Maria Leonor Veríssimo dos Santos

Licenciada em Biologia

A transcriptomic approach to the metabolism of porphyrin-like pigments in a marine Polychaeta (*Eulalia* sp.)

Dissertação para obtenção do Grau de Mestre em Genética Molecular e Biomedicina

Orientador: Prof. Doutor Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

Co-orientador: Prof. Doutor António Jorge Dias Parola, Professor Associado com Agregação, FCT/UNL

Júri:

Presidente: Prof.ª Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes, Professora Auxiliar, FCT/UNL

Arguente: Doutor Daniel Vieira Novo e Silva Sobral, Investigador, FCT/UNL

Vogal: Prof. Doutor Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro 2020

Maria Leonor Veríssimo dos Santos

Licenciada em Biologia

A transcriptomic approach to the metabolism of porphyrin-like pigments in a marine Polychaeta (*Eulalia* sp.)

Dissertação para obtenção do Grau de Mestre em Genética Molecular e Biomedicina

Orientador: Prof. Doutor Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

Co-orientador: Prof. Doutor António Jorge Dias Parola, Professor Associado com Agregação, FCT/UNL

Faculdade de Ciências e Tecnologia

Universidade Nova de Lisboa

Novembro 2020

A transcriptomic approach to the metabolism of porphyrin-like pigments in a marine Polychaeta (*Eulalia* sp.)

Agradecimentos

Ao Pedro Costa por toda a paciência para as inúmeras reuniões que tivemos ao longo deste ano, por todo o apoio dado ao longo da dissertação e por sempre me incentivar a não duvidar das minhas capacidades.

Ao Professor Jorge Parola pelo apoio prestado, e aos elementos da sua equipa, que me ajudaram durante o meu percurso no seu laboratório.

À equipa do SeaTox, Ana, Cátia, Carla, Mariaelena e Carolina que me ajudaram sempre que precisei e que tornaram a experiência de fazer parte de uma equipa de investigação a melhor possível.

À Márcia Grou por ter estado constantemente lá para mim ao longo destes dois anos de Mestrado, por todos os passeios, desabafos e crises existenciais que aguentou.

Ao André Gonçalves por estar sempre pronto para me fazer rir a qualquer altura, seja ela importuna ou não, e por todos os passeios e conversas sem fim.

Ao Duarte Seixas por me fazer rir constantemente, mesmo que a maior parte das vezes seja de mim própria, e por todos os interesses partilhados.

Aos meus amigos da FCUL, especialmente ao Mendes por todo o carinho e ao Afonso que me acompanhou desde o secundário até à faculdade e que sempre me apoiou durante estes anos todos.

Às amigas que a FCT me deu, Bruna e Carolina, por todas as conversas intermináveis, especialmente na reta final de entrega da tese.

Às minhas amigas, Joana, Bárbara, Sara e Catarina que me acompanham desde o secundário e que são amigas que vou levar para a vida toda.

Aos meus pais por serem o meu apoio principal, nos bons e maus momentos, por acreditarem em mim, e por ao longo da minha vida nunca demostrarem outra coisa que não orgulho em mim e em tudo o que já alcancei.

À minha irmã por todo o apoio e carinho nos dias mais complicados, e por estar sempre pronta para me fazer companhia.

À minha avó, que mesmo longe, apoia-me incondicionalmente e me deixa sempre bem-disposta independentemente de tudo.

Abstract

The last decade witnessed a growing interest on marine natural pigments for biotechnological and biomedical applications. One of the most abundant naturally occurring pigments are the tetrapyrroles which are prized targets due to their photodynamic properties. Their most notorious representatives are porphyrins. Animal porphyrins result from the breakdown of heme and are known as bile pigments, the best known of which are biliverdin and bilirubin. Because of their unique chemical structure, porphyrins can have several applications such as photosensitizers in photodynamic therapy as well as antioxidants or even antimicrobials. For such reasons, porphyrinoids are high-prized animal metabolites for biomedical research. Naturally, abundant sources of these compounds, particularly those offering a wide-variety of the compounds such as coastal marine invertebrates, yield high biotechnological potential. The Polychaeta *Eulalia* sp. is known for its bright green coloration provided by the multiple greenish and yellowish porphyrinoid pigments found in this worm, which turns this species into the perfect case study since the chemistry and biosynthetic process of heme-derived pigments in Polychaeta remains mainly unknown. The present study combined HPLC-DAD with a transcriptomic approach (RNA-Seq) on the main tissues of *Eulalia* sp. displaying pigmentation, the proboscis and epidermis, with the aim of understanding the diversity and origin of its complex pigmentation. The results showed that the endogenous pigments of this worm are seemingly heme-derived and have the necessary mechanism for conversion to bile pigments. Also, several protein variants of the heme biosynthetic pathway were found in the two organs, indicating the possible production of diverse heme-related products that can be then converted to products similar to biliverdin or bilirubin. The specific and common variants found in both organs can explain the similar and different pigmentation patterns between the proboscis and epidermis. Altogether, this species is indeed a prolific source of novel porphyrinoids.

Keywords: porphyrinoids; photosensitizers; heme; bile pigments; HPLC-DAD; RNA-Seq.

Resumo

Na última década surgiu um grande interesse por pigmentos de organismos marinhos devido às suas potenciais aplicações biotecnológicas. Uma das classes mais abundantes de pigmentos são os pigmentos tetrapirrólicos que são conhecidos pelas suas propriedades fotodinâmicas, sendo os seus maiores representantes as porfirinas. Nos animais, a maioria dos pigmentos porfirinóides resultam da degradação do hemo e são denominados de pigmentos biliares, como a biliverdina e a bilirrubina. Devido à estrutura única das porfirinas e das suas possíveis modificações, estes compostos podem ter várias aplicações, notoriamente como fotossensibilizadores em terapia fotodinâmica ou ainda como agentes antioxidantes ou antimicrobianos. Apesar de pouco estudados neste aspeto, os invertebrados marinhos apresentam uma grande variedade deste tipo de compostos e por isso têm um alto potencial biotecnológico. O poliqueta *Eulalia* sp. possui uma coloração características verde fornecida por vários pigmentos porfirinóides amarelos e verdes. Estes pigmentos tornam esta espécie o caso de estudo perfeito para analisar pigmentos porfirinóides derivados do hemo, especialmente visto que estes processos são pouco conhecidos em anelídeos. Este estudo combinou a técnica de HPLC-DAD com uma abordagem transcriptómica (utilizando RNA-Seq) para estudar os padrões pigmentares do proboscis e da epiderme desta minhoca. Segundo os resultados, os pigmentos são derivados do hemo e possuem o mecanismo necessário para a sua conversão em pigmentos biliares. Além disso, foram identificadas várias variantes de proteínas centrais na síntese do hemo nos dois órgãos. Estes resultados indicam que pode existir a produção de vários produtos semelhantes ao hemo que seguidamente são convertidos em produtos análogos à biliverdina e bilirrubina. Por último, foram encontradas variantes específicas para cada órgão, assim como variantes em comum nos dois órgãos, o que pode explicar as semelhanças e diferenças nos padrões pigmentares do proboscis e da epiderme. Estas conclusões confirmam que esta espécie é uma fonte de novos porfirinóides com potencial biotecnológico.

Palavras-chave: porfirinóides; fotossensibilizadores; hemo; pigmentos biliares; HPLC-DAD; RNA-Seq.

Index

Figure index

Table index

Abbreviation List

ABCB6 - ATP-binding cassette sub-family B member 6

ABCG2 - Broad substrate specificity ATP-binding cassette transporter

ACN - Acetonitrile

ALA - δ-aminolevulinic acid

ALAD - δ-aminolevulinic acid dehydratase

ALAS - δ-aminolevulinic acid synthase

BLAST - Basic Local Alignment Search Tool

BLVRA - Biliverdin reductase

CLPX - ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial

COX10 - Protoheme IX farnesyltransferase

COX15 - Cytochrome c oxidase assembly protein COX15 homolog

CPM - Counts per million

CPOX - Coproporphyrinogen oxidase

CPYP1A1 - Cytochrome P450 1A1

CPYP1A2 - Cytochrome P450 1A2

Ctrb1 - Peptidase S1 domain-containing protein

CYCS - Cytochrome c

cyp17a1 - Cytochrome P450 family 17 subfamily A member 1

CYP2C19 - Cytochrome P450 2C19

CYP2C9 - Cytochrome P450 2C9

CYP3A4 - Cytochrome P450 3A4

DEG - Differentially expressed gene

FC - Fold change

FDR - False discovery rate

FE^{2+} - Ferrous iron

FECH - Ferrochelatase

GAPDH - Glyceraldehyde-3-phosphate dehydrogenase

GEO - Gene Expression Omnibus

GSTA - Glutathione S-transferase A

HCL - Hydrochloric acid

HMB - Hydroxymethylbilane

HMBS - Hydroxymethylbilane synthase

HMOX - Heme oxygenase

HP - Haptoglobin

HPGDS - Hematopoietic prostaglandin D synthase

HPLC - High-performance liquid chromatography

LCB4 - Sphingoid long chain base kinase 4

MEGA - Molecular Evolutionary Genetics Analysis

MeOH - Methanol

ORF - Open reading frame

PBG - Porphobilinogen

PCR - Polymerase chain reaction

PPOX - Protoporphyrinogen oxidase

PRDX1 - Peroxiredoxin-1

ROS - Reactive oxygen species

RT-qPCR - Quantitative reverse transcription PCR

skpo-2 - ShK domain and Peroxidase domain containing protein

SRC - Proto-oncogene tyrosine-protein kinase Src

TDO2 - Tryptophan 2,3-dioxygenase

TPM - Transcripts per million

TSPO - Translocator protein

UGT2A3 - UDP-glucuronosyltransferase 2A3

UROD - Uroporphyrinogen decarboxylase

UROS - Uroporphyrinogen synthase

UV - Ultraviolet

1. Introduction

Natural pigments are normally metabolic products of biosynthetic or catalytic pathways. Over the last decade, marine natural pigments have gained attention in the biomedical field, as they can play central roles in cells as antioxidants and can even present antimicrobial and photoprotective properties (Bandaranayake, 2006; Manivasagan et al., 2018). Despite their acknowledged importance as a powerful source of natural pigments, there is still a lack of research on marine organisms, especially shallow-water invertebrates that are known for their bright colorations (Bandaranayake, 2006). A class of invertebrates that meets these criteria is the Polychaeta. However, biotechnology-oriented research of this class is still lagging behind other groups of invertebrates, aquatic or terrestrial, despite the growing evidence for a wide span of novel bioproducts (see Rodrigo & Costa, 2019, for a review).

One of the most abundant naturally occurring pigments are the tetrapyrroles (Milgrom, 1997). Tetrapyrroles are a class of compounds with four pyrrolic rings, generally arranged in a cyclic form. Their most well-known representatives are porphyrins and their analogs chlorins which include heme and chlorophyll, respectively (Moss, 1988). Not surprisingly, one of the first works on tetrapyrrolic pigments started with the Polychaeta (Echiura) *Bonellia viridis*. Its characteristic green pigment, now known as bonellin, is a chlorin-like pigment unrelated to chlorophyll for which photodynamic activity has been reported (Agius et al., 1979). Compounds with tetrapyrrolic structures such as chlorins and porphyrins are often photoactive and have been widely used as photosensitizers in photodynamic therapy (PDT) (Abrahamse & Hamblin, 2016). Photosensitizers are chemical entities that upon absorption of light react with oxygen and generate reactive oxygen species (ROS) such as singlet oxygen ($^1O_2$) (Berg et al., 2005). Therefore, PDT uses photosensitizers for targeted cytotoxic activity in malignant cells which explains the efficacy of this treatment for several different cancers, especially skin melanomas (Agostinis et al., 2011).

The most used photosensitizers in PDT are indeed porphyrins or porphyrin-like products (Berg et al., 2005). Porphyrins are compounds with unique photochemical properties. These compounds have a distinctive absorption spectrum, characterized by strong absorption near 380 and 500 nm, named the Soret band, followed by four weaker absorption bands between 500 and 750 nm, called the Q bands (Giovannetti, 2012). Because of these photodynamic properties and diverse chemical structures and modifications, these pigments can have a variety of functions. In marine invertebrates, not only the widely known role of heme in gas exchange is described, but also the synthesis of potent antimicrobial pigments (Bandaranayake, 2006).

Bile pigments are endogenous products that result from the metabolism of certain porphyrins. However, these compounds are arranged in linear structures rather than in the traditional cyclic configuration of porphyrins, making these pigments open-chain tetrapyrroles (Lester & Troxler, 1969). The best-known bile pigments are biliverdin and bilirubin, which result from the breakdown of heme (Schmid & McDonagh, 1975). Heme is a prosthetic group composed of iron and protoporphyrin IX. It is found in almost all life forms and is an essential component of biological structures such as hemoglobin, peroxidases, and cytochromes of the P450 family (Heinemann et al., 2008). The essential enzymatic reactions and intermediate metabolites involved in the heme biosynthetic pathway are illustrated in

Fig.1.1. The pathway represented can be found in animals, fungi, and α-proteobacteria, therefore illustrating that the metabolic pathway is well-conserved. The first reaction in the heme biosynthesis pathway occurs in the mitochondrial matrix with the condensation of glycine and succinyl-CoA by δ-aminolevulinic acid synthase (ALAS). The resulting intermediate, δ-aminolevulinic acid (ALA), is then exported to the cell's cytosol and two molecules of ALA are converted into porphobilinogen (PBG) by ALA dehydratase (ALAD). Four PBG molecules are transformed by hydroxymethylbilane synthase (HMBS) into the first cyclic tetrapyrrole, hydroxymethylbilane (HMB). Still in the cytosol, HMB is converted to uroporphyrinogen III by uroporphyrinogen synthase (UROS), which is then decarboxylated by uroporphyrinogen decarboxylase (UROD) to produce coproporphyrinogen III. Coproporphyrinogen III is then transported to the mitochondria, where it is decarboxylated by coproporphyrinogen oxidase (CPOX) to originate protoporphyrinogen IX, that is subsequently oxidized to protoporphyrin IX by protoporphyrinogen oxidase (PPOX). Lastly, ferrochelatase (FECH) inserts ferrous iron (Fe^{2+}) into protoporphyrin IX to form heme (see Heinemann et al., 2008 for detailed step-by-step biosynthesis of heme).

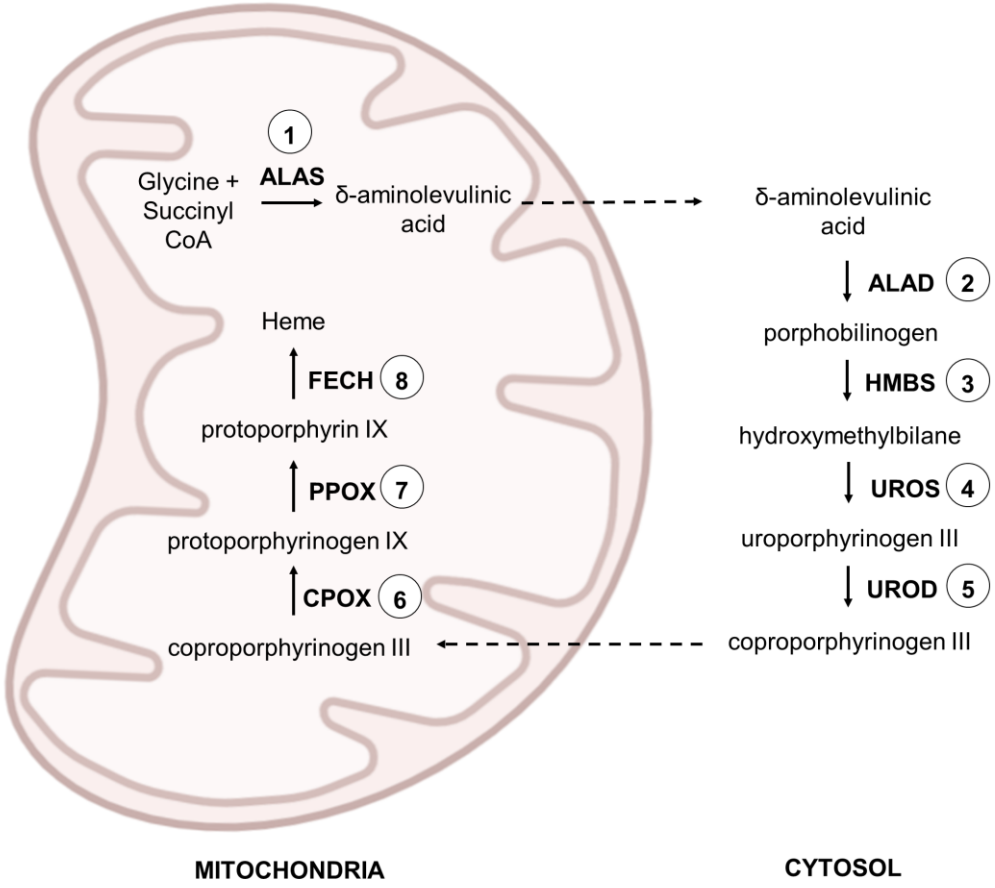


Fig.1.1. Schematic representation of the heme biosynthetic pathway. Heme biosynthesis is an eight-step enzymatic cascade that occurs in the mitochondria and cytosol. The enzymes by order of activity are 1) ALAS, δ-aminolevulinic acid synthase; 2) ALAD, δ-aminolevulinic acid dehydratase; 3) HMBS, hydroxymethylbilane synthase; 4) UROS, uroporphyrinogen synthase; 5) UROD, uroporphyrinogen decarboxylase; 6) CPOX, coproporphyrinogen oxidase; 7) PPOX, protoporphyrinogen oxidase; and 8) FECH, ferrochelatase. Created with BioRender.com. See Fujiwara and Harigae (2015) and Kim et al. (2015) for other schematic representations.

Due to the toxicity of heme biosynthesis intermediates, heme metabolism has to be tightly regulated to avoid intermediate accumulation (Hamza & Dailey, 2012). In most vertebrates, erythroid and non-erythroid cells are regulated by distinct mechanisms. For example, the enzyme ALAS has two different isoforms, ALAS1 that is encoded by a housekeeping gene, and ALAS2 that is expressed exclusively in erythroid cells (Yamamoto et al., 1985). It is also described that both heme and iron play central roles in the regulation of ALAS2 (Surinya et al.,1997). The origin of these two forms of ALAS in vertebrates is believed to arise from a gene duplication event that occurred between cephalochordates and the earliest vertebrates (Duncan et al., 1999). Even though the enzymes in heme biosynthesis have been identified, there are still gaps concerning the differences between vertebrate and invertebrate organisms.

The next step following heme biosynthesis is heme degradation which can occur through several mechanisms. One of the most well-known mechanism is in mammals, where the catabolism of heme takes place in the reticuloendothelial system (mostly in the liver and spleen) and involves the enzyme complex, heme oxygenase (HMOX). Heme is converted by HMOX to biliverdin, a green pyrrolic product, and is rapidly metabolized by biliverdin reductase (BLVRA) to bilirubin, the yellow pigment found in the bile of vertebrates. Usually, bilirubin is then excreted to the duodenum for elimination (Orten, 1971; Maines, 1999). Previous studies have described endogenous biliverdin in the intertidal Polychaeta *Hediste (Nereis) diversicolor* (Dales & Kennedy, 1954). However, the metabolic pathway of these heme-related pigments in Polychaeta is still mostly unknown.

Eulalia sp. is a marine Polychaeta mostly known for its bright green coloration. The species is mostly found in Atlantic rocky shores, usually seeking protection underneath mussel beds (Emson, 1977). It belongs to the Phyllodocidae family and, as in other known phyllodocids, it relies on a muscular eversible proboscis to collect food through suction (Tzetlin & Purschke, 2005). However, the densely muscular proboscis is devoid of jaws. Yet, feeding has been described as mostly carnivorous and its usual prey include mussels, barnacles, Polychaeta and even conspecifics (Emson, 1977; Morton, 2011). These unique features led to morphoanatomical studies of *Eulalia* sp. during which specialized pigment cells were identified in the proboscis and throughout the epidermis (Costa et al., 2013; Rodrigo et al., 2018). Additionally, pigment-containing granules were found in the intestine, suggesting that the pigments responsible for the worm's green coloration were, at least in part, metabolized and eliminated in the digestive epithelia (Rodrigo et al., 2018).

Despite the lack of focus given to *Eulalia* sp. pigments, the distribution and chemical source of the pigments found in this worm was recently explored by Martins et al. (2019). The endogenous pigments were identified as porphyrin-like pigments, most likely heme-derived, with greenish or yellowish colorations whose distribution and abundance differs throughout the organs. The main organs of *Eulalia* sp. are the proboscis, epidermis, intestine and oocytes, all displaying pigmentation. While the proboscis holds only yellow pigments, the epidermis has both types of pigmentation albeit the majority are green pigments. Contrarily, the intestine and oocytes only yield green pigmentation. Overall, the pigment distribution is similar to what is described for biliverdin in the intertidal Polychaeta, *Hediste diversicolor* (Dales & Kennedy, 1954). However, in *Eulalia* sp. the pigments are responsible for the organism's primary coloration, and in *Hediste diversicolor,* they are categorized as secondary metabolites. Lastly,

D’Ambrosio et al. (2020) demonstrated that crude pigment extracts from the proboscis and epidermis are capable of toxicity, especially the mixture of yellow pigments from the proboscis upon absorption of light. These preliminary results suggest the potential application of these pigments as photosensitizers. Yet, the chemical structure of the isolated pigments, their main potential, and the pathway in which they are formed remain mostly unknown.

1.1 Objectives

Natural pigments are usually products of complex metabolic pathways, with several enzymes involved. Understanding these metabolic pathways underlying the synthesis or *de novo* synthesis of pigments has many implications in the bioprospecting for novel compounds. Despite the countless pigment-rich invertebrates known to date, there is still little information regarding marine natural pigments and their respective biosynthetic pathways. Therefore, the porphyrinoid pigments of *Eulalia* sp. are the perfect case study for metabolic research and the main objectives of the present work are summarized as:

- To analyze the main porphyrinoid pigments in the proboscis, epidermis, intestine and oocytes of *Eulalia* sp.
- To provide a comparative assessment between the pigmentation patterns of the proboscis and epidermis, as the main pigment-bearing organs.
- Identify the main genes coding for enzymes involved in the metabolism of the pigments and analyze their differential gene expression and relative expression.
- Analyze the differences between the metabolic pathways of porphyrinoids in the proboscis and epidermis and relate such differences to their pigmentation pattern.
- Infer on the potential role of the pigments in the worm as a means to identify candidates for biotechnological applications.

2. Materials and Methods

2.1 Obtaining porphyrin-like pigments from *Eulalia* sp.

2.1.1 Animal collection

Adult *Eulalia* sp. were collected by hand from a rocky intertidal area in Avencas beach, Western Portugal (38°41'17.1"N 9°21'36.5"W, Fig. 2.1). Harvesting was done between fall and winter of 2019-2020 and animals ranged between 60 and 110 mm in total length and weighed around 280 mg. Individuals were maintained in a mesocosm environment recreating their natural habitat, as described elsewhere (Rodrigo et al., 2015). In brief, the mesocosm environment consisted of an aquarium with 7L of artificial saltwater (maximum capacity of 15L), protected from direct light and equipped with continuous aeration and recirculation. To provide both shelter and feeding, the aquarium was fitted with natural rocks and clumps of mussels, collected from Costa da Caparica beach, a clean area in Western Portugal (38°38'28.0"N 9°14'18.9"W, Fig. 2.1). Salinity, temperature, and photoperiod were restricted to 30, 18ºC, and 10:14h light:dark cycle, respectively.

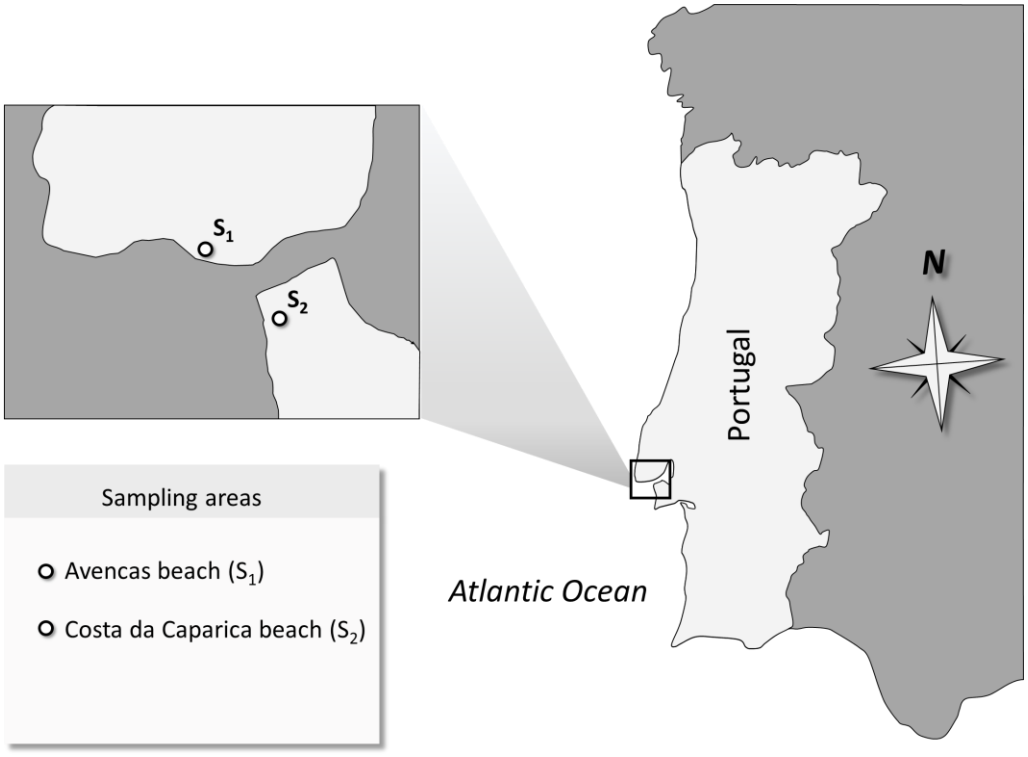


Fig.2.1. Map highlighting sampling locations. The sampling campaigns took place in Western Portugal, in Avencas beach (S1) to collect adult *Eulalia* sp. and in Costa da Caparica beach (S2) to gather natural rocks and clumps of mussels. The timeline of sampling was between fall and winter of 2019-2020. The coordinates for the sampling areas are S1 (38°41'17.1"N 9°21'36.5"W) and S2 (38°38'28.0"N 9°14'18.9"W).

2.1.2 Pigment extraction

Eulalia sp. were quickly euthanized by osmotic shock through immersion in ultrapure water. Specimens were immediately microdissected under an optical stereoscope and the target organs separated, namely the proboscis (Pr), epidermis (Ep), intestine (Int), and when available oocytes (Oo) from maturing

females. The extraction of porphyrin-like pigments from biological tissue was done according to the protocol described by Woods and Simmonds (2001) for tissue and fecal porphyrin extraction, with modifications by Martins et al. (2019). In brief, using a handheld homogenizer, the samples were homogenized in one volume of a mixture of 1 N hydrochloric acid (HCL) and acetonitrile (ACN) as the extraction solvent. After homogenization, the samples were centrifuged for 5 min at 10 000g (4ºC) to collect the supernatant (fraction containing the pigments). The extraction solvent was then added to the pellet (fraction with the organ sample), which was resuspended and homogenized, followed by centrifugation and supernatant collection. The process was repeated until a clear supernatant was obtained, which indicated that no more pigments were available in the sample. The pooled supernatants were stored at -20ºC until further analysis. Pigment extraction was performed with minimum exposure to light and air to protect porphyrins from photo-oxidation. A total of one hundred and one individuals were used for pigment extraction. The schematic representation of the protocol is shown in Fig. 2.2.

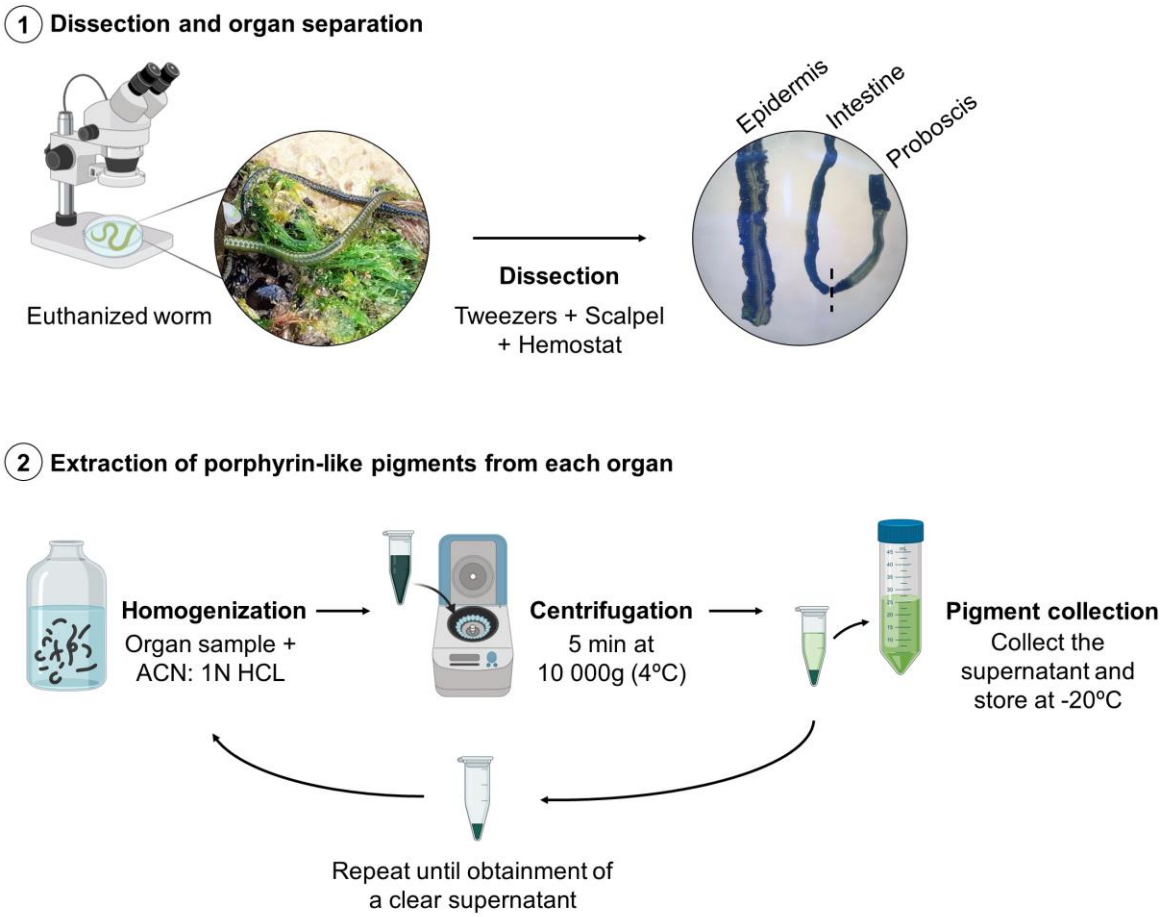


Fig.2.2. Schematic representation of the pigment extraction protocol. Summary of all the steps involved in pigment extraction according to the protocol described by Woods and Simmonds (2001) with modifications by Martins et al. (2019). The nomenclature “Epidermis” was used for the worm’s body-wall since the pigments are described to be contained in specialized epidermal cells and oocytes were not always present due to the low number of maturing females. Created with BioRender.com.

2.1.3 High-performance liquid chromatography (HPLC)

2.1.3.1 Analytical HPLC

Eulalia sp. pigment extracts with less than one week of storage, from the proboscis, epidermis, intestine, and oocytes, were fractionated by analytical HPLC to determine the most representative porphyrinoid pigments of each target organ. To this purpose, the organic phase of the extracts containing the pigments was filtered with a FILTER-LAB Polypropylene Syringe filter (PP) with a pore size of 0.45 µm before injection. High-performance liquid chromatography (HPLC) was performed with some optimizations according to the protocol for pigment chromatography by Martins et al. (2019) based on the separation and quantitation of porphyrins described by Woods and Simmonds (2001). Pigment analysis was conducted on a Merck-Hitachi instrument equipped with a diode array detector (DAD), scan range 200-800 nm (Merck-Hitachi L-4500 Diode Array Detector), using a reversed-phase analytical column (RP-HPLC, Phenomenex Onyx Monolithic C18 column, 100 mm × 4.6 mm i.d., 13 nm, and 2 µm). The injection volume was 20 µL of sample per extract. Optimal peak resolution was obtained with sodium phosphate buffer 10 mM, pH 3.5 (solvent A) and pure methanol (MeOH) (solvent B) at a flow rate of 2 mL/min using a 10 min linear gradient from 45% to 95% of solvent B and 55% to 5% of solvent A, followed by elution at 95% of solvent B for 2 min. The total run time excluding equilibration was 12 minutes, the temperature was maintained at 20ºC and the pressure around 66bar. The pigments were identified according to maximum absorbance registered in the chromatograms of each organ. For analysis, three wavelengths were selected: 280 nm for absorption in the UV region, 400 nm to detect yellowish pigments with absorption in the violet range, and 700 nm to detect greenish pigments that absorb in the red spectrum. Additionally, the absorption spectrum of each pigment was recorded. Before analysis, all data was normalized.

2.1.3.2 Preparative HPLC

Crude pigment extracts from the proboscis and epidermis were fractionated by preparative HPLC to isolate the pigments of interest, specifically Pr2, Ep2, Ep3, and Ep4. The extracts were from 44 pooled individuals stored between 12 to 15 weeks at -20ºC. The organic phase of the extracts was filtered with a FILTER-LAB Polypropylene Syringe filter (PP) with a pore size of 0.45 µm before injection. HPLC was performed with some optimizations according to the protocol for pigment chromatography by Martins et al. (2019) based on the separation and quantitation of porphyrins described by Woods and Simmonds (2001). Pigment analysis was carried on a Merck-Hitachi instrument equipped with a diode array detector (DAD), scan range 200-800 nm (Merck-Hitachi L-4500 Diode Array Detector), using a reversed-phase semi-preparative column (RP-HPLC, Phenomenex Onyx Monolithic Semi-prep C18 column, 100 mm × 10 mm i.d., 13 nm, and 2 µm). The injection volume varied from 100 to 200 µL of sample and the solvents used were sodium phosphate buffer 10 mM, pH 3.5 (solvent A) and pure MeOH (solvent B) at a flow rate of 9 mL/min using a 10 min linear gradient from 35% to 65% of solvent B and 65% to 35% of solvent A, followed by elution at 95% of solvent B for 5 min. The total run time excluding equilibration was 15 minutes and throughout the experiment, the temperature was maintained at 20ºC and the pressure at about 67bar. The collected individual pigments were subjected to a rotary evaporator

(BUCHI Rotavapor R-200) and vacuum at room temperature, to evaporate the HPLC solvents and obtain the pigments in the form of powder. The pigments were stored at -20ºC until further analysis (no further investigations were conducted due to the COVID-19 pandemic).

2.2 Bioinformatics and statistical approach

2.2.1 RNA sequencing and transcriptome assembly

Eulalia sp. raw (unannotated) transcriptome was retrieved from Gene Expression Omnibus (GEO), accession number GSE143954, having been deposited by Rodrigo et al. (2020) who extracted and sequenced RNA to assemble *Eulalia*'s transcriptome. In brief: total RNA was extracted from the proboscis and epidermis of three individuals of *Eulalia* sp. The extraction was done on samples stabilized with RNALater using the RNeasy Protect Mini Kit (from Qiagen, Hilden, Germany) for purification of total RNA from animal tissues. RNA fragmentation, cDNA synthesis, and cDNA library preparation were done using the Kapa Stranded mRNA Library Preparation Kit and the generated fragments were sequenced in an Illumina HiSeq 4000 platform, with 150bp paired-end reads. One sample from the proboscis and another from the epidermis were sequenced with high sequencing depth for transcriptome assembly (100M reads) and the remaining samples from the two biological replicates were sequenced for normal coverage (20M reads). Low-quality reads were removed from the RNA-seq data and the filtered high-depth sequenced samples were assembled with Trinity v2.8.4 (Grabherr et al., 2011). Coding regions of the assembled transcriptome were predicted using TransDecoder v5.5.0 (Haas et al., 2013). Lastly, all samples were mapped to *Eulalia's* predicted transcriptome with Kallisto v0.44.0 (Bray et al., 2016). Final data used in the present study consisted of translated transcripts abundance for the preceding samples (three from the proboscis and three from the epidermis).

2.2.2 Gene expression analysis

Statistical analysis was done using R 3.5.1 (Ihaka & Gentleman, 1996), through packages edgeR and limma. RNA-Seq data normalization was performed by creating a DGEList object with the abundance results from the proboscis and epidermis. A linear model was fitted on the transformed data, designed to contrast the transcript's expression in the proboscis relative to the epidermis. Log fold change (log_2FC) was chosen to display the expression differences. Proboscis differentially expressed transcripts relative to the epidermis were selected with FDR (False Discovery Rate) adjusted p-value ≤ 0.05 and $log_2FC \geq 2$ for over-expressed transcripts and $log_2FC \leq -2$ for under-expressed transcripts. Relative expression within each organ was assessed through analysis of logTPM (transcripts per million). Expression levels were measured relative to the average expression of all transcripts in the sample. Therefore, a threshold of relative expression was set and transcripts with logTPMs ≥ 2 were considered with high levels of expression and transcripts with logTPMs ≤ -2 were considered with low levels of expression.

2.2.3 Sequence identification

To identify the proteins in the proboscis and epidermis associated with porphyrin-like pigments, *Eulalia* sp. translated transcripts were annotated based on sequence homology against UNIPROT protein databases with BLASTP v2.9.0 (Gish & States, 1993), having set a maximum e-value of 10^{-5}. Each database consisted of sequences of proteins associated with the search terms: “chlorins”, “porphyrin Eumetazoa”, “heme biosynthesis” and “heme degradation”, resulting in the construction of four customized subsets with matches between the sequence of a known protein and a *Eulalia* sp. translated transcript. After protein identification, the main represented heme biosynthesis enzymes in both the proboscis and epidermis were validated by sequence comparison to other Metazoa organisms with MrBayes 3.2.7a (Huelsenbeck & Ronquist, 2001). The settings used were the Maximum Likelihood method and LG gamma distribution model, with 100 0000 iterations. Sequence alignment (by ClustalW) and best protein model choice were conducted in MEGA X (Kumar et al., 2018). Metazoa sequences were obtained from UNIPROT.

2.2.4 Gene ontology and network analysis

Eulalia sp. protein-protein interactions related to heme metabolism were inferred with STRING Database v11.0 (Szklarczyk et al., 2019). For each subset, porphyrin Eumetazoa, heme biosynthesis, and heme degradation, the highly expressed proteins specifically for the proboscis and epidermis, and in both organs, were selected to infer differences in the enzymatic process of heme metabolism in both organs. To do so, the organism selected to detect the proteins in the subsets was *Homo sapiens* due to the high level of genomic annotation. Proteins with no known interactions to other proteins as well as no association with heme metabolism were removed from the networks.

2.2.5 Validation by RT-qPCR

The RNA-Seq data results were validated by quantitative reverse transcription PCR (RT-qPCR) for mRNAs coding for three relevant heme biosynthesis proteins. In brief, total RNA was extracted from the proboscis and epidermis of three individuals. The extraction was done on samples stabilized with RNALater using the RNeasy Protect Mini Kit (Qiagen, Hilden, Germany) for purification of total RNA from animal tissues. Synthesis of cDNA was done by reverse-transcription using the First-Strand cDNA Synthesis Kit (NZYTech), according to manufacturer’s instructions. Primers were designed with Primer Blast against *Eulalia*’s sequences of best-matched target mRNAs and verified with Oligo Analyzer (see Appendix Table 7.1 for primer sequences). The RT-qPCR was done in a Rotor-Gene 6000 thermal cycler (Corbett Research) with the NZY qPCR Green Master Mix (NZYTech), following manufacturer instructions. The programmed settings were, with 45 cycles per run: denaturation at 94ºC for 45 seconds, annealing at 55ºC for 35 seconds, and extension at 72ºC for 30 seconds. As a calibrator, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used, as suggested by Thiel et al. (2017). Relative expression was determined by the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001). Statistical analysis was conducted with Student's t-test and Mann-Whitney U test as parametric and non-parametric procedures, respectively. Normality was analyzed through Shapiro’s test and

homoscedasticity with Levene's test. Statistics were performed with R. Melting curve analysis of the RT-qPCR results is demonstrated in the Appendix (Fig.7.2).

3. Results

3.1 Characterization of *Eulalia* sp. main pigments

Chromatograms of the extracted pigments from the proboscis, epidermis, intestine, and oocytes are shown in Fig.3.1. The results revealed multiple maximum peaks registered and each absorbance peak represents the presence of a pigment in the sample. On total, twelve main individual pigments were identified. The oocytes yielded four pigments, followed by the epidermis and intestine both with three major pigments, and lastly, the proboscis with two identified pigments. As shown in Fig.3.1, the pigments display strong absorption in the UV zone and high absorption in the violet and red regions of the visible light spectrum, which confirms the presence of yellowish and greenish pigments, respectively.

The pigments from the proboscis bore a yellowish tone and were termed Pr1 and Pr2. Both had absorption maxima at 280 nm and high absorption at 400 nm, albeit lower absorption by Pr1 when in comparison to Pr2 (Fig.3.1a). The retention time of Pr1 and Pr2 was 1.04 to 1.63 min and 3.69 to 4.34 min, respectively. The chromatogram from the epidermis also demonstrated a yellow pigment, termed Ep2, with similar retention time (3.85 to 4.25 min) and absorbance comparatively to Pr2 (Fig.3.1b). Two additional green pigments named Ep3 and Ep4 were identified in the epidermis, with similar retention times (from 6.36 to 6.76 min). Both these pigments yielded absorption maxima at 280 and high absorption at 400 and 700 nm, albeit Ep4 displayed higher absorbance at these wavelengths (Fig.3.1b).

The intestine and oocytes revealed several greenish pigments with absorption maxima at 280 nm, and high absorption at 400 and 700 nm (Fig.3.1c,d). The intestine pigments, named Int1, Int3, and Int4, had retention times between 4.01 and 6.72 min. In turn, the oocytes displayed four major pigments identified (Oo2, Oo3, Oo4, and Oo5) with retention times between 4.70 and 6.70 min. Green pigments from the epidermis, intestine, and oocytes revealed high absorbance at 700 nm, which explains their greenish coloration since green pigments absorb higher within the red component of the visible light spectrum. In turn, the yellow pigments in the proboscis and epidermis displayed higher absorbance at 400 nm since they absorb more blue light.

The yellow pigments, Pr2 from the proboscis and Ep2 from the epidermis, in addition to having similar retention times, the pigments also display identical absorption spectra (Fig.3.2a) which indicates that they are the same pigment. Similarly, the green pigments, Int1 from the intestine and Oo3 from the oocytes, also display similar retention times and matching absorption spectra (Fig.3.2b). A similar result was found for the green pigments Int3, Oo5, and Ep4 from the intestine, oocytes, and epidermis, respectively (Fig.3.2c). The summary of these results is shown in Table 3.1.

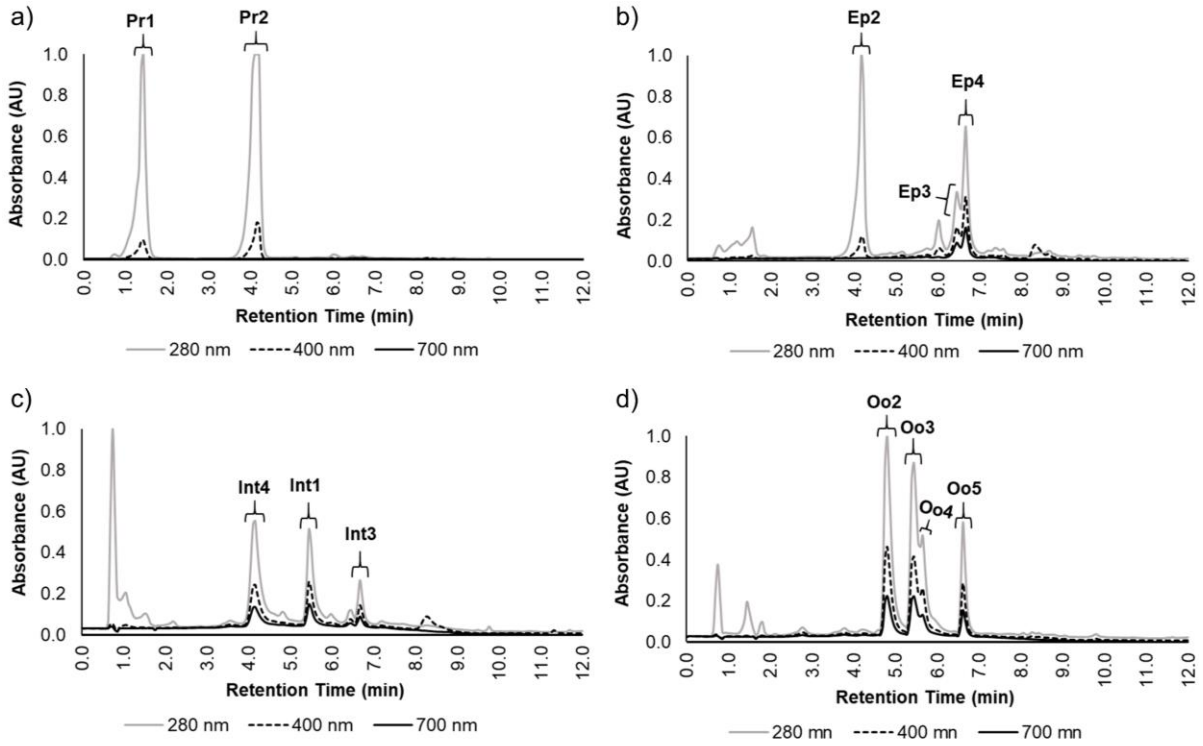


Fig.3.1. Main pigments in the multiple organs of *Eulalia* sp. The data was retrieved from a high-performance liquid chromatography with a detector diode array (HPLC-DAD). The absorbance (AU) and retention time (min) are illustrated for each chromatogram of the a) proboscis (Pr), b) epidermis (Ep), c) intestine (Int1), and d) oocytes (Oo). Each pigment was identified with the organ abbreviation and number identification. The wavelengths selected were 280, 400, and 700 nm which match the absorbance in the range of UV, violet (to detect yellow pigments), and red (to detect green pigments) light.

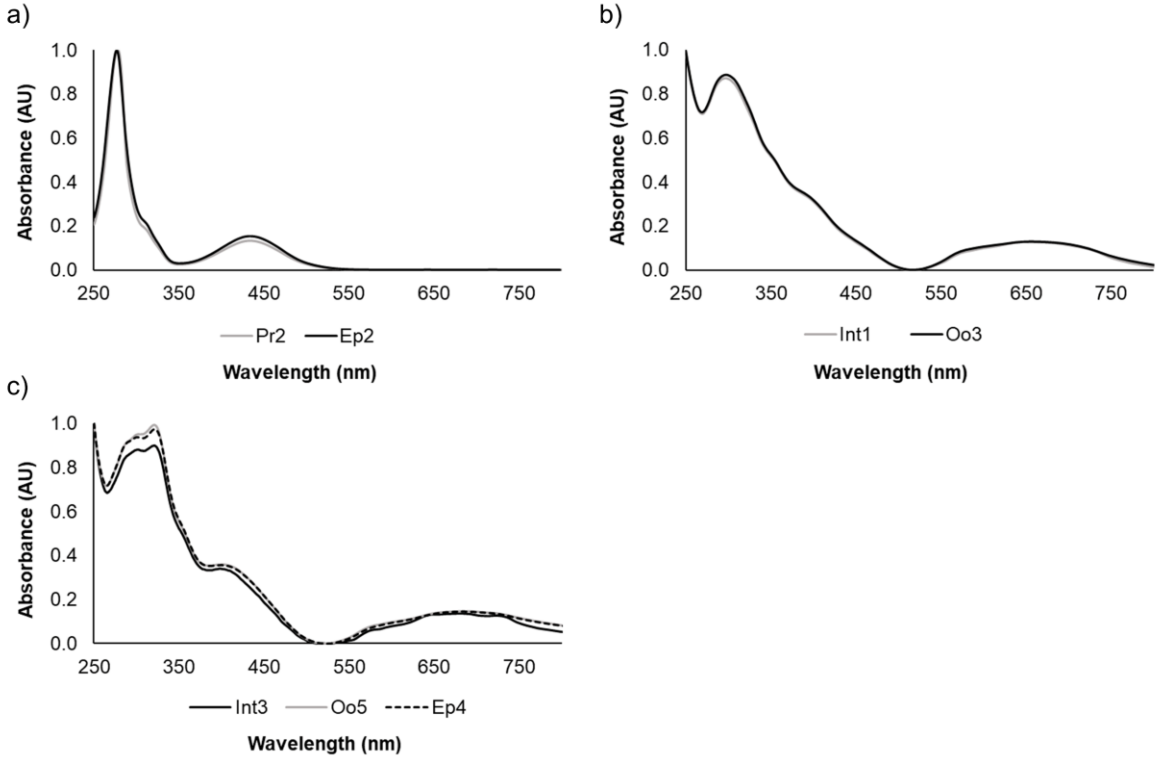


Fig.3.2. Common pigments between the different organs of *Eulalia* sp. The data was retrieved from a high-performance liquid chromatography with a detector diode array (HPLC-DAD). The comparisons are a) two yellow pigments, Pr2 from the proboscis and Ep2 from the epidermis, b) two green pigments, Int1 from the intestine and Oo3 from the oocytes, and c) three green pigments, Int3 from the intestine, Oo5 from the oocytes and Ep4 from the epidermis.

Table 3.1. Main pigments of *Eulalia* sp. per organ. Each pigment was selected according to pigment maxima absorbance registered. Retention time (min) corresponds to the timeframe in which the pigment was eluted from the HPLC column. Pigment color was determined by visual inspection and absorption at 400 nm (for yellow pigments) and 700 nm (for green pigments). Similar pigments are identified as a,b, or c (see Fig.3.2).

Organ	Pigment ID	Color	Retention time (min)	
Proboscis	Pr1	Yellow	1.04 – 1.63	
	Pr2	Yellow	3.69 – 4.34	a
Epidermis	Ep2	Yellow	3.85 – 4.25	a
	Ep3	Green	6.36 – 6.51	
	Ep4	Green	6.61 – 6.76	c
Intestine	Int1	Green	5.41 – 5.56	b
	Int3	Green	6.63 – 6.72	c
	Int4	Green	4.01 – 4.32	
Oocytes	Oo2	Green	4.70 – 5.01	
	Oo3	Green	5.37 – 5.53	b
	Oo4	Green	5.64 – 5.74	
	Oo5	Green	6.56 – 6.70	c

The porphyrinoid nature of the main pigments from the multiple organs was verified by the presence of the Soret and Q bands (Fig.3.3). The main pigment in each target organ was selected for representation and all spectra exhibit the two features, the Soret band between 380 and 500 nm, followed by the four weaker absorption Q bands, between 500 and 750 nm. These features have distinguished resolution on the yellow representative pigment Pr2 from the proboscis (Fig.3.3a). The green pigments, Ep4 from the epidermis, Int1 from the intestine, and Oo2 from oocytes, also present the Soret and Q bands, although less resolved in comparison with the representative yellow pigment. In contrast, in the green pigments the Soret band is less evident and only one broader band can be observed in the Q band region instead of the four typical bands (Fig.3.3b,c,d). The absorption spectra of the remaining pigments show the same differences between yellow and greenish pigments (Appendix Fig.7.1).

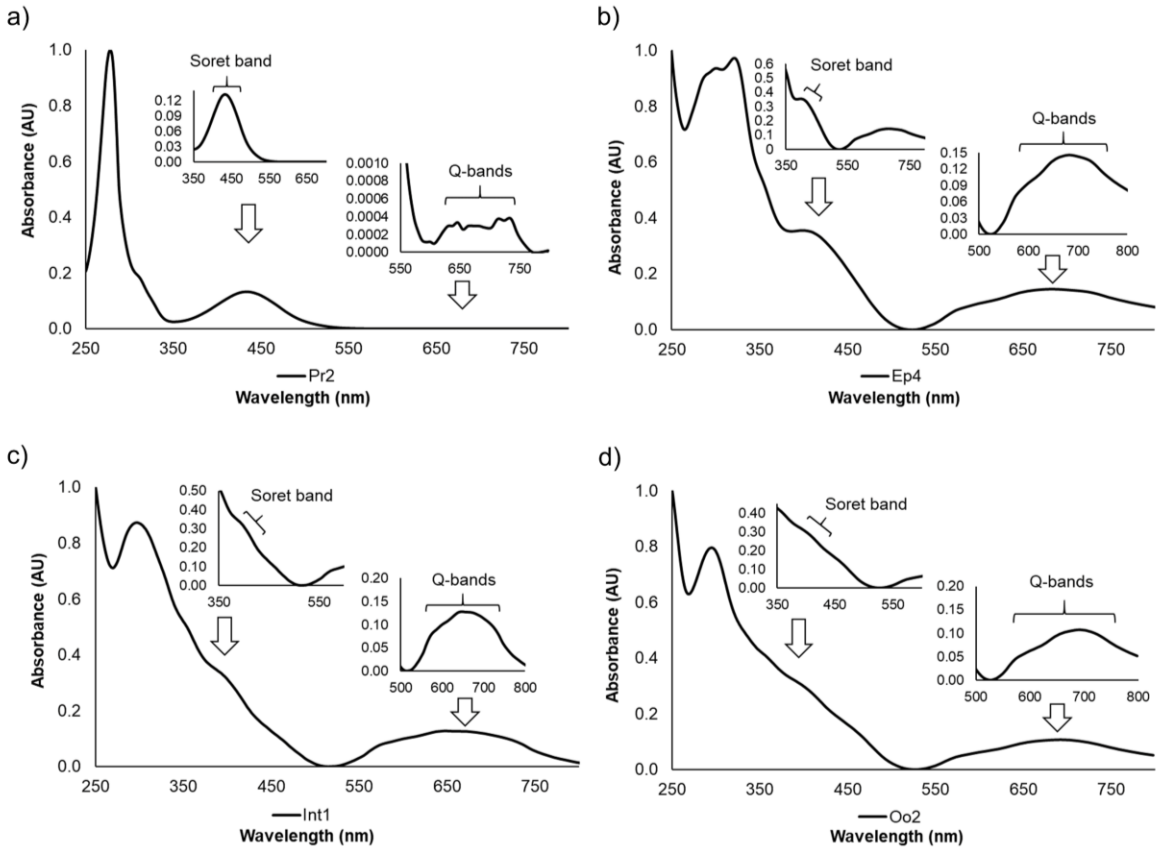


Fig.3.3. Representative pigment absorption spectra highlighting Soret and Q bands typical of porphyrinoids. The data was retrieved from high-performance liquid chromatography with a detector diode array (HPLC-DAD). Each graph displays a Soret band (380-500 nm) and Q bands (500-750 nm) with different degrees of resolution. a) Pr2, a yellow pigment from the proboscis; b) Ep4, a green pigment from the epidermis; c) In1, a green pigment from the intestine; and d) Oo2, a green pigment from the oocytes. The retention time corresponding to each spectrum is 4.05 min for Pr2, 6.63 min for Ep4, 5.47 min for Int1, and 4.89 min for Oo2.

3.2 RNA-Seq data analysis

3.2.1 Differential gene expression between the proboscis and epidermis

After trimming of low-quality reads, the *Eulalia* sp. transcriptome yielded 55 508 individual open reading frames (ORF), each encoding for a different protein or different variants of the same protein. Of these ORFs, 734 were over-expressed in the proboscis in comparison to the epidermis and 1 350 were under-expressed in the proboscis relative to the epidermis. The remaining 53 424 ORFs yielded no differential gene expression between the proboscis and epidermis (Fig.3.4).

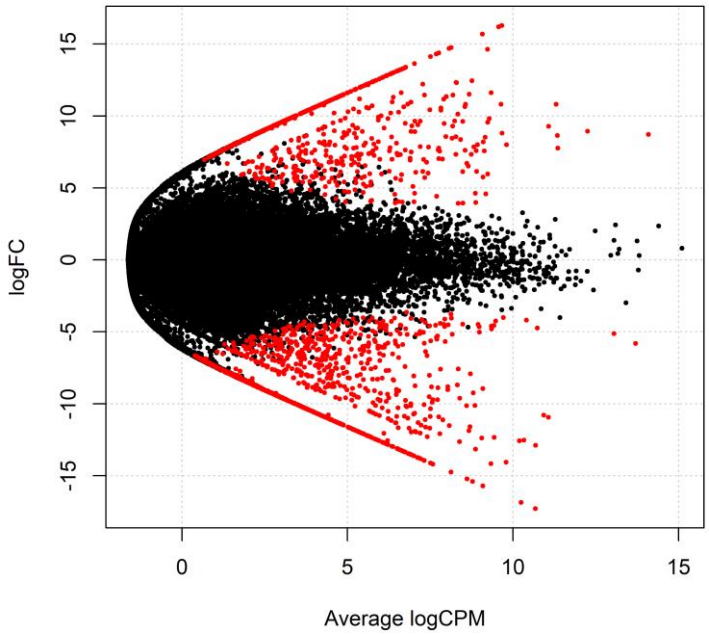


Fig.3.4. Smear plot showing an overview of the differential expressed genes in the proboscis relative to the epidermis. Each dot represents an ORF of *Eulalia*'s transcriptome, in red are the differentially expressed ORFs in the proboscis when in comparison to the epidermis, and in black are the ORFs without differential expression. The criteria for differential gene expression was adjusted p-value (FDR) ≤ 0.05 and $\log_2$FC ≥ 2 for over-expression and $\log_2$FC ≤ -2 for under-expression out of a total of 55 508 ORFs.

Eulalia's predicted translated transcripts were annotated based on homology matching against to four sets of proteins associated to the search terms: "chlorins", "porphyrin Eumetazoa", "heme biosynthesis", and "heme degradation". The subsets with the most matches were heme biosynthesis and heme degradation with 7 198 and 4 651 proteins, respectively, followed by porphyrin Eumetazoa with 2 294 proteins, and lastly, chlorins with only 32 identified proteins (Table 3.2). Of these last 32 proteins, 17 were within the | logTPM | ≥ 2 threshold (Fig.3.5b). These findings led to the discarding of the chlorins subset. On the other hand, the preceding subsets, heme biosynthesis, heme degradation and porphyrin Eumetazoa yielded 5 577, 3 671, and 1 804 proteins within the threshold, respectively, most of which yielded low levels of expression (logTPM ≤ -2) (Fig.3.5a,c,d). These proteins were distributed in the proboscis, epidermis or both organs according to their relative expression (Fig.3.5).

Table 3.2. Homology-matching between *Eulalia* sp. translated transcripts and database proteins. *Eulalia* sp. translated transcripts were annotated based on homology against UNIPROT protein sequences associated with the search terms “chlorins”, “porphyrin Eumetazoa”, “heme biosynthesis”, and “heme degradation”, generating four independent subsets. The obtained matches correspond to the number of translated transcripts with homology to a known database protein of the total 55 508 translated transcripts from *Eulalia’s* predicted transcriptome, having set a maximum e-value of 10^{-5}.

Subset	Number of matches
Chlorins	32
Porphyrin Eumetazoa	2 294
Heme biosynthesis	7 198
Heme degradation	4 651

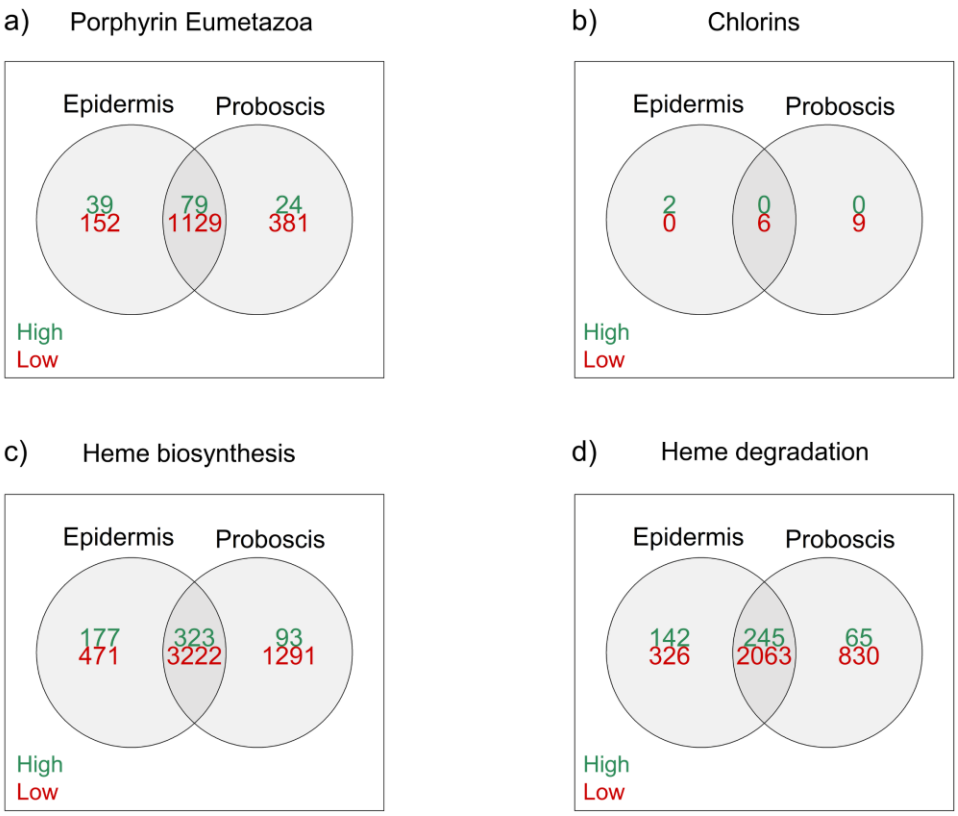


Fig.3.5. Distribution of the proteins in each subset according to their relative expression in *Eulalia* sp. Each Venn diagram represents a customized subset a) porphyrin Eumetazoa, b) chlorins, c) heme biosynthesis and d) heme degradation. The proteins are distributed according to their expression values: with high levels of expression (“high” in green, logTPM ≥ 2) and with low levels of expression (“low” in red, logTPM ≤ -2) for the proboscis, epidermis, and even both organs.

The three main subsets, porphyrin Eumetazoa, heme biosynthesis, and heme degradation, yielded conspicuous patterns of differentially expressed genes (DEGs) between the proboscis and epidermis (adjusted p-value (FDR) ≤ 0.05 and | $\log_2$FC | ≥ 2) (Fig.3.6). Each subset revealed the same protein identification for different ORFs of *Eulalia*'s translated transcriptome indicating that these proteins are protein variants. DEGs cluster analysis revealed the expected clustering of the replicate samples from the proboscis (in gray) and from the epidermis (in red), albeit with some intraspecific variation since each sample is from a different individual. All subsets yielded more up-regulated proteins in the epidermis than in the proboscis (Fig.3.6). The subsets with the most differentially expressed proteins were heme biosynthesis and heme degradation with 289 and 248 proteins, respectively (Fig.3.6b,c). The porphyrin Eumetazoa subset had 54 DEGs and several up-regulated ALAD protein variants can be observed in the proboscis and in lower number in the epidermis. On the other hand, more FECH protein variants were found up-regulated in the epidermis than in the proboscis (Fig.3.6a).

The top 5 up-regulated proteins in the proboscis relative to the epidermis of the subset porphyrin Eumetazoa were ALAD, broad substrate specificity ATP-binding cassette transporter (ABCG2), and two variants of the protein FECH. The top 5 down-regulated proteins in the proboscis relative to the epidermis (indicating up-regulation in the epidermis) were ATP-binding cassette sub-family B member 6 (ABCB6), and four variants of the protein FECH (Table 3.3). In the heme biosynthesis subset, the top up-regulated proteins in the proboscis were ShK domain and PerOxidase domain containing protein (skpo-2), a conidial pigment biosynthesis oxidase arb2 brown2 (FLAG1_09148), and three variants of the cytochrome P450 family 17 subfamily A member 1 (cyp17a1). In turn, in the epidermis, the main proteins were a plant peroxidase (CCACVL1_26323), a peptidase S1 domain-containing protein (ctrb1), and two variants of cyp17a1 (Table 3.4). In the heme degradation subset, the top up-regulated proteins in the proboscis were sphingoid long chain base kinase 4 (LCB4), proto-oncogene tyrosine-protein kinase Src (SRC), haptoglobin (HP), and two variants of the protein tryptophan 2,3-dioxygenase (TDO2). In the epidermis, the main protein was TDO2 with at least five variants (Table 3.5).

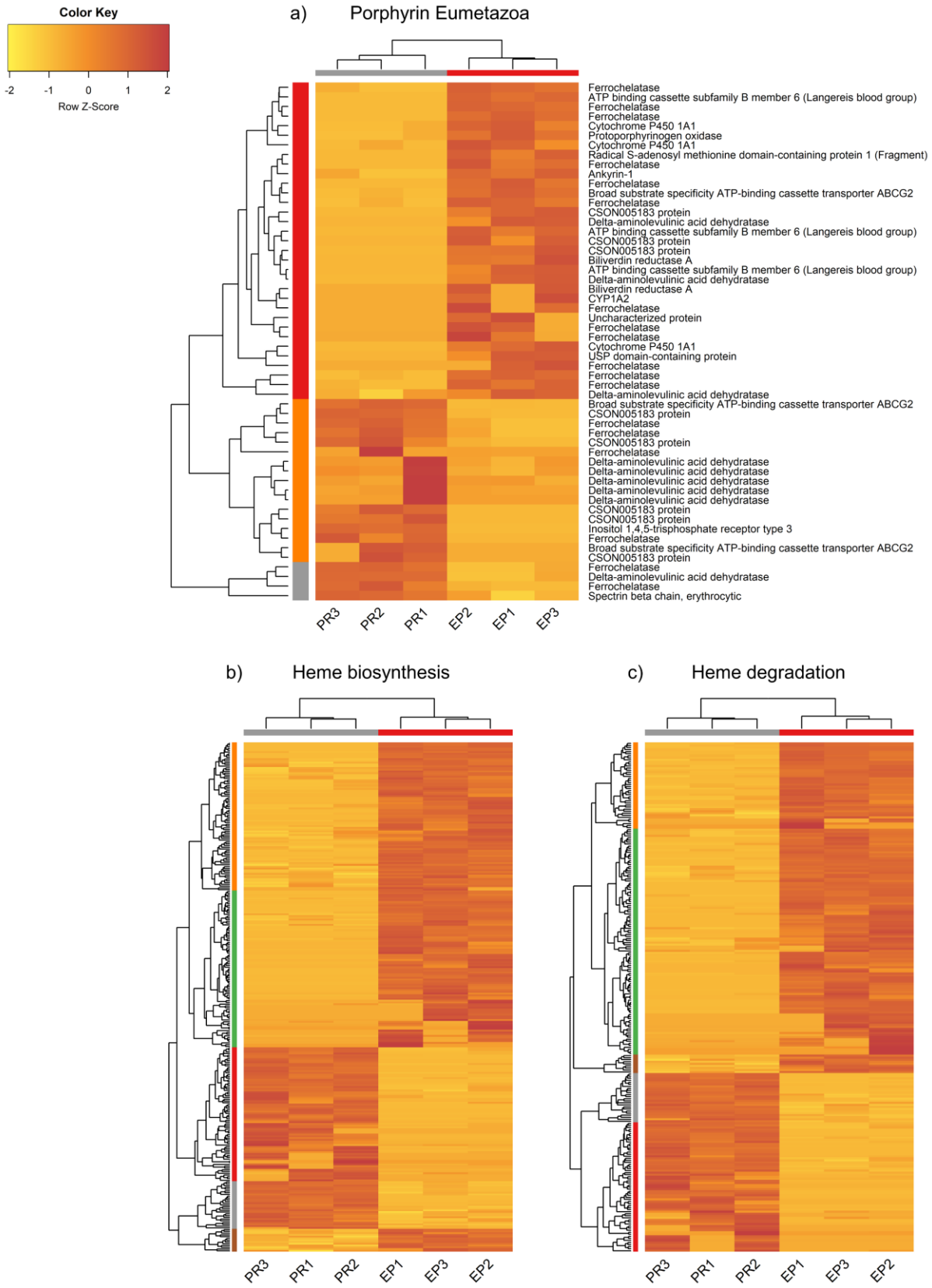


Fig.3.6. Heatmaps illustrating gene relative expression between the proboscis and epidermis. Bellow the horizontal dendrogram is the association between the proboscis (PR) and epidermis (EP), both with three independent replicates, and next to the vertical dendrogram is the associations between proteins. The criteria for protein selection was adjusted p-value (FDR) ≤ 0.05 and | $\log_2$FC | ≥ 2. Expression levels of individual replicates are given as counts per million (CPM) and high protein expression corresponds to a red color key and low protein expression to a yellow color key. Cluster analysis was based on Euclidean distances as metric and dendrograms were built using complete linkage.

Table 3.3. Top ten DEGs in the proboscis relative to the epidermis from the porphyrin Eumetazoa subset. The criteria for protein selection was lowest adjusted p-value of the selected proteins with p-value (FDR) ≤ 0.05 and $log_2FC \geq 2$ for up-regulated and $log_2FC \leq -2$ for down-regulated proteins in the proboscis relative to the epidermis.

Protein	E-Value	Regulation	Log_2FC	Adjusted p-value (FDR)
FECH	9.30E-12	Up	11.2	7.23E-06
FECH	8.97E-12	Up	10.8	8.05E-06
ALAD	1.22E-10	Up	9.4	6.82E-05
ABCG2	8.94E-19	Up	10.8	7.40E-05
CSON005183	4.70E-06	Up	8.6	6.08E-04
FECH	5.95E-06	Down	-10.9	8.09E-07
FECH	4.52E-06	Down	-11.5	2.06E-06
FECH	2.59E-07	Down	-11.1	2.71E-05
ABCB6	4.36E-67	Down	-10.9	3.82E-05
FECH	2.57E-07	Down	-10.9	3.90E-05

FECH: Ferrochelatase; ALAD: δ-aminolevulinic acid dehydratase; ABCG2: broad substrate specificity ATP-binding cassette transporter; CSON005183: similar to 5-aminolevulinate synthase; ABCB6: ATP-binding cassette sub-family B member 6.

Table 3.4. Top ten DEGs in the proboscis relative to the epidermis from the heme biosynthesis subset. The criteria for protein selection was lowest adjusted p-value of the selected proteins with p-value (FDR) ≤ 0.05 and $log_2FC \geq 2$ for up-regulated and $log_2FC \leq -2$ for down-regulated proteins in the proboscis relative to the epidermis.

Protein	E-Value	Regulation	Log_2FC	Adjusted p-value (FDR)
skpo-2	2.00E-07	Up	12.3	8.09E-07
cyp17a1	6.88E-12	Up	10.4	1.35E-06
cyp17a1	2.78E-17	Up	13.1	2.18E-06
cyp17a1	1.23E-12	Up	9.5	3.50E-06
FLAG1_09148	7.22E-14	Up	12.2	3.54E-06
cyp17a1	5.90E-06	Down	-13.9	7.27E-07
cyp17a1	1.46E-14	Down	-10.9	8.09E-07
CCACVL1_26323	8.06E-23	Down	-10.9	9.44E-07
ctrb1	1.65E-26	Down	-9.5	1.31E-06
cyp17a1	3.73E-12	Down	-11.5	2.06E-06

skpo-2: ShK domain and PerOxidase domain containing protein; cyp17a1: cytochrome P450 family 17 subfamily A member 1; CCACVL1_26323: conidial pigment biosynthesis oxidase arb2 brown2; ctrb1: peptidase S1 domain-containing protein.

Table 3.5. Top ten DEGs in the proboscis relative to the epidermis from the heme degradation subset. The criteria for protein selection was lowest adjusted p-value of the selected proteins with p-value (FDR) ≤ 0.05 and $log_2FC \geq 2$ for up-regulated and $log_2FC \leq -2$ for down-regulated proteins in the proboscis relative to the epidermis.

Protein	E-Value	Regulation	Log_2FC	Adjusted p-value (FDR)
TDO2	7.93E-15	Up	10.4	1.35E-06
TDO2	5.14E-14	Up	9.5	3.50E-06
LCB4	4.59E-09	Up	11.0	4.11E-06
SRC	3.69E-14	Up	12.2	4.45E-06
HP	4.35E-21	Up	9.3	4.56E-06
TDO2	3.86E-12	Down	-15.4	3.95E-08
TDO2	1.55E-09	Down	-14.1	3.25E-07
TDO2	7.10E-16	Down	-13.2	7.83E-07
TDO2	3.06E-08	Down	-11.0	8.09E-07
TOD2	8.87E-09	Down	-11.0	8.20E-07

TDO: tryptophan 2,3-dioxygenase; LCB4: sphingoid long chain base kinase 4; SRC: proto-oncogene tyrosine-protein kinase Src; HP: haptoglobin.

3.2.2 Metabolic pathway differences in the proboscis and epidermis of *Eulalia* sp.

Regardless of their expression, the eight fundamental enzymes involved in the canonical heme biosynthesis pathway were identified in the transcriptome of *Eulalia* sp. (Fig.3.7). By order of activity, the enzymes are ALAS1, ALAD, HMBS, UROS, UROD, CPOX, PPOX, and FECH. Also, RT-qPCR validated the detection of the proteins ALAD, UROS and PPOX (Fig.3.8).

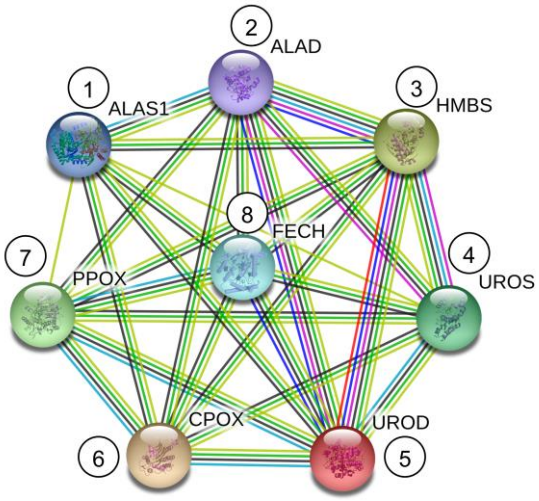


Fig.3.7. Canonical interaction network of the enzymes in heme biosynthesis found in *Eulalia* sp. The enzymes are, by order of activity, 1) ALAS1, δ-aminolevulinic acid synthase 1; 2) ALAD, δ-aminolevulinic acid dehydratase; 3) HMBS, hydroxymethylbilane synthase; 4) UROS, uroporphyrinogen synthase; 5) UROD, uroporphyrinogen decarboxylase; 6) CPOX, coproporphyrinogen oxidase; 7) PPOX, protoporphyrinogen oxidase; and 8) FECH, ferrochelatase. The data was retrieved from STRING database with information from the heme biosynthesis subset.

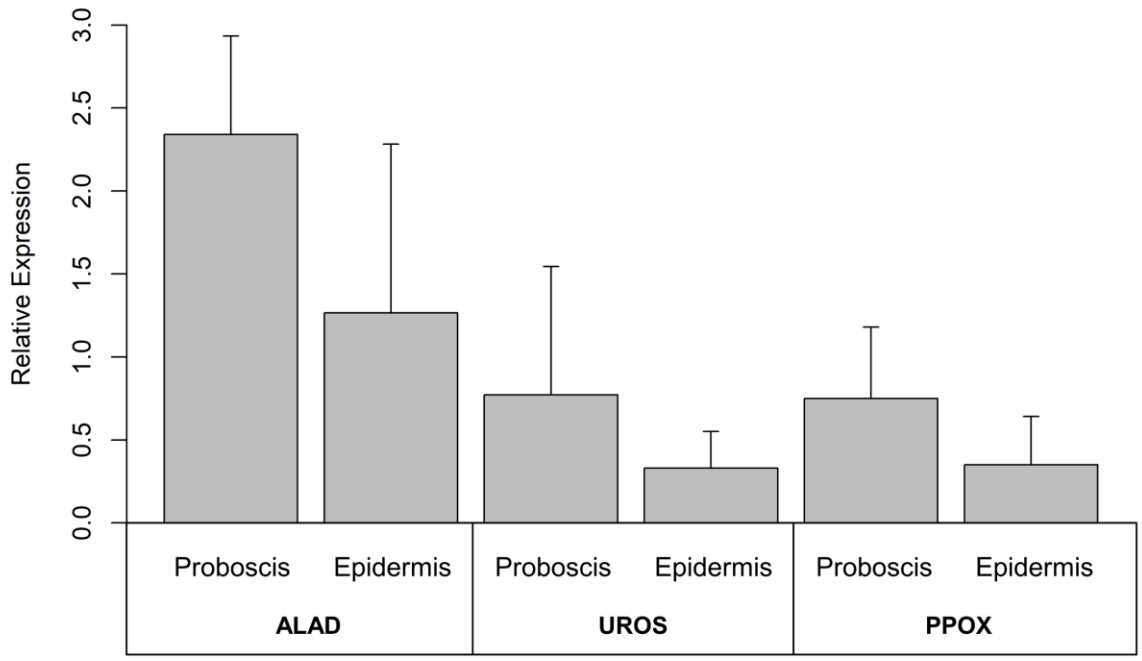


Fig.3.8. Validation of genes encoding for heme biosynthesis proteins by RT-qPCR. Detection was done for mRNAs coding for δ-aminolevulinic acid dehydratase (ALAD), uroporphyrinogen synthase (UROS) and protoporphyrinogen oxidase (PPOX). Relative expression was determined by $2^{-\Delta\Delta Ct}$. The results are presented as mean ± the standard deviation. No statistical differences were found between proboscis and epidermis.

Metabolic pathway analysis was done for the proteins in the proboscis and epidermis yielding high expression values (logTPM ≥ 2) in each subset. The enriched pathways or functions of the proteins were the same between the proboscis and epidermis (Table 3.6). The main proteins in the porphyrin Eumetazoa and heme biosynthesis subsets were found to be part of general metabolism and porphyrin biosynthesis. On the other hand, the heme degradation proteins were mainly part of drug metabolism and heme-binding compounds. The preceding subsets, heme biosynthesis and degradation, also displayed proteins within cellular oxidant detoxification. Each protein network revealed a higher number of highly expressed proteins in the epidermis than in the proboscis (Fig.3.9-3.11). Also, different variants of the same proteins were found highly expressed in the proboscis and epidermis, and also variants with high expression values in both organs were identified (Table 3.7).

The porphyrin Eumetazoa subset revealed highly expressed ALAD, UROD, and FECH proteins with different variants for the proboscis and epidermis (Fig.3.9). Another protein with different highly expressed variants in each organ is the cytochrome P450 1A1 (CYP1A1). Contrarily, the proteins cytochrome P450 1A2 (CYP1A2) and cytochrome c oxidase assembly protein COX15 homolog (COX15) were only found highly expressed in the epidermis. The overall view of the proboscis and epidermis networks, not only with specific organ variants, yielded in both organs the same three highly expressed proteins, ALAD, UROD, and FECH with specific and common variants for the proboscis and epidermis (Table 3.7). Both organs also have the following heme-related proteins highly expressed, ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial (CLPX), broad substrate

specificity ATP-binding cassette transporter (ABCG2), translocator protein (TSPO), and cytochrome c (CYCS).

The highly expressed proteins in the heme biosynthesis subset also revealed different protein variants of UROD and FECH specific for either the proboscis or epidermis (Fig.3.10). Both organs also yielded different highly expressed variants of the proteins glutathione S-transferase A1 (GSTA1) and peroxiredoxin-1 (PRDX1). On the other hand, the protein COX15 was only found highly expressed in the epidermis. In the overall view, with all the highly expressed proteins in the proboscis and epidermis, both organs displayed the same three major proteins ALAD, UROD, and FECH, having the last two organ-specific and non-specific variants (Table 3.7). The proteins heme oxygenase 2 (HMOX2) and protoheme IX farnesyltransferase (COX10) were also found highly expressed in both organs.

Lastly, the heme degradation subset yielded different highly expressed variants for the proboscis and epidermis of GSTA1, cytochrome P450 2C9 (CYP2C9), cytochrome P450 3A4 (CYP3A4), hematopoietic prostaglandin D synthase (HPGDS), UDP-glucuronosyltransferase 2A3 (UGT2A3), and haptoglobin (HP) (Fig.3.11). However, only the epidermis displayed organ-specific highly expressed variants for glutathione S-transferase A2 (GSTA2), glutathione S-transferase A3 (GSTA3), and cytochrome P450 2C19 (CYP2C19). Nonetheless, in the wide-ranging view of the proboscis and epidermis pathways, the same core proteins are highly expressed, HMOX2, GSTA1, GSTA2, GSTA3, CYP2C9, CYP3A4, HPGDS, and UGT2A3.

Table 3.6. Significant enriched pathways and functions in the proboscis and epidermis of *Eulalia* sp. Enriched pathways were selected by the highly expressed proteins in each subset: porphyrin Eumetazoa, heme biosynthesis and heme degradation. The proteins were selected according to LogTPM≥2 threshold. The proteins were divided into four categories: highly expressed proteins/protein variants only found in the proboscis or epidermis (Specific) and all highly expressed proteins/protein variants in the proboscis or epidermis (All). The significance values represent the adjusted p-value (FDR) for each pathway or function. The data was retrieved from STRING database.

Subset	Enriched pathways or functions	Proboscis		Epidermis	
		Specific	All	Specific	All
Porphyrin Eumetazoa	Metabolism	7.8E-04	6.41E-05	8.95E-06	1.17E-06
	Porphyrin biosynthesis	1.43E-08	5.42E-07	3.10E-07	1.14E-06
	Porphyrin-containing compound metabolic process	4.73E-09	2.46E-09	7.16E-16	3.17E-14
Heme biosynthesis	Metabolism	8.3E-04	4.81E-21	4.71E-10	3.80E-24
	Porphyrin biosynthesis	1.3E-04	2.90E-05	5.50E-04	4.08E-05
	Cellular oxidant detoxification	8.72E-06	7.96E-05	2.00E-03	9.69E-06
Heme degradation	Cellular oxidant detoxification	8.52E-07	3.58E-06	1.45E-06	2.82E-07
	Heme binding	2.49E-11	1.70E-18	1.96E-14	2.36E-23
	Drug metabolism - cytochrome P450	1.48E-08	4.09E-10	1.22E-14	3.94E-11

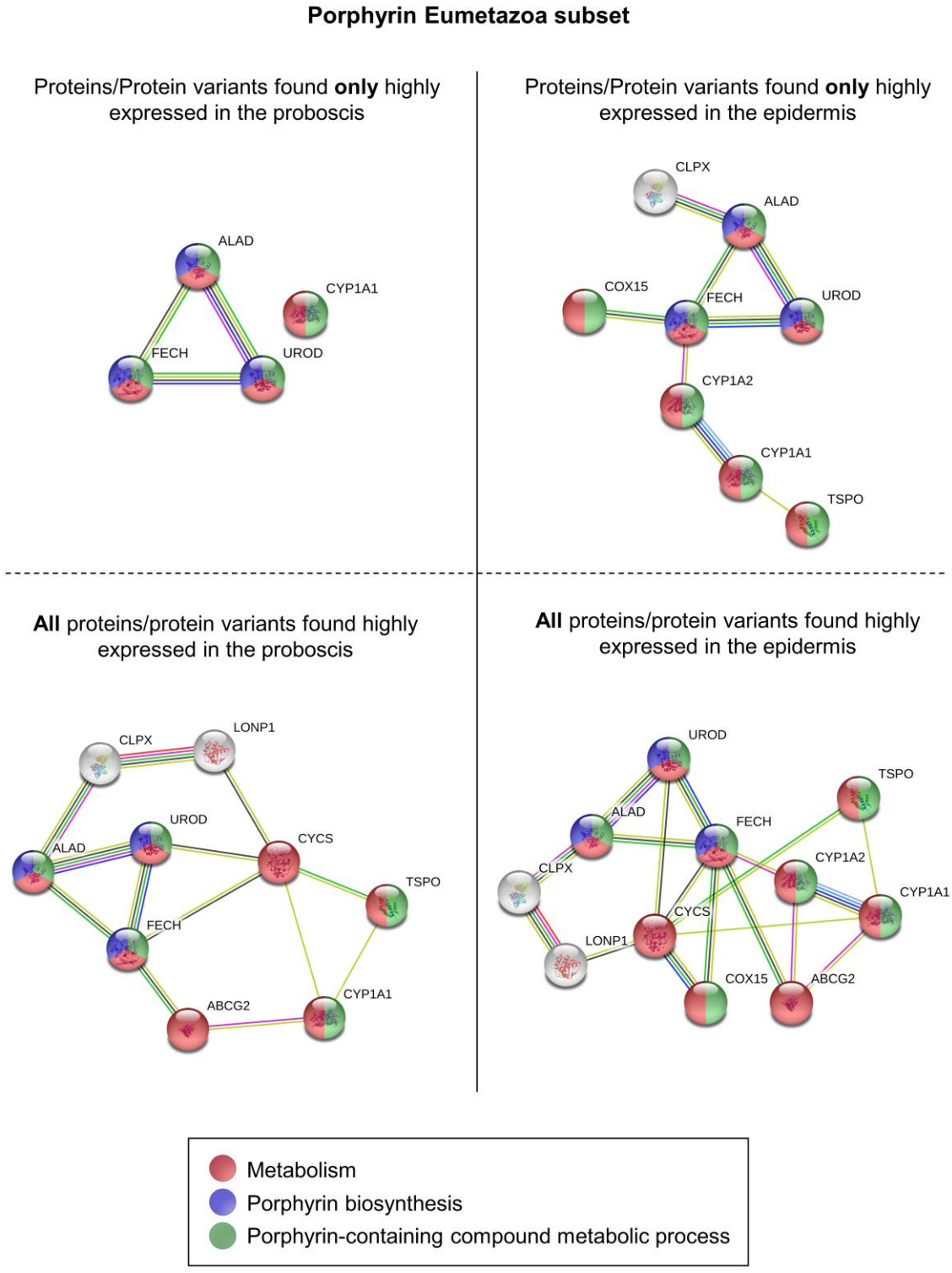


Fig.3.9. Protein-protein connections and respective pathways or functions of the highly expressed proteins in the porphyrin Eumetazoa subset. Schematic representation of the highly expressed proteins in the porphyrin Eumetazoa subset. The proteins are divided into highly expressed proteins/protein variants only found in the proboscis or epidermis and all highly expressed proteins in each organ. The all highly expressed proteins in each organ, include specific and common variants. The criteria for highly expressed protein was logTPMs≥2. The data was retrieved from STRING database.

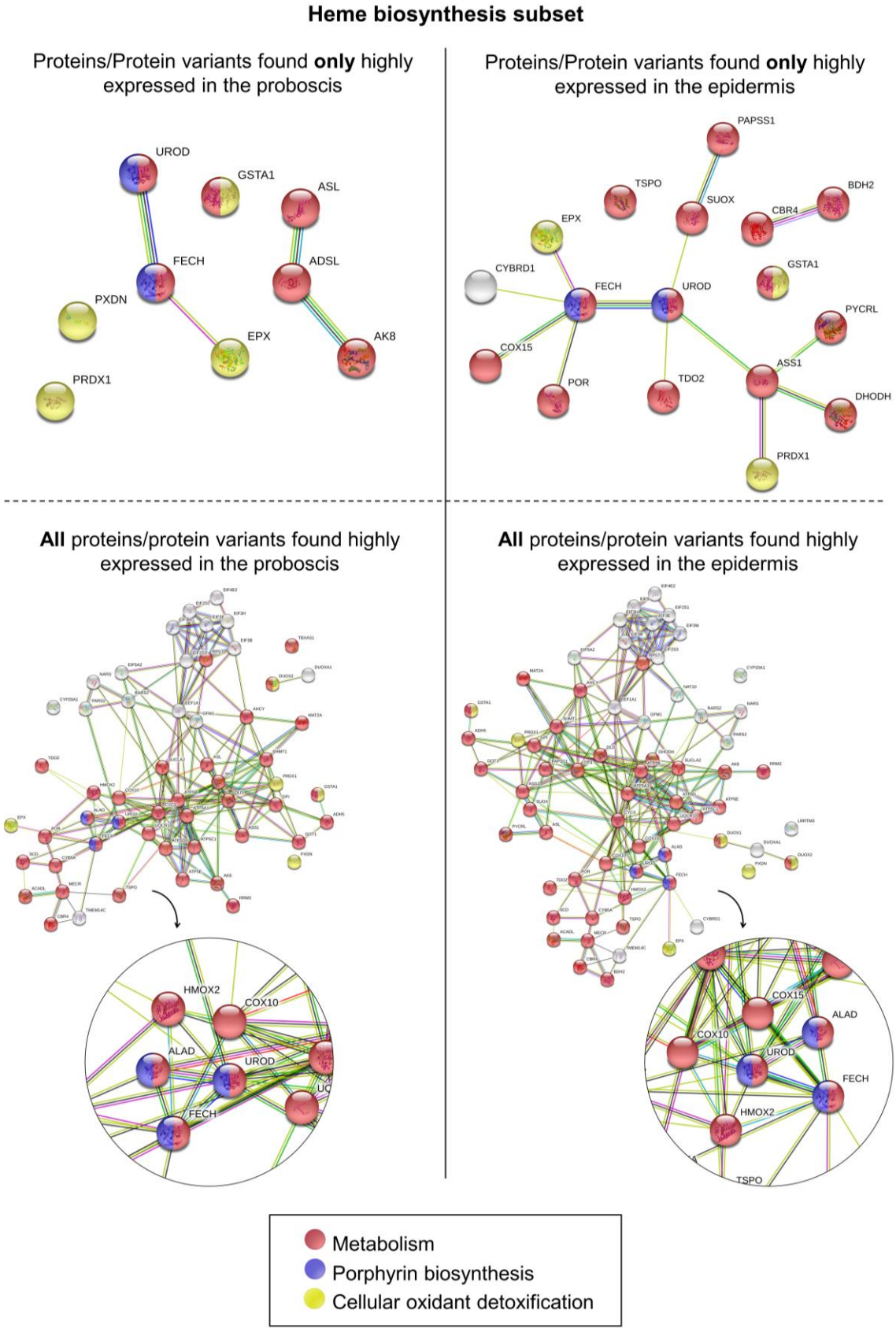


Fig.3.10. Protein-protein connections and respective pathways or functions of the highly expressed proteins in the heme biosynthesis subset. Schematic representation of the highly expressed proteins in the heme biosynthesis subset. The proteins are divided into highly expressed protein variants only found in the proboscis or epidermis and all highly expressed proteins in each organ. The all highly expressed proteins in each organ, include specific and common variants. The criteria for highly expressed protein was logTPMs≥2. The data was retrieved from STRING database.

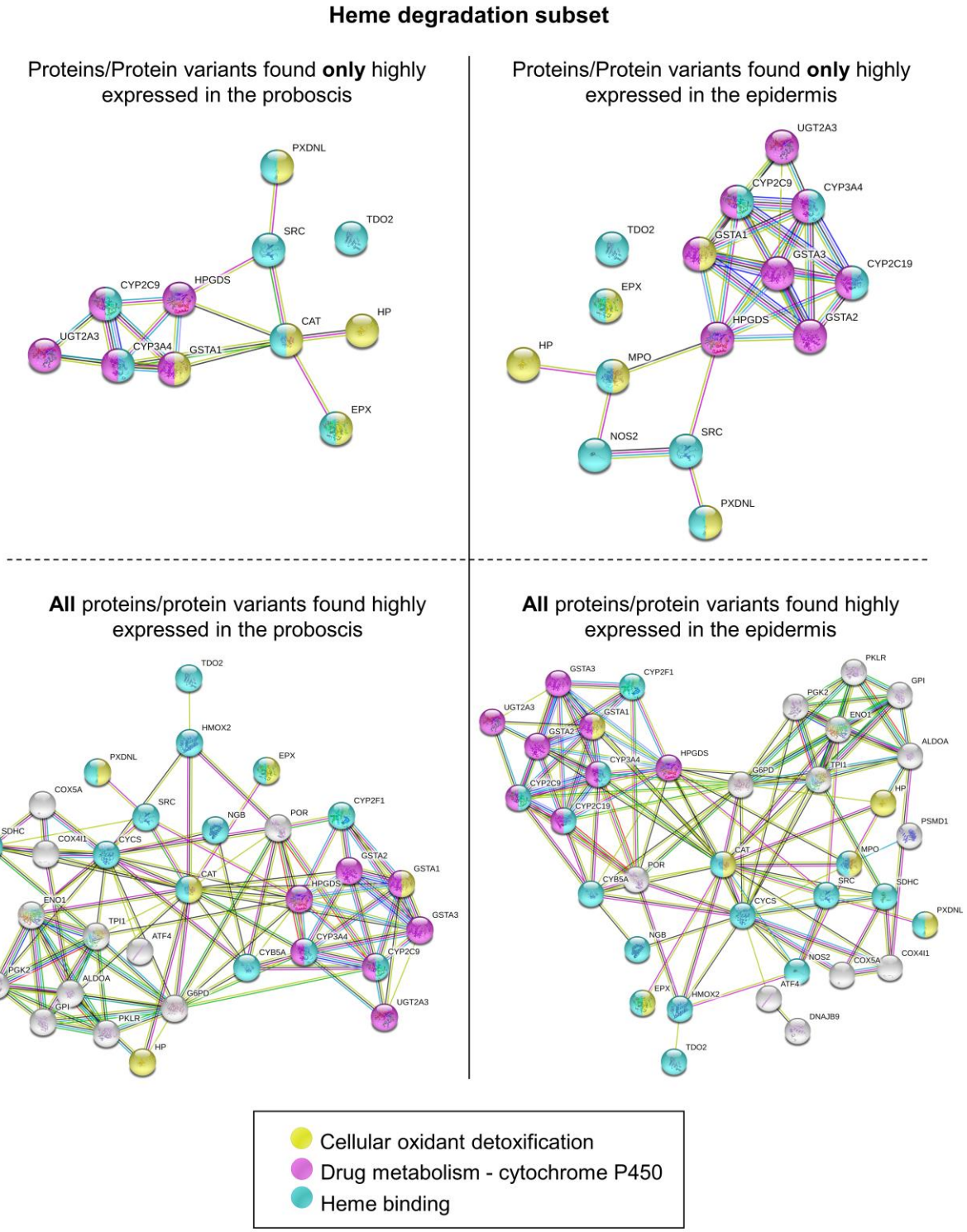


Fig.3.11. Protein-protein connections and respective pathways or functions of the highly expressed proteins in the heme degradation subset. Schematic representation of the highly expressed proteins in the heme degradation subset. The proteins are divided into highly expressed protein variants only found in the proboscis or epidermis and all highly expressed proteins in each organ. The all highly expressed proteins in each organ, include specific and common variants. The criteria for highly expressed protein was logTPMs≥2. The data was retrieved from STRING database.

Table 3.7. Protein variants highly expressed in the proboscis and epidermis of *Eulalia* sp. For each subset, porphyrin Eumetazoa, heme biosynthesis and heme degradation, the number of highly expressed variants only found in the proboscis, epidermis or even in both organs were registered for heme metabolism related enzymes.

Subset	Proteins	Proboscis	Epidermis	Proboscis and Epidermis
Porphyrin Eumetazoa	ALAD	2	2	16
	UROD	2	2	3
	FECH	8	5	8
Heme biosynthesis	ALAD	0	0	1
	UROD	2	4	4
	FECH	8	6	15
	HMOX2	0	0	1
Heme degradation	HMOX2	0	0	1

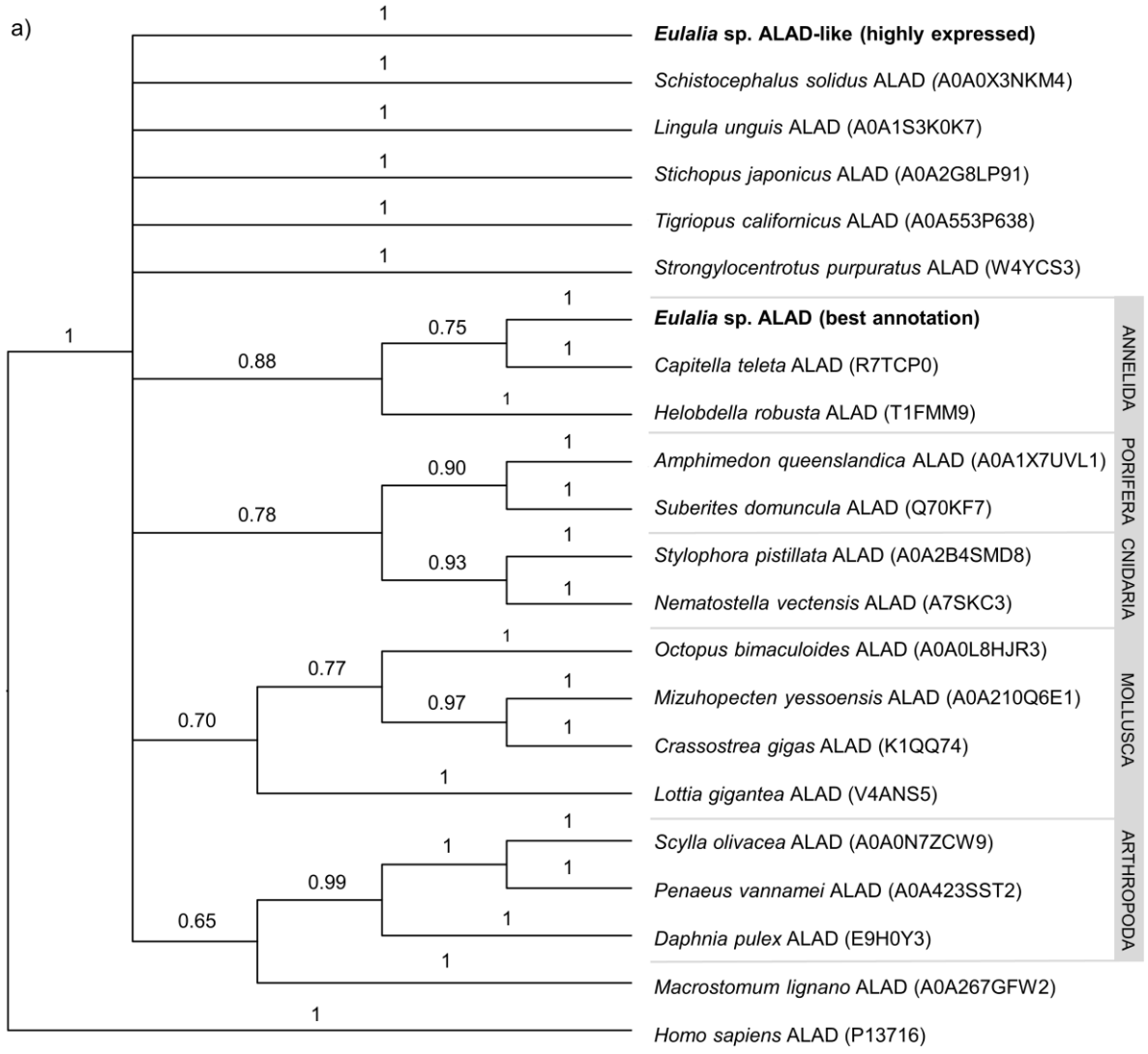
ALAD, δ-aminolevulinic acid dehydratase; UROD, uroporphyrinogen decarboxylase; FECH, ferrochelatase; and HMOX2, heme oxygenase 2.

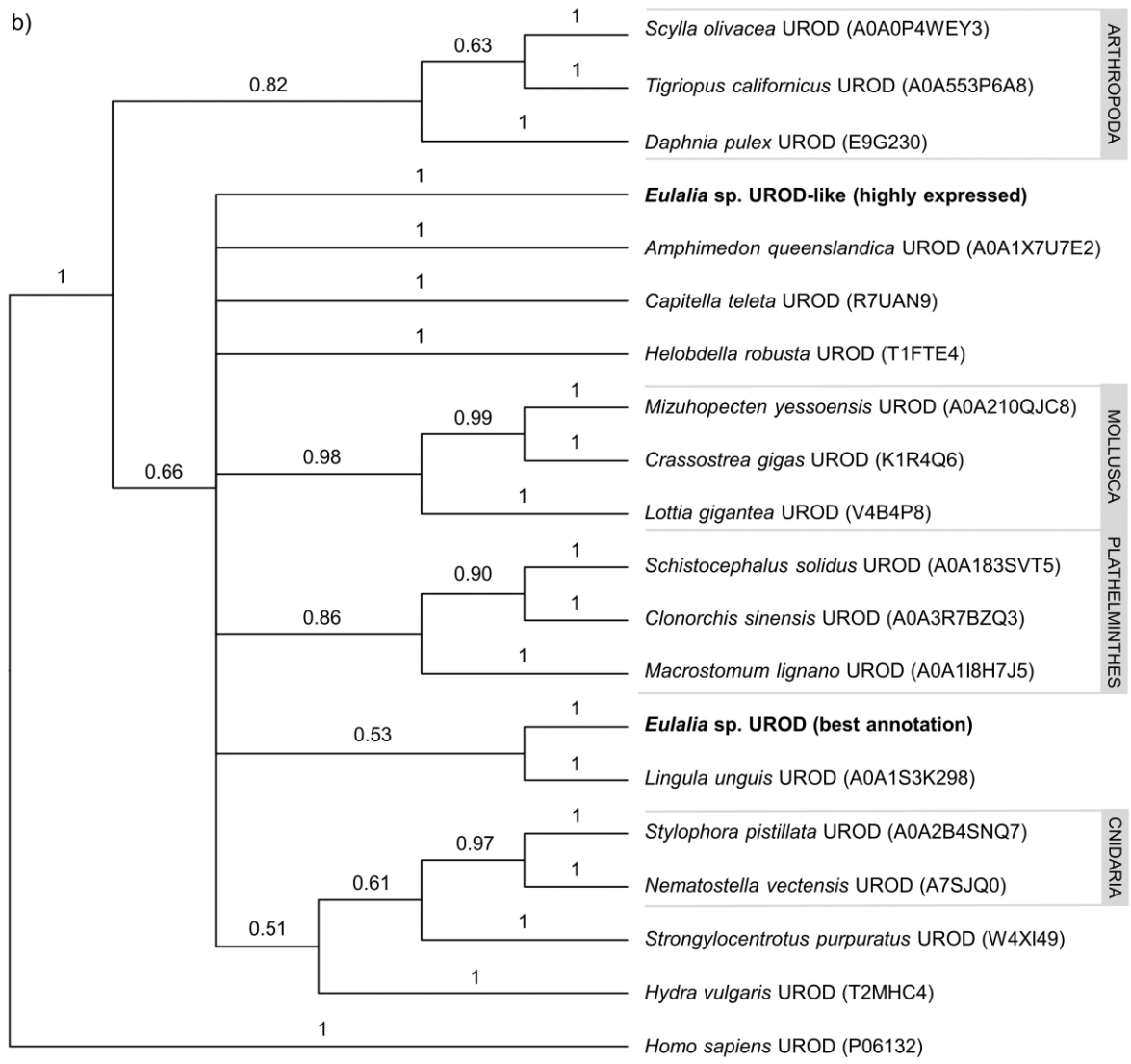
3.2.3 Enzyme variants and homologies to other Metazoa organisms

Three core heme biosynthesis proteins, ALAD, UROD, and FECH were found highly expressed in the proboscis and epidermis with different and common variants (recall Table 3.7). The sequences pertaining to ALAD, UROD, and FECH from *Eulalia* were compared to annotated Metazoa UNIPROT sequences of marine invertebrates for validation (Fig.3.10). The *Eulalia* sp. sequences selected were, the best annotated highly expressed sequence, that ended up being a variant present in both the proboscis and epidermis, and the best annotated sequence regardless of expression. Best annotation was selected by the lowest e-value in both conditions mentioned and indicated better annotation of sequences regardless of expression than with significant expression (data not showed). This observation is in accordance with the results of the phylogenetic trees (Fig.3.10). The ALAD, UROD, and FECH highly expressed sequences were identified as ALAD-like, UROD-like, and FECH-like since they displayed weak homologies within other Metazoa sequences. On the contrary, the sequences with the best annotations and no significant expression yielded good homologies to other marine invertebrates. Nonetheless, some of the Metazoa sequences themselves do not display the expected clustering of organisms from the same phylum.

The ALAD phylogenetic tree yielded a clear taxonomical organization between Metazoa sequences, except for some poorly annotated organisms such as the highly expressed protein variant of *Eulalia* (ALAD-like) (Fig.3.10a). On the contrary, the best annotated *Eulalia* sequence had 75% homology with *Capitella teleta*, another member of the Polychaeta class, and both have homology with *Helobdella robusta*, all members of the phylum Annelida. The UROD and FECH phylogenetic trees (Fig.3.10b and c, respectively) revealed less evident homologies between taxonomically-related sequences, indicating a worse annotation in general. Both *Eulalia*'s best annotated sequence of UROD and FECH revealed

better homology than the highly expressed sequence, but with organisms unrelated to Polychaeta or Annelida. The best annotated *Eulalia* UROD protein yielded higher similarity to *Lingula unguis* from the Phylum Brachiopoda and FECH with *Halichondria panicea* belonging to Phylum Porifera.




b)
Scylla olivacea UROD (A0A0P4WEY3)
Tigriopus californicus UROD (A0A553P6A8)
Daphnia pulex UROD (E9G230)
ARTHROPODA
Eulalia sp. UROD-like (highly expressed)
Amphimedon queenslandica UROD (A0A1X7U7E2)
Capitella teleta UROD (R7UAN9)
Helobdella robusta UROD (T1FTE4)
Mizuhopecten yessoensis UROD (A0A210QJC8)
Crassostrea gigas UROD (K1R4Q6)
Lottia gigantea UROD (V4B4P8)
MOLLUSCA
Schistocephalus solidus UROD (A0A183SVT5)
Clonorchis sinensis UROD (A0A3R7BZQ3)
Macrostomum lignano UROD (A0A1I8H7J5)
PLATHELMINTHES
Eulalia sp. UROD (best annotation)
Lingula unguis UROD (A0A1S3K298)
Stylophora pistillata UROD (A0A2B4SNQ7)
Nematostella vectensis UROD (A7SJQ0)
CNIDARIA
Strongylocentrotus purpuratus UROD (W4XI49)
Hydra vulgaris UROD (T2MHC4)
Homo sapiens UROD (P06132)
1
0.63
0.82
0.99
0.98
0.66
0.90
0.86
0.53
0.97
0.61
0.51

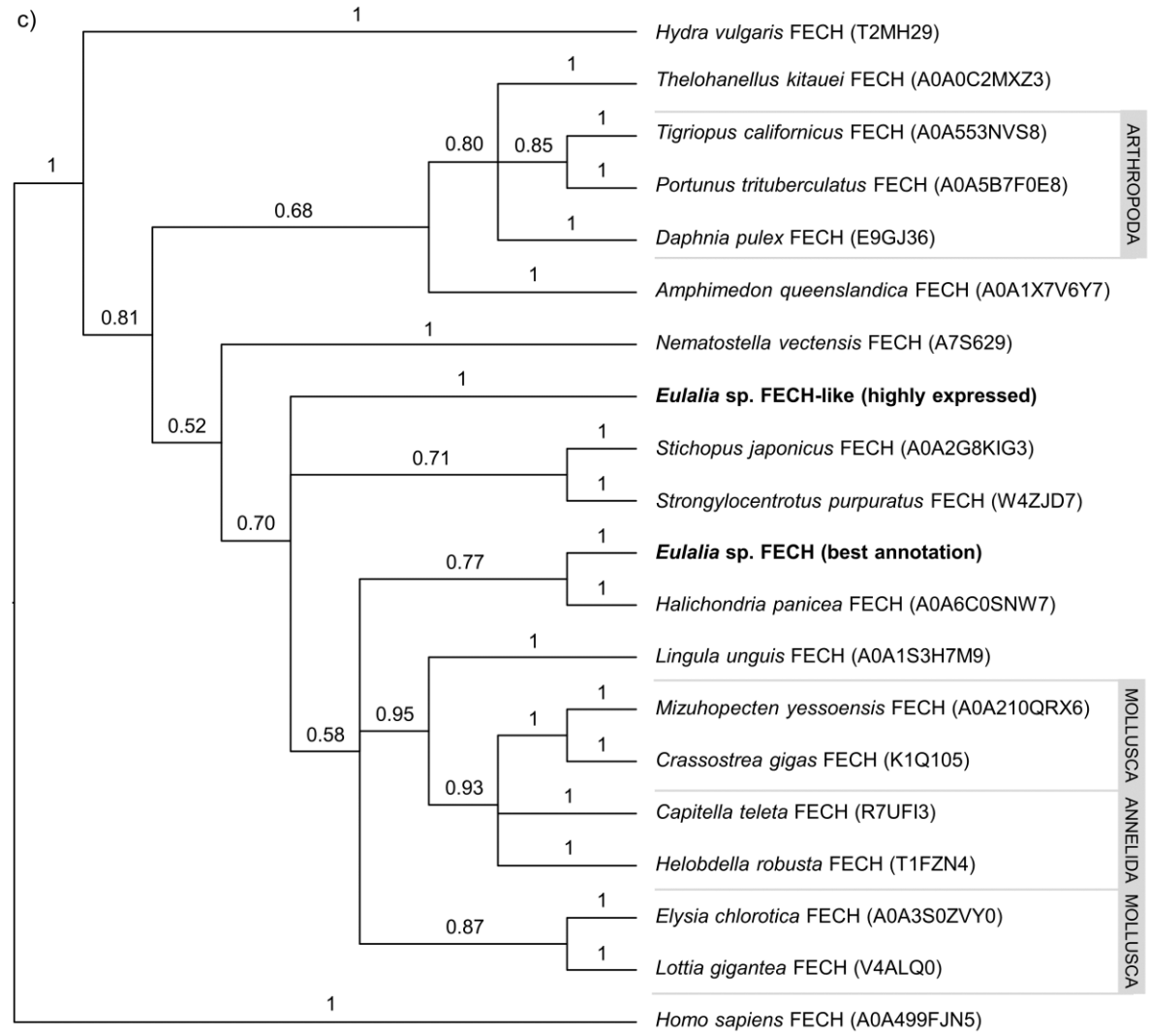


Fig.3.12. Phylogenetic trees comparing the sequences of three main heme biosynthesis proteins. Each dendrogram was inferred by the Maximum Likelihood method and LG gamma distribution model (consensus tree obtained from 100 0000 iterations). The sequences used were from *Eulalia* sp. predicted transcriptome and Metazoa sequences from UNIPROT for a) ALAD, δ-aminolevulinic acid dehydratase: b) UROD, uroporphyrinogen decarboxylase; and c) FECH, ferrochelatase. The *Eulalia* sp. sequences selected were the best annotated highly expressed sequence (highly expressed) and the best annotated sequence regardless of expression (best annotation). Analyses were conducted in MrBayes.

4. Discussion

Eulalia sp. is known for its bright green coloration, which results from the multiple greenish and yellowish pigments found in this worm. These pigments are detected by their strong absorption in the UV zone and the visible light spectrum, mostly in the violet range (400 nm) for yellow pigments and in the red region (700 nm) for green pigments. Each organ, namely the proboscis, epidermis, intestine, and oocytes have different pigment signatures as well as some closely related pigments between organs. Greenish pigments are found in almost every organ, while the yellowish pigments are mostly restricted to the proboscis and somewhat the epidermis. Overall, the green pigments are more diverse and widespread than the yellow pigments. However, the porphyrinoid nature of the pigments is more noticeable in the yellow than in the green pigments.

The major pigments in each organ are classified as green or yellow according to absorption maxima and emitted color. On total, three yellow pigments and nine green pigments were identified. The two main yellow pigments identified in the proboscis (Pr2) and epidermis (Ep2) display the same retention times and matching absorbance spectra. These findings indicate that these pigments are most likely the same compound or in some way chemically-related. In turn, the green pigments display a wider variety and also higher interindividual variability when in comparison to the yellow pigments. However, there are also similarities between green pigments from the epidermis, intestine, and oocytes since the results support that the compounds termed Int1 (intestine) and Oo3 (oocytes) are most likely the same pigments, similarly to pigments Ep4, Int3, and Oo5 (from the epidermis, intestine and oocytes, respectively). The existence of common pigments between organs can indicate translocation, especially between the intestine and oocytes which have very similar chromatograms and individual pigment spectra.

The porphyrinoid nature of the pigments was confirmed by the presence of the Soret and Q bands. Also, no resemblances were found between the absorbance spectra of *Eulalia*'s pigments and chlorophylls or carotenoids which represent the most widely known greenish and yellowish pigments in plants and animals, respectively (see Linschitz & Sarkanen, 1958 and Pereira et al., 2014 for absorption spectra of chlorophylls and carotenoids, respectively). Altogether, the present findings suggest that the endogenous pigments of *Eulalia* sp. are most likely heme-derived metabolites since it is the most notorious known porphyrin. In fact, when comparing the absorbance spectra of *Eulalia*'s pigments to the heme byproducts biliverdin and bilirubin (van Dijk et al., 2017), some similarities were found. The yellow pigments have the same absorbance peak as bilirubin near 435 nm and the green pigments have similar absorbance peaks to biliverdin around 400 nm and 670 nm. Therefore, *Eulalia*'s green and yellow pigments are most likely analogous to biliverdin and bilirubin, respectively.

The current findings support the original study by Martins et al. (2019), where the major porphyrin-like pigments of *Eulalia* sp. were identified. This study also highlighted pigment distribution in each organ and revealed that the pigments are mainly stored in non-fluorescent pigment granules. In the proboscis and epidermis, these granules are located in specialized pigment cells, being more abundant in the epidermis than in the proboscis. Additionally, the stability of the pigments was also investigated, and the pigments were found to be stable after several weeks of storage in freezing conditions (- 20ºC). In

particular, the major pigments from the proboscis and epidermis which has important implications for the potential applications of these compounds. The pigment ID classification by Martins et al. (2019) was used to identify the pigments in the current study, but some differences were found, showing some degree of natural variability that must be considered in future studies. The following three pigments could not be identified in the current study: Ep1, a yellow pigment from the epidermis (similar to Pr1 from the proboscis), Int2, a green pigment from the intestine, and Oo1, a green pigment from the oocytes. However, Martins et al. (2019) also reported that these pigments were not always found in the sample extracts, mostly because of their low concentrations. Furthermore, the retention times of all the identified pigments were earlier by a few minutes, probably due to the alterations made in the chromatography settings. Also, in both studies, the pigments' maximum peaks registered in the chromatograms were almost identical for the proboscis and epidermis but distinct in the intestine and oocytes. This indicates that the pigments' abundance in the proboscis and epidermis seems to be stable. However, in the intestine and oocytes, their abundance varies constantly.

According to Rodrigo et al. (2015), who studied the structure of the gut epithelia along the digestive cycle, the several pigments found throughout the digestive epithelia are seemingly metabolized and eliminated in this tissue which can explain the variability of the pigments found in this organ. In fact, quite a similar process is described in the Polychaeta *H. diversicolor* where the epidermal granules of biliverdin seem to be removed by coelomic cells to the gut for excretion (Dales & Kennedy, 1954). If a similar process occurs in *Eulalia* sp. it can explain the possible translocation of pigments from the epidermis to the intestine. Also, coelomocytes in this same species are known to transfer nutrients to the oocytes during vitellogenesis (Olive, 1975), from which it may be expected pigment transfer as well.

Besides the seemingly variation of the pigments from the intestine and oocytes, the latter were difficult to retrieve due to the low number of maturing females. It must be noticed that worm harvesting was done in the fall-winter period whereas vitellogenesis takes place in early spring, as described elsewhere (Olive, 1975). Hence, pigment isolation in the present work was chiefly done for the proboscis and epidermis which yielded less pigment variation. The isolated pigments were the major yellow pigment in the proboscis (Pr2), and its less abundant correspondent in the epidermis (Ep2), followed by the two main green pigments in the epidermis (Ep3 and Ep4). However, the Ep3 and Ep4 pigments have very close retention times which makes it difficult to obtain the individual pigments without contamination from one another. Thus, the yellow pigment, identified as Pr2 in the proboscis and Ep2 in the epidermis, is the easiest pigment to isolate, as well as the pigment with the most defined Soret and Q bands which confirms its porphyrinoid nature. Altogether, these characteristics make this pigment a promising candidate for further biotechnological research.

Masking of the Soret and Q bands was observed in all green pigments of *Eulalia* sp. which can be explained by porphyrin agglomeration. A recent study by Zannotti et al. (2018) demonstrated that low porphyrin concentration can induce aggregation (possibly in the form of dimers) and result in spectral changes, like widening and reduced definition of the Soret and Q bands. Even though the concentrations of the pigments were not estimated, the differences between the magnitude of the pigments' maxima absorbance registered suggest that the concentration of the yellow pigments is superior to the green

pigments. This can explain the added difficulty in retrieving the standard porphyrinoid signature in the green pigments of *Eulalia* sp. Also, a study by D'Ambrosio et al. (2020) demonstrated that crude pigment extracts from the proboscis, which is rich in yellow pigments, displayed higher light-mediated toxicity in comparison to the extracts from the epidermis that bears mostly green pigments. This report confirms the porphyrinoid properties, namely photoactivation, found in the yellow pigments that may be diminished in green pigments.

The strong absorption of UV light (~280 nm) in the spectra of all the identified pigments suggests that these pigments have a role in protecting the worm against UV-induced damage, similar to melanin's (Boulton, 1990). Also, most porphyrinoid pigments are known for their photodynamic properties and can lead to the production of reactive oxygen species (ROS) upon irradiation with UV or visible light (Afonso et al., 1999). In fact, *Eulalia* sp. usually seeks protection from light underneath mussel beds (Emson, 1977), showing that this adaptive feature to actively avoid damaging radiation while roaming the intertidal is probably due to its photodynamic pigments. However, it must be noted that porphyrinoids like bile pigments are not only known to be strongly photodynamic and therefore phototoxic but can also possess antioxidant and anti-mutagenic properties (Bulmer et al., 2008). Therefore, the porphyrinoid pigments of *Eulalia* sp. might have other important properties besides photoactivation.

Since the last century, human and other vertebrates' metabolism involving heme formation and its degradation to bile pigments has received much attention in biomedical research (Lester & Troxler, 1969). However, little is known about the metabolism of porphyrins in marine invertebrates, especially in the Class Polychaeta. Unlike Polychaeta and other marine invertebrates, mammals possess specific organs and tissues for heme synthesis and degradation such as bone marrow, spleen, and liver (Orten, 1971; Ajioka et al., 2006). Yet, *Eulalia* sp. has specialized pigment cells in the proboscis and epidermis that according to the transcriptomic analysis done in this study are capable of producing heme or heme-derived metabolites since all eight canonical enzymes involved in this process were successfully identified. These results indicate that the porphyrinoid pigments of *Eulalia* sp. are in fact heme by-products, which is in accordance with the chemical characterization of the pigments. Also, no significant proteins related to chlorins were found, the other major class of tetrapyrroles analogous to porphyrins that are associated with chlorophylls.

Different gene expression patterns were found between the proboscis and epidermis, which contributes to explaining the differences between pigment metabolites from the two organs. However, it is important to mention that when searching for homologies between the translated transcripts of *Eulalia* sp. and known protein sequences related to heme metabolism, some challenges were encountered. Even though proteins related to heme metabolism were identified, so were many proteins related to general metabolism. Since the epidermis of *Eulalia* sp. is a relatively extensive and exposed tissue in comparison to the proboscis, which has a more structural role for predation and sensing (see Rodrigo et al., 2018), this can explain why more proteins were found significantly expressed and with a more complex protein network in the epidermis. Therefore, in the present study, the major differences found truly related to heme metabolism were not in the protein networks of each organ, but in the different variants of the heme biosynthesis proteins found highly expressed in the proboscis and epidermis.

The major heme biosynthesis enzymes found highly expressed in the proboscis and epidermis were proteins similar to δ-aminolevulinic acid dehydratase (ALAD), uroporphyrinogen decarboxylase (UROD) and ferrochelatase (FECH), with different and common variants for the two organs. These highly expressed variants were classified as similar proteins since the best annotated variant revealed weak homology to other marine organisms. On the other hand, some variants of these enzymes, with low expression levels, yielded good matching to ALAD, UROD, and FECH as they demonstrated good homologies to the same proteins in other marine invertebrates, which can be verified by the phylogenetic models (recall Fig. 3.12). It is the special case of ALAD, which was found to be included in the same cluster as other members of Annelida. These three proteins appear to be central in the metabolism of heme since they correspond to the beginning (ALAD), middle (UROD), and end (FECH) of the enzymatic cascade known to form heme (recall Fig.1.1). Also, these results indicate that *Eulalia* sp. might not only have the canonical forms of these proteins for heme biosynthesis but also specific variants for the production of different heme-like products.

The main source of heme for bile pigment synthesis in most vertebrates and some invertebrates, like *H. diversicolor*, derives from the catabolism of hemoglobin (Dales & Kennedy, 1954; Orten, 1971). However, in *Eulalia* sp. few blood vessels are observed (Rodrigo et al., 2018). This evidence and the results of this study suggest that *Eulalia* sp. produces heme for hemoglobin on a lower scale when in comparison with the amount of heme-like products spent in green and yellow pigment production. Since the enzyme heme oxygenase 2 (HMOX2) was found highly expressed in both the proboscis and epidermis, this observation suggests that these heme-like products are then converted to biliverdin or a similar compound. Usually, biliverdin is then rapidly converted to bilirubin by biliverdin reductase (BLVRA), and although this enzyme was not found within the most highly expressed proteins, at least one enzyme with a good annotation value was found with a basal expression in both organs (data not shown). This indicates that both organs may indeed produce bilirubin or similar pigments.

The enzymes responsible for the conversion of heme to heme O (COX10) and sequentially heme A (COX15) were found highly expressed (logTPM ≥ 2) in both organs and epidermis, respectively. Even though COX15 was not identified within the highly expressed proteins of the proboscis, it yielded a logTPM value of 1.6 (data not shown). Therefore, both organs can produce heme A which is a prosthetic group of the protein cytochrome c oxidase. Also, both organs have the protein cytochrome c (CYCS) highly expressed. These are both conserved proteins in eukaryotic cells and are crucial for mitochondrial respiration (Kim et al., 2012).

The protein networks of the proboscis and epidermis also highlighted other processes within heme metabolism, like pathway regulation and transportation of intermediates. For example, the enzyme ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial (CLPX) was found highly expressed in both organs. This enzyme has an important role in the activation of δ-aminolevulinic acid synthase (ALAS) (Kardon et al., 2015) which is the first enzyme in the heme biosynthetic pathway. Also, genes coding for proteins involved in porphyrin transportation, namely Broad substrate specificity ATP-binding cassette transporter (ABCG2) and translocator protein (TSPO) were also found with high levels of expression in both organs. These enzymes have central roles in regulating the accumulation of the

last intermediate of heme biosynthesis, protoporphyrin IX, which is an endogenous photosensitizer (Kobuchi et al., 2012; Veenman et al., 2016). Thus, these two enzymes avert the production of reactive oxygen species (ROS) and subsequent metabolic deregulation. Other enzymes with protective roles against oxidative stress were also found highly expressed in both organs such as glutathione S-transferase A (GSTA1) and peroxiredoxin-1 (PRDX1) (see for instance, Suvakov et al., 2013 and Ding et al., 2017 for further details on eukaryotic antioxidant enzymes). Since the deregulation of the heme biosynthetic pathway can lead to the production of free radicals, ROS included, both GSTA1 and PRDX1 might play a regulatory role in heme biosynthesis. Lastly, several cytochromes of the P450 family have a significant representation in the proboscis and epidermis networks which makes sense, since heme is an essential prosthetic group of all P450s (Correia et al., 2011), therefore benefitting from heme recycling and biosynthesis as well. The major highly expressed enzymes of this family identified were CYP1A1 for both organs and CYP1A2 for the epidermis which are both possibly related to bilirubin degradation (Sinal & Bend, 1997).

Overall, the present findings indicate that the proboscis and epidermis of *Eulalia* sp. have conserved heme metabolic pathways. These include the canonical proteins involved in heme biosynthesis and degradation, as well as proteins for pathway regulation and transportation, described for higher-order Metazoa organisms, particularly vertebrates. This outcome highlights that porphyrinoid metabolism is well conserved among the animal kingdom. A major novelty found in this study was the several variants of the heme biosynthesis proteins for the proboscis and epidermis. The fact that this species has different protein variants can explain the variety of yellow and green pigments found in this worm, rather than the typical biliverdin and bilirubin. Also, having specific variants mostly expressed in one organ in relation to the other as well as common variants can explain the differences and similarities between pigment signatures found in the proboscis and epidermis. Altogether, it is clear that *Eulalia* is well-equipped to metabolize and recycle heme into a far wider range of porphyrinoids than mammals and other vertebrates as well as confer the species with adaptive leverage towards its specific environment.

One of the main questions that remains unanswered is why the yellow pigmentation is mostly restricted to the proboscis and the green pigmentation to the epidermis. Previous studies have described that biliverdin accumulation is usually associated to the lack of BLVRA (Langille & Youson, 1983; Hongo et al., 2017). Most likely, in *Eulalia* sp., the heme-related products in the proboscis are rapidly converted to compounds similar to biliverdin and then to bilirubin. On the other hand, in the epidermis, even though the enzyme responsible for bilirubin production was also detected, it seems that its expression might be absent in the majority of the epidermal cells, resulting in the accumulation of green pigments in the uncanny specialized pigment cells described by Rodrigo et al. (2018) and Martins et al. (2019). The function of these specialized cells is not yet fully understood, it can be related to UV protection, chemical defense and even photosensing.

5. Conclusion

Marine natural pigments hold diverse properties with interest in the biomedical field. The Polychaeta *Eulalia* sp. is no exception with its uncanny green pigmentation given by the multiple greenish and yellowish porphyrinoid pigments many of which seemingly photoactive. The species' endogenous pigments are heme-derived products as indicated by chemical characterization and the transcriptomic approach on the worm's main organs displaying pigmentation, the proboscis and epidermis. The several protein variants of the heme biosynthesis pathway found in these two organs are reflected on diverse heme-related products that are then converted to pigments similar to biliverdin or bilirubin. Also, having found specific and common variants of these proteins can explain the similar and different pigmentation patterns between the proboscis and epidermis. As for the remaining organs of *Eulalia* sp., the intestine is most likely the main tissue for excretion, even though this process remains mainly obscure.

Altogether, even though marine animals still have poor genomic annotation, which is itself both a hindrance and a challenge to transcriptomics, omics can offer new perspectives on the networks underlying the synthesis of secondary metabolites bearing high biotechnological value. In the specific case, porphyrinoids are prized targets due to their value as photosensitizers in photodynamic therapy. Such advantages apply especially to the specific yellowish pigment identified in the proboscis and epidermis (here identified as Pr2 and Ep2, respectively) although less concentrated in the latter. Besides being the easiest pigment to extract, it seems to be stable throughout different studies and has the strongest porphyrinoid properties. Therefore, further studies should focus on isolating this candidate pigment as well as one of the green pigments for comparison. Not only to assess the potential photodynamic properties of these pigments but also to elucidate the exact chemical structure of these compounds.

6. References

Abrahamse, H., & Hamblin, M. R. (2016). New photosensitizers for photodynamic therapy. *Biochemical Journal*, 473(4), 347-364. doi:1042/BJ20150942

Afonso, S. G., Enriquez de Salamanca, R., & Batlle, A. D. C. (1999). The photodynamic and non-photodynamic actions of porphyrins. *Brazilian journal of medical and biological research*, 32(3), 255-266. doi:10.1590/S0100-879X1999000300002

Agius, L., Jaccarini, V., Ballantine, J. A., Ferrito, V., Pelter, A., Psaila, A. F., & Zammit, V. A. (1979). Photodynamic action of bonellin, an integumentary chlorin of *Bonellia viridis*, Rolando (Echiura, Bonelliidae). *Comparative biochemistry and physiology. B, Comparative biochemistry*, 63(1), 109-117. doi:10.1016/0305-0491(79)90242-6

Agostinis, P., Berg, K., Cengel, K. A., Foster, T. H., Girotti, A. W., Gollnick, S. O., ... & Korbelik, M. (2011). Photodynamic therapy of cancer: an update. *CA: a cancer journal for clinicians*, 61(4), 250-281. doi:10.3322/caac.20114

Ajioka, R. S., Phillips, J. D., & Kushner, J. P. (2006). Biosynthesis of heme in mammals. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1763(7), 723-736. doi:10.1016/j.bbamcr.2006.05.005

Bandaranayake, W. M. (2006). The nature and role of pigments of marine invertebrates. *Natural Product Reports*, 23(2), 223-255. doi:10.1039/B307612C

Berg, K., Selbo, P. K., Weyergang, A., Dietze, A., Prasmickaite, L., Bonsted, A., ... & Høgset, A. (2005). Porphyrin-related photosensitizers for cancer imaging and therapeutic applications. *Journal of microscopy*, 218(2), 133-147. doi:10.1111/j.1365-2818.2005.01471.x

Boulton, M., Docchio, F., Dayhaw-Barker, P., Ramponi, R., & Cubeddu, R. (1990). Age-related changes in the morphology, absorption and fluorescence of melanosomes and lipofuscin granules of the retinal pigment epithelium. *Vision research*, 30(9), 1291-1303. doi:10.1016/0042-6989(90)90003-4

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5), 525-527. doi:10.1038/nbt.3519

Bulmer, A. C., Ried, K., Blanchfield, J. T., & Wagner, K. H. (2008). The anti-mutagenic properties of bile pigments. *Mutation research/Reviews in mutation research*, 658(1-2), 28-41. doi:10.1016/j.mrrev.2007.05.001

Correia, M. A., Sinclair, P. R., & De Matteis, F. (2011). Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug metabolism reviews*, 43(1), 1–26. doi: 10.3109/03602532.2010.515222

Costa, P. M., Carrapiço, F., de Matos, A. A., & Costa, M. H. (2013). A microscopical study of the "chlorophylloid" pigment cells of the marine polychaete *Eulalia viridis* (L.). *Microscopy and Microanalysis*, *19*(S4), 15-16. doi:10.1017/S143192761300069X

D'Ambrosio, M., Santos, A. C., Alejo-Armijo, A., Parola, A. J., & Costa, P. M. (2020). Light-Mediated Toxicity of Porphyrin-Like Pigments from a Marine Polychaeta. *Marine Drugs*, 18(6), 302. doi:10.3390/md18060302

Dales, R. P., & Kennedy, G. Y. (1954). On the diverse colours of *Nereis diversicolor*. *Journal of the Marine Biological Association of the United Kingdom*, 33(3), 699-708. doi:10.1017/S0025315400026977

Ding, C., Fan, X., & Wu, G. (2017). Peroxiredoxin 1–an antioxidant enzyme in cancer. *Journal of cellular and molecular medicine*, 21(1), 193-202. doi:10.1111/jcmm.12955

Duncan, R., Faggart, M. A., Roger, A. J., & Cornell, N. W. (1999). Phylogenetic analysis of the 5-aminolevulinate synthase gene. *Molecular biology and evolution*, 16(3), 383-396. doi:10.1093/oxfordjournals.molbev.a026119

Emson, R. H. (1977). The feeding and consequent role of *Eulalia viridis* (OF Müller)(Polychaeta) in intertidal communities. *Journal of the Marine Biological Association of the United Kingdom*, 57(1), 93-96. doi:10.1017/S0025315400021251

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *evolution*, 39(4), 783-791. doi:10.2307/2408678

Fujiwara, T., & Harigae, H. (2015). Biology of heme in mammalian erythroid cells and related disorders. *BioMed research international*, 2015. doi:10.1155/2015/278536

Giovannetti, R. (2012). The use of spectrophotometry UV-Vis for the study of porphyrins. *Macro to nano spectroscopy*, 1, 87-108. doi:10.5772/38797

Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature genetics*, 3(3), 266-272. doi:10.1038/ng0393-266

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, 29(7), 644. doi:10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), 1494-1512. doi:10.1038/nprot.2013.084

Hamza, I., & Dailey, H. A. (2012). One ring to rule them all: trafficking of heme and heme synthesis intermediates in the metazoans. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research*, 1823(9), 1617-1632. doi: 10.1016/j.bbamcr.2012.04.009

Heinemann, I. U., Jahn, M., & Jahn, D. (2008). The biochemistry of heme biosynthesis. *Archives of biochemistry and biophysics*, 474(2), 238-251. doi:10.1016/j.abb.2008.02.015

Hongo, Y., Yasuda, N., & Nagai, S. (2017). Identification of genes for synthesis of the blue pigment, biliverdin IXα, in the blue coral *Heliopora coerulea*. The Biological Bulletin, 232(2), 71-81. doi:10.1086/692661

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755. doi:10.1093/bioinformatics/17.8.754

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3), 299-314.

Kardon, J. R., Yien, Y. Y., Huston, N. C., Branco, D. S., Hildick-Smith, G. J., Rhee, K. Y., ... & Baker, T. A. (2015). Mitochondrial ClpX activates a key enzyme for heme biosynthesis and erythropoiesis. *Cell*, 161(4), 858-867. doi:10.1016/j.cell.2015.04.017

Kim, H. J., Khalimonchuk, O., Smith, P. M., & Winge, D. R. (2012). Structure, function, and assembly of heme centers in mitochondrial respiratory complexes. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1823(9), 1604-1616. doi:10.1016/j.bbamcr.2012.04.008

Kim, H. R., Won, S. J., Fabian, C., Kang, M. G., Szardenings, M., & Shin, M. G. (2015). Mitochondrial DNA aberrations and pathophysiological implications in hematopoietic diseases, chronic inflammatory diseases, and cancers. *Annals of Laboratory Medicine*, 35(1), 1-14. doi:10.3343/alm.2015.35.1.1

Kobuchi, H., Moriya, K., Ogino, T., Fujita, H., Inoue, K., Shuin, T., ... & Utsumi, T. (2012). Mitochondrial localization of ABC transporter ABCG2 and its function in 5-aminolevulinic acid-mediated protoporphyrin IX accumulation. *PloS one*, 7(11), e50082. doi:10.1371/journal.pone.0050082

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6), 1547-1549. doi:10.1093/molbev/msy096

Langille, R. M., & Youson, J. H. (1983). Biliverdin accumulation in the caudal intestinal segment of juvenile adult lampreys, *Petromyzon marinus* L. *Canadian journal of zoology*, 61(8), 1824-1834. doi:10.1139/z83-235

Lester, R., & Troxler, R. F. (1969). Recent advances in bile pigment metabolism. *Gastroenterology*, 56(1), 143-169.

Linschitz, H., & Sarkanen, K. (1958). The absorption spectra and decay kinetics of the metastable states of chlorophyll A and B1. *Journal of the American Chemical Society*, *80*(18), 4826-4832. doi:10.1021/ja01551a018

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *methods*, 25(4), 402-408. doi:10.1006/meth.2001.1262

Maines, M. D. (1999). Overview of heme degradation pathway. *Current protocols in toxicology*, (1), 9-1. doi:10.1002/0471140856.tx0901s00

Manivasagan, P., Bharathiraja, S., Santha Moorthy, M., Mondal, S., Seo, H., Dae Lee, K., & Oh, J. (2018). Marine natural pigments as potential sources for therapeutic applications. *Critical reviews in biotechnology*, 38(5), 745-761. doi:10.1080/07388551.2017.1398713

Martins, C., Rodrigo, A. P., Cabrita, L., Henriques, P., Parola, A. J., & Costa, P. M. (2019). The complexity of porphyrin-like pigments in a marine annelid sheds new light on haem metabolism in aquatic invertebrates. *Scientific Reports*, *9*(1), 1-11. doi:10.1038/s41598-019-49433-1

Milgrom, L. R. (1997). *The colours of life: an introduction to the chemistry of porphyrins and related compounds*. Oxford: Oxford University Press.

Morton, B. (2011). Predator–prey-scavenging interactions between *Nucella lapillus*, *Carcinus maenas* and *Eulalia viridis* all exploiting *Mytilus galloprovincialis* on a rocky shore recovering from tributyl-tin (TBT) pollution. *Journal of Natural History*, *45*(39-40), 2397-2417. doi:10.1080/00222933.2011.596637

Moss, G. P. (1988). Nomenclature of tetrapyrroles: Recommendations 1986. *European journal of biochemistry*, 178(2), 277-328. doi:10.1515/iupac.59.0018

Olive, P. J. W. (1975). Reproductive biology of *Eulalia viridis* (Müller)(Polychaeta: Phyllodocidae) in the north eastern UK. *Journal of the Marine Biological Association of the United Kingdom*, 55(2), 313-326. doi:10.1017/S0025315400015964

Orten, J. M. (1971). Metabolism of hemoglobin and bile pigments. *Annals of Clinical & Laboratory Science*, *1*(2), 113-124.

Pereira, D. M., Valentão, P., & Andrade, P. B. (2014). Marine natural pigments: Chemistry, distribution and analysis. *Dyes and Pigments*, 111, 124-134. doi:10.1016/j.dyepig.2014.06.011

Rodrigo, A. P., & Costa, P. M. (2019). The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. *Environmental research*, 173, 270-280. doi:10.1016/j.envres.2019.03.048

Rodrigo, A. P., Costa, M. H., de Matos, A. P. A., Carrapiço, F., & Costa, P. M. (2015). A study on the digestive physiology of a marine polychaete (*Eulalia viridis*) through microanatomical changes of epithelia during the digestive cycle. *Microscopy and Microanalysis*, 21(1), 91-101. doi:10.1017/S143192761401352X

Rodrigo, A. P., Grosso, A. R., Baptista, P. V., Fernandes, A. R., & Costa, P. M. submitted (2020). A transcriptomic approach to the recruitment of venom proteins in a marine Polychaeta.

Rodrigo, A. P., Martins, C., Costa, M. H., Alves de Matos, A. P., & Costa, P. M. (2018). A morphoanatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). *Journal of anatomy*, *233*(5), 567-579. doi:10.1111/joa.12870

Schmid, R., & McDonagh, A. F. (1975). The enzymatic formation of bilirubin. *Annals of the New York Academy of Sciences*, 244(1), 533-552. doi:10.1111/j.1749-6632.1975.tb41553.x

Sinal, C. J., & Bend, J. R. (1997). Aryl hydrocarbon receptor-dependent induction of cyp1a1 by bilirubin in mouse hepatoma hepa 1c1c7 cells. *Molecular pharmacology*, 52(4), 590-599. doi:10.1124/mol.52.4.590

Surinya, K. H., Cox, T. C., & May, B. K. (1997). Transcriptional regulation of the human erythroid 5-aminolevulinate synthase gene identification of promoter elements and role of regulatory proteins. *Journal of Biological Chemistry*, 272(42), 26585-26594. doi:10.1074/jbc.272.42.26585

Suvakov, S., Damjanovic, T., Stefanovic, A., Pekmezovic, T., Savic-Radojevic, A., Pljesa-Ercegovac, M., ... & Ivanisevic, J. (2013). Glutathione S-transferase A1, M1, P1 and T1 null or low-activity genotypes are associated with enhanced oxidative damage among haemodialysis patients. *Nephrology Dialysis Transplantation*, 28(1), 202-212. doi:10.1093/ndt/gfs369

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & Jensen, L. J. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1), D607-D613. doi:10.1093/nar/gky1131

Thiel, D., Hugenschütt, M., Meyer, H., Paululat, A., Quijada-Rodriguez, A. R., Purschke, G., & Weihrauch, D. (2017). Ammonia excretion in the marine polychaete *Eurythoe complanata* (Annelida). *Journal of experimental biology*, 220(3), 425-436. doi:10.1242/jeb.145615

Tzetlin, A., & Purschke, G. (2005). Pharynx and intestine. In *Morphology, molecules, evolution and phylogeny in polychaeta and related taxa* (pp. 199-225). Springer, Dordrecht. doi:10.1007/s10750-004-1431-z

van Dijk, R., Aronson, S. J., de Waart, D. R., van de Graaf, S. F., Duijst, S., Seppen, J., ... & Bosma, P. J. (2017). Biliverdin Reductase inhibitors did not improve severe unconjugated hyperbilirubinemia in vivo. *Scientific reports*, 7(1), 1-9. doi:10.1038/s41598-017-01602-w

Veenman, L., Vainshtein, A., Yasin, N., Azrad, M., & Gavish, M. (2016). Tetrapyrroles as endogenous TSPO ligands in eukaryotes and prokaryotes: Comparisons with synthetic ligands. *International journal of molecular sciences*, 17(6), 880. doi:10.3390/ijms17060880

Woods, J. S., & Simmonds, P. L. (2001). HPLC methods for analysis of porphyrins in biological media. *Current Protocols in Toxicology*, *7*(1), 8-9. doi:10.1002/0471140856.tx0809s07

Yamamoto, M., Yew, N. S., Federspiel, M., Dodgson, J. B., Hayashi, N., & Engel, J. D. (1985). Isolation of recombinant cDNAs encoding chicken erythroid delta-aminolevulinate synthase. *Proceedings of the National Academy of Sciences*, 82(11), 3702-3706. doi:10.1073/pnas.82.11.3702

Zannotti, M., Giovannetti, R., Minofar, B., Řeha, D., Plačková, L., D'Amato, C. A., ... & Minicucci, M. (2018). Aggregation and metal-complexation behaviour of THPP porphyrin in ethanol/water solutions as function of pH. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 193, 235-248. doi:10.1016/j.saa.2017.12.021

7. Appendix

7.1 *Eulalia* sp. pigment absorption spectra

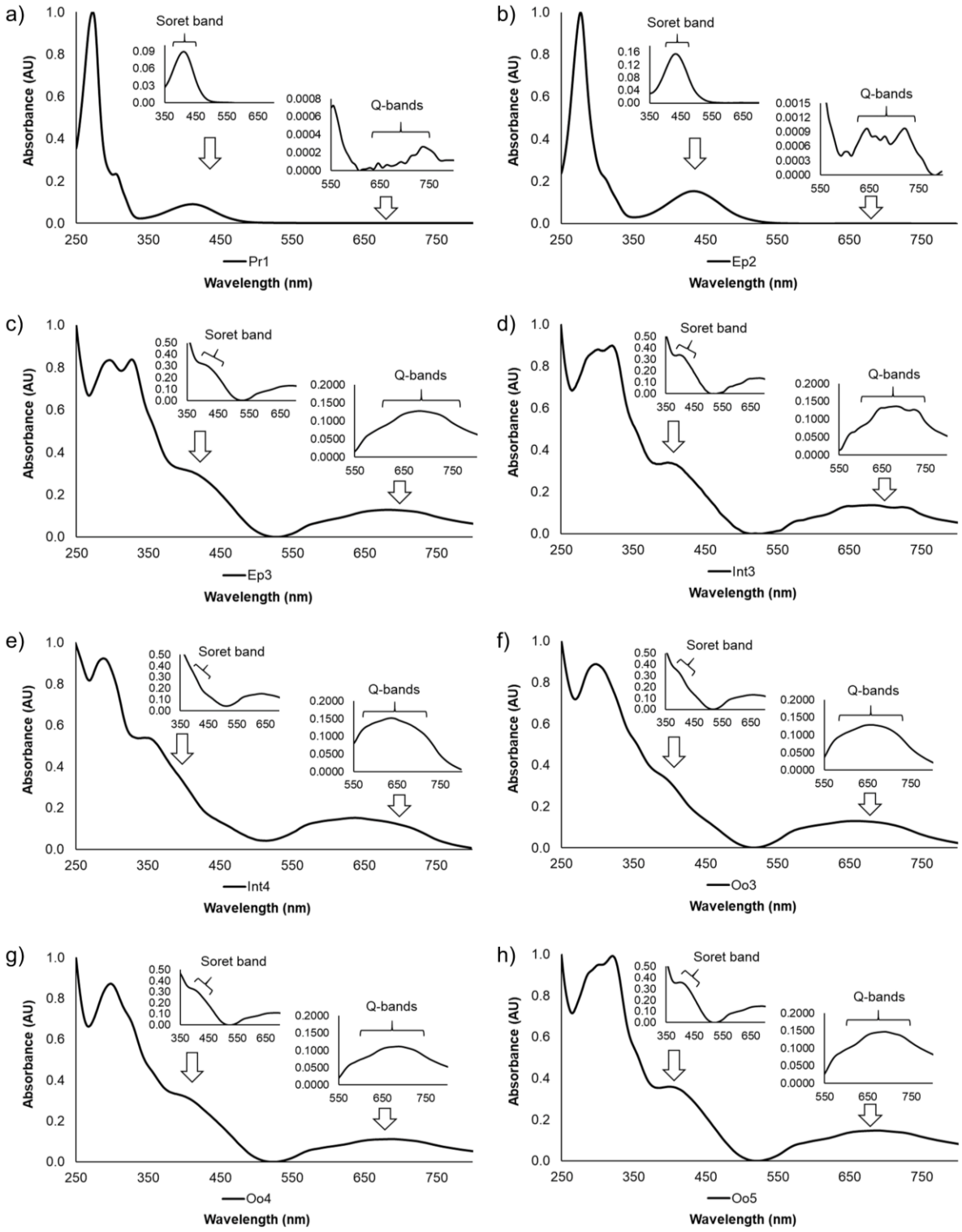
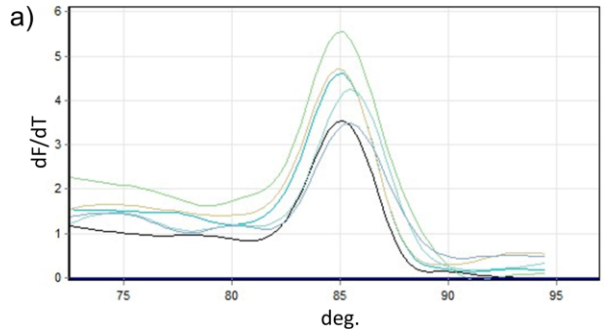
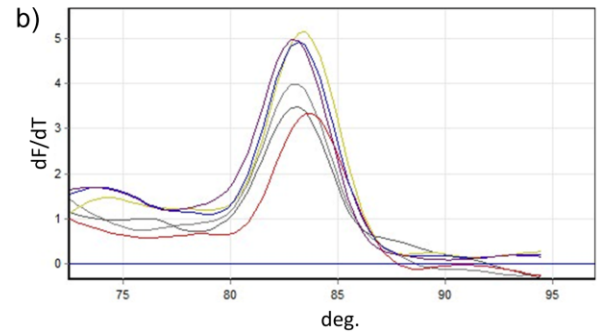


Fig.7.1. Pigment absorption spectra of *Eulalia* sp. organs. The data was retrieved from a high-performance liquid chromatography with a detector diode array (HPLC-DAD). Each graph displays a Soret band (380-500 nm) and Q bands (500-750 nm). The pigments represented are a) Pr1, a yellow pigment from the proboscis; b) Ep2, a yellow pigment from the epidermis; c) Ep3, a green pigment from the epidermis; d) Int3, a green pigment from the intestine; e) Int4, a green pigment from the intestine; f) Oo3, a green pigment from the oocytes; e) Oo4, a green pigment from the oocytes; and f) Oo5, a green pigment from the oocytes. The retention time corresponding to each spectrum is 1.31 min for Pr1, 4.16 min for Ep2, 6.46 for Ep3, 6.68 min for Int3, 4.18 min for Int4, 5.45 min for Oo3, 5.64 min for Oo4, and 6.61 for Oo5.

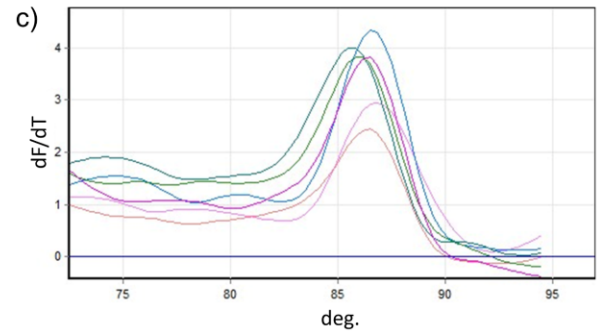
7.2 Melting curve analysis of RT-qPCR results



No.	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3	Peak 4
43		ALAD_EP1		78.0	85.0	89.7	
44		ALAD_PR1		77.0	85.0	93.2	
45		ALAD_EP2		74.3	85.0	93.5	
46		ALAD_PR2		85.0			
47		ALAD_EP3		74.7	80.3	85.5	
48		ALAD_PR3		74.0	80.0	85.5	93.0



No.	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3
31		UROS_EP1		83.0		
32		UROS_PR1		76.0	83.0	
33		UROS_EP2		78.7	83.7	
34		UROS_PR2		74.3	83.5	89.3
35		UROS_EP3		74.0	83.2	89.5
36		UROS_PR3		73.5	83.0	93.2



No.	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3
37		PPOX_EP1		73.0	78.0	86.8
38		PPOX_PR1		74.8	80.3	86.5
39		PPOX_EP2		74.2	85.7	90.7
40		PPOX_PR2		86.5		
41		PPOX_EP3		75.5	78.8	86.0
42		PPOX_PR3		77.2	86.5	

Fig.7.2. RT-qPCR melting curve analysis. The data was retrieved from a real time PCR assay in a Rotor-Gene 6000 thermal cycler. Detection was done for mRNAs coding for a) δ-aminolevulinic acid dehydratase (ALAD), b) uroporphyrinogen synthase (UROS), and c) protoporphyrinogen oxidase (PPOX), in three replicate samples of the proboscis and epidermis. Each panel displays a colored curve representing a sample indicated in the table with the respective temperature peaks. dF/dT means change in fluorescence/change in temperature and is plotted against temperature in degrees (deg).

Table 7.1. RT-qPCR primer sequences. The primers were selected for mRNAs coding for a) δ-aminolevulinic acid dehydratase (ALAD), b) uroporphyrinogen synthase (UROS), and c) protoporphyrinogen oxidase (PPOX). Primers were designed with Primer Blast against *Eulalia*'s sequences of best-matched target mRNAs and verified with Oligo Analyzer.

Name	Forward primer	Reverse primer
ALAD	ACAAGCACACGTCACAGACA	CCGTCCTGCTCTTTGGAGTT
UROS	ACACCTGACGGGCTGTAGTA	CAACTGCGCTTGCAGAGAAA
PPOX	GGACTCATCCTCGACTTCGC	AGAAGATCCACCGACTCCCA

7.3 R programming

7.3.1 RNA-Seq data processing and analysis

```
##Choosing the directory of the folders with the abundance results of 3 samples from the proboscis (PR1,PR2,PR3) and 3 samples from the epidermis (PL1,PL2,PL3)
folder<-"C:\\Read data\\"
samples<-dir(folder)
file<-c(paste(sep="",folder,samples,"/","abundance.tsv"))
names(file)=samples
names(file)

##Importing the abundance results from the several files
library(tximport)
EVtsv <- tximport(file, type = "kallisto", txOut = TRUE,countsFromAbundance = "lengthScaledTPM")
head(EVtsv$counts)

##File with 6 columns (PL1,PL2,PL3,PR1,PR2,PR3) and 3936346 contigs
dim(EVtsv$counts)
colnames(EVtsv$counts)

##File with the trinity IDs and respective peptide sequence (only ORF regions) as a reference proteome
library(seqinr)
dir<-"C:\\Transdecoder Trinity\\"
file<-"Trinity.fasta.transdecoder.pep"
EVfa<-read.fasta(paste(sep="",dir,file),as.string = TRUE)
head(EVfa)

##Combine the ID with the respective sequence (only ORF regions)
EVtxID<-getName(EVfa)
ID<-as.data.frame(sapply(strsplit(as.character(EVtxID), split="::", fixed=TRUE), function(x) (x[2])))
EVtxSEQ<-unlist(getSequence(EVfa,as.string=TRUE))
EVtxIDxSEQ<-as.data.frame(cbind(ID,EVtxSEQ))
head(EVtxIDxSEQ)
colnames(EVtxIDxSEQ)[1]<-"ID"
row.names(EVtxIDxSEQ)<-EVtxIDxSEQ$ID
nrow(EVtxIDxSEQ)

##Merge the data frame with the abundance results and the peptide sequence by Trinity ID
##The 2 tables are merged by the smallest number of lines, leaving only the transcripts with the ORF sequences
EVc<-cbind(as.data.frame(row.names(EVtsv$counts)),EVtsv$counts)
colnames(EVc)[1]<-"ID"
EV.fulldata <- merge(EVc,EVtxIDxSEQ,by="ID")
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)
write.table(EV.fulldata, "EV.fulldata.csv", sep=";", col.names=NA)

##Normalization
##Create a DGEList object from a table of counts with a group indicator for each column
colnames(EV.fulldata)<-c("ID","PL1","PL2","PL3","PR1","PR2","PR3","SEQ")
EVGroups<- c("PL","PL","PL","PR","PR","PR")
library(edgeR)
library(limma)
EVData<-DGEList(counts=EV.fulldata[,2:7], group=EVGroups,genes=EV.fulldata[,1])

##Calculate normalization factors
EVData<-calcNormFactors(EVData)
```

```
##Linear model
EVDesign<-model.matrix(~0+EVGroups,data=EVData$samples)
colnames(EVDesign) <- levels(EVData$samples$group)
EVData<-estimateDisp(EVData,EVDesign)
EVfit <- glmFit(EVData,EVDesign)
PRvPL <- makeContrasts(PR-PL, levels=EVDesign)

##The function glmLRT estimates the relative expression for each comparison (PR-PL)
EVlrtPR <- glmLRT(EVfit,contrast=PRvPL)
topTags(EVlrtPR)

##File with ID, SEQ, log2FC, logCPM and p-value
EVResults<-cbind(EV.fulldata$ID,EV.fulldata$SEQ,EVlrtPR$table)
colnames(EVResults)<-c("ID", "SEQ", "log2FC", "logCPM","LR","PValue")
head(EVResults)
write.table(EVResults,"EVResults.csv",sep=";",col.names=NA)

##Decide Tests organizes the expression values in 1,0,-1.
##0 means no differential gene expression; 1 PR gene is over-expressed in comparison
to PL; and -1 PR gene is under-expressed in comparison to Pl
##With adjusted p-value (FDR) ≤ 0,05 and |log2FC| ≥ 2
expressionTable<-decideTests(EVResults$PValue,coeffi-
cients=EVResults$log2FC,lfc=2,adjust.method="fdr")
colnames(expressionTable)<-"PR-PL"
head(expressionTable)
nrow(expressionTable)

##EVResults with the expression values (0,1,-1)
EVet<-cbind(EVResults,p.adjust(EVResults$PValue,"fdr"),expressionTable)
head(EVet)
write.table(EVet,"EVet.csv",sep=";",col.names=NA)

##Smear plot with all the contigs of Eulalia
##Differentially expressed contigs in red in comparison with the rest of the contigs
in black
degTable<-EVet[which(abs(EVet[,8]) == 1),]
head(degTable)
nrow(degTable)
windows(15,15)
plotSmear(EVlrtPR,de.tags=rownames(degTable),cex=0.4)
fileName<-"SmearPlot_ALL.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST    e    Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 15,
          units = 'cm',
          type="windows",
          res=600)

##Download protein sequences related to a keyword (using uniprot)
##Keywords: Porphyrin Eumetazoa, Heme Biosynthesis, Heme Degradation and Chlorins
##The next coding lines were conducted with command line and not R
##Create the database porphyrin Eumetazoa
makeblastdb -in uniprot-porphyrin+eumetazoa.fasta -dbtype prot -out uniprotPorphy-
rinEum.fa
##Matching between sequences from the database and ALLORFs (file with all the se-
quences of Eulalia) to create the subset porphyrin Eumetazoa
blastp -query AllORFs.fasta -db uniprotPorphyrinEum.fa -max_target_seqs 1 -max_hsps
1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5
-num_threads 20 > PorphyrinEumetazoa.csv

##Create the database heme biosynthesis
makeblastdb -in uniprot-heme+biosynthesis.fasta -dbtype prot -out uniprothemebiosyn-
thesis.fa
```

```
##Matching between sequences from the database and ALLORFs (file with all the sequences of Eulalia) to create the subset heme biosynthesis
blastp -query AllORFs.fasta -db uniprothemebiosynthesis.fa -max_target_seqs 1 -max_hsps 1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -num_threads 20 > HemeBiosynthesis.csv

##Create the database heme degradation
makeblastdb -in uniprot-heme+degradation.fasta -dbtype prot -out uniprothemedegradation.fa
##Matching between sequences from the database and ALLORFs (file with all the sequences of Eulalia) to create the subset heme degradation
blastp -query AllORFs.fasta -db uniprothemedegradation.fa -max_target_seqs 1 -max_hsps 1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -num_threads 20 > HemeDegradation.csv

##Create the database chlorin
makeblastdb -in uniprot-chlorin.fasta -dbtype prot -out uniprotclorin.fa
##Matching between sequences from the database and ALLORFs (file with all the sequences of Eulalia)to create the subset chlorins
blastp -query AllORFs.fasta -db uniprotchlorin.fa -max_target_seqs 1 -max_hsps 1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -num_threads 20 > Chlorins.csv


##In loop: add column names to each subset + results from the comparison between Proboscis VS Epiderme
rownames(EVet)<-EVet[,1]
EVet<-EVet[,-1]

DBfiles<-c("PorphyrinEumetazoa.csv","HemeBiosynthesis.csv","Chlorins.csv","HemeDegradation.csv")
DBfiles2<-c("PorphyrinEumetazoa_EVet.csv","HemeBiosynthesis_EVet.csv","Chlorins_EVet.csv","HemeDegradation_EVet.csv")

for (i in 1:length(DBfiles))
{
  pData<-read.csv(DBfiles[i],sep=";",header=FALSE,row.names=1)
  colnames(pData)<-c("sseqid","pident","nident","length","evalue","stitle")
  pData2<-merge(pData, EVet,by="row.names")
  write.table(pData2,DBfiles2[i],sep=";",row.names=FALSE)
}

##Merge the databases with the expression values (logTPMs) for each organ
##With logTPMs≥ 2 as 1,  2>logTPMs>-2 as 0 and logTPMs≤-2 as -1
fileName<-"Eulalia_allORFs.RData"
load(paste(sep="",fileName))
EA_Exp<-eulalia_allORFs[,1:2]
EA_Exp[EA_Exp>-2&EA_Exp<2]<- 0
EA_Exp[EA_Exp<=-2]<- -1
EA_Exp[EA_Exp>=2]<- 1
DBfileslog<-c("PorphyrinEumetazoa.csv","HemeBiosynthesis.csv","Chlorins.csv","HemeDegradation.csv")
DBfileslog2<-c("PorphyrinEumetazoa_logTPMs.csv","HemeBiosynthesis_logTPMs.csv","Chlorins_logTPMs.csv","HemeDegradation_logTPMs.csv")

for (i in 1:length(DBfileslog))
{
  pData<-read.csv(DBfileslog[i],sep=";",header=FALSE,row.names=1)
  colnames(pData)<-c("sseqid","pident","nident","length","evalue","stitle")
  pData2<-merge(pData, EA_Exp,by="row.names")
  write.table(pData2,DBfileslog2[i],sep=";",row.names=FALSE)
}

##Venn diagram for proteins with high and low levels of expression in the proboscis and epidermis
##Porphyrin Eumetazoa
PE_LogTPM<-read.csv("PorphyrinEumetazoa_logTPMs.csv", TRUE, ";",row.names=1)
degTablePE<-PE_LogTPM[which(abs(PE_LogTPM[,7]) == 1 | abs(PE_LogTPM[,8]) == 1),]
```

```
##Heme Biosynthesis
HB_LogTPM<-read.csv("HemeBiosynthesis_logTPMs.csv", TRUE, ";",row.names=1)
degTableHB<-HB_LogTPM[which(abs(HB_LogTPM[,7]) == 1 | abs(HB_LogTPM[,8]) == 1),]
##Heme Degradation
HD_LogTPM<-read.csv("HemeDegradation_logTPMs.csv", TRUE, ";",row.names=1)
degTableHD<-HD_LogTPM[which(abs(HD_LogTPM[,7]) == 1 | abs(HD_LogTPM[,8]) == 1),]
##Chlorins
CH_LogTPM<-read.csv("Chlorins_logTPMs.csv", TRUE, ";",row.names=1)
degTableCH<-CH_LogTPM[which(abs(CH_LogTPM[,7]) == 1 | abs(CH_LogTPM[,8]) == 1),]

library(limma)
##PE, HB, HD and CH Venn Diagrams
windows(40,40)
par(mfrow=c(2,2))
vennDiagram(degTablePE[,7:8],
            include=c("up", "down"),
            counts.col=c("seagreen","red3"),
            circle.col=c("black","black"),
            show.include=TRUE,
            cex=c(2.5,1,1),
            main="Porphyrin Eumetazoa", cex.main=2.5, line=-5,
            names=c("Epidermis","Proboscis"))
vennDiagram(degTableCH[,7:8],
            include=c("up", "down"),
            counts.col=c("seagreen","red3"),
            circle.col=c("black","black"),
            show.include=TRUE,
            cex=c(2.5,1,1),
            main="Chlorins", cex.main=2.5, line=-5,
            names=c("Epidermis","Proboscis"))
vennDiagram(degTableHB[,7:8],
            include=c("up", "down"),
            counts.col=c("seagreen","red3"),
            circle.col=c("black","black"),
            show.include=TRUE,
            cex=c(2.5,1,1),
            main="Heme Biosynthesis", cex.main=2.5, line=-5,
            names=c("Epidermis","Proboscis"))
vennDiagram(degTableHD[,7:8],
            include=c("up", "down"),
            counts.col=c("seagreen","red3"),
            circle.col=c("black","black"),
            show.include=TRUE,
            cex=c(2.5,1,1),
            main="Heme Degradation", cex.main=2.5, line=-5,
            names=c("Epidermis","Proboscis"))

fileName<-"VeenDiagALLBD.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 40,
          width = 40,
          units = 'cm',
          type="windows",
          res=600)

##Heatmaps with the differentially expressed genes for each subset
cpms <- cpm(EVData, normalized.lib.sizes=TRUE, log=T)
row.names(cpms)<-EV.fulldata$ID
library(gplots)
library(RColorBrewer)
display.brewer.all()

##Porphyrin Eumetazoa
PE<-read.csv("PorphyrinEumetazoa_Evet.csv", TRUE, ";",row.names=1)
PE_Heat_Norm<-merge(PE, cpms,by="row.names")
```

```
degPE<-PE_Heat_Norm[which(abs(PE_Heat_Norm[,14]) == 1),]
PE_gene<-as.data.frame(sapply(strsplit(as.character(degPE[,7]),     split="     OS=",
fixed=TRUE), function(x) (x[1])))
PE_gene2<-as.data.frame(sapply(strsplit(as.character(PE_gene[,1]),  split=degPE[,2],
fixed=TRUE), function(x) (x[2])))
HeatDataPE_Exp<-as.matrix(degPE[,15:20])
rownames(HeatDataPE_Exp)<-PE_gene2[,1]
colnames(HeatDataPE_Exp)<-c("EP1","EP2","EP3","PR1","PR2","PR3")

##Heatmap
clustFunction <- function(x) hclust(x, method="complete")
distFunction <- function(x) dist(x,method="euclidean")

##clusterint (sidebars)
colCutoff = 40
cCol<-colorRampPalette(brewer.pal(9,"Set1"))
colFit<-clustFunction(distFunction(t(HeatDataPE_Exp)))
cClusters<-cutree(colFit,h=colCutoff)
cHeight<-length(unique(as.vector(cClusters)));
colHeight = cCol(cHeight)
rowCutoff = 15
cRow<-colorRampPalette(brewer.pal(9,"Set1"))
rowFit<-clustFunction(distFunction(HeatDataPE_Exp))
rClustersPE<-cutree(rowFit,h=rowCutoff)
rHeight<-length(unique(as.vector(rClustersPE)));
rowHeight = cRow(rHeight)
heatColour<-colorRampPalette(c("#fff04a", "#f28e2b", "#c23e40"))(n = 100)
windows(20,15)
windowsFonts(A = windowsFont("Arial"))
heatmap.2(HeatDataPE_Exp,
          hclust=clustFunction,
          distfun=distFunction,
          ColSideColors=colHeight[cClusters],
          RowSideColors=rowHeight[rClustersPE],
          density.info="none",
          col=heatColour,
          trace="none",
          scale="row",
          cexRow=0.95,
          cexCol=1.2,
          lhei = c(0.5,9,0.04),
          lwid = c(1.1,5.7,0.5),
          offsetRow = 0,
          offsetCol = 0,
          key = FALSE,
          srtCol = 45,
          key.title=NA,
          margins = c(3,29))

fileName<-"Heatmap_PE_EXP.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST    e    Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 20,
          units = 'cm',
          type="windows",
          res=600)

##Heatmap with the color key
windowsFonts(A = windowsFont("Arial"))
windows(25,15)
par(family="A")
heatmap.2(HeatDataPE_Exp,
          scale="row",
          trace="none",
          density.info="none",
```

```
          col=heatColour,
          margins=c(5,10),
          main="Porphyrin Eumetazoa",  cexRow = 0.5,
          labRow ="")

fileName<-"ColorKey.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 25,
          units = 'cm',
          type="windows",
          res=600)

##Heme Biosynthesis
HB<-read.csv("HemeBiosynthesis_EVet.csv", TRUE, ";",row.names=1)
HB_Heat_Norm<-merge(HB, cpms,by="row.names")
degHB<-HB_Heat_Norm[which(abs(HB_Heat_Norm[,14]) == 1),]
HB_gene<-as.data.frame(sapply(strsplit(as.character(degHB[,7]), split="  OS=",
fixed=TRUE), function(x) (x[1])))
HB_gene2<-as.data.frame(sapply(strsplit(as.character(HB_gene[,1]), split=degHB[,2],
fixed=TRUE), function(x) (x[2])))
HeatDataHB_Exp<-as.matrix(degHB[,15:20])
rownames(HeatDataHB_Exp)<-HB_gene2[,1]
colnames(HeatDataHB_Exp)<-c("EP1","EP2","EP3","PR1","PR2","PR3")


##Heatmap
clustFunction <- function(x) hclust(x, method="complete")
distFunction <- function(x) dist(x,method="euclidean")

##clusterint (sidebars)
colCutoff = 40
cCol<-colorRampPalette(brewer.pal(9,"Set1"))
colFit<-clustFunction(distFunction(t(HeatDataHB_Exp)))
cClusters<-cutree(colFit,h=colCutoff)
cHeight<-length(unique(as.vector(cClusters)));
colHeight = cCol(cHeight)
rowCutoff = 15
cRow<-colorRampPalette(brewer.pal(9,"Set1"))
rowFit<-clustFunction(distFunction(HeatDataHB_Exp))
rClustersHB<-cutree(rowFit,h=rowCutoff)
rHeight<-length(unique(as.vector(rClustersHB)));
rowHeight = cRow(rHeight)
heatColour<-colorRampPalette(c("#fff04a", "#f28e2b", "#c23e40"))(n = 100)
windows(10,15)
windowsFonts(A = windowsFont("Arial"))
heatmap.2(HeatDataHB_Exp,
          hclust=clustFunction,
          distfun=distFunction,
          ColSideColors=colHeight[cClusters],
          RowSideColors=rowHeight[rClustersHB],
          density.info="none",
          col=heatColour,
          trace="none",
          scale="row",
          cexRow=0.1,
          cexCol=1.2,
          lhei = c(0.5,9,0.04),
          lwid = c(1.1,5.7,0.5),
          offsetRow = 0,
          offsetCol = 0,
          key = FALSE,
          srtCol = 45,
          key.title=NA,
          margins = c(5,5),
```

```
          labRow="")

fileName<-"Heatmap_HB_EXP.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 10,
          units = 'cm',
          type="windows",
          res=600)

##Heme Degradation
HD<-read.csv("HemeDegradation_Evet.csv", TRUE, ";",row.names=1)
HD_Heat_Norm<-merge(HD, cpms,by="row.names")
degHD<-HD_Heat_Norm[which(abs(HD_Heat_Norm[,14]) == 1),]
HD_gene<-as.data.frame(sapply(strsplit(as.character(degHD[,7]), split=" OS=",
fixed=TRUE), function(x) (x[1])))
HD_gene2<-as.data.frame(sapply(strsplit(as.character(HD_gene[,1]), split=degHD[,2],
fixed=TRUE), function(x) (x[2])))
HeatDataHD_Exp<-as.matrix(degHD[,15:20])
rownames(HeatDataHD_Exp)<-HD_gene2[,1]
colnames(HeatDataHD_Exp)<-c("EP1","EP2","EP3","PR1","PR2","PR3")

##Heatmap
clustFunction <- function(x) hclust(x, method="complete")
distFunction <- function(x) dist(x,method="euclidean")

##clusterint (sidebars)
colCutoff = 40
cCol<-colorRampPalette(brewer.pal(9,"Set1"))
colFit<-clustFunction(distFunction(t(HeatDataHD_Exp)))
cClusters<-cutree(colFit,h=colCutoff)
cHeight<-length(unique(as.vector(cClusters)));
colHeight = cCol(cHeight)
rowCutoff = 15
cRow<-colorRampPalette(brewer.pal(9,"Set1"))
rowFit<-clustFunction(distFunction(HeatDataHD_Exp))
rClustersHD<-cutree(rowFit,h=rowCutoff)
rHeight<-length(unique(as.vector(rClustersHD)));
rowHeight = cRow(rHeight)
heatColour<-colorRampPalette(c("#fff04a", "#f28e2b", "#c23e40"))(n = 100)
windows(10,15)
windowsFonts(A = windowsFont("Arial"))
heatmap.2(HeatDataHD_Exp,
          hclust=clustFunction,
          distfun=distFunction,
          ColSideColors=colHeight[cClusters],
          RowSideColors=rowHeight[rClustersHD],
          density.info="none",
          col=heatColour,
          trace="none",
          scale="row",
          cexRow=0.1,
          cexCol=1.2,
          lhei = c(0.5,9,0.04),
          lwid = c(1.1,5.7,0.5),
          offsetRow = 0,
          offsetCol = 0,
          key = FALSE,
          srtCol = 45,
          key.title=NA,
          margins = c(5,5),
          labRow="")

fileName<-"Heatmap_HD_EXP.tif"
```

```
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expres-
sion\\Resultados\\Tratadas\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 10,
          units = 'cm',
          type="windows",
          res=600)

##Chlorins
##Zero genes with differentially expression between the proboscis and epidermis
CH<-read.csv("Chlorins_EVet.csv", TRUE, ";",row.names=1)
CH_Heat_Norm<-merge(CH, cpms,by="row.names")
degCH<-CH_Heat_Norm[which(abs(CH_Heat_Norm[,14]) == 1),]

##Pathway Analysis
##Retrieving the names of the differentially expressed genes
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\String, Uniprot e Da-
vid\\Genes Names\\"
getwd()
setwd(dir)

##Heatmaps deg genes
##PE
listClustersPE<-as.matrix(rClustersPE)
nClustersPE<-nlevels(as.factor(listClustersPE[,1]))
PE_Clusters<-c("HeatPE_C1.txt","HeatPE_C2.txt","HeatPE_C3.txt")
for (i in 1:nClustersPE)
{
  HeatPE_Cluster<-subset(listClustersPE, listClustersPE[,1]==i, drop=FALSE)
  write.table(HeatPE_Cluster[,0], file = PE_Clusters[i], sep = "\t",
              quote = FALSE)
}

##HB
listClustersHB<-as.matrix(rClustersHB)
nClustersHB<-nlevels(as.factor(listClustersHB[,1]))
HB_Clusters<-
c("HeatHB_C1.txt","HeatHB_C2.txt","HeatHB_C3.txt","HeatHB_C4.txt","HeatHB_C5.txt")
for (i in 1:nClustersHB)
{
  HeatHB_Cluster<-subset(listClustersHB, listClustersHB[,1]==i, drop=FALSE)
  write.table(HeatHB_Cluster[,0], file = HB_Clusters[i], sep = "\t",
              quote = FALSE)
}

##HD
listClustersHD<-as.matrix(rClustersHD)
nClustersHD<-nlevels(as.factor(listClustersHD[,1]))
HD_Clusters<-
c("HeatHD_C1.txt","HeatHD_C2.txt","HeatHD_C3.txt","HeatHD_C4.txt","HeatHD_C5.txt")
for (i in 1:nClustersHD)
{
  HeatHD_Cluster<-subset(listClustersHD, listClustersHD[,1]==i, drop=FALSE)
  write.table(HeatHD_Cluster[,0], file = HD_Clusters[i], sep = "\t",
              quote = FALSE)
}

##Retrieving the names of the proteins highly expressed in the subsets
##PE
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)

PE_LogTPM<-read.csv("PorphyrinEumetazoa_logTPMs.csv", TRUE, ";",row.names=1)
```

```
PElogTPM_gene<-as.data.frame(sapply(strsplit(as.character(PE_LogTPM[,6]),    split="
OS=", fixed=TRUE), function(x) (x[1])))
PElogTPM_gene2<-as.data.frame(sapply(strsplit(as.character(PElogTPM_gene[,1]),
split=PE_LogTPM[,1], fixed=TRUE), function(x) (x[2])))

PE_LogTPM_exp<-as.matrix(PE_LogTPM[,7:8])
rownames(PE_LogTPM_exp)<-PElogTPM_gene2[,1]

dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\String, Uniprot e Da-
vid\\Genes Names\\"
getwd()
setwd(dir)

##Higly expressed PE protein names
VennPE_sobEP<-subset(PE_LogTPM_exp,  PE_LogTPM_exp[,1]==1  &  PE_LogTPM_exp[,2]!=1,
drop = FALSE)
write.table(VennPE_sobEP[,0], file = "VennPE_sobEP.txt", sep = "\t", quote = FALSE)

VennPE_sobPR<-subset(PE_LogTPM_exp,  PE_LogTPM_exp[,2]==1  &  PE_LogTPM_exp[,1]!=1,
drop = FALSE)
write.table(VennPE_sobPR[,0], file = "VennPE_sobPR.txt", sep = "\t", quote = FALSE)

VennPE_sobPRnEP<-subset(PE_LogTPM_exp, PE_LogTPM_exp[,1]==1 & PE_LogTPM_exp[,2]==1,
drop = FALSE)
write.table(VennPE_sobPRnEP[,0], file = "VennPE_sobPRnEP.txt", sep = "\t", quote =
FALSE)

##HB
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)

HB_LogTPM<-read.csv("HemeBiosynthesis_logTPMs.csv", TRUE, ";",row.names=1)

HBlogTPM_gene<-as.data.frame(sapply(strsplit(as.character(HB_LogTPM[,6]),    split="
OS=", fixed=TRUE), function(x) (x[1])))
HBlogTPM_gene2<-as.data.frame(sapply(strsplit(as.character(HBlogTPM_gene[,1]),
split=HB_LogTPM[,1], fixed=TRUE), function(x) (x[2])))

HB_LogTPM_exp<-as.matrix(HB_LogTPM[,7:8])
rownames(HB_LogTPM_exp)<-HBlogTPM_gene2[,1]

dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\String, Uniprot e Da-
vid\\Genes Names\\"
getwd()
setwd(dir)

##Higly expressed HB protein names
VennHB_sobEP<-subset(HB_LogTPM_exp,  HB_LogTPM_exp[,1]==1  &  HB_LogTPM_exp[,2]!=1,
drop = FALSE)
write.table(VennHB_sobEP[,0], file = "VennHB_sobEP.txt", sep = "\t", quote = FALSE)

VennHB_sobPR<-subset(HB_LogTPM_exp,  HB_LogTPM_exp[,2]==1  &  HB_LogTPM_exp[,1]!=1,
drop = FALSE)
write.table(VennHB_sobPR[,0], file = "VennHB_sobPR.txt", sep = "\t", quote = FALSE)

VennHB_sobPRnEP<-subset(HB_LogTPM_exp, HB_LogTPM_exp[,1]==1 & HB_LogTPM_exp[,2]==1,
drop = FALSE)
write.table(VennHB_sobPRnEP[,0], file = "VennHB_sobPRnEP.txt", sep = "\t", quote =
FALSE)

##HB
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)

HD_LogTPM<-read.csv("HemeDegradation_logTPMs.csv", TRUE, ";",row.names=1)
```

```
HDlogTPM_gene<-as.data.frame(sapply(strsplit(as.character(HD_LogTPM[,6]),    split="
OS=", fixed=TRUE), function(x) (x[1])))
HDlogTPM_gene2<-as.data.frame(sapply(strsplit(as.character(HDlogTPM_gene[,1]),
split=HD_LogTPM[,1], fixed=TRUE), function(x) (x[2])))

HD_LogTPM_exp<-as.matrix(HD_LogTPM[,7:8])
rownames(HD_LogTPM_exp)<-HDlogTPM_gene2[,1]

dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\String, Uniprot e Da-
vid\\Genes Names\\"
getwd()
setwd(dir)

## Higly expressed HD protein names
VennHD_sobEP<-subset(HD_LogTPM_exp,  HD_LogTPM_exp[,1]==1  &  HD_LogTPM_exp[,2]!=1,
drop = FALSE)
write.table(VennHD_sobEP[,0], file = "VennHD_sobEP.txt", sep = "\t", quote = FALSE)

VennHD_sobPR<-subset(HD_LogTPM_exp,  HD_LogTPM_exp[,2]==1  &  HD_LogTPM_exp[,1]!=1,
drop = FALSE)
write.table(VennHD_sobPR[,0], file = "VennHD_sobPR.txt", sep = "\t", quote = FALSE)

VennHD_sobPRnEP<-subset(HD_LogTPM_exp, HD_LogTPM_exp[,1]==1 & HD_LogTPM_exp[,2]==1,
drop = FALSE)
write.table(VennHD_sobPRnEP[,0], file = "VennHD_sobPRnEP.txt", sep = "\t", quote =
FALSE)

##Search for the 8 enzymes in heme metabolism in the subset Heme Biosynthesis
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)

HB_LogTPM<-read.csv("HemeBiosynthesis_logTPMs.csv", TRUE, ";",row.names=1)

nLinhasGenes<-grep("5-aminolevulinate synthase|Delta-aminolevulinic acid dehydra-
tase|hydroxymethylbilane synthase |Uroporphyrinogen-III synthase|Uroporphyrinogen
decarboxylase|Coproporphyrinogen|Protoporphyrinogen oxidase|Ferrochelatase",HB_Lo-
gTPM$stitle,ignore.case=TRUE)
HB_LogTPM_Genes<-HB_LogTPM[nLinhasGenes,]

HB_heme<-as.data.frame(sapply(strsplit(as.character(HB_LogTPM_Genes[,6]),    split="
OS=", fixed=TRUE), function(x) (x[1])))
HB_heme2<-as.data.frame(sapply(strsplit(as.character(HB_heme[,1]),
split=HB_LogTPM_Genes[,1], fixed=TRUE), function(x) (x[2])))

dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\String, Uniprot e Da-
vid\\Genes Names\\"
getwd()
setwd(dir)

write.table(HB_heme2[,1], file = "HB_heme.txt", sep = "\t", quote = FALSE, row.names
= FALSE, col.names = FALSE)


##Add the sequences of each protein in the subset Porphyrin Eumetazoa and Heme bio-
synthesis
##In order to obtain the protein sequences for phylogeny analysis

##PE +Seq
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)

LogTPMs<-read.csv("logTPMs.csv", TRUE, ";",row.names=1)
PElogTPMS<-read.csv("PorphyrinEumetazoa_logTPMs.csv", TRUE, ";",row.names=1)
PElogTPMSeq<-merge(PElogTPMS, LogTPMs,by="row.names")

dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\Mr Bayes\\Alinhamentos\\"
```

```
getwd()
setwd(dir)

write.table(PElogTPMSeq,"PE_logTPM_Seq_Final.csv",sep=";",row.names=FALSE)

##HB + Seq
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\RNA-Seq\\BLAST e Expression\\"
getwd()
setwd(dir)
HBlogTPMS<-read.csv("HemeBiosynthesis_logTPMs.csv", TRUE, ";",row.names=1)
HBlogTPMSeq<-merge(HBlogTPMS, LogTPMs,by="row.names")
dir<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\Mr Bayes\\Alinhamentos\\"
getwd()
setwd(dir)

write.table(HBlogTPMSeq,"HB_logTPM_Seq_Final.csv",sep=";",row.names=FALSE)
```

7.3.2 RT-qPCR statistical analysis

```
Path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\Real time PCR\\Análise estatís-
tica\\"
FileName<-"qPCR_ALL.csv"
Data<-read.table(paste(sep="",Path,FileName),sep=";",header=TRUE)
Data
Categorias<-levels(Data[,3])
NCategorias<-nlevels(Data[,3])

##Normality test
for(i in 1:NCategorias)
{
  print(paste(sep = " ", "Sample:",Categorias[i]))
  Média>dia=format(mean(as.numeric(Data[Data$Sample==Categorias[i],4])),dig-
its=2,nsmall=2)
  DesvPad=format(sd(as.numeric(Data[Data$Sample==Categorias[i],4])),dig-
its=2,nsmall=2)
  print(paste(sep=" ","Mean",Categorias[i],"(SD):",Média,"+/-",DesvPad))
  print(shapiro.test(as.numeric(Data[Data$Sample==Categorias[i],4])))
}

##All of the samples have p-value > 0.05, meaning that you cannot reject the null
hypothesis (the data came from a normally distributed population)
##Except for UROS

##Homoscedasticity test
library(car)
print(leveneTest(Data[,4]~as.factor(Data[,3]),center=median))
##The population variances are equal

##T-test for ALAD (parametric test)
print(t.test(Data[Data$Sample=="ALAD_PR",4],Data[Data$Sample=="ALAD_EP",4]))

##T-test for PPOX (parametric test)
print(t.test(Data[Data$Sample=="PPOX_PR",4],Data[Data$Sample=="PPOX_EP",4]))

##Mann whitney test for UROS (non-parametric test)
print(wilcox.test(Data[Data$Sample=="UROS_PR",4],Data[Data$Sample=="UROS_EP",4]))

##Graph
MeanTable<-aggregate(Data[,4],list(Data[,3]),mean,na.rm=TRUE)
MeanTable<-MeanTable[c(2,1,6,5,4,3),]
SDTable<-aggregate(Data[,4],list(Data[,3]),sd,na.rm=TRUE)
SDTable<-SDTable[c(2,1,6,5,4,3),]
graphics.off()
windows()
Graph<-barplot(names.arg=MeanTable[,1],
               MeanTable[,2],
```

```
                col=c("gray"),
                ylab="Relative Expression",
                cex.lab=1,
                ylim=c(0,3))
abline(h=0)
arrows(Graph[,1],MeanTable[,2],Graph[,1],MeanTable[,2]+SDTable[,2],an-
gle=90,length=0.05,code=3)

fileName<-"qPCRfinal.tif"
path<-"C:\\Users\\maria\\Documents\\FCT\\2Ano\\Tese\\Real time PCR\\Análise estatís-
tica\\"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 13,
          width = 20,
          units = 'cm',
          type="windows",
          res=600)
```

7.3.3 Phylogenetic tree construction with MrBayes

```
##ALAD phylogenetic tree
set autoclose=yes nowarn=yes;
execute /data/seatox/Output/leonor/ALAD_LeonorV4.nexus
charset ALAD = 1-324;
partition favored = 1: ALAD;
set partition=favored;
prset app=(1) aamodelpr=fixed(lg)
lset app=(1) rates=gamma nst=6;
lset coding = variable;
unlink revmat=(all) pinvar=(all) shape=(all) statefreq=(all);
prset ratepr=variable;
mcmc ngen=1000000 samplefreq=100 printfreq=100 diagnfreq=1000
sump
sumt

##UROD phylogenetic tree
set autoclose=yes nowarn=yes;
execute /data/seatox/Output/leonor/UROD_LeonorV4.nexus
charset UROD = 1-401;
partition favored = 1: UROD;
set partition=favored;
prset app=(1) aamodelpr=fixed(lg)
lset app=(1) rates=gamma nst=6;
lset coding = variable;
unlink revmat=(all) pinvar=(all) shape=(all) statefreq=(all);
prset ratepr=variable;
mcmc ngen=1000000 samplefreq=100 printfreq=100 diagnfreq=1000
sump
sumt

##FECH phylogenetic tree
set autoclose=yes nowarn=yes;
execute /data/seatox/Output/leonor/FECH_LeonorV4.nexus
charset FECH = 1-343;
partition favored = 1: FECH;
set partition=favored;
prset app=(1) aamodelpr=fixed(lg)
lset app=(1) rates=gamma nst=6;
lset coding = variable;
unlink revmat=(all) pinvar=(all) shape=(all) statefreq=(all);
prset ratepr=variable;
mcmc ngen=1000000 samplefreq=100 printfreq=100 diagnfreq=1000
sump
sumt
```