

A functional metagenomics approach to identify novel ultraviolet resistance genes

Marta Cortesão

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Mestrado em Biologia Celular e Molecular
2015

MSc

2.º
CICLO

FCUP
2015

U. PORTO

A functional metagenomics approach to identify
novel ultraviolet resistance genes

Marta Cortesão

FC





A functional metagenomics approach to identify novel ultraviolet resistance genes

Marta Cortesão

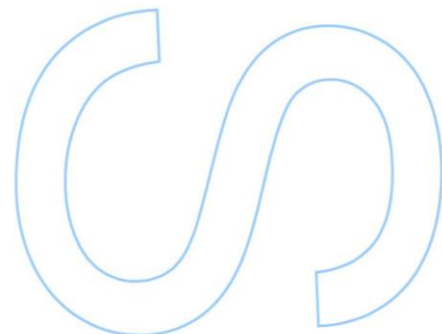
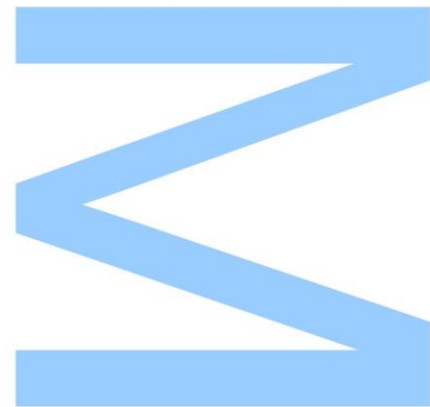
Mestrado em Biologia Celular e Molecular
Departamento de Biologia
2015

Orientador

José Eduardo González Pastor, Lab Chief, Department of Molecular Ecology, Centro de Astrobiología (CAB), CSIC-INTA, Madrid

Coorientador

Olga Maria Oliveira da Silva Lage, Professora Auxiliar, Faculdade de Ciências da Universidade do Porto (FCUP)

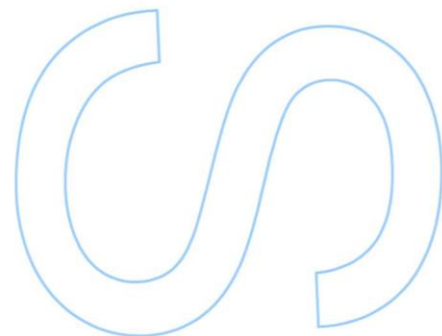
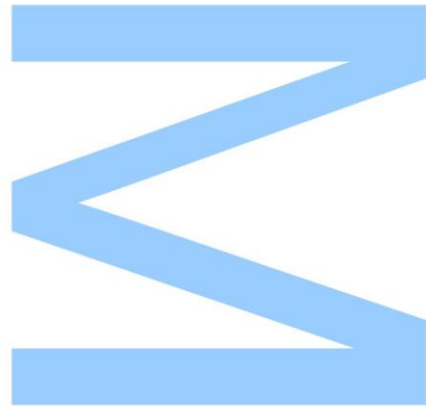




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____ / ____ / ____



Acknowledgements

Someone once told “If you want to go fast, you go alone. If you want to go far, you go together”. This was what I learned, above all, during this past year – teamwork is what counts to achieve more and better.

It wasn't just in terms of lab work, but also regarding the European students' association I am proudly part of - BEST. This amazing NGO has taught me very much, not only to develop myself, but also to learn how to help others to develop themselves. In the past four years I have managed to gain strengths to come by many challenges, and believe me, it wouldn't be possible to do it all alone.

Por isso, obrigada a todos os membros do BEST Porto por me fazerem crescer de baby, a full e a possivelmente não-old bastard.

Thanks to BEST in general for raising my awareness for so many issues, for so many cultures, social activities, and for making me want to be an actual part of the world.

Gracias a todos de BEST Madrid, al CAB y todas las personas simpáticas por la calle, tiendas, restaurantes, porque al final de todo los Españoles han sido justo mis hermanos. Un muchas gracias a todos los del Invernadero: Eduardo, Maria, Caro, Ale, Salvador, Sara, Javi, Kiko, Stefano, Maca por mucha ayuda, mucho trabajo y muchas risas. Gracias a los aparatos que no han fallado cuando los necesitaba, dias de sol, lluvia, nieve, viento fuerte y calor (pero todos muy felices) que he pasado por Madrid.

Obrigada Professora Olga por me ter apoiado nesta aventura, e ao LEMUP em geral por me acolher tão bem. Obrigada Mafalda por seres uma colega, amiga, angel e confidente e pêras durante já alguns anitos. E obrigada coleguinhas de Biologia por fazerem estes 5 anos serem ainda mais especiais.

Um obrigada aos meus avós, Maria, tios, primos, pai, mãe e mirma por me aturarem e me fazerem seguir em frente, seja qual for o caminho que eu escolha. E por fim, obrigada Tiago por nunca desistires de mim e me ajudares a seguir os meus sonhos.

Resumo

O conhecimento de estratégias moleculares e de mecanismos responsáveis pela resistência de microorganismos a ambientes extremos é fundamental para descobrir os limites da vida. As técnicas independentes de cultivo fornecem informação sobre os microorganismos não cultiváveis, desempenhando um papel fulcral no estudo da biodiversidade terrestre. Sabe-se que a Terra primordial tolerou condições extremas de exposição aos raios ultravioletas (UVs), e que estes incidem actualmente sobre diversos ambientes tanto em superfícies de naves espaciais ou corpos planetários como em ambientes de elevada altitude na Terra. Apesar de os raios UVB e UVC serem prejudiciais à vida, existem microorganismos capazes de sobreviver à sua irradiação, surgindo o interesse pela evolução e adaptação da vida aos UVs. Deste modo, o presente estudo consiste na aplicação de metagenómica funcional na identificação de novos genes responsáveis pela resistência aos Uvs.

Neste trabalho, foram construídas e analisadas três bibliotecas metagenómicas a partir de comunidades microbianas expostas aos UVs em ambientes hipersalinos (lagos nos Andes, Argentina e uma salina, Mallorca, Espanha), usando *E. coli DH10B* como hospedeiro. Cada biblioteca foi rastreada para resistência aos UVB e UVC, permitindo a identificação de clones recombinantes com fragmentos de DNA ambiental que conferem resistência aos UVs. No total foram identificados cinco clones resistentes: pML5, pML6, pML56, pML84 (lagos), and pML105 (salina), com uma taxa de sobrevivência aos UVB cerca de 15% superior ao controlo *E. coli DH10B*.

Para cada clone, os fragmentos de DNA ambiental foram sequenciados e as respectivas Open Reading Frames (ORFs) identificadas. O clone pML84-orf1 codifica para um domínio C-terminal de uma proteína, o clone pML56-orf1 codifica para uma ribonuclease III enquanto o pML56-orf2 codifica para um factor de transcrição. Por sua vez, o clone pML5-orf1 codifica para a proteína RecA - uma recombinase previamente identificada e que se mostra envolvida na resistência aos UV através da *SOS response* na reparação do DNA. Tanto o clone pML6-orf1 como o pML105-orf1, originários de locais geograficamente distantes, codificam para proteínas hipotéticas que partilham 32% de identidade.

De modo a entender o mecanismo de resistência, cada clone foi tratado com 4-nitroquinolina 1-óxido (um composto que imita o efeito da radiação UV no DNA),

mostrando taxas de sobrevivência cerca de 16% superiores ao controlo *E. coli* DH10B. Isto sugere o seu envolvimento na reparação de DNA.

Contudo, uma caracterização posterior destes genes permitirá um melhor conhecimento sobre os mecanismos moleculares e as vias metabólicas por detrás da resistência aos UVs.

Palavras-chave: metagenómica, radiação ultravioleta, ambientes extremos, *recA*, reparação de DNA, astrobiologia, mecanismos de adaptação

Abstract

To disclose the limits of life, it is fundamental to study the molecular strategies and adaptation mechanisms of microorganisms to extreme environments on Earth. Culture independent techniques have recently unveiled information on the resistance mechanisms of uncultured organisms, correcting our biased understanding of Earth's biodiversity. Extreme ultraviolet (UV) radiation exposure conditions are believed to have existed on early Earth, and are currently affecting surfaces of spacecrafts and planetary bodies as well as high altitude environments on Earth. Although UVB and UVC are harmful to life, microorganisms have been found thriving under high doses of UV radiation, triggering the curiosity of scientists on how life on Earth has evolved to adapt and resist to such conditions. In this project a functional metagenomic approach was used to identify novel genes responsible for UV-resistance.

Three metagenomic libraries were constructed and analysed using *E.coli DH10B* as a host from microbial communities highly exposed to UV radiation in hypersaline environments (two Andean ponds in Argentina and a saltern in Spain). Each library was screened for resistance to UVB and UVC, allowing the identification of recombinant clones harbouring an environmental DNA fragment conferring UV-resistance. In total, five resistant clones were identified: pML5, pML6, pML56, pML84 (Andean ponds), and pML105 (saltern), with survival rates around 15% higher than the control *E. coli* DH10B.

The environmental DNA fragments in these clones were sequenced and the open reading frames (ORF) were identified and annotated. The clone pML84 showed to harbour a single ORF each, encoding a C-terminal domain protein. The clone pML56 was shown to harbour two ORFs encoding a ribonuclease III and a transcription factor. In turn, the pML5 contains an ORF encoding for the RecA protein, a recombinase previously identified as involved in UV-resistance through DNA repair, mainly within the SOS response. Interestingly, the clones pML6 (Andean pond) and pML105 (saltern), from a distant geographical origin, encode each for hypothetical proteins sharing 32% identity.

To elucidate the mechanism of resistance, the clones were treated with 4-nitroquinoline 1-oxide, a compound that mimics the effect of UV radiation on DNA. In the presence of this compound, the survival rates of the clones were around 16% higher than those of the control, suggesting their direct involvement in DNA repair.

Nevertheless, further characterization of the identified UV-resistance genes will improve the knowledge of the molecular mechanisms and metabolic pathways behind them.

Key-words: metagenomics, ultraviolet radiation, extreme environments, *recA*, DNA repair, astrobiology, adaptation mechanisms

List of tables and Figures

Table 1. Extremophiles and their characteristics. Adapted from Pearce (2012)..... 15

Table 2. Characteristics of the Metagenomic libraries 35

Table 3. Characteristics of the identified clones and their DNA fragments 38

Table 4. DNA-binding predicted results for each identified ORF 41

Figure 1. The three domains of life: *Bacteria*, *Archaea* and *Eucarya*, defined by Woese (1987) 16

Figure 2. Effects triggered by UV irradiation on bacterial cells and respective DNA repair mechanisms. (UV) Ultra-violet radiation; (IR) Ionizing radiation; (ROS) Reactive Oxygen Species; (SSB) Single-strand break; (DSB) Double-strand break; (CPD) Cyclobutane dimer; (AP-site) Apurinic/aprimidinic site ; (BER) Base excision repair; (SP-BER) short-patch BER; (LP-BER) Long patch BER; (HR) Homologous recombination; (NHEJ) Non-homologous end joining; (NER) Nucleotide excision repair; (TC-NER) Transcription-coupled NER; (GG-NER) Global genome NER; (MMR) Mismatch repair. Adapted from (Rastogi et al. 2010). 18

Figure 3 - Overview of SOS response mechanism in *E. coli*. Adapted from Rastogi et al. (2010). When there is DNA damage, there are several DNA repair mechanisms that ensure survival. These can be (NER) Nucleotide Excision repair, (BER) Base excision repair; (MMR) Mismatch repair, (HR) Homologous recombination; or (NHEJ) Non-homologous end joining, among others. But when these mechanisms fail to repair the damage, there is a genomic wide mechanism that comes into action, known as SOS response. The SOS response activates a series of genes through the interaction of two main proteins: *lexA* and *RecA*. So when there is a failure of DNA repair mechanisms the *RecA* protein cleaves *LexA*, which will no longer repress the SOS regulon, leading to the expression of damage induced genes that assure the repair of the DNA. In case the damage is too much, then SOS response skips from the initial repair stage to the mutagenesis stage, where *RecA* activates the DNA Polymerase V that is able to perform Translesion synthesis, ensuring the replication and survival of the cell. 20

Figure 4 – Diagram of Metagenomics techniques. Function-based screening is marked in yellow. Adapted from Simon & Daniel (2011). 23

Figure 5. Scheme of the construction of metagenomic libraries. a) Environmental sample from three different environments; b) DNA isolation; c) DNA fragmentation – digestion with <i>Sau3AI</i> enzyme resulting in the optimal insert size 1-8 kb; d) vector preparation – the vector used was <i>pSKII+</i> digested with <i>BamHI</i> ; e) ligation – <i>pSKII+</i> with the fragmented DNA in order to originate a recombinant plasmid; f) electroporation; g) library titting; h) library amplification; i) functional screening – in this case, the screening was exposure to UVB radiation (312 nm) for 150 seconds.	26
Figure 6. Map of <i>pSKII+</i> (Stratagene) with the multiple cloning site region.	27
Figure 7. Steps of the UV screening. In general for 1 ml aliquot the procedure was the following: the initial screening consisted in spreading 10^8 cells in LB-Ampicillin plates, and irradiate them with UVB for 150 seconds. The plates were incubated ON at 37°C, and the clones that grow are UV resistant. The resistant clones were grown in patch in an LB+Ap+X-gal in order to select them for transformed cells and recombinant plasmids. The white ones were selected for further re-transformation of the plasmid in <i>E.coli</i> to assure the resistance was due to the environmental DNA and not due to spontaneous mutations. The re-transformed clones were screened by Drop Assay to confirm their resistance to UV radiation.....	30
Figure 8. (1.) Drop assay results. <i>E. coli</i> DH10B does not grow when exposed to UVB radiation for 80 seconds or to UVC radiation for 15 seconds. In turn, the recombinant clones are shown to be resistant, up to 10^{-2} dilution. (2.) Graphic representation of the results of the functional screenings of OS1(a) and D1 (b) metagenomic libraries by methodological stages (Initial, Patch, Re-transformation, Drop Assay and UVC).....	37
Figure 9. Representation of the organisation of the identified genes within the plasmid <i>pSKII+</i> . E-values in brakets. (HP) Hypothetical protein.	39
Figure 10. Alignment of <i>pML6-orf1</i> and <i>pML105-orf1</i> . The two HPs share 32% of identity, revealing potential common function in UV-resistance.....	40
Figure 11. Survival rate to 120 s of UVB (312 nm). *** indicate the statistical significance of Tukey test at a level of 0.05.....	42
Figure 12. Survival rate of the five resistant-clones to 60' of incubation with 4NQO (50µM). * indicates the significance of Dunnett test at a level of 0.05.....	43
Figure 13. Biochemically documented template-dependent DNA polymerases found in <i>E. coli</i> (Sutton & Walker 2001).	45

Abbreviations

°C	Degrees Celsius
µL	Microlitre
µM	Micromolar
4NQO	4-Nitroquinoline 1-oxide
6-4PPs	Pyrimidine 6-4 pyrimidone photoproducts
aa	Amino acid
Ap	Ampicillin
AP site	Apurinic/aprimidinic site
APE-1	AP endonuclease-1
ATP	Adenosine triphosphate
ATPase	Adenosinetriphosphatase
BER	Base excision repair
BLAST	Basic Local Alignment Search Tool
bp	Base-pair
CDS	Coding sequence (nucleotide)
CPDs	Cyclobutane pyrimidine dimers
din	Damage induced genes
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
DSBs	Double Stranded breaks
GC%	Guanine/Cytosine content
GFP	Green Fluorescent Protein
GG-NER	Global genome NER
h	Hour
HP	Hypothetical protein
HR	Homologous recombination
IPT	Immunoglobulin-like fold shared by plexins and transcription factors domains
kb	Kilo-base-pair
kV	Kilovolts
LB	Luria Bertrani
mRNA	Messenger RNA
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining repair

ON	Overnight
ORF	Open Reading Frame
PCD	Programed cell dead
PCR	Polymerase Chain Reaction
pSKII ⁺	pBluescript SKII ⁺
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
rpm	Revolutions per minute
s	Seconds
SOC	Super Optimal broth with Catabolite repression
SSB	Single Stranded breaks
TAE	Tris-acetate-Ethylenediaminetetraacetic acid
TBP	TATA-box binding protein
TC-NER	Transcription-coupled NER
TFIIH	Transcription factor-IIH
TLS	Translesion DNA synthesis
T _m	Melting temperature
UV	Ultraviolet radiation
UVA	Ultraviolet-A
UVB	Ultraviolet-B
UVC	Ultraviolet-C
V	Volts
v/v	Volume/volume
w/v	Weight/volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Table of contents

Acknowledgements.....	1
Resumo.....	2
Abstract	4
List of tables and Figures.....	6
Abreviations	9
Table of contents.....	11
1. Introduction	13
1.1. Astrobiology	13
1.2. Limits of Life	14
1.3. UV Radiation.....	17
1.4. Functional Metagenomics.....	22
1.5. Objectives.....	24
2. Materials and Methods.....	25
2.1. Bacterial strains, media and culture conditions.....	25
2.2. Construction of metagenomic libraries.....	25
2.2.1. DNA isolation and extraction	25
2.2.2. DNA fragmentation	25
2.2.3. Vector preparation.....	27
2.2.4. Ligation.....	28
2.2.5. Electroporation.....	28
2.2.6. Library titting.....	28
2.2.7. Library amplification.....	29
2.3. Functional screening of UV-resistance.....	30
2.3.1. UV irradiation	30
2.3.2. Initial plate assay.....	30
2.3.3. Drop Assay.....	31
2.3.4. Average insert size characterization	32
2.3.5. UV survival rate	33
2.4. Identification and analysis of resistance-conferring genes.....	34
2.4.1. Primer design and PCR reactions	34
2.4.2. Gene identification and <i>in silico</i> analysis	34
2.4.3. 4NQO survival rate	34

3.	Results	35
3.1.	Metagenomic libraries	35
3.2.	Functional Screening	36
3.3.	Gene identification and <i>in silico</i> analysis	38
3.4.	UV survival rate	42
3.5.	4NQO survival rate	43
4.	Discussion	44
4.1.	Gene identification	44
4.2.	pML5	44
4.3.	pML6 and pML105	45
4.4.	pM56	46
4.5.	pML84	46
4.6.	UV survival rate and 4NQO survival rate	47
5.	Conclusion and Future perspectives	48
6.	References	49
	Appendix	56
1.	Protein sequences of the identified ORFs	56

1. Introduction

1.1. Astrobiology

Astrobiology might strike as a new science but the possibility of life beyond Earth is an ancient Human question. Yet, it was the 50s' space age that revolutionized the modern debate over life on other worlds. In 1960 the term "exobiology" was first introduced by Joshua Lederberg as the study of life beyond Earth (Chyba & Hand 2005). The term "bioastronomy" was also used. By the end of the 20th century the term "astrobiology" took place, defined as a broader and multidisciplinary field addressing life in the context of its planetary history (Dick 2006). Today, Astrobiology is defined as the study of the origins, evolution, distribution, and future of life in the universe (Des Marais et al. 2008; Billings et al. 2006).

The NASA Astrobiology Roadmap emphasizes the multidisciplinary aspect of astrobiology, both in its content and execution. It requires the sciences of molecular biology, ecology, geology, planetary science, astronomy, information science, space exploration technologies, and others, to work together to (i) understand the origin and emergence of life in the universe, (ii) to know the limits to the Earth's biosphere, (iii) to understand the evolution of life on Earth, and (iv) to discover if life on Earth is unique and if there are other forms of intelligences in the universe (Des Marais et al. 2008).

Generally an agreed-upon definition of life is still lacking, but astrobiology depends on some kind of life definition which allows the differentiation of living beings when searching for life in other planets (Chodasewicz 2014). Definitions of life should be open and not limited by our current state of knowledge, thus establishing a framework from diverse perspectives (from the origin of life on Earth, to synthetic biology, to the search for extra-terrestrial life) rather than a precise definition (Tirard 2010). The definition accepted by NASA nowadays is that life is a self-sustained chemical system capable of undergoing Darwinian evolution, although this view has been a target of criticism (Chodasewicz 2014).

1.2. Limits of Life

Life on Earth is, by now, our only analogue to what we might expect to find in other worlds. This makes Earth analogy studies to be our most important methods when speculating about extraterrestrial environments. The study of the Earth's biogeochemistry and microbiome that characterise extreme environments of our planet are of great importance to the design and development of robotic systems that will be needed to search for biosignatures and evidence of past or present life on Mars, Europa, Titan and other prime targets for Astrobiology research (Pikuta et al. 2007; Gleeson et al. 2012). A biosignature is defined by Des Marais et al. (2008) as "an object, substance and/or pattern whose origin specifically requires a biological agent", and it can be, for instance, organic biomarker compounds, stable isotopes, atmospheric gases or biosedimentary structures (Mustard et al. 2013).

On Earth, the lack of liquid water is likely to represent the most fundamental constraint for life, but combinations of other stressors (e.g., radiation and vacuum conditions) may also control the habitability of extraterrestrial environments. In fact, desiccation and radiation were found to exert the strongest deleterious impacts on bacterial survival, acting as dominant selective pressure on microorganisms exposed to the terrestrial stratosphere and thermosphere (Harm 1980).

On the search for establishing the limits of life on Earth, every year scientists identify and characterize extremophiles - organisms that thrive in extreme environments – as shown in Table 1. Extremophiles reveal the ranges of habitability and possibility of life beyond Earth, demonstrating the incredible adaptability of life. There are several examples of extremophiles. Archaea, for instance, are known as extreme-lovers, including hyperthermophilic (high temperatures), halophilic (high salt concentrations), acidophilic (low pH) and UV-resistant microorganisms (Leuko et al. 2014; Mastascusa et al. 2014). Cyanobacteria can also cooperate with extreme temperatures and radiation exposure, even though they are photosynthetic organisms (Cowan et al. 2015). *Bacillus subtilis* has been identified to resist several extreme conditions, mainly due to endospore formation (Nicholson et al. 2000). And *Deinococcus radiodurans* (Battista 1997), is one of the most well-known species capable of surviving extreme radiation, withstanding lethal and mutagenic effects of DNA damage.

Table 1. Extremophiles and their characteristics. Adapted from Pearce (2012).

<i>Physico-chemical factor</i>	<i>Descriptive term</i>	<i>Genus/species</i>	<i>Lineage</i>	<i>Min</i>	<i>Opt</i>	<i>Max</i>	<i>Reference</i>
Temperature high	Hyperthermophile	<i>Methanopyrus kandleri</i> strain 116	Archaea	90°C	105°C	122°C	(Takai, et al., 2008)
Temperature low	Psychrophile	<i>Psychromonas ingrahamii</i>	Bacteria	-12°C	5°C	10°C	(Auman, et al., 2006)
pH low	Acidophile	<i>Picrophilus oshimae</i>	Archaea	pH - 0.06	pH 0.7	pH 4	(Schleper, et al., 1995)
pH high	Alkaliphile	<i>Alkaliphilus transvaalensis</i>	Bacteria	pH 8.5	pH 10	pH 12.5	(Takai, et al., 2001)
Hydrostatic pressure	Piezophile	<i>Pyrococcus CH1</i>	Archaea	20 MPa	52 MPa	120 MPa	(Zeng, et al., 2009)
Salt (NaCl)	Halophile	<i>Halobacterium salinarum</i>	Archaea	15% NaCl	25% NaCl	32% NaCl	(Grant, et al., 2001)
Water activity	Xerophile	<i>Xeromyces bispora</i>	Fungus	aw ^a 0.61	aw 0.82	aw 0.92	(Hocking & Pitt, 1999)
Radiation	Radioresistant	<i>Deinococcus radiodurans</i>	Bacteria	D ₃₇ ^b = 550 J/m ²		D ₃₇ = 600 J/m ²	(Battista 1997)

^a water activity^b dose required to kill 63% of cells

Knowledge of Earth's microbial diversity has influenced taxonomic classification over the years from early approaches in the 18th century based on physical traits (Paterlini 2007), towards recent advances in molecular phylogenetics (Fox et al. 1977), metagenomics (Handelsman 1998; Chistoserdova 2010), metaproteomics (Wilmes & Bond 2006) and single-cell sequencing (Raghunathan et al. 2005; McLean & Lasken 2014) that allowed direct access to microbial genomes and made it possible to rectify the bias of known biodiversity (Woese et al. 1990). Nowadays, the tree of life contains three primary lines of descent referred to as "domains" by Carl Woese: *Bacteria*, *Archaea* and *Eukarya* (Figure 1).

Understanding the physico-chemical limits of life on Earth, made it possible to turn out to the solar system, which is considered as of exobiological interest. Our solar system has been evaluated in terms of habitable zone and potential for planets other than Earth to harbour life. Defining reliable habitable zones (Kasting et al. 2013), can eventually lead to the detection of habitable exoplanets, and perhaps to evidence of extraterrestrial life. In the 1990's, the confirmation of a thick sheet of ice covering Europa, and the possibility that an ocean exists beneath it, focused renewed attention on the possibility of life beyond the presumed habitable zone of our solar system (Khurana et al. 1998). Moreover, the existence of organic molecules in the atmosphere of the Saturnian moon Titan (Lorenz et al. 2008), fostered the interest on the origin of life, and the limits of life as we know it (Ali et al. 2015; Lorenz et al. 2008).

Nevertheless, it will be crucial to evaluate whether any biological material potentially found beyond Earth represents an independent origin or another branch in the family tree of Earth life (Worth et al. 2013).

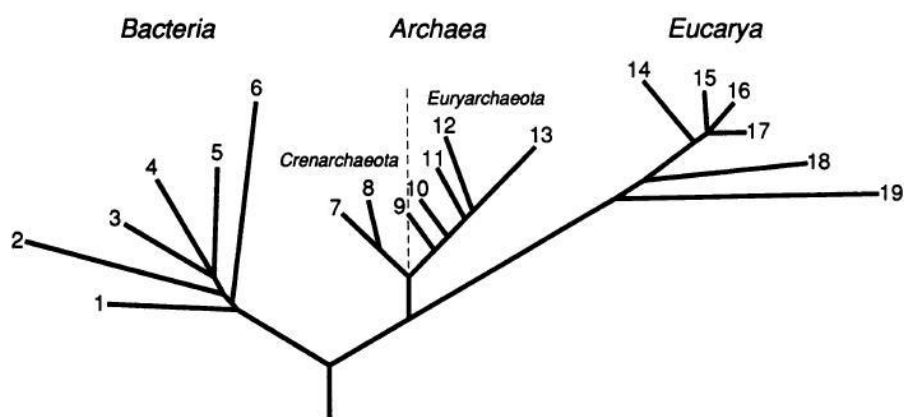


Figure 1. The three domains of life: *Bacteria*, *Archaea* and *Eucarya*, defined by Woese (1987)

1.3. UV Radiation

UV radiation is divided in three types: ultraviolet-A (UVA) (315-400 nm), ultraviolet-B (UVB) (280–315 nm) and ultraviolet-C (UVC) (280-100 nm), as defined by the International Commission on Illumination (Slinney 2007). On earth, most of the UVB radiation is attenuated by the ozone layer. Nevertheless, the exposure of short-wavelengths of ultraviolet radiation UVB and UVC causes direct and indirect damage to cells, particularly in DNA and photosystems (Sinha & Häder 2002; Cadet et al. 2014). High UV radiation exposure is believed to have existed in early Archaean Earth, posing a challenge to life's protection and repair processes (Cockell 1998) and is currently prevailing in space environments such as surfaces of spacecrafts and planetary bodies (Mancinelli & Klovstad 2000). Nevertheless microorganisms such as *Bacillus subtilis* or *Deinococcus radiodurans*, and others such as planctomycetes have been found thriving under these extreme conditions (Nicholson et al. 2000; Battista 1997). This has triggered scientists to question how life has evolved to resist to UV-light (Cockell 2000).

There are several strategies of UV radiation mitigation and UV screening providing a first line of defence against UV radiation. These can be followed by repair processes used to deal with damage. In fact there is a sequence of effects triggered by UV irradiation on bacterial cells such as mutagenesis and reactive oxygen species (ROS) formation that mainly affect DNA (Figure 2). UV radiation is responsible for most of the mutagenesis due to a process of DNA translesion synthesis (TLS) in which a polymerase encounters a noncoding or miscoding lesion, inserts an incorrect nucleotide opposite the lesion and then continues elongation. UV can also originate cyclobutane pyrimidine dimers (CPDs), pyrimidine 6-4 pyrimidone photoproducts (6-4PPs) and their Dewar isomers, apurinic/apyrimidinic sites (AP site), single-stranded breaks (SSB) and double-stranded breaks (DSB) (Rastogi et al. 2010). Consequently, there are several tolerance and damage-control mechanisms that allow organisms to cope with UV radiation (Figure 2).

These damage-control mechanisms can be: Nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), non-ending homologous recombination (NEHJ) and photoreactivation, acting in a lesion-directed level. Whereas base excision repair (BER) and SOS response act in a general line of DNA damage repair (Harm, 1980). Post-replication repair, de novo synthesis of proteins and lipids and programmed cell death (PCD) or apoptosis may also become effective for the

recovery of genome integrity (Cockell & Knowland 1999; Cohen & Walker 2011; Rastogi et al. 2010).

Regarding AP-site repair, the AP-site is removed by the action of AP endonuclease-1 (APE-1) along with phosphodiesterase that breaks the DNA strand along 5' or 3' to the AP site, respectively, and subsequently the gap is filled by a repair DNA polymerase and the strand is joined by a DNA ligase (Rastogi et al. 2010).

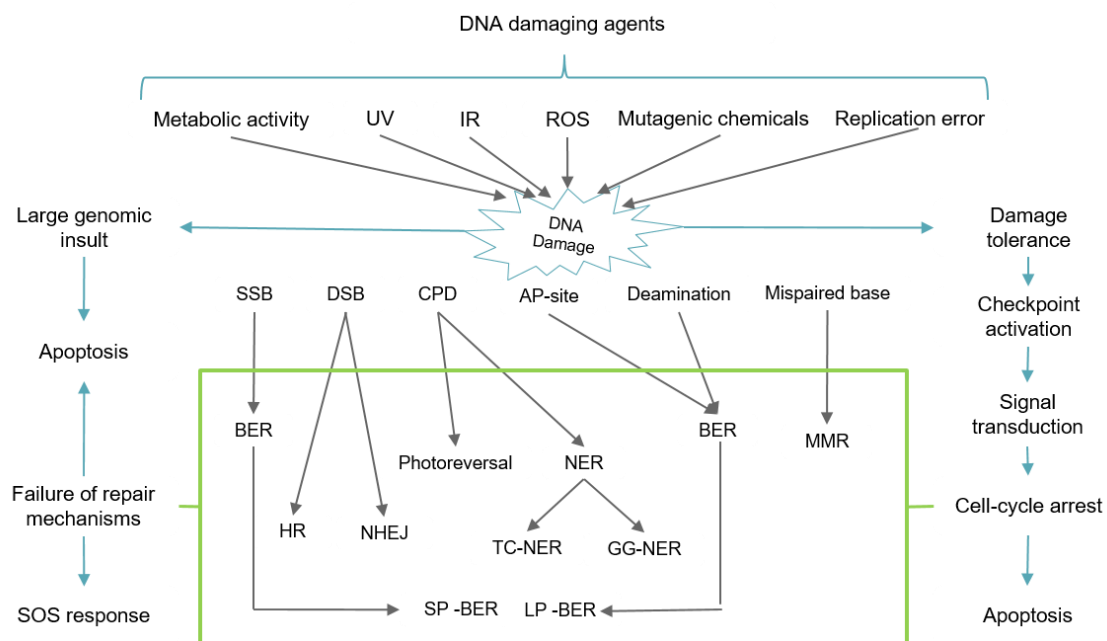


Figure 2. Effects triggered by UV irradiation on bacterial cells and respective DNA repair mechanisms. (UV) Ultra-violet radiation; (IR) Ionizing radiation; (ROS) Reactive Oxygen Species; (SSB) Single-strand break; (DSB) Double-strand break; (CPD) Cyclobutane dimer; (AP-site) Apurinic/aprimidinic site; (BER) Base excision repair; (SP-BER) short-patch BER; (LP-BER) Long patch BER; (HR) Homologous recombination; (NHEJ) Non-homologous end joining; (NER) Nucleotide excision repair; (TC-NER) Transcription-coupled NER; (GG-NER) Global genome NER; (MMR) Mismatch repair. Adapted from (Rastogi et al. 2010).

NER is one of the most versatile and flexible repair systems found in most organisms, being highly conserved in eukaryotes. Discovery of NER was first described in *E. coli* where about six proteins such as UvrA, B, and C (known as ABC-complex, which shows exonuclease activity), UvrD (helicase II), DNA polymerase I (pol. I), and DNA ligase are recruited to complete the repair (Rastogi et al. 2010). NER is critically important in the repair of UV-induced DNA lesions, sorting out a wide range of DNA damages such as CPDs and 6-4PPs. Depending on the damage location, two sub-pathways of NER coexist and are mechanistically conserved from prokaryotic to eukaryotic cells: global genome NER (GG-NER) removes lesions anywhere in the genome, whereas transcription-coupled NER (TC-NER) specifically removes lesions

from the transcribed strand of active genes (Alekseev & Coin 2015; Logette et al. 2011).

In turn, the BER pathway against UV radiation is done via generation of ROS which efficiency and specificity are determined by several forms of DNA glycosylase. These glycosylases remove different types of modified base by cleaving the N-glycosidic bond between the abnormal base and deoxyribose creating either an abasic site or an SSB which is further processed by short-patch repair or long-patch repair (Krokan & Bjøra 2013; Rastogi et al. 2010).

On one hand, homologous recombination can help repair double-stranded breaks, by using foreign DNA as a template (Bernstein et al., 1981). The strand-exchange reaction is mediated mainly by RecA family proteins both in *Bacteria* and *Eukarya*. Evidence of homologous recombination can be found in multilocus sequence typing studies (Didelot et al. 2010). On the other hand, non-homologous end joining (NHEJ) is an evolutionarily conserved repair system that helps to maintain the gene integrity for DNA damage (Ochi et al. 2014).

The process of DNA repair by photoreactivation is done through photoreactivating enzymes known as photolyases, which are conserved and can be found throughout the three domains of life, but seem to be absent or non-functional in humans. These enzymes are present in archaea suggesting their role as ancient repair proteins, which may have helped in the evolution of the earliest organisms on primordial Earth (Sinha & Häder 2002).

The SOS response is activated when there is severe DNA damage impeding DNA replication and repair from proceeding effectively, as it is exemplified in Figure 3. The primary task of the SOS response is to restart replication before the cell dies. This system is regulated by both the LexA transcriptional repressor and the RecA recombinase. In *E. coli*, the UV radiation firstly damages the DNA by creating lesions that mechanically interrupt the process of DNA duplication by stalling the DNA-polymerase (Pol III) in a moving replication fork. This results in the production of single-stranded DNA breaks (SSB). These breaks are recognised by the protein RecA involved in the non-mutagenic coating of SSB via homologous recombination. RecA also catalyses the cleavage of LexA diminishing its levels. The cleaved LexA de-represses the regulon involved in the SOS response – *din* (*damage induced*) genes -

and allows expression of SOS responding genes (Krishna et al. 2007; Rastogi et al. 2010).

The UmuD protein, which is part of the UmuC/UmuD₂ complex of the error-prone DNA polymerase (Pol V), is also cleaved by RecA, directly resulting in mutagenesis in UV-irradiated *E. coli* cells. Pol V inserts several random base pairs in the DNA strand directly opposite a lesion, thus helping a replication fork to quickly bypass the lesion (TLS) after which Pol III can take over and continue replication. DNA polymerase IV (*dinB*) is also involved in TLS in *E. coli* (Rastogi et al. 2010).

SOS response is complex, in fact SOS regulon is known to comprise more than 40 genes, including those encoding the mutagenesis proteins UmuD and UmuC, RecA, and LexA. Also part of the SOS regulon are genes encoding UvrA, B, C—a group of nucleotide excision repair (NER) proteins that locate and excise damaged regions from the DNA (Shah & He 2015; Aksenov 1999; Janion 2008).

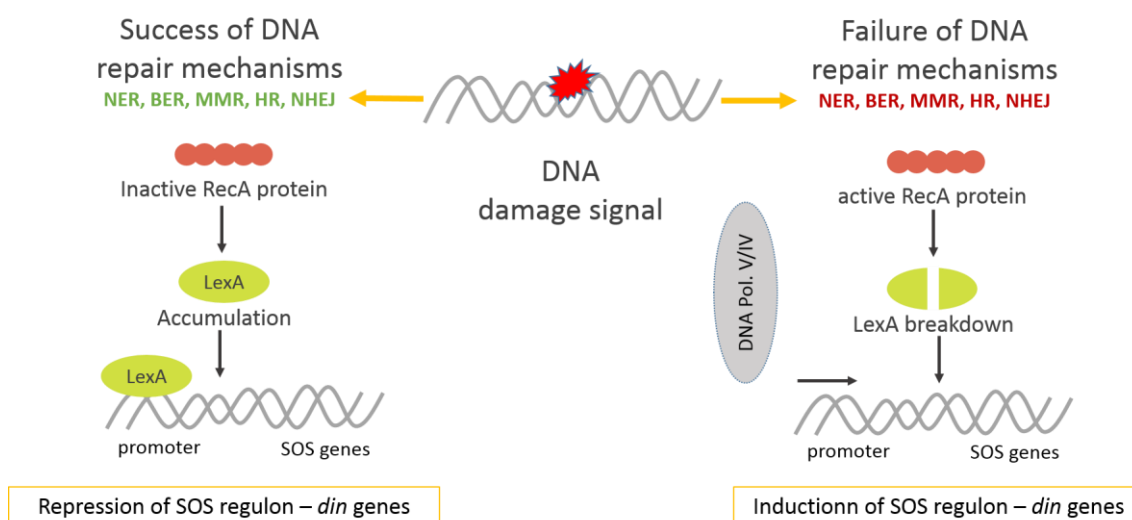


Figure 3 - Overview of SOS response mechanism in *E. coli*. Adapted from Rastogi et al. (2010). When there is DNA damage, there are several DNA repair mechanisms that ensure survival. These can be (NER) Nucleotide Excision repair, (BER) Base excision repair; (MMR) Mismatch repair, (HR) Homologous recombination; or (NHEJ) Non-homologous end joining, among others. But when these mechanisms fail to repair the damage, there is a genomic wide mechanism that comes into action, known as SOS response. The SOS response activates a series of genes through the interaction of two main proteins: *lexA* and *RecA*. So when there is a failure of DNA repair mechanisms the *RecA* protein cleaves *LexA*, which will no longer repress the SOS regulon, leading to the expression of damage induced genes that assure the repair of the DNA. In case the damage is too much, then SOS response skips from the initial repair stage to the mutagenesis stage, where *RecA* activates the DNA Polymerase V that is able to perform Translesion synthesis, ensuring the replication and survival of the cell.

Although SOS response has been studied for decades, new discoveries are being made every year. However, most of the currently known mechanisms of adaptation to UV radiation are based on studies that address less than 1% of the total amount of microorganisms that exist on Earth, the ones that can be cultured and maintained in a laboratory. There is still 99% that remain unstudied, giving us scientists a biased understanding of the microbial and functional diversity (Staley & Konopka 1985).

Thus, to better explore the diversity of adaptation mechanisms, there is a need to study the other 99% of uncultured microorganisms. This can be done by using culture independent techniques, associated with earths' extreme environments.

1.4. Functional Metagenomics

Metagenomic strategies comprise analysis of both DNA – construction of metagenomics libraries, sequencing of DNA, sequencing of phylogenetic markers (such as 16S rRNA) – and RNA – analysis of mRNA or rRNA, as shown in Figure 4 (Simon & Daniel 2011). The direct sequence analysis of metagenomic DNA is presently considered the most accurate method for assessing the structure of an environmental microbial community, since it does not involve any selection (by cultivation and/or enrichment) and minimizes technical biases, as the ones introduced by PCR amplification of the 16S rRNA (Lewin et al. 2013). Previously identified genes found through metagenomics include chaperons, transporters, DNA-binding proteins, (Chistoserdova 2010; Shestakov 2012), bacteriorhodopsins (Beja et al. 2000), among others.

Function-based metagenomics, or functional metagenomics, has become an efficient tool to find novel genes involved in stress adaptation processes by expressing the metagenome of a microbial community in a foreign host (Ekkers et al. 2012). Metagenomic libraries can thus be screened for genes coding for proteins with a particular function: antibiotics (Gillespie et al. 2002), enzymes of commercial interest (Jeon et al. 2009), quorum sensing inhibitors or inducers (Williamson et al. 2005), aromatic compounds (Uchiyama & Miyazaki 2013), or compounds conferring resistance to extreme conditions (Allen et al. 2009). Moreover, hypothetical and unknown genes not previously assigned to confer resistant to these conditions, can be annotated with the help of functional metagenomics.

By assuring that uncultured microorganisms are taken into account as an important source of unknown genes, metagenomics plays a key-role in understanding microbial adaptation mechanisms and shows great potential in the area of biotechnology. In fact, the use of metagenomics in the search for novel molecular mechanisms of adaptation of the microorganisms to extreme conditions has been popular in recent studies, contributing to uncover the limits of life (Cowan et al. 2015). These studies involve resistance to toxic metals (González-Pastor & Mirete 2010), acidic pH (Guazzaroni et al. 2013), low and high temperatures (Lewin et al. 2013) or high salt concentrations (Culligan et al. 2013).

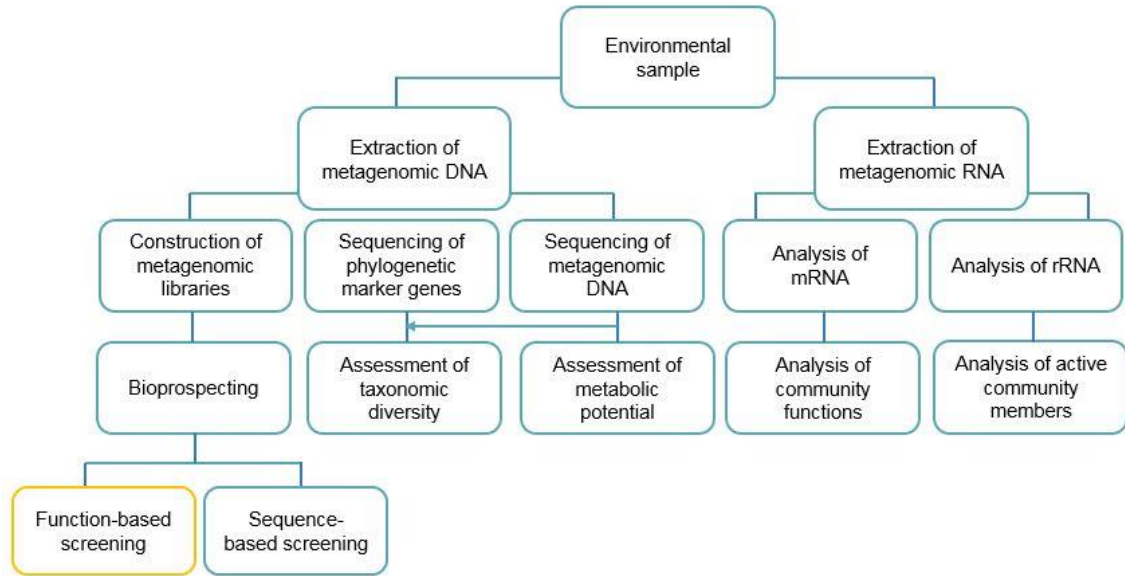


Figure 4 – Diagram of Metagenomics techniques. Function-based screening is marked in yellow. Adapted from Simon & Daniel (2011).

1.5. Objectives

The current project is set within the goal number five of the Astrobiology Roadmap, which strives to “Understand the evolutionary mechanisms and environmental limits of life. Determine the molecular, genetic, and biochemical mechanisms that control and limit evolution, metabolic diversity, and acclimatization of life” (Des Marais et al. 2008). In this way, this project aims to:

1. Identify novel genes responsible for UV-resistance from the metagenome of different environmental samples (representing different pools of microorganisms adapted to both hypersaline and UV-exposure environments) by using a functional metagenomic approach;
2. Characterize these genes and their involvement in the adaptation mechanisms of microorganisms to UV radiation.

The main goal of this 10-month project was to improve the knowledge of microbial molecular strategies and metabolic function of adaptation mechanisms involved in UV-resistance, taking uncultured microorganisms as an important source of yet unknown genes. Enlightening the molecular basis of adaptation of microorganisms to UV radiation will allow us to disclose aspects of the evolution of life on Earth and other planets, and like other metagenomic studies, it will give us the chance to apply the knowledge to biotechnological applications (Shestakov 2012; Simon & Daniel 2011).

2. Materials and Methods

2.1. Bacterial strains, media and culture conditions

Escherichia coli DH10B was grown in Luria-Bertani (LB) medium, over-night (ON) at 37 °C. The growth LB medium for recombinant *E. coli* strains was supplemented with 80 µg/ml ampicillin (Ap).

2.2. Construction of metagenomic libraries

The construction of metagenomic libraries was performed, following the González-Pastor & Mirete protocol (González-Pastor & Mirete 2010), consisting of: DNA isolation and extraction, DNA fragmentation, vector preparation, ligation, electroporation, library titting and library amplification. A schematic overview of metagenomic library construction is shown in Figure 5.

2.2.1. DNA isolation and extraction

Environmental samples were previously collected from several high-altitude hypersaline environments. These were two Andean ponds from Argentina (Diamante at 4600 m and Ojo seco at 3900 m), and one Spanish saltern in Mallorca (Es Trenc). Each of these samples contain a different pool of microorganisms highly exposed to UV radiation. Metagenomic DNA was isolated using BIO101 FastDNA spin kit (Qbiogene, Carlsbad, CA, USA) assessing the quality, quantity and concentration by spectrophotometer (dilution 1/12). To extract the genomic DNA the Wizard R Genomic DNA purification Kit (Promega) was used, following the manufacturer's instructions.

2.2.2. DNA fragmentation

To obtain DNA fragments between 1-8 kb which ensure the presence of one or more genes, the metagenomic DNA was digested using the frequent-cutting restriction enzyme *Sau3AI* at a concentration that permitted a partial digestion of the DNA.

Digestion was analysed in a low melting point agarose gel electrophoresis (1%, TAE), which allowed the recovery of intact DNA fragments after electrophoresis. DNA was purified using QIAquickR Gel Extraction kit, Qiagen, and DNA sample was concentrated on SpeedVac. DNA concentration values were determined using a spectrophotometer.

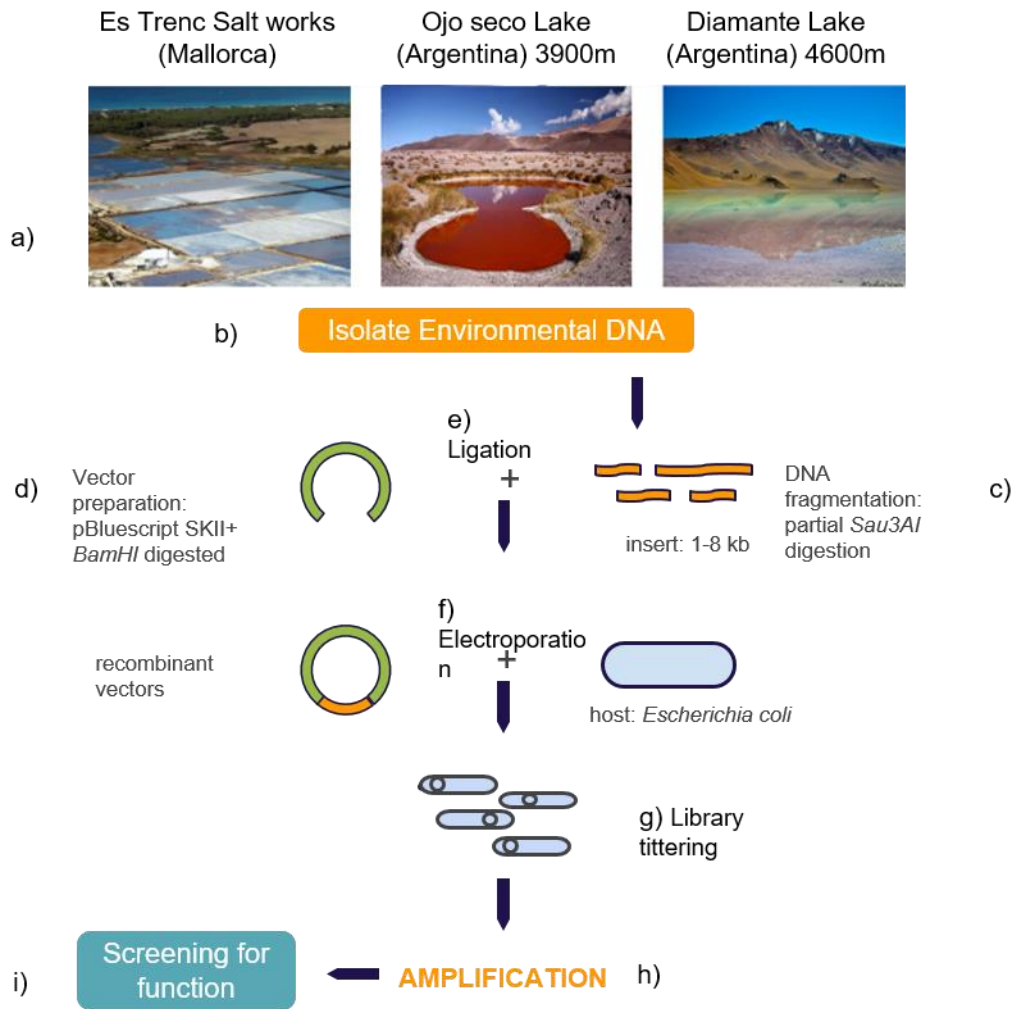


Figure 5. Scheme of the construction of metagenomic libraries. a) Environmental sample from three different environments; b) DNA isolation; c) DNA fragmentation – digestion with *Sau3AI* enzyme resulting in the optimal insert size 1-8 kb; d) vector preparation – the vector used was pSKII+ digested with *Bam*HI; e) ligation – pSKII+ with the fragmented DNA in order to originate a recombinant plasmid; f) electroporation; g) library titting; h) library amplification; i) functional screening – in this case, the screening was exposure to UVB radiation (312 nm) for 150 seconds.

2.2.3. Vector preparation

The vector pBluescript SKII⁺ (pSKII⁺) was prepared for cloning by BamHI digestion, which provides a restriction site compatible with that of the *Sau3A*I enzyme. To access the information contained in the metagenome, the environmental DNA fragments were inserted in pSKII⁺ by ligation. The resulting recombinant plasmids were used to transform the *Escherichia coli* DH10B strain, allowing their expression and subsequent screening for UV-resistance. By using pSKII⁺ the selection of correct incorporation of the plasmid in the *E. coli* strain is enabled by ampicillin resistance and alpha-complementation (Langley et al. 1975). In this way, pSKII⁺ was digested with restriction enzymes allowing further correct cloning of DNA fragments.

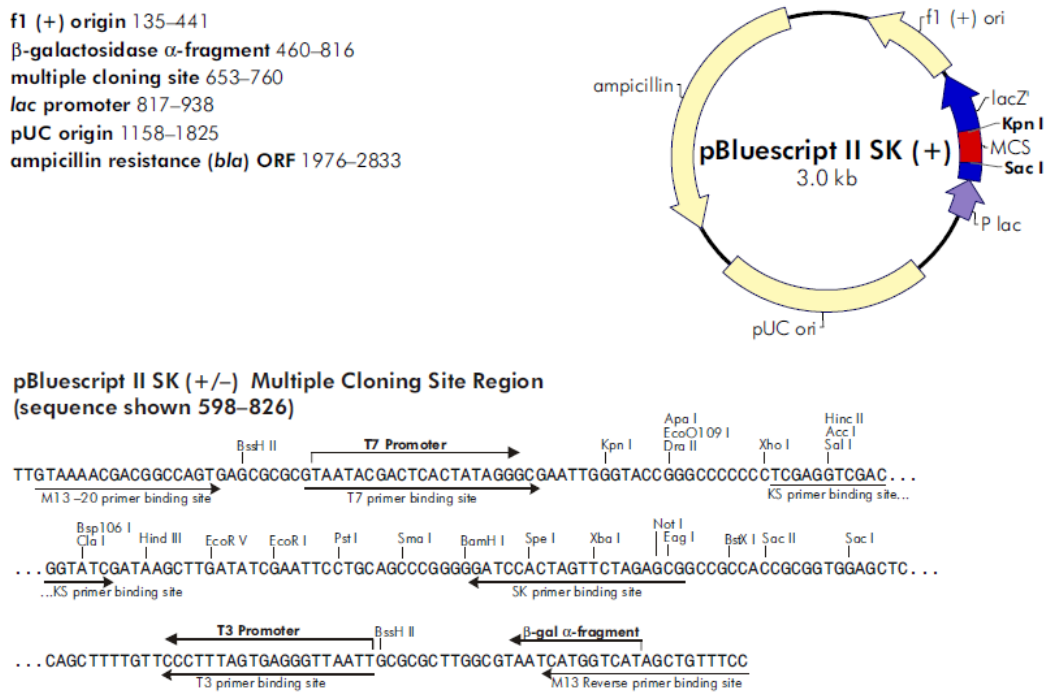


Figure 6. Map of pSKII⁺ (Stratagene) with the multiple cloning site region.

2.2.4. Ligation

Ligation of the metagenomic DNA fragments (100 ng) and pSKII⁺ (100 ng) was done using 1 µL of T4 DNA ligase (Roche, Mannheim, Germany), and 1 µL of T4 DNA ligase buffer (10x) in a total volume of 10 µL. For each library, independent ligation reactions were set up (x3) and incubated ON at 16 °C in a heating block. Ligase was inactivated at 65 °C for 15 min in a heating block. The three ligation reactions were transferred by pipetting to a fresh microcentrifuge tube and water was added up to 200 µL. DNA was precipitated following the protocol described in (González-Pastor & Mirete 2010) except at the final dissolution, done by adding 10 µL of distilled water and incubated at 55 °C for 5 min.

2.2.5. Electroporation

Transformation of 3 µl of recombinant plasmid (ligation) in 20 µl of electrocompetent *E.coli* DH10B cells (Invitrogen, Carlsbad, CA, USA) was performed by electroporation using Micropulser (Bio-Rad, Hemel Hempstead, UK) at 1.8 kV according to the manufacturer's instructions. DNA, electroporation cuvettes and filter tips were previously cooled on ice or on fridge for at least 30 minutes. To assure enough transformants each electroporation was performed 3 times. After electroporation, 1ml S.O.C medium (Hanahan 1983) was added and samples were incubated at 37 °C for 1:15 h at 225 rpm. 1/20 dilutions in LB-Ap (80mg/ml) and X-gal (40mg/ml) were plated and incubated at 37 °C ON.

2.2.6. Library titting

Library titting was done by preparing serial dilutions of 100 µL of the metagenomic library, up to 10⁻⁸ dilution. 100 µL of the last 3 dilutions were plated in LB-Ap, and incubated at 37°C ON. The titting value allows us to determine the precise number of cells to be exposed to UV treatment. In this case, 10⁸ cells/plate were used in the UV-screening. Average insert size of the metagenomic library was determined by isolating and digesting 10 random clones with restriction enzymes flanking the insert: XhoI and XbaI (Roche) [1 µg de plasmid, 1,5 µl buffer H (Roche), 0,5 µl XbaI, 0,5 µl XhoI, adding water up to 15 µl]. Incubate at 37 °C for 3h in the heating block. Fragment length polymorphism was analysed with a 1% agarose gel TAE, stained with Pronasafe. Control used was non-recombinant vector pSKII⁺ digested with XhoI and XbaI.

2.2.7. Library amplification

To assure enough library sample for multiple functional screenings, library amplification must be performed, either in solid or liquid medium, ending with up to 10^8 - 10^{10} cfu/ml. A metagenomic library aliquot of 20 μ l was kept at -80°C for further studies.

- i) Plate amplification: the metagenomic library was directly plated on LB-Ap solid medium. To get the final number of plates to be used, the following formula was applied: total number of clones (titering value) \times 3 / colonies per plate. Plates with LB-Ap were incubated at 37°C ON. Afterwards, cells from each plate were mixed with 3,5 ml LB and 10% glycerol (w/v), being pooled in a flask with cells from the same library, mixed again. The pool was divided in several 1ml/ependorf aliquots which were stored at -80°C .
- ii) Liquid amplification: the metagenomic library was centrifuged at 4°C , 8000 rpm for 15 min. The supernatant was discarded and the remaining sample was resuspended in 1 ml of LB+Ap. The sample was added to 50 ml LB+Ap (80 $\mu\text{L}/\mu\text{g}$) and incubate at 37°C for 6 h in the water bath with agitation (no measurements to avoid loss of sample). When the culture reached $A_{600}=2$ glycerol was added at 10% (w/v) and aliquots of 1 ml/ependorf were stored at -80°C .

2.3. Functional screening of UV-resistance

2.3.1. UV irradiation

In order to select the clones harbouring genes conferring UV-resistance, the metagenomic libraries were exposed to UV radiation. The UV lamp used is VL6-MC from Vilber Lourmat, and can irradiate at 254 nm (UVC) or at 312 nm (UVB). The UV irradiation chamber has an upper aperture, where the lamp fits, and a side aperture where the Petri dish is placed. The distance from the UV lamp and the plate is approximately 24 cm, which gives us the chance to irradiate with 1.56 W m^{-2} at 254 nm and 2.26 W m^{-2} at 312 nm. A general scheme of the steps of the UV-screening is presented below in Figure 7.

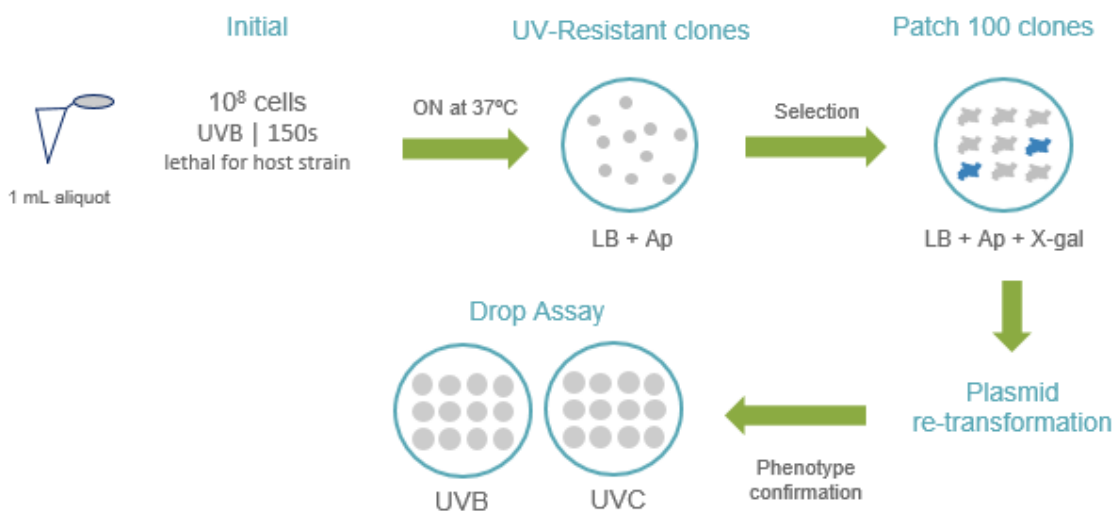


Figure 7. Steps of the UV screening. In general for 1 ml aliquot the procedure was the following: the initial screening consisted in spreading 10^8 cells in LB-Ampicillin plates, and irradiate them with UVB for 150 seconds. The plates were incubated ON at 37°C, and the clones that grow are UV resistant. The resistant clones were grown in patch in an LB+Ap+X-gal in order to select them for transformed cells and recombinant plasmids. The white ones were selected for further re-transformation of the plasmid in *E.coli* to assure the resistance was due to the environmental DNA and not due to spontaneous mutations. The re-transformed clones were screened by Drop Assay to confirm their resistance to UV radiation.

2.3.2. Initial plate assay

Each metagenomic library was plated on LB-Ap solid medium (10^8 cells) and exposed to UV radiation, 312 nm for 150 s, conditions that we previously observed that were lethal to the *E. coli* DH10B strain, host of the metagenomic libraries. The UV resistant clones were isolated and patched on LB agar medium supplemented with Ap and X-gal

and incubated ON at 37 °C. These two supplements originate a selective medium for an antibiotic (only transformed cells can grow) and alpha-complementation selectivity by blue and white colonies (the colonies harbouring an environmental DNA fragment cloned in the pSKII⁺ will be white, whereas the colonies without DNA fragment will be blue) (Langley et al. 1975).

To rule out potential chromosomal spontaneous mutants conferring UV-resistance, the resistant colonies were pooled by adding 3,5 mL of LB-Ap to each plate and collecting the clones into several aliquots of 1mL. Plasmids were isolated from the aliquots, and used to transform DH10B cells by the heat shock method. 1/10 and 1/20 dilutions of the transformations were plated on LB-agar supplemented with Ap and X-gal and incubated ON at 37 °C. The resulting colonies will be first streaked on LB-Ap-Xgal and incubated ON at 37 °C, and kept at 4 °C until the Drop Assay.

2.3.3. Drop Assay

To select the clones harbouring a recombinant vector containing an environmental DNA fragment conferring UV resistance, a drop assay was used. For that, a culture is established from a single colony, using Ap as the selective medium, and incubated ON at 37 °C, 140 rpm. The cultures were normalized to get an A600=1, to exclude false resistance phenotypes due to the presence of higher amounts of cells in some cultures. Afterwards, 10 µL of each culture and the negative control (*E. coli* DH10B) were spread on LB-Ap-X-gal plates, and exposed to UV radiation (312 nm) for 80 s, and incubated ON at 37 °C. These UV exposure conditions are lethal for *E. coli*.

To exclude clone resistance conferred by different cell growth or higher amounts of cells, cultures from each resistant clones at A600=1 were serially diluted up to 10⁻⁵, and drops from each dilution were exposed to UVB radiation at 312 nm, 80 s and to UVC radiation at 254 nm, 15 s. In this assay, two controls were included: i) a non-irradiated growth of the dilutions, to evaluate if the differences in UV resistance are due to differences in growth of the clones, and ii) dilutions of a culture of the strain host of the library (DH10B) at A600=1.

2.3.4. Average insert size characterization

To characterize the insert size of the resistant clones, they were grown in liquid LB-Ap-X-gal and incubated ON at 37 °C. This culture allowed plasmid extraction using QIAprep R Spin Miniprep Kit (250) (Quiagen), determination of plasmid concentration using Nanodrop (Thermo), and insert size characterization by performing a double digestion of the plasmid with XhoI and XbaI enzymes, followed by result analysis in 0,8% TAE agarose gel electrophoresis, ran at 80V for 45 min. A small amount of culture of each clone was stored in 15% (w/v) glycerol at -80 °C for further analysis. Unique clones were recovered by plasmid digestion with XhoI, XbaI [2 µL DNA, 1.5 µL buffer H (Roche), 0,5 µL XhoI, 0.5 µL XbaI, adding H₂O up to a total of 15 µL] and HindIII [2 µL of DNA, 1.5 µL of Buffer B (Roche), 0,5 µL of HindIII and H₂O up to a total of 15 µL]. Pattern analysis was done through in 0,8% TAE agarose gel electrophoresis, ran at 80 V for 45 min.

2.3.5. UV survival rate

To compare the survival rate among the resistant clones and the negative control the *E. coli* strain DH10B, a culture of each clone was grown in LB-broth-Ap and incubated ON at 37 °C 140 rpm until A600=1. 100 µL of each culture on Eppendorf lids attached to a petri dish were exposed to 120 s of UVB (312 nm). This was followed by serial dilutions up to 10⁻⁶. Dilutions 10⁻⁴, 10⁻⁵ and 10⁻⁶ were plated on solid LB-Ap. Controls were assured by serial dilutions up to 10⁻⁷ of unexposed cultures, having the 10⁻⁵, 10⁻⁶ and 10⁻⁷ dilutions plated in LB-Ap. All plates were incubated at 37 °C ON and survival rate was further calculated. Each experiment was performed at least three times for each clone.

2.4. Identification and analysis of resistance-conferring genes

2.4.1. Primer design and PCR reactions

Once selected, the environmental DNA fragments from the resistant clones were sequenced by primer walking to search for the genes conferring resistance. For that, plasmid sequencing (STAB vida) using M13 forward and reverse primers was performed. This allowed the design of specific primers to “walk” along the sequence, using DNA start and ApE software, to obtain a trustful sequence of the insert, which in turn will guarantee the right *in silico* analysis of the insert.

2.4.2. Gene identification and *in silico* analysis

In silico analysis mainly consisted on the identification and characterization of putative Open Reading Frames (ORFs) using ORF finder and fgenesB (Solovyev & Salamov, 2011) and further similarity check using Basic Local Alignment Search Tool (BLAST) (Altschul, et al., 1997). Protein characterization and function determination were carried out using bioinformatics tools, such as to find protein domains with NCBI conserved domain search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer et al. 2014) as well as to assess DNA-binding properties with iDNA-Prot|dis (http://bioinformatics.hitsz.edu.cn/iDNA-Prot_dis/) (Liu et al. 2014), iDNA-prot (<http://www.jci-bioinfo.cn/iDNA-Prot>) (Lin et al. 2011), DNABinder (<http://www.imtech.res.in/raghava/dnabinder/submit.html>) (Kumar et al. 2007) and DNABIND (<http://dnabind.szilab.org/>) (Szilágyi & Skolnick 2006).

2.4.3. 4NQO survival rate

To further characterize the role of each DNA fragment in UV-resistance, the survival rate clones and *E. coli* DH10B cells were analysed when exposing 100 µL of A600=1 cells to 100 µM of 4-Nitroquinoline 1-oxide (4NQO) for 60 minutes, in comparison with unexposed cells. The treatment was performed in a ventilated reaction chamber using the appropriate equipment. Serial dilutions of the exposed 100 µL of A600=1 cells were performed up to 10⁻⁶. Dilutions, 10⁻⁴, 10⁻⁵ and 10⁻⁶ were plated in LB-Ap. Controls were assured by serial dilutions up to 10⁻⁷ of unexposed cultures, having the 10⁻⁵, 10⁻⁶ and 10⁻⁷ dilutions plated in LB-Ap. All plates were incubated at 37 °C ON and survival rate was further calculated. Each clone's treatment with 4NQO was repeated at least three times.

3. Results

3.1. Metagenomic libraries

In this project, two metagenomic libraries were constructed, from DNA extracted from planktonic microorganisms from brines of two hypersaline environments: Ojo Seco (OS1) and Diamante (D1) as described in material and methods. The metagenomic library from Mallorca (PMB) was previously constructed by Salvador Mirete, being analysed in this study.

Characteristics of the Metagenomic libraries analysed in this study are shown in Table 2. Approximately 240,000 clones were obtained from PMB, 150,000 from OS and 170,000 from D, and an estimation of 685, 541 and 170 MB of environmental DNA respectively was cloned in each library. Library amplification was done as described in material and methods. Average insert size was calculated as described in Materials and Methods resulting in 2.9 kb for PMB, 1.9 kb for OS1 and 1.0 kb for D1.

Table 2. Characteristics of the Metagenomic libraries

Source	Name	No. of recombinant clones	Average insert size (kb)	Library size (Mb)	Library titting
<i>Es Trenc saltern Mallorca</i>	PMB	236.250	2,9	865	10 ¹⁰ cell/ml
<i>Ojo seco Lake Argentina 3900 m</i>	OS1	150.000	1,9	541,5	10 ¹⁰ cell/ml
<i>Diamante Lake Argentina 4600 m</i>	D1	170.000	1,0	170	10 ¹⁰ cell/ml

3.2. Functional Screening

The functional screening of the PMB metagenomic library was done previously to this work by Maria Lamprecht Grandio (unpublished data). Only one clone (pML105) was shown to harbour a recombinant plasmid and resist to both UVB and UVC radiation using Drop Assay as described in Materials and Methods and depicted in Figure 8.1.

The screening of the metagenomic library of OS (Figure 8.2.a)) has resulted in few clones (13 in total), each of which was isolated in patch in LB-AP-X-gal plates to check for the ones carrying the recombinant plasmid. 6 out of the 13 clones grew and were selected for re-transformation of *E. coli*. This re-transformation assured that the resistance was conferred by the environmental DNA fragment and not by spontaneous mutations caused by UVB exposure. After re-transformation, only 1 clone (pML6) was shown to harbour the recombinant plasmid and resist to both UVB and UVC radiation using Drop Assay as described in Materials and Methods (Figure 8.1).

The initial screening of the metagenomic library of Diamante (Figure 8.2.b)) has resulted in 76 resistant clones, each of which was isolated in patch in LB-AP-X-gal plates to check for the ones carrying the recombinant plasmid. The 67 grown patches were gathered in several 1 mL pools of LB-Ap, being afterwards re-transformed in *E. coli*. This re-transformation assured that the resistance was due to the environmental fragment and not to spontaneous mutations caused by UVB exposure. After re-transformation, only 26 clones were shown to still harbour the recombinant plasmid. 17 out of the 36 were shown to survive to UVB exposure conditions. Out of these 17, 3 of them (pML5, pML56 and pML84), were also shown to resist to UVC exposure using Drop Assay (Figure 8.1).

In total, there were 4 UV-resistant clones identified in this study: pML5 (D), pML6 (OS), pML56 (D), pML84 (D). And 1 clone from a previous study was further analysed pML105 (PMB).

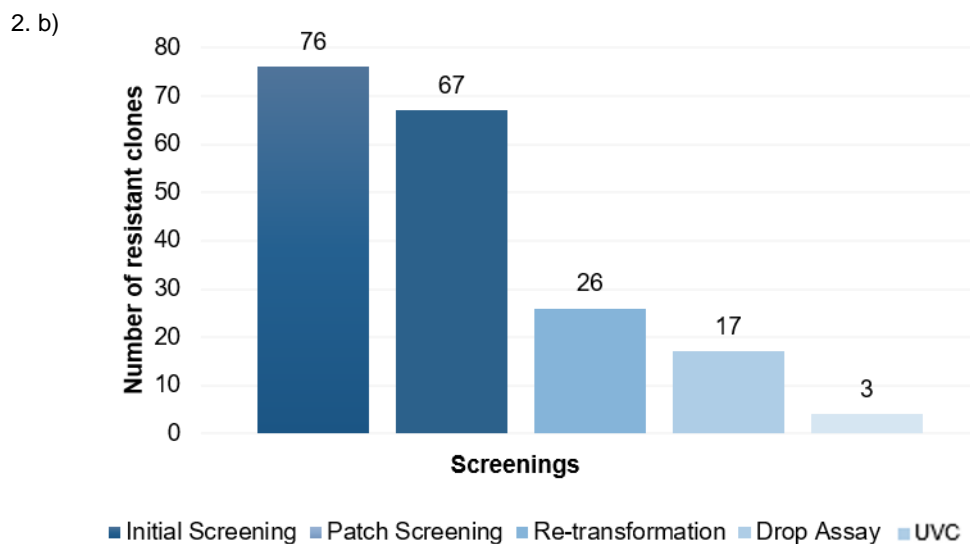
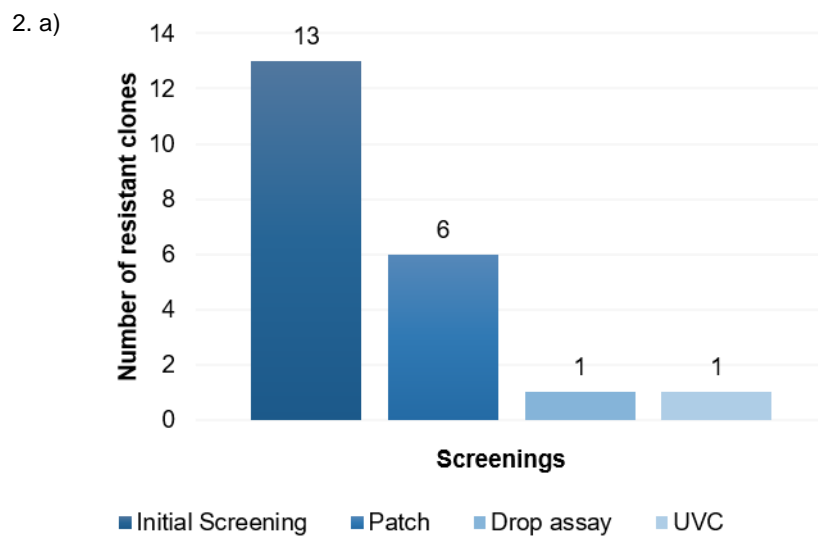
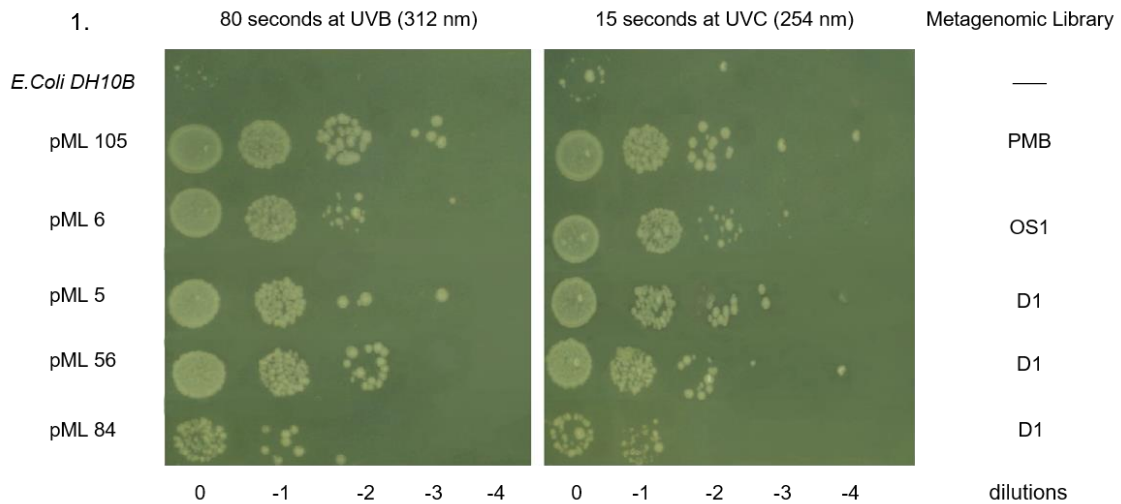


Figure 8. (1.) Drop assay results. *E. coli* DH10B does not grow when exposed to UVB radiation for 80 seconds or to UVC radiation for 15 seconds. In turn, the recombinant clones are shown to be resistant, up to 10⁻² dilution. (2.) Graphic representation of the results of the functional screenings of OS1(a) and D1 (b) metagenomic libraries by methodological stages (Initial, Patch, Re-transformation, Drop Assay and UVC).

3.3. Gene identification and *in silico* analysis

Each of the 5 clones identified in this study harbours an environmental DNA fragment, conferring resistance to UV radiation. Each fragment contained one or more ORFs which sequences were ran through BLAST in order to identify the genes involved in UV-resistance. The BLAST results are presented in Table 3 and the genetic organization of the 5 clones is presented in Figure 9 along with a short description of the best blast hit for the identified ORFs.

Table 3. Characteristics of the identified clones and their DNA fragments

ID clones [Length (bp)] %G+C	ORF	Truncated ORFs	Length (aa)	Closest similar protein (Microorganism)	% identity/ E-value	Length (aa)	GenBank Accession No.
pML5 (1544) 59%	1	No	352	recA protein Marinobacter nanhaiticus	87% (0.0)	348	WP_004578747.1
pML6 (863) 46%	1	No	158	Hypothetical protein (<i>Rickettsia</i> endosymbiont of <i>Ixodes scapularis</i>)	35% (8.00e ⁻¹⁸)	182	EER22366.1
	2	C-terminal	108	DNA polymerase (<i>Thioalkalivibrio</i> sp ALgr3)	31% (0,55)	1078	WP_018870135.1
pML56 (812) 49%	1	C-terminal	118	Putative Ribonuclease III (<i>Sulfitobacter</i> sp. 20_GPM-1509m)	36% (1.4)	228	WP_028957201.1
	2	No	70	Transcription factor (<i>Haloferax volcanii</i>)	78% (4e ⁻²⁹)	180	WP_004042154.1
pML84 (1640) 43%	1	C-terminal	279	T9SS C-terminal target domain-containing protein (<i>Hymenobacter swuensis</i>)	33% (3e ⁻³³)	1319	WP_044001462.1
pML105 (1547) 48%	1	No	189	Hypothetical protein (<i>Moritella dasanensis</i>)	43% (3e ⁻³⁷)	157	WP_01722522.1
	2	No	233	Putative hypothetical protein PFICI_078 47 (<i>Pestalotiopsis fici</i> W106-1)	29% (0.46)	1192	XP_007834619.1
	3	C-terminal	90	Hypothetical protein (<i>Mesorhizobium huakuii</i>)	52% (0.22)	102	WP_038652848.1

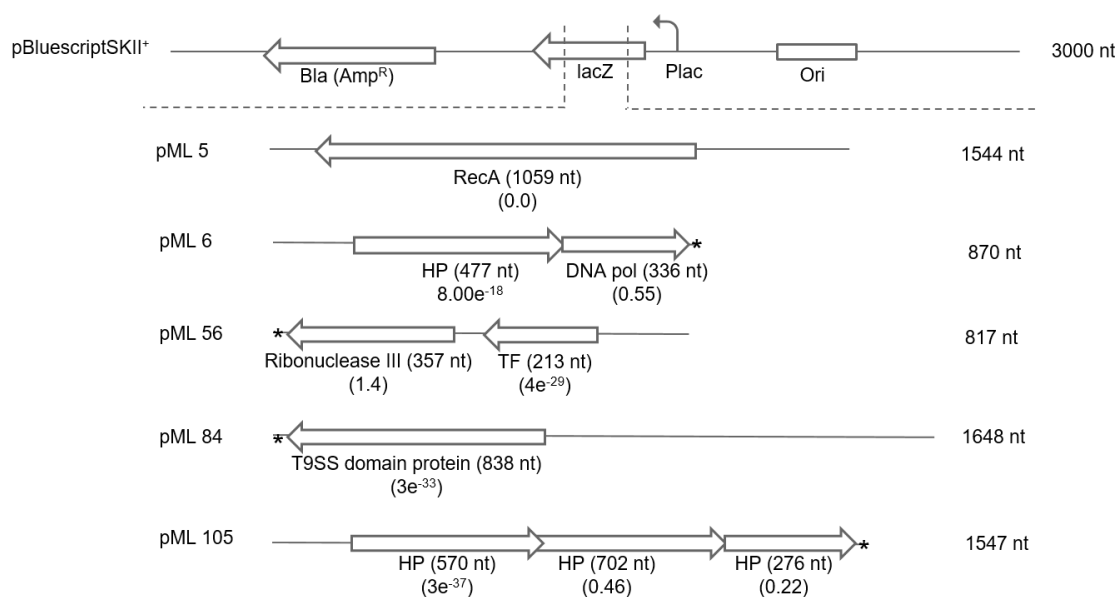


Figure 9. Representation of the organisation of the identified genes within the plasmid pSKII⁺. E-values in brackets. (HP) Hypothetical protein.

ORF1 closest BLAST hit was RecA protein from *Marinobacter nanhaiticus*, with an identity of 87% and an E-value of 0.0. The closest BLAST hit for pML6-orf1 was shown to be a hypothetical protein (HP) from *Rickettsia* endosymbiont of *Ixodes scapularis*, with an identity of 35% and an E-value of 8.00e⁻¹⁸. A lower BLAST hit for pML6-orf1 showed a HP from *Moritella dasanensis* with 31% identity and an E-value of 2e⁻¹¹ (WP_017221522.1). Interestingly pML105-orf1 closest BLAST hit was a HP from *M. dasanensis* with an identity of 43% and an E-value of 3e⁻³⁷. Further analysis between the latter two ORFs was made by alignment in NCBI, revealing a 32% identity between the two HPs (see Figure 10).

The BLAST hit of pML6-orf2 was shown to be a DNA polymerase from *Thioalkalivibrio* sp *ALgr3*, a Gammaproteobacteria, with an identity of 31% and an E-value of 0,55. In turn, the closest BLAST hit for pML105-orf2 was a putative HP PFICI_078 47 (*Pestalotiopsis fici* W106-1). For pML105-orf3 the closest BLAST hit was a HP (*Mesorhizobium huakuii*).

pML56-orf1 was shown to be a putative ribonuclease III *Sulfitobacter* sp (20_GPM-1509m) whereas pML56-orf 2 closest BLAST hit was a transcription factor from *Haloferax volcanii* with an identity of 78% and an E-value of 4e⁻²⁹. Interestingly the second BLAST hit showed to be a TATA-binding transcription initiation factor from *Haloquadratum walsbyi* with an identity of 58% and an E-value of 5e⁻²³. Several

conserved domain hits were found for pML56-orf2, including archaeal TATA box binding protein (TBP) and Transcription factor TFIID (or TATA-binding protein, TBP).

pML84-orf1 closest BLAST hit was a T9SS C-terminal target domain-containing protein from the organism *Hymenobacter swuensis*, with an identity of 33% and E-value of $3e^{-33}$. A conserved domain was found to be a non-specific hit with E-value of $1.88e^{-04}$. Pfam accession number is 07610, representing a protein of unknown function (DUF1573). This domain is present in bacteria such as *Rhodopirellula baltica*, *Bacteroides thetaiotaomicron*, and *Porphyromonas gingivalis* which share a region of conserved sequence towards their N-termini.

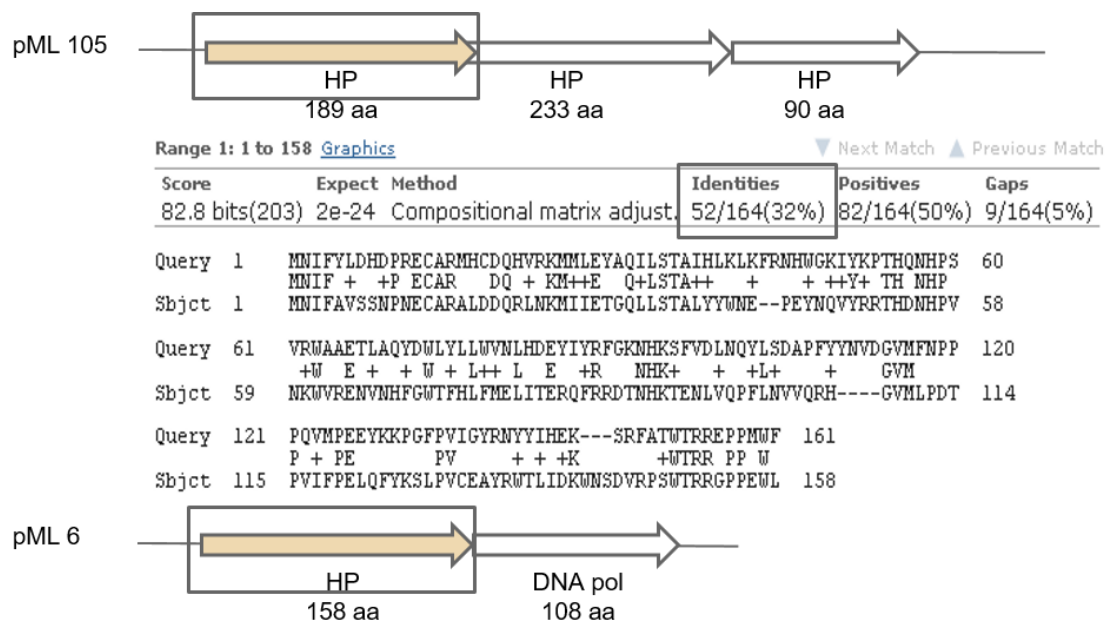


Figure 10. Alignment of pML6-orf1 and pML105-orf1. The two HPs share 32% of identity, revealing potential common function in UV-resistance.

To assess the DNA binding proteins of each identified ORF, the four programmes described in Materials and Methods section were used to analyse the protein sequences (Appendix 1) and results are depicted in Table 4. It was found that both pML6-orf1 is consistently predicted as DNA-binding proteins on the four analysis, and that pML56-orf2 and pML105-orf2 were predicted as DNA-binding proteins in three out of the four analysis.

Table 4. DNA-binding predicted results for each identified ORF

	<i>iDNA-Protdis</i>	<i>iDNA-prot</i>	<i>DNAbinder</i>	<i>DNABIND</i>
<i>pML5-orf1</i>	DNA-binding protein	-	-	-
<i>pML6-orf1</i>	DNA-binding protein	DNA-binding protein	DNA-binding protein	DNA-binding protein
<i>pML6-orf2</i>	-	-	DNA-binding protein	DNA-binding protein
<i>pML56-orf1</i>	-	-	-	-
<i>pML56-orf2</i>	DNA-binding protein	DNA-binding protein	DNA-binding protein	-
<i>pML84-orf1</i>	-	-	-	-
<i>pML105-orf1</i>	DNA-binding protein	-	DNA-binding protein	DNA-binding protein
<i>pML105-orf2</i>	-	DNA-binding protein	DNA-binding protein	DNA-binding protein
<i>pML105-orf3</i>	DNA-binding protein	-	DNA-binding protein	-

3.4. UV survival rate

In order to quantitatively determine the levels of resistance conferred by the selected clones (represented in Table 3) their survival rate under UV exposure was assayed as described in Materials and Methods section. This experiment also ruled out the possibility of resistance due to variation of cell numbers. When compared to the *E.coli* DH10B survival when exposed to 120 s of UVB radiation (5,3%), all the clones have shown an average of around 15% more of survival rate than the control: pML5 (20%), pML6 (19,7%), pML56 (18,1%), pML84 (19,3%), pML105 (27,3%) as can be seen in Figure 11. The results show values that are statistically significance, according to the Tukey test, at a level of 0.05.

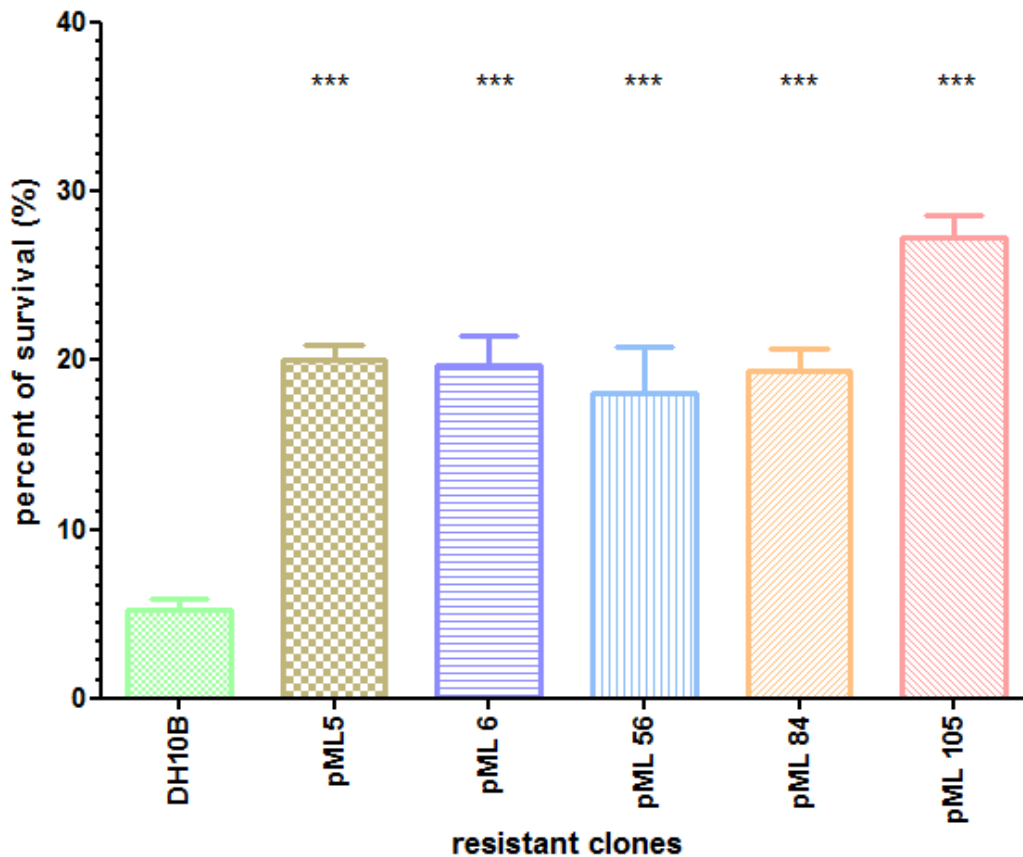


Figure 11. Survival rate to 120 s of UVB (312 nm). *** indicate the statistical significance of Tukey test at a level of 0.05

3.5. 4NQO survival rate

In order to explore the molecular mechanism of resistance of each gene in UV resistance, the clones were tested for their survival rate when exposed to 4NQO, a compound that only induces DNA lesions mimicking the effect of UV radiation on DNA. Thus, if the retrieved genes conferred also resistance to 4NQO means that they could have a specific role in DNA repair. The treatment was done as described in Materials and Methods.

E.coli DH10B survival rate, when exposed to 50 µM of 4NQO for one hour was shown to be 18,1%. All the clones have shown an average of around 16% survival rate above the one from *E.coli*: pML5 (31,9%), pML6 (34,1%), pML56 (38,1%), pML84 (38,3%), pML105 (30,7%) as can be seen in Figure 12. The results show values that are statistically significance, according to the Dunnett test, at a level of 0.05.

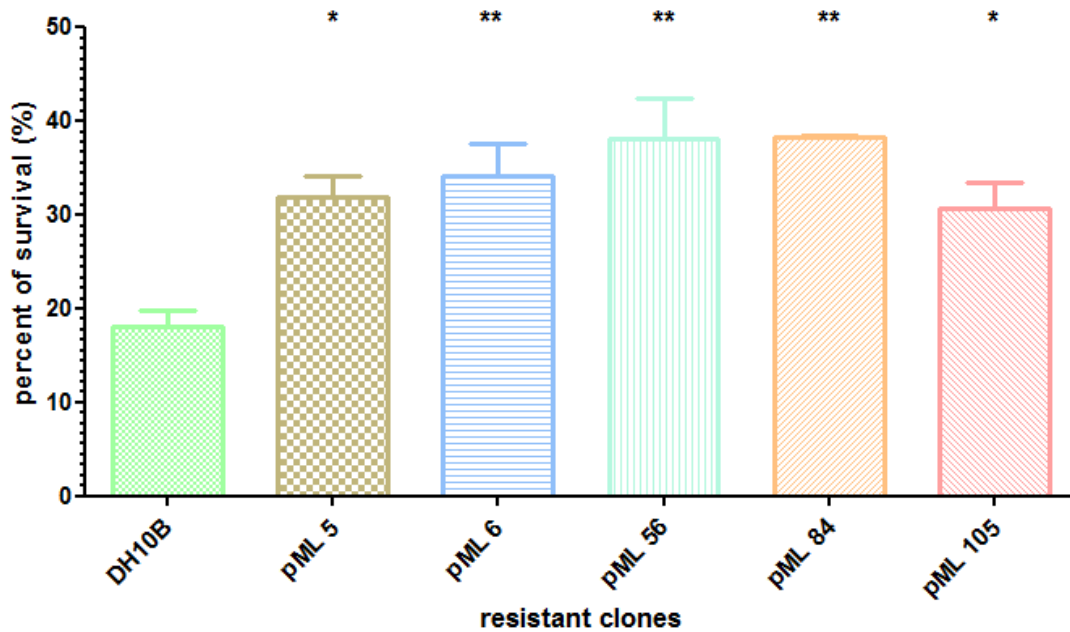


Figure 12. Survival rate of the five resistant-clones to 60' of incubation with 4NQO (50µM). * indicates the significance of Dunnett test at a level of 0.05.

4. Discussion

4.1. Gene identification

One limitation for the functional screening of metagenomic libraries relies in the inappropriate expression of heterologous proteins in *E. coli*. Nevertheless, a total of 9 ORFs were identified from three geographically distant environments.

The low sequence identities of most of the genes identified in this work highlight the novelty of these proteins. From a total of nine identified ORF products involved in UV-resistance, six presented less than 50% of amino acid identity with their closest BLAST hit (Table 3). The lack of sequence conservation with previously described proteins supports the role of culture-independent techniques in correctly assessing biodiversity. It supports that functional screenings could identify genes that would likely be missed in culture-based analyses.

4.2. pML5

pML5-orf1 closest BLAST hit was RecA protein, as expected, since the screening was performed in a *recA*-mutant *E. coli*. This protein was initially identified and characterized genetically in a screen for recombination deficient mutants of *E. coli* and was quickly found to have multiple roles in recombination and repair (Patel et al. 2010). RecA has four known functions in the *E. coli* cell: catalysing the DNA strand exchange reaction in the context of recombinational DNA repair, induction of the SOS response by promotion of the autocatalytic cleavage of the LexA repressor, activation of UmuD' by mediating autocatalytic cleavage of UmuD, and direct participation in SOS mutagenesis by activation of DNA polymerase V.

In SOS-uninduced cells, RecA is present at less than 10,000 monomers per cell, but upon SOS induction the level of RecA can increase to over 70,000 monomers per cell (Krishna et al. 2007). The bacterial RecA protein is a DNA-dependent Adenosinetriphosphatase (ATPase). The RecA of *Escherichia coli* hydrolyses ATP (Adenosine triphosphate) at a rate of 20–30 per minute, depending on the nature of the

bound DNA. RecA binds to DNA as a nucleoprotein (np) filament that forms most rapidly onto single-stranded DNA (Patel et al. 2010).

4.3. pML6 and pML105

After analysis of both pML6-orf1 and pML105-orf1 HPs, it was shown that they share a 32% of identity (see Figure 10). Interestingly these two clones come from two distant geographical sites: pML6 from Ojo Seco Andean Lake in Argentina, and pML105 from Es Trenc saltern, in Mallorca, Spain. Besides that, both ORFs were predicted as DNA-binding proteins, supporting their role in DNA repair in the context of UV resistance (Table 4).

The three ORFs of pML105 were previously separately subcloned (unpublished data) in order to understand which one was responsible for resistance-conferring. It was found that pML105 orf1 was the one conferring UV resistance. Thus we were drawn to believe it is the pML6-orf1, which also confers UV-resistance in pML6. Nevertheless further subcloning of pML6 is needed to confirm this assumption.

The BLAST hit of pML6-orf2 was shown to be a DNA polymerase from *Thioalkalivibrio* sp *ALgr3*, a Gammaproteobacteria. DNA polymerases exist with a vast range of roles, including enzymes that are capable of replicating imperfect DNA templates such as UmuC-DinB-Rad30-Rev1 superfamily involved in TLS. In fact, *E. coli* contains DNA polymerases known to participate in DNA replication, repair, NER, TLS and cell cycle progression as seen in Figure 13 (Sutton & Walker 2001).

Since both ORFs of pML6 appear to have potential functions related to UV-resistance, it enhances the need for subcloning pML6 and correct identification of the resistance-conferring ORF.

<i>E. coli</i>			
Pol I		A	Maturation of Okazaki fragments; nucleotide excision repair
Pol II	DinA	B	Induced replisome reactivation; TLS
Pol III		C	DNA replication; nucleotide excision repair
Pol IV	DinB	UmuC	Adaptive mutagenesis, TLS
Pol V	UmuC or UmuD ₂ C	UmuC	TLS ("SOS mutagenesis")

Figure 13. Biochemically documented template-dependent DNA polymerases found in *E. coli* (Sutton & Walker 2001).

4.4. pM56

pML56-orf1 closest BLAST hit was shown to be a putative ribonuclease III *Sulfitobacter sp* (20_GPM-1509m). RNase III is a global regulator of gene expression in *Escherichia coli*, being involved in the maturation of ribosomal and other structural RNAs. RNase III can affect gene expression in a variety of different ways, such as changing the structure of an mRNA, promoting its degradation and thereby reducing gene expression, or altering the structure of an mRNA such that it is translated more efficiently by ribosomes. This supports RNase III involvement in UV adaptation and resistance mechanisms by contributing to an efficient regulation of gene expression.

pML56-orf2 closest BLAST hit was a transcription factor from *Haloferax volcanii*, although the second BLAST hit showed to be a TATA-binding transcription initiation factor from *Haloquadratum walsbyi*, and fifth BLAST hit was shown to be a TATA-box binding protein (TBP), component of TFIID and TFIIB from *Haloquadratum walsbyi* J07HQW1. pML56-orf2 being predicted as a transcription factor, is in agreement with DNA-binding protein prediction results shown in Table 4., revealing its role in DNA repair. In fact, it is known that in the repair of AP-sites, a greater binding effect of APE1 enzyme on TATA sequence has been suggested, identifying a prominent role of the TATA sequence in DNA repair (Singh et al. 2014). The unwinding of DNA double helix at the site of lesion takes place by the components of multi-subunit transcription factor-IIH (TFIIH), which is usually involved in initiation of RNA Pol II transcription, but upon DNA damage can be employed in cell cycle regulation and NER (both in global genome and TC-NER) (Alekseev & Coin 2015).

4.5. pML84

pML84-orf1 closest BLAST hit was a T9SS C-terminal target domain-containing protein from the organism *Hymenobacter swuensis*, with an identity of 33% and E-value of $3e^{-33}$. A conserved domain was found to be a non-specific hit with E-value of $1.88e^{-04}$, representing a protein of unknown function (DUF1573), present in bacteria that share a region of conserved sequence towards their N-termini.

The T9SS is a type IX secretion system, and its protein substrates have conserved C-terminal domains. These domains are important for secretion, posttranslational modification, and cell surface attachment (Wang et al. 2014). The domains associated

with the BLAST hit *H. swuensis* (a red-pink pigmented and radiation-resistant bacterium (Lee et al. 2014)) are Immunoglobulin-like fold shared by plexins and transcription factors domains (IPTs), also known as Transcription factor ImmunoGlobin (TIG) domains. These domains are present in intracellular transcription factors and cell surface receptors (such as plexins) (Bork et al. 1999), suggesting their involvement in UV-resistance by regulation and/or compound sorting to the extracellular environment. There are known proteins with IPT domains that act in the disintegration of the actin filament, suggesting that pML84 might be involved in apoptosis induction as a mechanisms assuring genome integrity (Hung & Terman 2011). These IPT domains have also been found in thermophilic archaea and bacteria, supporting the importance of these domains for survival in extreme environments (Bertoldo & Antranikian 2002).

4.6. UV survival rate and 4NQO survival rate

UV radiation causes damage at different levels and 4NQO is a compound that mimics the effect of UV radiation on DNA. Thus, if the retrieved genes confer also resistance to 4NQO it means that they might confer UV-resistance by being involved in DNA repair. However, when assessing the survival rate of the five identified clones there is the need to keep in mind that each gene might be expressed at a different level in *E. coli*. Thus, a bigger higher survival rate can be due to a higher expression of the resistance conferring gene and not necessarily due to a better fitted resistance mechanism. These experiments were made with the full environmental DNA fragment, not having into account the role of each ORF (which can be done by further subcloning). Nevertheless, there might be cases where only one gene is responsible for the resistance within the fragment. In this sense, analysing the whole fragment constitutes a proper first assessment of the UV and 4NQO survival rate that they provide. Variation in expression levels and whole fragment experiments can explain why pML105 and pML6, present a 7,6% difference while sharing 32% identity between the ORF1 of each clone. When comparing between UV and 4NQO survival rates, no further assessments can be made, since respective controls have different base survival rates, and since varying the 4NQO levels on the experiment can lead to higher or lower survival rates of each clone.

In summary, all the five clones were shown to develop statistically significant results when assessing their survival rate indicating the involvement of the harboured genes in UV resistance and DNA damage repair.

5. Conclusion and Future perspectives

Culture independent techniques are important to unveiled information on the resistance mechanisms of uncultured organisms and correct our biased understanding of Earth's biodiversity, and metagenomic is one of the key technologies to access and explore this potential. Study of microorganisms' adaptation to extreme environments is essential to a better knowledge of the limits of life on Earth and beyond. Recently, the continuous depletion of the ozone layer and the extensive use of radioactive elements and compounds for energy, in medicine, research, and industry has greatly influenced the amount of UV radiation in Earth's biosphere (Singh & Gabani 2011). Nevertheless, there are microorganisms that thrive under extreme radiation conditions, being referred to as radiation-resistant extremophiles. These have been found in wide environmental niches such as high altitude areas and open fields where UV levels are high. Molecular and physiological responses and resistance mechanisms to UV radiation have been previously investigated in cultivable microorganisms such as *E. coli* and other microorganisms (Santos et al. 2012; Ordoñez et al. 2009; Kurth et al. 2015). However, to the best of our knowledge functional metagenomics has not ever been applied to explore the mechanisms that confer UV-resistance to microorganisms. Thus this represents one of the first approaches in which this methodology has been applied successfully in the identification UV-resistance genes.

Through the functional screening of metagenomic libraries from microorganisms exposed to high UV radiation conditions we have obtained five clones that confer resistance. Within the environmental DNA fragments of these clones, 9 ORFs were identified: an expected RecA protein (the screening was performed in a *recA*-mutant *E. coli*), four hypothetical proteins, one DNA polymerase, one ribonuclease III, one transcription factor and one T9SS C-terminal target domain-containing protein. Interestingly, two of the identified HPs were found to be similar, although they come from geographically distant environments, suggesting a relatively conserved UV resistance mechanism in microorganisms. Nevertheless further *in silico* and *in vivo* characterization of the identified genes will improve the understanding of the molecular mechanisms and metabolic pathways behind UV-resistance. Subcloning each ORF followed by drop assay and survival rate experiments will give a more detailed knowledge on which ORF is responsible for the UV-resistance. Also, subcellular location, using GFP-fusions of each expressed protein would help to determine their precise role and level of involvement in UV-resistance mechanisms.

6. References

- Aksenov, S. V., 1999. Induction of the SOS Response in Ultraviolet-Irradiated *Escherichia coli* Analyzed by Dynamics of LexA, RecA and SulA Proteins. *Journal of Biological Physics*, 25, pp.263–277.
- Alekseev, S. & Coin, F., 2015. Orchestral maneuvers at the damaged sites in nucleotide excision repair. *Cellular and Molecular Life Sciences*. Available at: <http://link.springer.com/10.1007/s00018-015-1859-5>.
- Ali, a. et al., 2015. Organic chemistry in Titan's upper atmosphere and its astrobiological consequences: I. Views towards Cassini plasma spectrometer (CAPS) and ion neutral mass spectrometer (INMS) experiments in space. *Planetary and Space Science*, 109-110, pp.46–63. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0032063315000161>.
- Allen, H.K. et al., 2009. Functional metagenomics reveals diverse β -lactamases in a remote Alaskan soil. *The ISME Journal*, 3, pp.243–251.
- Battista, J.R., 1997. Against all odds: the survival strategies of *Deinococcus radiodurans*. *Annual review of microbiology*, 51, pp.203–224.
- Beja, O. et al., 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, 289.
- Bertoldo, C. & Antranikian, G., 2002. Starch-hydrolyzing enzymes from thermophilic archaea and bacteria. *Current Opinion in Chemical Biology*, 6(2), pp.151–160.
- Billings, L. et al., 2006. The astrobiology primer: an outline of general knowledge--version 1, 2006. *Astrobiology*, 6(5), pp.735–813.
- Bork, P. et al., 1999. Domains in plexins: Links to integrins and transcription factors. *Trends in Biochemical Sciences*, 24(7), pp.261–263.
- Cadet, J., Grand, A. & Douki, T., 2014. Solar UV Radiation-Induced DNA Bipyrimidine Photoproducts: Formation and Mechanistic Insights. *Top Curr Chem*, 11(1), pp.13–35.
- Chistoserdova, L., 2010. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology Letters*, 32(10), pp.1351–1359.
- Chodasewicz, K., 2014. Evolution, reproduction and definition of life. *Theory in Biosciences*, 133(1), pp.39–45.

- Chyba, C.F. & Hand, K.P., 2005. A STROBIOLOGY : The Study of the Living Universe. *Annual Reviews of Astronomy & Astrophysics*, 43(1960), pp.31–74. Available at: <http://www.annualreviews.org/doi/abs/10.1146/annurev.astro.43.051804.102202>.
- Cockell, C.S., 1998. Biological effects of high ultraviolet radiation on early earth--a theoretical evaluation. *Journal of theoretical biology*, 193(4), pp.717–729.
- Cockell, C.S., 2000. The ultraviolet history of the terrestrial planets — implications for biological evolution. *Planetary and Space Science*, 48(2-3), pp.203–214.
- Cockell, C.S. & Knowland, J., 1999. Ultraviolet radiation screening compounds. *Biological reviews of the Cambridge Philosophical Society*, 74(3), pp.311–345.
- Cohen, S.E. & Walker, G.C., 2011. New discoveries linking transcription to DNA repair and damage tolerance pathways. *Transcription*, 2(1), pp.37–40.
- Cowan, D. et al., 2015. Metagenomics of extreme environments. *Current Opinion in Microbiology*, 25, pp.97–102. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1369527415000569>.
- Culligan, E.P. et al., 2013. Functional environmental screening of a metagenomic library identifies stIA; a unique salt tolerance locus from the human gut microbiome. *PLoS ONE*, 8(12), pp.1–15.
- Dick, S.J., 2006. NASA and the search for life in the universe. *Endeavour*, 30(2), pp.71–75.
- Didelot, X. et al., 2010. Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*, 186(4), pp.1435–1449.
- Ekkers, D.M. et al., 2012. The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, 93(3), pp.1005–1020.
- Fox, G.E. et al., 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences of the United States of America*, 74(10), pp.4537–4541.
- Gillespie, D.E. et al., 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology*, 68(9), pp.4301–4306.
- Gleeson, D.F. et al., 2012. Biosignature Detection at an Arctic Analog to Europa. *Astrobiology*, 12(2), pp.135–150.
- González-Pastor, J.E. & Mirete, S., 2010. Novel Metal Resistance Genes from Microorganisms: A Functional Metagenomic Approach. In *Metagenomics*:

Methods and Protocols, Methods in Molecular Biology. Springer Science+Business Media, LLC, pp. 273–285.

- Guazzaroni, M.E. et al., 2013. Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environmental Microbiology*, 15(4), pp.1088–1102.
- Hanahan, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, 166(4), pp.557–580. Available at: <http://www.sciencedirect.com/science/article/pii/S0022283683802848> [Accessed January 7, 2015].
- Handelsman, J., 1998. Molecular biological access to the chemistry of unknown soil microbes. *Chemistry & Biology*, 5(10), pp.245–249.
- Harm, W., 1980. *Biological Effects of Ultraviolet Radiation*, Akademie-Verlag. Available at: <http://dx.doi.org/10.1002/jobm.19810210917>.
- Hung, R.-J. & Terman, J.R., 2011. Extracellular Inhibitors, Repellents, and Semaphorin/Plexin/ MICAL-mediated Actin Filament Disassembly. *Cytoskeleton (Hoboken)*, 68((8)), pp.415–433.
- Janion, C., 2008. Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. *International Journal of Biological Sciences*, 4(6), pp.338–344.
- Jeon, J.H. et al., 2009. Characterization and its Potential Application of Two Esterases Derived from the Arctic Sediment Metagenome. *Mar Biotechnology*, 11, pp.307–316.
- Kasting, J.F. et al., 2013. Remote life-detection criteria, habitable zone boundaries, and the frequency of Earth-like planets around M and late K stars. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35), pp.1–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24277805>.
- Khurana, K.K. et al., 1998. Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature*, 395(6704), pp.777–780. Available at: <http://dx.doi.org/10.1038/27394>.
- Krishna, S., Maslov, S. & Sneppen, K., 2007. UV-induced mutagenesis in *Escherichia coli* SOS response: A quantitative model. *PLoS Computational Biology*, 3(3), pp.0451–0462.
- Krokan, H.E. & Bjørn, M., 2013. Base Excision Repair. *Cold Spring Harb Perspect Biol*, 5, pp.1–22.
- Kumar, M., Gromiha, M.M. & Raghava, G.P.S., 2007. Identification of DNA-binding proteins using support vector machines and evolutionary profiles. *BMC bioinformatics*, 8, p.463.

- Kurth, D. et al., 2015. Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile *Acinetobacter* sp. Ver3. *Frontiers in Microbiology*, 6, pp.1–18. Available at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00328/abstract>.
- Langley, K.E. et al., 1975. Molecular basis of beta-galactosidase alpha-complementation. *Proceedings of the National Academy of Sciences of the United States of America*, 72(4), pp.1254–1257.
- Lee, J.J. et al., 2014. *Hymenobacter swuensis* sp. nov., a gamma-radiation-resistant bacteria isolated from mountain soil. *Current Microbiology*, 68(3), pp.305–310.
- Leuko, S. et al., 2014. On the Response of Halophilic Archaea to Space Conditions. *Life*, 4(1), pp.66–76. Available at: <http://www.mdpi.com/2075-1729/4/1/66/>.
- Lewin, A., Wentzel, A. & Valla, S., 2013. Metagenomics of microbial life in extreme temperature environments. *Current Opinion in Biotechnology*, 24(3), pp.516–525. Available at: <http://dx.doi.org/10.1016/j.copbio.2012.10.012>.
- Lin, W.-Z. et al., 2011. iDNA-Prot: Identification of DNA Binding Proteins Using Random Forest with Grey Model. *PLoS ONE*, 6(9), p.e24756.
- Liu, B. et al., 2014. iDNA-Prot|dis: Identifying DNA-Binding Proteins by Incorporating Amino Acid Distance-Pairs and Reduced Alphabet Profile into the General Pseudo Amino Acid Composition. *PLoS ONE*, 9(9), p.e106691. Available at: <http://dx.plos.org/10.1371/journal.pone.0106691>.
- Logette, E. et al., 2011. PIDD orchestrates translesion DNA synthesis in response to UV irradiation. *Cell death and differentiation*, 18(6), pp.1036–1045.
- Lorenz, R.D. et al., 2008. Titan's rotation reveals an internal ocean and changing zonal winds. *Science (New York, N.Y.)*, 319(5870), pp.1649–1651.
- Mancinelli, R.L. & Klovstad, M., 2000. Martian soil and UV radiation: microbial viability assessment on spacecraft surfaces. *Planetary and Space Science*, 48(11), pp.1093–1097.
- Des Marais, D.J. et al., 2008. The NASA Astrobiology Roadmap. *Astrobiology*, 8(4), pp.715–730.
- Marchler-Bauer, a. et al., 2014. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43(D1), pp.D222–D226. Available at: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1221>.
- Mastascusa, V. et al., 2014. Extremophiles Survival to Simulated Space Conditions: An Astrobiology Model Study. *Origins of Life and Evolution of Biospheres*, 44(3), pp.231–237. Available at: <http://link.springer.com/10.1007/s11084-014-9397-y>.

- McLean, J.S. & Lasken, R.S., 2014. Single cell genomics of bacterial pathogens: outlook for infectious disease research. *Genome Medicine*, 6(11), pp.1–3. Available at: <http://genomemedicine.com/content/6/11/108>.
- Mustard, J.F. et al., 2013. Appendices to the Report of the Mars 2020 Science Definition Team. *Mars Exploration Program Analysis Group (MEPAG)*, (CI), pp.155–205. Available at: http://mepag.jpl.nasa.gov/reports/MEP/Mars_2020_SDT_Report_Appendix.pdf.
- Nicholson, W.L. et al., 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and molecular biology reviews: MMBR*, 64(3), pp.548–572.
- Ochi, T., Wu, Q. & Blundell, T.L., 2014. The spatial organization of non-homologous end joining: From bridging to end joining. *DNA Repair*, 17, pp.98–109. Available at: <http://dx.doi.org/10.1016/j.dnarep.2014.02.010>.
- Ordoñez, O.F. et al., 2009. Extremophile culture collection from Andean lakes: Extreme pristine environments that host a wide diversity of microorganisms with tolerance to UV radiation. *Microbial Ecology*, 58(3), pp.461–473.
- Patel, M. et al., 2010. A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. *Critical reviews in biochemistry and molecular biology*, 45(3), pp.171–184.
- Paterlini, M., 2007. There shall be order. The legacy of Linnaeus in the age of molecular biology. *EMBO reports*, 8(9), pp.814–816.
- Pearce, D.A., 2012. Extremophiles in Antarctica: life at low temperatures. In H. Stan-Lotter & S. Fendrihan, eds. *Adaptation of Microbial Life to Environmental Extremes*. SpringerWienNewYork, pp. 87–118. Available at: http://ezproxy.upm.edu.my:2135/chapter/10.1007/978-3-211-99691-1_5.
- Pikuta, E. V, Hoover, R.B. & Tang, J., 2007. Microbial extremophiles at the limits of life. *Critical reviews in microbiology*, 33(3), pp.183–209.
- Raghunathan, A. et al., 2005. Genomic DNA Amplification from a Single Bacterium. *Applied and environmental microbiology*, 71(6), pp.3342–3347.
- Rastogi, R.P. et al., 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of nucleic acids*, 2010, p.592980.
- Santos, A.L. et al., 2012. Contribution of reactive oxygen species to UV-B-induced damage in bacteria. *Journal of Photochemistry and Photobiology B: Biology*, 117, pp.40–46. Available at: <http://dx.doi.org/10.1016/j.jphotobiol.2012.08.016>.

- Shah, P. & He, Y.-Y., 2015. Molecular Regulation of UV-Induced DNA Repair. *Photochemistry and Photobiology*, p.n/a–n/a. Available at: <http://doi.wiley.com/10.1111/php.12406>.
- Shestakov, S. V., 2012. Impact of metagenomics on biotechnological development. *Applied Biochemistry and Microbiology*, 48(9), pp.705–715. Available at: <http://link.springer.com/10.1134/S0003683812090050> \nhttp://download.springer.com/static/pdf/361/art:10.1134/S0003683812090050.pdf?auth66=1400114472_48df4767a5a8aafbac4db98a1a5ff220&ext=.pdf.
- Simon, C. & Daniel, R., 2011. Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4), pp.1153–1161.
- Singh, O. V. & Gabani, P., 2011. Extremophiles: Radiation resistance microbial reserves and therapeutic implications. *Journal of Applied Microbiology*, 110(4), pp.851–861.
- Singh, V. et al., 2014. Direct observation of preferential processing of clustered abasic DNA damages with APE1 in TATA box and CpG island by reaction kinetics and fluorescence dynamics. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 766-767, pp.56–65. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0027510714001109>.
- Sinha, R.P. & Häder, D.P., 2002. UV-induced DNA damage and repair: a review. *Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology*, 1(4), pp.225–236.
- Sloney, D.H., 2007. Radiometric quantities and units used in photobiology and photochemistry: recommendations of the Commission Internationale de L'Eclairage (International Commission on Illumination). *Photochemistry and photobiology*, 83(2), pp.425–432.
- Staley, J.T. & Konopka, a, 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology*, 39, pp.321–346.
- Sutton, M.D. & Walker, G.C., 2001. Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp.8342–8349.
- Szilágyi, A. & Skolnick, J., 2006. Efficient Prediction of Nucleic Acid Binding Function from Low-resolution Protein Structures. *Journal of Molecular Biology*, 358(3), pp.922–933.
- Tirard, S., 2010. Origin of life and definition of life, from Buffon to Oparin. *Origins of Life and Evolution of Biospheres*, 40(2), pp.215–220.

- Uchiyama, T. & Miyazaki, K., 2013. Metagenomic Screening for Aromatic Compound-Responsive Transcriptional Regulators. *PLoS ONE*, 8(9), pp.1–8.
- Wang, Y. et al., 2014. FLP-FRT-based method for unmarked deletions of CHU_3237 (porU) and large genomic fragments of *Cytophaga hutchinsonii*. *Applied and environmental microbiology*, 80(July), pp.6037–6045. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25063660>.
- Williamson, L.L. et al., 2005. Intracellular Screen To Identify Metagenomic Clones That Induce or Inhibit a Quorum-Sensing Biosensor. *Applied and Environmental Microbiology*, 71(10), pp.6335–6344.
- Wilmes, P. & Bond, P.L., 2006. Metaproteomics: Studying functional gene expression in microbial ecosystems. *Trends in Microbiology*, 14(2), pp.92–97.
- Woese, C.R., 1987. Bacterial Evolution Background. *Microbiology*, 51(2), pp.221–271. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC373105/pdf/microrev00049-0051.pdf>.
- Woese, C.R., Kandler, O. & Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12), pp.4576–4579.
- Worth, R.J., Sigurdsson, S. & House, C.H., 2013. Seeding life on the moons of the outer planets via lithopanspermia. *Astrobiology*, 13(12), pp.1155–65. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3870607&tool=pmcentrez&rendertype=abstract>.

Appendix

1. Protein sequences of the identified ORFs

>pML5_ORF1_ORF:1..1059 Frame -3

ATGGACGACAACCGTAAAAAGGCGCTGGGCGCTGCACTGTACAGATTGAGCGGCAATTCGGGAAGGGCG
 CCATCATGAGGATGGGCGATCAGCCCCGGGAAGCCATGCCGGCGGTATCCACGGCTCACTGGGACTGGA
 TGTGGCCCTGGGGATTGGCGGACTGCCACGGGTCGGGTGGTTGAGATATTTGGTCCGGAGAGTTCTGGC
 AAGACCACCTGACGCTTGAAGTCATCGCGCAGGCACAGCGTGTGGCAAGACCTGTGCCTTTGTGGATG
 CGGAGCACGCACTGGACCCGAATTACGCTGAGAAGCTTGGGGTCGATGTGAATGAGCTGCTGGTGTCA
 GCCGGATAACCGGTGAACAGGCTCTGGAGATCTGCGACATGCTGGTGCCTCCGGCGCTGTGGACGTGGTC
 ATTGTGGACTCGGTTGCGGCACTGACACCCAGGCGGAGATTGAGGGCGAAATGGGCGACTCCCATGTGC
 GCCTCCAGGCCCGTCTCATGTCCCAGGCGCTGCGCAAGCTTACGGCCAGTGTAAAACCTCAGGCACCAT
 GCTGGTGTTCATCAACCAGATCCGGATGAAAATCGGTGTGATGTTCCGTAACCCGAAACCACAACCGGC
 GGTAACGCGCTCAAGTTCTACTCCTCGGTGCGCCTGGACATACGCCGCATCGGTTCCGTTAAGGAAGGTG
 ATGAAGCGGTGGGCAATGAGACCCGCGTGAAAGTGGTCAAGAACAAGGTGGCACCGCCTTCCGTCAGGC
 GGAATTCAGATCATGTACGGCACCGGTATCAACCGGCTGGCCGAGGTTATTGATATGGGTGTGAAAGAA
 GGCTTTGTTGAAAGTCCGGCGCCTGGTACAGCTACAACGGTGACAAGATCGGACAGGGCAAGGCCAATGC
 CAGCAAGTTCCTGGCGGATAATCCCAGAAATGGCAGCGGAAATGAAGGCAAGATCCGCGACAAGCTCATG
 CCCAAACCGGGGGCGGCCAGTGAGGGCGATGGCAAGGCTGCGGACGGCAGTGAGGCCGAGGAAACCAGCG
 AGGATCTGA

>pML6_ORF1 ORF:4..480 Frame +1

MNIFAVSSNPNECARALDDQRLNKMI IETGQLLSTALYYWNEPEYNQVYRRTHDNHPVNKQVRENVNHF
 WTFHLFME LITERQFRDTHNKTENLVQPFLNVVQRHGVMLPDTPIVIFPELQFYKSLPVCEAYRWTLIDK
 WNSDVRPSWTRRGPPPEWL*

>pML6_ORF2 ORF:538..873 Frame +1

MFARHGLDVDRLNGYENGAQGLCAFPRGYPMGLYRVVSLKGTQRHTATEESIMNVREMPNQCIIIVLGDF
 PDSPDVPGGPKFSPDWDKKRTKIERRNWERDIDRFLSEKIH*

>pML56_ORF1 ORF:3..359 Frame -1

MFNDAITTESVLLDNDTDGESGVVDLSLTESGDITGDVTGDVEPSETVEESFGQVTGSQPPQFDEFHGET
 RDGVPVTLEKVIITSTEFDGRDGTTRVTGITPSVVVFDSDAPDNAFVGEDV

>pML56_ORF2 ORF:564..776 Frame -1

MNLNALAIGLGLQVEYEPEQFPGLIYRPESAEGVLLFSNGRVVITGCQSIDAAEKIFSRLIETVSDLP
 *

>pML84_ORF1 ORF:2..838 Frame -2
MLMKLAILSYILCFLLLPVQQLSEDFESGSKSGYAAGDVELESGSWHFHEALIGDDDRDRTNGTQSA
RLDNRDDNGYIRMNFDKSDGADEISFYAANADFSGDDGGALQVAYSTDEGSSWNELGDEIQLDDEFTEYL
LEARIGGNIRFEIRTTDGGRVSVDDVTVDYVEAADEPTIQVRLDNSTLESGYNYEFSSIVSGDSRTTVF
EIRNQGNEPLSVTDIETSGDDVFSVMDIPDEELEFGESEADFNVNFEPETIDEFEGSVTIHSNADNIPEF

>pML105_ORF1 ORF:5..574 Frame +2
MNIFYLDHDPRECARMHCDQHVRKMMLEYAQILSTAIHLKLFKRNHWGKIYKPTHQNHPSVRWAAETLAQ
YDWLYLLWVNLHDEYIYRFGKNHKSFVDLNQYLSDAPFYYNVDGVMFNPPPQVMPEEYKKPGFPVIGYRN
YYIHEKSRFATWTRREPPMWFLQGKKNKYKLPEDIRIEGNAPASRRAG*

>pML105_ORF2 ORF:507..1208 Frame +3
MDTNYLRTFELKEMLQHPDARDEDGMLLLAAKLRELPEENRARFYFDRYCSLFEQGLDRYLSADELLES
NIELAEDTRLINKEIRAI IAAAGWYAYFNTXGKMKRTDWLTDDEVKNHRVLLYIEDVRNALSEVFPREN
LLGYFVQLIGTERARYFTPKEISLRKLPYPPEEHKQLGIIVDYVLVIEVESSDEGGGIYAVSMEHLNLSLE
EIEWSIPADKNTLXGGGKHQRPK*

>pML105_ORF3 ORF:1132..1404 Frame +1
MKKSNGQSPQIKTPXVVVENTNAPSEANLPYFITS LGVTSNPQVHEISGELDSSLVEQYTYPSEHEYLEI
YLRKRTLLSDLKESEVILKGDP