–442.

Vercammen, D. *et al.* (2004) ‘Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine’, *The Journal of biological chemistry*, 279(44), pp. 45329–45336.

Vercammen, D. *et al.* (2007) ‘Are metacaspases caspases?’, *Journal of Cell Biology*, 179(3), pp. 375–380.

Wagner, A. *et al.* (2017) ‘Mechanisms of gene flow in archaea’, *Nature Reviews Microbiology*, 15(8), pp. 492–501.

Walker, N.P. *et al.* (1994) ‘Crystal structure of the cysteine protease interleukin-1 beta-converting enzyme: a (p20/p10)2 homodimer’, *Cell*, 78(2), pp. 343–352.

Wang, H. *et al.* (2018) ‘CpMCA, a novel metacaspase gene from the harmful dinoflagellate Cochlodinium polykrikoides and its expression during cell death’, *Gene*, 651, pp. 70–78.

Watanabe, N. and Lam, E. (2005) ‘Two Arabidopsis Metacaspases AtMCP1b and AtMCP2b Are Arginine/Lysine-specific Cysteine Proteases and Activate Apoptosis-like Cell Death in Yeast’, *Journal of Biological Chemistry*, 280(15), pp. 14691–14699.

Watanabe, N. and Lam, E. (2011) ‘Calcium-dependent Activation and Autolysis of Arabidopsis Metacaspase’, *Journal of Biological Chemistry*, 286(12), pp. 10027–10040.

Waterhouse, A.M. *et al.* (2009) ‘Jalview Version 2-A multiple sequence alignment editor and analysis workbench’, *Bioinformatics*, 25(9), pp. 1189–1191.

Wei, Y. *et al.* (2000) ‘The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity’, *Chemistry & Biology*, 7(6), pp. 423–432.

Weisburg, W.G. *et al.* (1991) ‘16S ribosomal DNA amplification for phylogenetic study’, *Journal of bacteriology*, 173(2), pp. 697–703.

West, S., Griffin, A. and Gardner, A. (2007) ‘Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection’, *Journal of evolutionary biology*, 20(2), pp. 415–432.

West, S.A. *et al.* (2015) ‘Major evolutionary transitions in individuality’, *Proceedings of the National Academy of Sciences*, 112(33), pp. 10112–10119.

Wheeler, T.J., Clements, J. and Finn, R.D. (2014) ‘Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models’, *BMC Bioinformatics*, 15(1), p. 7.

Wiesmann, C. *et al.* (2012) ‘Structural Determinants of MALT1 Protease Activity’, *Journal of Molecular Biology*, 419(1), pp. 4–21.

Wilson, K.P. *et al.* (1994) ‘Structure and mechanism of interleukin-1 beta converting enzyme’, *Nature*, 370(6487), pp. 270–275.

Yang, J. and Zhang, Y. (2015) ‘Protein Structure and Function Prediction Using I-TASSER’, *Current protocols in bioinformatics*, 52, pp. 5.8.1-5.8.15.

Yu, J.W. *et al.* (2011) ‘Crystal structure of the mucosa-associated lymphoid tissue lymphoma translocation 1 (MALT1) paracaspase region’, *Proceedings of the National Academy of Sciences*, 108(52), pp. 21004–21009.

Zaremba-Niedzwiedzka, K. *et al.* (2017) ‘Asgard archaea illuminate the origin of eukaryotic cellular complexity’, *Nature*, 541(7637), pp. 353–358.

Zhang, Z. *et al.* (2020) ‘Estimate of the sequenced proportion of the global prokaryotic genome’,

Microbiome, 8(1), p. 134.

APPENDIX

List of Appendices

Figure A1 *H. salinarum* culture at stationary phase after 36 hours of growth with OD_{600} = 2.0.

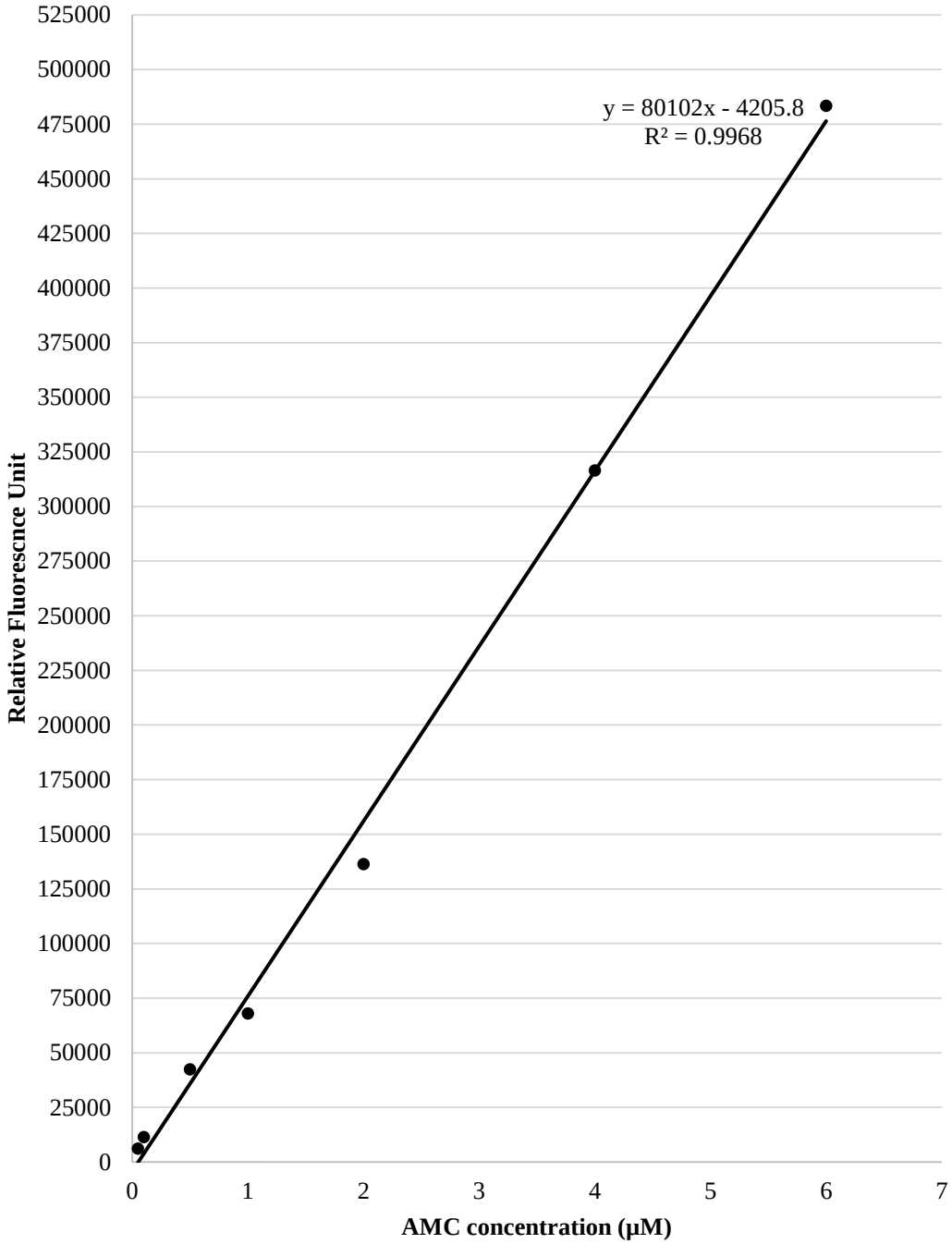

Figure A2 AMC standard curve for CA and MCA fluorometric enzyme assays.

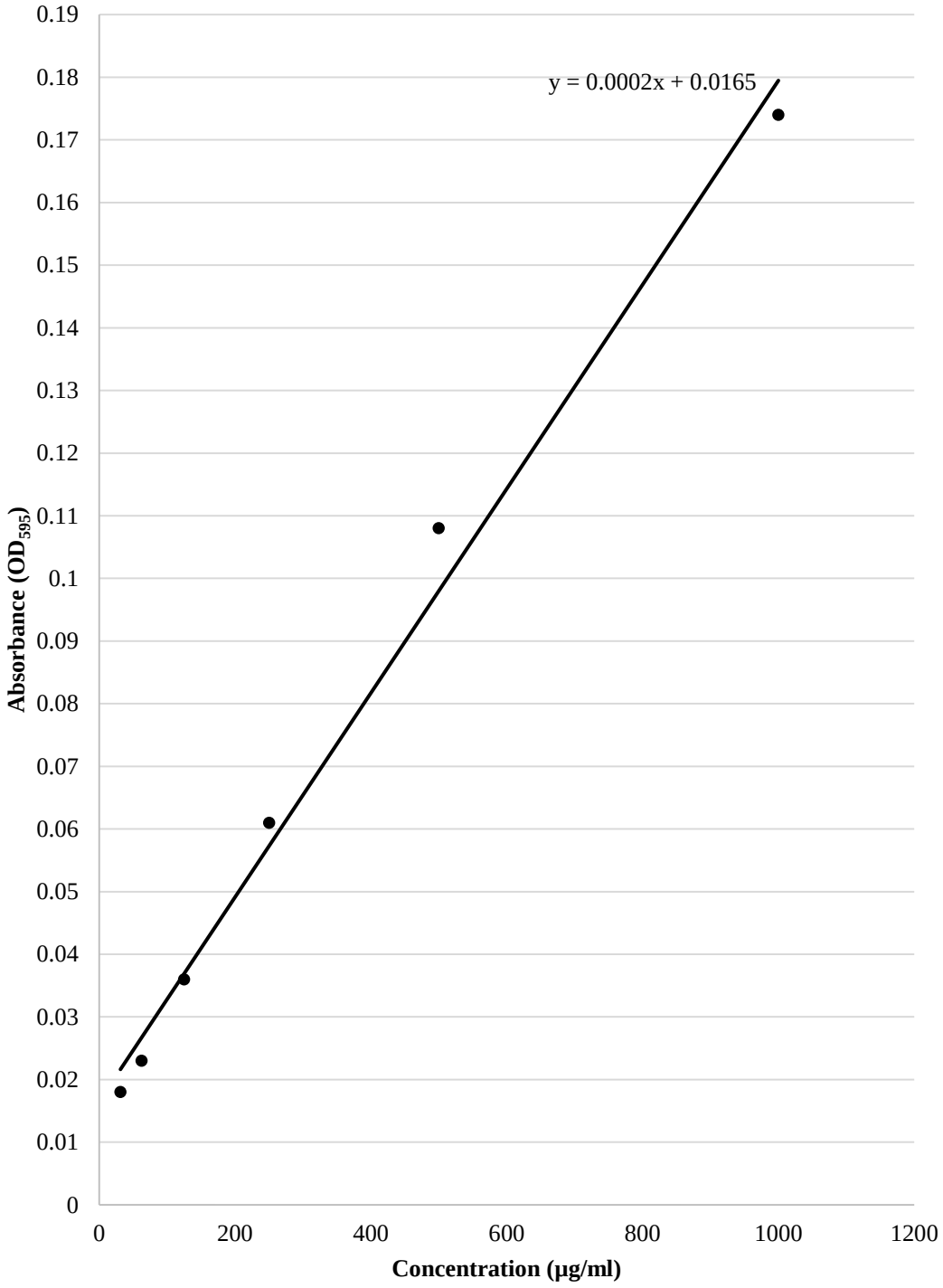

Figure A3 BSA standard curve.

Table A1 Full BLASTn hits for the taxonomic identification of the *H. salinarum* sample using the Archaea_BAK-R_E07_13 primer.

Description	**Scientific Name**	**E-value**	**Per. ident**	**Acc. Len**
Halobacterium salinarum NBRC 14715 gene for 16S rRNA, partial sequence	*Halobacterium salinarum*	0E00	99.66	1439
TPA: *Halobacterium salinarum* NRC-1, complete genome	*Halobacterium salinarum* NRC-1	0E00	99.66	2014239
Halobacterium salinarum strain 91-R6 chromosome, complete genome	*Halobacterium salinarum*	0E00	99.66	2178608
Halobacterium salinarum strain NES7 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1360
Halobacterium salinarum strain NES5 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1228
Halobacterium salinarum strain NES2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1232
Halobacterium salinarum strain NES1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1391
Halobacterium salinarum strain 67A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1342
Halobacterium salinarum strain 66A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1323
Halobacterium salinarum strain 64A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1330

Halobacterium salinarum strain 58A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1348
Halobacterium salinarum strain 56A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1326
Halobacterium salinarum strain 54A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1347
Halobacterium salinarum strain 49A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1474
Halobacterium salinarum strain 39A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1346
Halobacterium salinarum strain 35A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1347
Halobacterium salinarum strain 32A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1353
Halobacterium salinarum strain 29A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1474
Halobacterium salinarum strain 28A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1331
Halobacterium salinarum strain 26A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1339
Halobacterium salinarum strain 17A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1347
Halobacterium salinarum strain 12A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1348
Halobacterium salinarum strain 8A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1357
Halobacterium salinarum strain 7A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1347
Halobacterium salinarum strain 6A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1351
Halobacterium salinarum strain 4A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1474

Halobacterium salinarum strain 3A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1353
Halobacterium salinarum strain 403 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1360
Halobacterium sp. strain S7FS1 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.*	0E00	99.66	1473
Halobacterium salinarum strain ETD19 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1360
Halobacterium sp. AMS4 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* AMS4	0E00	99.66	1473
Halobacterium salinarum partial 16S rRNA gene, isolate AT2RS16	*Halobacterium salinarum*	0E00	99.66	1363
Halobacterium salinarum strain H11 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1420
Halobacterium salinarum strain H8 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1417
Halobacterium salinarum strain H5 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1425
Halobacterium salinarum strain H4 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1415
Halobacterium salinarum strain H2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1393
Halobacterium salinarum strain N1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1365
Halobacterium salinarum strain BL1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	977
Halobacterium salinarum strain LXP1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1385

Halobacterium sp. NRC-34001 chromosome, complete genome	*Halobacterium sp.* NRC-34001	0E00	99.66	2012898
Halobacterium sp. BOL4-2 chromosome, complete genome	*Halobacterium sp.* BOL4-2	0E00	99.66	1989773
Halobacterium salinarum strain 28 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1393
Halobacterium salinarum strain OL1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	951
Halobacterium sp. GSL-19 chromosome, complete genome	*Halobacterium sp.* GSL-19	0E00	99.66	1987132
Halobacterium piscisalsi strain C37 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1463
Halobacterium salinarum strain ARJ-RS5 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1363
Halobacterium salinarum strain 6T2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1343
Halobacterium salinarum strain 3S1 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1335
Halobacterium salinarum strain 2m-40-15-R2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1440
Halobacterium salinarum strain AR2-2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 2 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 9 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 10 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472

Halobacterium salinarum strain 20 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 22 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 16 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 18 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 21 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 17 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 14 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 19 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 11 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 8 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 7 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 13 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 12 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 15 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1473
Halobacterium salinarum strain 5 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 6 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472

Halobacterium salinarum strain 4 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Halobacterium salinarum strain 3 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1472
Archaeon RC38 16S ribosomal RNA gene, partial sequence	archaeon RC38	0E00	99.66	1347
Halobacterium sp. NRC-1 gene for 16S rRNA, partial sequence, strain: JCM 11081	*Halobacterium salinarum* NRC-1	0E00	99.66	1449
Halobacterium sp. P102070208-3R 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* P102070208-3R	0E00	99.66	1334
Halobacterium sp. P102070208-3O 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* P102070208-3O	0E00	99.66	1346
Halobacterium salinarum strain 91-R6 16S ribosomal RNA, partial sequence	*Halobacterium salinarum*	0E00	99.66	1434
Halobacterium salinarum R1 complete genome	*Halobacterium salinarum* R1	0E00	99.66	2000962
Halobacterium sp. NRC-1, complete genome	*Halobacterium salinarum* NRC-1	0E00	99.66	2014239
Halobacterium salinarum gene for 16S ribosomal RNA, complete sequence	*Halobacterium salinarum*	0E00	99.66	1473
Halobacterium salinarum 16S rRNA gene, type strain DSM 3754T	*Halobacterium salinarum* DSM 3754	0E00	99.66	1473
Halobacterium salinarum strain Cer6a 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.66	1376
Halobacterium sp. HM01 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* HM01	0E00	99.66	1376
Halobacterium salinarum strain JCM 14661 16S ribosomal RNA, partial	*Halobacterium salinarum*	0E00	99.66	1473

sequence				
Halobacterium sp. HSC3 partial 16S rRNA gene, strain HSC3	*Halobacterium sp.* HSC3	0E00	99.77	1353
Halobacterium salinarum strain 75A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1344
Halobacterium salinarum strain 72A 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1344
Halobacterium salinarum 16S ribosomal RNA and tRNA-Ala genes, complete sequence; 23S ribosomal RNA genes, partial sequence; and 5S ribosomal RNA and tRNA-Cys genes, complete sequence	*Halobacterium salinarum*	0E00	99.55	3823
Halobacterium salinarum strain H6 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1421
Halobacterium salinarum strain ETD8 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1442
Halobacterium salinarum strain P-1-S8 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1440
Halobacterium salinarum strain J-1-S4 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.55	1440
Halobacterium sp. FIC145_2 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC145_2	0E00	99.55	1421
Halobacterium sp. FIC145_1 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC145_1	0E00	99.55	1413
Halobacterium sp. FIC146_4 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC146_4	0E00	99.55	1416

Halobacterium sp. FIC146_3 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC146_3	0E00	99.55	1418
Halobacterium sp. FIC146_2 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC146_2	0E00	99.55	1421
Halobacterium sp. FIC146_1 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* FIC146_1	0E00	99.55	1426
Halobacterium sp. HM13 16S ribosomal RNA gene, partial sequence	*Halobacterium sp.* HM13	0E00	99.55	1373
Halobacterium salinarum strain HAL1RS5 16S ribosomal RNA gene, partial sequence	*Halobacterium salinarum*	0E00	99.77	1277

Table A2 Measured CA3-, CA8- and MCA-like enzyme activity (µM.AMC µg protein^{-1} min^{-1}) in *H. salinarum* of each treatment group after one hour of exposure to H_2O_2 and the results of One-way ANOVA analysis. The mean value of each treatment group with standard deviation is indicated.

C14 peptidase	Concentration of H2O2	No inhibitor (µmol AMC µg protein^{-1} min^{-1}) (Mean ± SD)	Inhibitor (µmol AMC µg protein^{-1} min^{-1}) (Mean ± SD)	P-value
CA3	0 mM	0.013 (±0.0046)	0.016 (±0.0096)	0.57
CA3	10 mM	0.0089 (±0.0076)	0.0026 (±0.0047)	0.29
CA3	25 mM	0.029 (±0.020)	0.022 (±0.0072)	0.61
CA8	0 mM	0.027 (±0.016)	0.022 (±0.013)	0.71
CA8	10 mM	0.039 (±0.0049)	0.043 (±0.0054)	0.43
CA8	25 mM	0.010 (±0.0031)	0.021 (±0.014)	0.28
MCA	0 mM	0.036 (±0.015)	0.017 (±0.0083)	0.13
MCA	10 mM	0.019 (±0.010)	0.011 (±0.025)	0.13
MCA	25 mM	0.033 (±0.011)	0.022 (±0.0046)	0.19

Table A3 Measured CA3-, CA8- and MCA-like enzyme activity (µM.AMC µg protein^{-1} min^{-1}) in *H. salinarum* of each treatment group after three hours of exposure to H_2O_2 and the results of One-way ANOVA analysis. The mean value of each treatment group with standard deviation is indicated.

C14 peptidase	Concentration of H2O2	No inhibitor (µmol AMC µg protein^{-1} min^{-1}) (Mean ± SD)	Inhibitor (µmol AMC µg protein^{-1} min^{-1}) (Mean ± SD)	P-value
CA3	0 mM	0.041 (±0.015)	0.033 (±0.017)	0.0051
CA3	10 mM	0.0074 (±0.037)	0.038 (±0.022)	0.29
CA3	25 mM	-0.00890 (±0.0092)	0.031 (±0.055)	0.53
CA8	0 mM	0.012 (±0.028)	0.011 (±0.012)	0.97
CA8	10 mM	0.043 (±0.0045)	0.0069 (±0.017)	0.023
CA8	25 mM	-0.0061 (±0.037)	0.0083 (±0.040)	0.67
MCA	0 mM	0.00078 (±0.025)	0.046 (±0.015)	0.057
MCA	10 mM	0.0065 (±0.018)	0.027 (±0.014)	0.19
MCA	25 mM	0.048 (±0.016)	0.041 (±0.0037)	0.55