



EVOLUTION OF PSEUDOMONAS BALEARICA DSM 6083T PROMOTER FROM RANDOM SEQUENCES

By

Sharlene Usman (G1625552T) – BBTD1 1711A

University ID.: 16045266

Supervised By

Dr. Maurice Ling

A Final Year Project Report Submitted in Partial Fulfillment of the  
 Requirements for the Degree of  
Bachelor of Science (Hons) in Biotechnology

University of Northumbria & Management Development Institute of Singapore

Faculty of Health and Life Sciences

Academic Year 2019/2020

**Table of Contents**

Abstract

Acknowledgement

It is a genuine pleasure to express my sincere gratitude and thankfulness upon the successful completion of this final year project report with the title of “….” This project report would not be able to accomplish without the special people who guided, assisted, challenged, and supported along the circuitous way.

First and foremost, I would like to express the deepest appreciation to my parents for the continuous encouragement, support, inspiration, and strength. Thank you for believing and keep motivating me to achieve good grades throughout this academic journey.

Second, I owe a deep sense of gratitude for the opportunity of having Dr. Maurice Ling as my project supervisor, thank you for the endless dedications, advices, feedbacks, supports, and references, and always be patient in guiding and assisting us throughout every stage of the project report. His prompt enthusiasm and timely suggestions with kindness have empowered me to complete my project journey.

In addition, it is my privilege to thank all of MDIS lecturers for the kind supports and advices whenever I was having a difficult time throughout the degree course. Last but not least, a big applause to my colleagues who stuck with me along the way, all of my friends, and relatives for the consent and the unending encouragement.

**Chapter 1: Introduction**

* 1. **Origins of Functional DNA**

For decades, the genes with no detectable homology that emerge from previously noncoding genomic regions had been considered once as an extremely implausible event (Jacob, 1977). Instead, gene fusion and fission, gene duplication and divergence, and exon shuffling were believed to be the merely possible sources of *de novo* genes (McLysaght and Guerzoni, 2015). However, as the availability of the full genomic sequence of yeast (Dujon, 1996) came into this picture, it concludes that the *de novo* evolution of complex traits can be considered as possible (Jacob, 1977). About one third of the complete set of yeast genes has no similarity in the genes sequence to the genes from other organisms (Dujon, 1996). There were no identified ancestors from these new orphan genes (or ORFans in the microbial world) (Fischer and Eisenberg, 1999). Therefore, a combination of genetic changes is required before a beneficial function can be obtained from a *de novo* evolution (Jacob, 1977). The evolutionary path in such cases cannot be negligible, as a trivial selective advantage of first mutations may prevent them from spreading in the population. The probability of achieving multiple desired mutations simultaneously is very small, especially in asexual populations such as bacteria. This organism is incompetent to combine mutations that were required in different individuals (Dujon, 1996).

Indeed, *de novo* evolution of a new promoter are not clearly understood (Yona, Alm and Gore, 2018). However, further studies in evolutionary step of promoters that have been done by genomic analysis (Manson and Church, 2000) and experimental interactions between protein and DNA achieved potential results (Cho *et al.*, 2009). The evolutionary-refined promoters; such as, wild-type promoters and their derivatives; have built mostly the promoter libraries (Kinney *et al.*, 2010). The extensive studies in the activation (Ochman, Lawrence and Groisman, 2000) or inactivation of new genes (Rutledge and Challis, 2015) in cells suggested that *de novo* promoters may not be accessible evolutionary if activation of the genes was done by copying the existing promoters. Copying the existing promoters is predominant in evolution presumably because multiple mutations are required; hence, longer time are required to obtain desired promoters (Yona, Alm and Gore, 2018).

* 1. **Recent and Current Status of *De Novo* Evolution**

In the latest years, some evidences regarding de novo gene origination started to accumulate (McLysaght and Guerzoni, 2015). It was believed that novel peptide sequences were intensely related to the pre-existing gene, mostly being an extension of coding sequence into an intron, or, more fundamentally, translating all or part of a sequence in an alternate reading frame of the mRNA in so-called ‘overlapping’ or ‘overprinting’ gene (Keese and Gibbs, 1992). Until quite recently, it has now proved that the de novo evolution of genes are also possible from non-coding DNA. It is also known as one of the consistent features of eukaryotic genomes, such as yeast, plants, mammals, primates, flies, and even in the human evolution (McLysaght and Guerzoni, 2015).

In 2006, Begun and colleagues’ works contributed evidences for *de novo* genes in *Drosophila yakuba* and *D. erecta* accessory gland (Begun *et al.*, 2005; Levine *et al.*, 2006). Later, in 2008, Cai *et al.* finally presented the first functional characterization of a budding yeast gene by showing that *BSC4* in *Saccharomyces cerevisiae* has a DNA repair function and is a synthetic lethal (Cai et al., 2008). Several eukaryotes *de novo* genes such as flies and mammals’ genes have been showed that selection or mutation could contribute to the *de novo* genes even in the absence of precise functional annotation. This could be a definite sign that they are contributing directly to fitness (Zhao et al., 2014). Several studies and experiments regarding the *de novo* gene origination are continuously arising as a phenomenon in recent prokaryotic and eukaryotic evolution (Guerzoni and McLysaght, 2011), and yet, the extent of its influences are still remained to be discovered.

* 1. ***Pseudomonas balearica***

*Pseudomonas balearica* which was isolated from the Mediterranean Sea (Ramos, 2004), is the species of true marine denitrifying microorganism (Bennasar-Figueras *et al.*, 2016). *P. balearica* is a Gram-negative, short (0.3 – 0.5 pm wide) and straight (1.5 – 3.0 pm long), rod-shaped, and nonfluorescent bacterium. It is an environmental microorganism that belongs to the authentic genus *Pseudomonas* in the gamma subclass of the *Proteobacteria* (Stackebrandt, Murray and Truper, 1988). This organism is motile with a single polar flagellum, strictly oxidative, and widely distributed in polluted environments all over the world (Bennasar *et al.*, 1996). Hence, it acts as a vigorous denitrifier that have the capabilities to degrade copious amounts of nitrogen gas from nitrate. Based on 16s rRNA analysis (Rossello *et al.*, 1991), *P. balearica* is classified as a new species of the *P. stutzeri* group (81.2% with *P. stutzeri* ATCC 17588T) (Bennasar-Figueras *et al.*, 2016).

*P. balearica* shares some similar phenotypic and morphological characteristics as *P.stutzeri* including amylase and maltose positive, arginine dihydrolase and gelatinase negative, and a wrinkly dry colony morphology after a fresh isolation. Some distinct characteristics that differentiated *P. balearica* from *P. stutzeri* including its ability to grow at 46, physiological tolerance with the presence of 8.5% NaCl, and capability to make use of xylose, ethylene glycol, mannitol, suberate, and 4-aminobutyrate as sole carbons and energy sources. The genome size varies between 4.2 and 4.4 Mb with a G+C content that ranges from 64.1 to 64.4 mol% as identified using high-performance liquid chromatography. The type strain of *P. balearica* is SP1402 which is also known as DSM 6083 (Rossello *et al.*, 1991).

As a novel species (Bennasar *et al.*, 1996), *P. balearica* has several strain types including strains LS401 (CCUG 66666) (Rossello *et al.*, 1991), ST101 (CCUG 66667) (Daane *et al.*, 2001), and SP1402 (CCUG 44595) (Rossello *et al.*, 1991). *P. balearica* strain LS401 was isolated as a naphthalene degrader from a polluted marine sediment in Barcelona, Spain (Rossello *et al.*, 1991), while, *P. balearica* strain st101 was isolated as a naphthalene as well as phenanthrene degrader from a Spartina patens rhizosphere sample at an oil refinery site in New York, United States of America (Daane *et al.*, 2001). In addition, strains SP1402 was isolated as a robust denitrifier from the wastewater treatment plant by enrichment with 2-methylnaphthalene at 40 (Bennasar-Figueras *et al.*, 2016). Furthermore, other strains of *P. balearica* were also obtained from marine samples (Rossello *et al.*, 1991) and salt marshes (Mulet *et al.*, 2008).

As *P. balearica* is well known for its biodegradation capabilities and ecological implications (Salvà-Serra *et al.*, 2017), a paper by Zerrad *et al.* also suggested that a marine microbe such as P. balearica may offer great opportunities for biodiscovery (Zerrad *et al.*, 2014) and provide a new resource for structurally diverse secondary metabolites such as melanin (Glöckner and Joint, 2010). Melanins are amorphous black pigments (Nosanchuk and Casadevall, 2006) formed by oxidative polymerization of phenolic compounds within animals, bacteria, plants, and fungi (Riley, 1997). The study reported that the *P. balearica* strain U7 which was isolated from the marine alga Ulva lactuca, produces high level of melanin using L-tyrosine as precursor in a minimal medium (Zerrad *et al.*, 2014). Therefore, the melanin-producing *P. balearica* strain U7 have a very effective radical scavenging capacity (Sichel *et al.*, 1991) to enhance microorganisms’ ability to survive in unfavorable environmental conditions, such as UV radiation (Sarna, Menon and Sealy, 1984) and gamma rays in cancer therapy (Kunwar *et al.*, 2012). They were also applied in cosmetics productions for protection against the noxious effects of UV radiation, and in bio-electric sciences development as semiconductors (Mostert *et al.*, 2012).

* 1. **Reclassification of *Pseudomonas balearica***

During the last few decades, several studies have been performed to demonstrate that *P. stutzeri*, as one of the most controversial bacterium species, contains a heterogenous set of strains that could be categorized in more than one species (Mandel, 1966; Palleroni *et al.*, 1973). Despite of the extreme phenotypic diversity of the presumptive species (Mandel, 1966; Stanier, Palleroni and Doudoroff, 1966; Palleroni *et al.*, 1973), the bacterial phylogenetic relationships were successfully estimated by utilizing the small-subunit rRNA, 16s rRNA in a journal article by Bennasar and colleagues. The 16s rRNA gene sequences of 14 strains represented the seven *P. stutzeri* genomovars were compared in order to clarify the taxonomic positions and the phylogenetic relationships (Bennasar *et al.*, 1996). As *P. stutzeri* has been reported to have at least seven genomic groups without taxonomic status, it has been known as genomovars (Rossello *et al.*, 1991). The sequence analysis was achieved by using conserved primary sequence and secondary-structure characteristics as references (Woese *et al.*, 1983). The results revealed signature nucleotide positions for each *P. stutzeri* genomovar. They were confirmed that strain ZoBell (= ATCC 14405) is a member of genomovar 2 and identified that genomovar 6 strain SP1402T (T= type of strain) is sufficiently different from the other type strain of *P. stutzeri*. Therefore, this organism should be considered as a member of a new species, Pseudomonas balearica SP1402 (=DSM 6083) (Bennasar *et al.*, 1996).

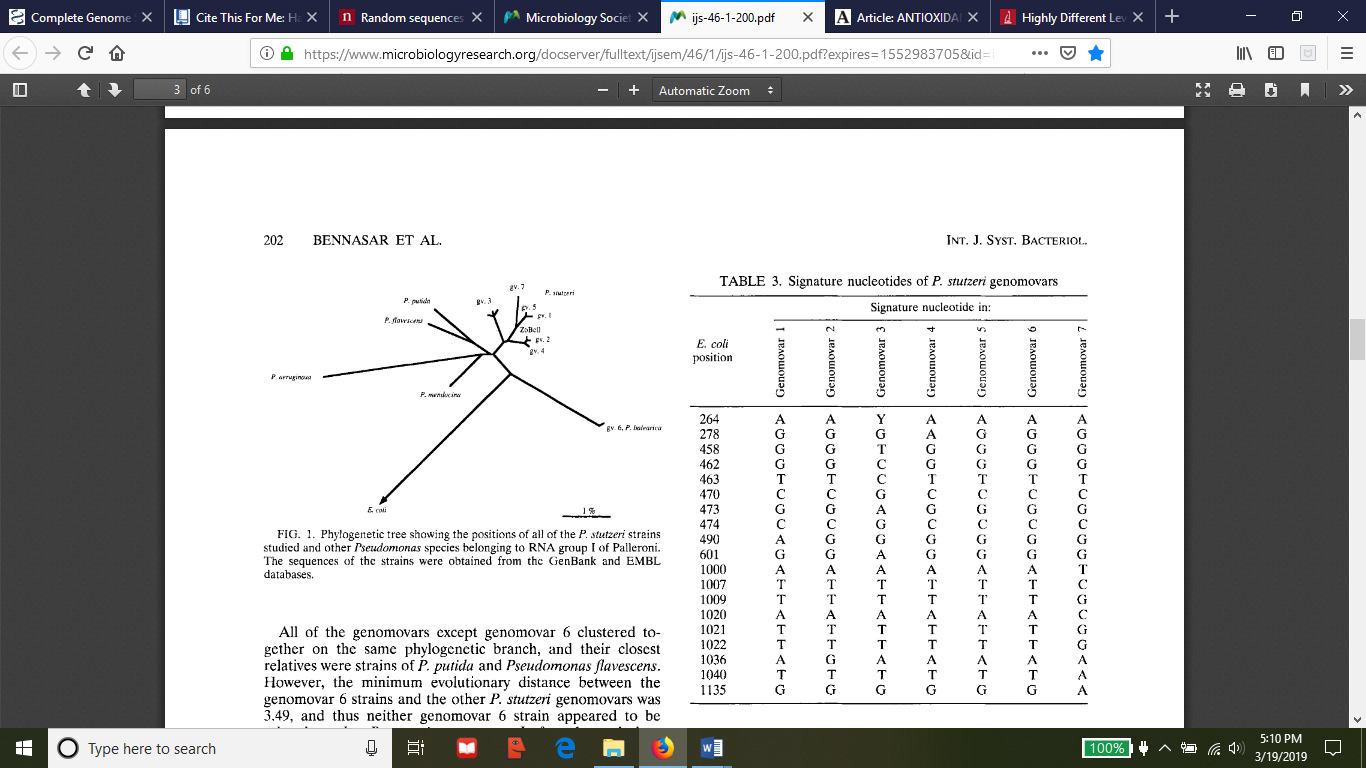


Figure 1. Phylogenetic tree showing the positions of all *P. stutzeri* strains studied and other *Pseudomonas* species belonging to RNA group I of Palleroni. The sequences of the strains were obtained from the GenBank and EMBL databases (Bennasar *et al.*, 1996).

* 1. **Possible Origins of Functional Promoters**

Ever since the pre-molecular era, how functional promoters, or even ribosome binding sites and first peptides arise randomly from scratch has been considered as a fundamental question in the evolution biology. Under this question, how many mutations are required to create a functional promoter ultimately from a specific length of random sequences? A study by Kaessmann reported that the birth of new protein-coding and RNA genes can evolve from “scratch” or nonfunctional genomic sequences, various types of gene fusion, and even RNA intermediates (Kaessmann, 2010). This study is supported by articles from Schlötterer and Andersson, Jerlström-Hultqvist, and Näsvall which emphasized that a pre-existing gene can be emerged by modification of the existing gene by divergence or evolved de novo from noncoding DNA (Andersson, Jerlström-Hultqvist and Näsvall, 2015; Schlötterer, 2015).

Decades ago, an article by Horwitz and Loeb reported a technique to create several unusual promoter recognition sequences. they found out that these promoter recognition sequences aid in the transcription and replacement of the promoter recognition sequences will produce promoter substitutions (Horwitz and Loeb, 1986). These techniques are further applied in a study by Yona, Alm, and Gore, which showed that evolution of E. coli promoters can be achieved by replacing the lac promoter with several random sequences of the same length (~100 bp) in the presence of lactose. The results revealed that ~10% of random DNA sequences can emerge into functional promoters without evolution, while, another ~60% of random sequences require at least one mutation to evolve into active promoters (Yona, Alm and Gore, 2018).

Studying the de novo evolution of promoters systematically is ought to start from the non-functional sequences. The concept of random sequences lies on the sequences that contains of Adenyl, Cytosine, Guanosine, and Thymine in equal probabilities composed of no information and represented the non-active sequence space without prejudices. De novo evolution of promoters is prevalently using purely random sequences with genomes preferably contain ~50% GC content, such as the *E. coli* genome with 50.8% GC content. Such genomes are advantageous because random sequences serve as a null model in the functionality test without proposing any perplexing factors due to diverging from the natural GC content of the studied genome (Yona, Alm and Gore, 2018). Therefore, the promoter of *P. balearica* DSM 6083T was considered as a worthy subject to investigate further since it contains 64.1 to 64.4% GC content (Rossello et al., 1991). Promoter is a 10-100 base pairs long region of DNA that can be located either on the upstream DNA (near the 3’ of the anti-sense strand) or towards the transcription start sites of genes (Eckstein and Lilley, 1997). The *P. balearica* promoter represents a complex sequence feature that consists of different elements which perform collectively to initiate transcription of a particular gene (Yona, Alm and Gore, 2018).

* 1. **Aims and Objectives**

In experimental and quantitative terms, a remarkable question of how many mutations are required to successfully change random sequences into a functional promoter is not yet answered. As the study of de novo evolution of a promoter is not fully understood and discovered (Yona, Alm and Gore, 2018), this project is aimed to investigate further on how a functional promoter *P. balearica* DSM 6083T can emerge *de novo* from generated random sequences. specifically, this project aims:

* To analyze the sequence variations in *P. balearica* DSM 6083T promoter sequences.
* To examine the possibility of generating randomly sequences for *P. balearica* DSM 6083T promoter activity.
* To analyze and compare the possible evolutionary results of *P. balearica* DSM 6083T after performing silico mutations and computational-based evolutionary experiments.

**Chapter 2: Materials and Methods**

* 1. **Strains**

The first version of the complete genome sequences of *P. balearica* DSM 6083T has been annotated successfully in DDBJ/ENA/GenBank under the accession number CP007511. This naphthalene degrader was isolated originally from the water of a lagooning wastewater treatment plant within the condition of rich 2-methylnaphthalene at 40 (Bennasar-Figueras *et al.*, 2016). The strain promoters have been predicted by using an analysis tool in Berkeley Drosophila Genome Project. The promoter predictor by M. G. Reese, a neural network-based program, helps to locate the possible transcription promoters. The whole 4,383,480 bp DNA sequences of *P. balearica* was inserted in the promoter predictor separately as the capacity of the program is only 100,000 bp. The program was set with prokaryote as the type of organism and the minimum promoter score is 0.8.

**References**

Andersson, D., Jerlström-Hultqvist, J. and Näsvall, J. (2015). Evolution of New Functions De Novo and from Preexisting Genes. *Cold Spring Harbor Perspectives in Biology*, 7(6), p.a017996.

Begun, D., Lindfors, H., Thompson, M. and Holloway, A. (2005). Recently Evolved Genes Identified From Drosophila yakuba and D. erecta Accessory Gland Expressed Sequence Tags. *Genetics*, 172(3), pp.1675-1681.

Bennasar, A., Rossello-mora, R., Lalucat, J. and Moore, E. (1996). 16S rRNA Gene Sequence Analysis Relative to Genomovars of Pseudomonas stutzeri and Proposal of Pseudomonas balearica sp. nov. *International Journal of Systematic Bacteriology*, 46(1), pp.200-205.

Bennasar-Figueras, A., Salvà-Serra, F., Jaén-Luchoro, D., Seguí, C., Aliaga, F., Busquets, A., Gomila, M., Moore, E. and Lalucat, J. (2016). Complete Genome Sequence of Pseudomonas balearica DSM 6083 T. *Genome Announcements*, 4(2), pp.217.

Cai, J., Zhao, R., Jiang, H. and Wang, W. (2008). De Novo Origination of a New Protein-Coding Gene in Saccharomyces cerevisiae. *Genetics*, 179(1), pp.487-496.

Cho, B., Zengler, K., Qiu, Y., Park, Y., Knight, E., Barrett, C., Gao, Y. and Palsson, B. (2009). The transcription unit architecture of the Escherichia coli genome. *Nature Biotechnology*, 27(11), pp.1043-1049.

Daane, L., Harjono, I., Zylstra, G. and Haggblom, M. (2001). Isolation and Characterization of Polycyclic Aromatic Hydrocarbon-Degrading Bacteria Associated with the Rhizosphere of Salt Marsh Plants. *Applied and Environmental Microbiology*, 67(6), pp.2683-2691.

Dujon, B. (1996). The yeast genome project: what did we learn? *Trends Genet*, 12(7), pp.263–270.

Eckstein, F. and Lilley, D. (1997). *Mechanisms of Transcription*. Berlin, Heidelberg: Springer.

Fischer, D. and Eisenberg, D. (1999). Finding families for genomic ORFans. *Bioinformatics*, 15(9), pp.759–762.

Glöckner, F. and Joint, I. (2010). Marine microbial genomics in Europe: current status and perspectives. *Microbial Biotechnology*, 3(5), pp.523-530.

Guerzoni, D. and McLysaght, A. (2011). De Novo Origins of Human Genes. *PLoS Genetics*, 7(11), p.e1002381.

Horwitz, M. and Loeb, L. (1986). Promoters selected from random DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 83(19), pp.7405-7409.

Jacob, F. (1977). Evolution and tinkering. *Science*, 196 (4295), pp.1161–1166.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Research*, 20(10), pp.1313-1326.

Keese, P. and Gibbs, A. (1992). Origins of genes: "big bang" or continuous creation?. *Proceedings of the National Academy of Sciences*, 89(20), pp.9489-9493.

Kinney, J., Murugan, A., Callan, C. and Cox, E. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences*, 107(20), pp.9158-9163.

Kunwar, A., Adhikary, B., Jayakumar, S., Barik, A., Chattopadhyay, S., Raghukumar, S. and Priyadarsini, K. (2012). Melanin, a promising radioprotector: Mechanisms of actions in a mice model. *Toxicology and Applied Pharmacology*, 264(2), pp.202-211.

Levine, M., Jones, C., Kern, A., Lindfors, H. and Begun, D. (2006). Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. *Proceedings of the National Academy of Sciences*, 103(26), pp.9935-9939.

Mandel, M. (1966). Deoxyribonucleic Acid Base Composition in the Genus Pseudomonas. *Journal of General Microbiology*, 43(2), pp.273-292.

Manson, A. and Church, G. (2000). Predicting regulons and their cis-regulatory motifs by comparative genomics. *Nucleic acids research*, 28(22), pp.4523-4530.

McLysaght, A. and Guerzoni, D. (2015). New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1678), p.20140332.

Mostert, A., Powell, B., Pratt, F., Hanson, G., Sarna, T., Gentle, I. and Meredith, P. (2012). Role of semiconductivity and ion transport in the electrical conduction of melanin. *Proceedings of the National Academy of Sciences*, 109(23), pp.8943-8947.

Mulet, M., Gomila, M., Gruffaz, C., Meyer, J., Palleroni, N., Lalucat, J. and Garcia-Valdes, E. (2008). Phylogenetic analysis and siderotyping as useful tools in the taxonomy of Pseudomonas stutzeri: description of a novel genomovar. *International Journal of Systematic and Evolutionary Microbiology*, 58(10), pp.2309-2315.

Nosanchuk, J. and Casadevall, A. (2006). Impact of Melanin on Microbial Virulence and Clinical Resistance to Antimicrobial Compounds. *Antimicrobial Agents and Chemotherapy*, 50(11), pp.3519-3528.

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

Palleroni, N., Kunisawa, R., Contopoulou, R. and Doudoroff, M. (1973). Nucleic Acid Homologies in the Genus Pseudomonas. *International Journal of Systematic Bacteriology*, 23(4), pp.333-339.

Ramos, J. (2004). *Pseudomonas*. 3rd ed. Boston: Kluwer Academic, pp.195-196.

Riley, P. (1997). Melanin. *The International Journal of Biochemistry & Cell Biology*, 29(11), pp.1235-1239.

Rossello, R., Garcia-Valdes, E., Lalucat, J. and Ursing, J. (1991). Genotypic and Phenotypic Diversity of Pseudomonas stutzeri. *Systematic and Applied Microbiology*, 14(2), pp.150-157.

Rutledge, P. and Challis, G. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, 13(8), pp.509-523.

Salvà-Serra, F., Jakobsson, H., Busquets, A., Gomila, M., Jaén-Luchoro, D., Seguí, C., Aliaga-Lozano, F., García-Valdés, E., Lalucat, J., Moore, E. and Bennasar-Figueras, A. (2017). Genome Sequences of Two Naphthalene-Degrading Strains of Pseudomonas balearica, Isolated from Polluted Marine Sediment and from an Oil Refinery Site. *Genome Announcements*, 5(14).

Sarna, T., Menon, I. and Sealy, R. (1984). Photoinduced oxygen consumption in melanin systems-II, action spectra, and quantum yields for pheomelanins. *Photochemistry and Photobiology*, 39(1), pp.805-809.

Schlötterer, C. (2015). Genes from scratch – the evolutionary fate of de novo genes. *Trends in Genetics*, 31(4), pp.215-219.

Sichel, G., Corsaro, C., Scalia, M., Di-Bilio, A. and Bonomo, R. (1991). In vitro scavenger activity of some flavonoids and melanins against O2−dot. *Free Radical Biology and Medicine*, 11(1), pp.1-8.

Stackebrandt, E., Murray, R. and Truper, H. (1988). Proteobacteria classis nov., a Name for the Phylogenetic Taxon That Includes the "Purple Bacteria and Their Relatives." *International Journal of Systematic Bacteriology*, 38(3), pp.321-325.

Stanier, R., Palleroni, N. and Doudoroff, M. (1966). The Aerobic Pseudomonads a Taxonomic Study. *Journal of General Microbiology*, 43(2), pp.159-271.

Woese, C., Gutell, R., Gupta, R. and Noller, H. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological reviews*, 47(4), pp.621-69.

Yona, A., Alm, E. and Gore, J. (2018). Random sequences rapidly evolve into de novo promoters. *Nature Communications*, 9(1), pp.1-10.

Zerrad, A., Anissi, J., Ghanam, J., Sendide, K. and El-Hassouni, M. (2014). Antioxidant and antimicrobial activities of melanin produced by a Pseudomonas balearica strain. *Journal of Biotechnology Letters*, 5(1), pp.87-94.

Zhao, L., Saelao, P., Jones, C. and Begun, D. (2014). Origin and Spread of de Novo Genes in Drosophila melanogaster Populations. *Science*, 343(6172), pp.769-772.