ACKNOWLEDGES

I thank Dr. Agostinho Antunes for give me the opportunity to be part of the Project PTDC/MAR/68106/2006 - “A modelação das vias de sinalização do ácido retinóico por contaminantes ambientais em teleósteos” and in Project PTDC/MAR/115199/2009 -"Estudo da modelação das vias de sinalização dos retinóides e das hormonas da tiróide, por disruptores endócrinos, em invertebrados"; for his support and advices.

I am grateful to Barbara Frazão for her helps in the integration in the lab work and for share her knowledge about the protocols implemented; to Anoop Alex for his suggestions and discussions, which helped me to increase the knowledge about molecular techniques and to Cidália Gomes for her PCR recommendations and suggestions for the master thesis. I am grateful to all my lab mates of the molecular biology for the help in the handling of many tools and specially Pedro Leão, Joana Martins and Raquel Branco for their help in the TA cloning; to all my lab mates of the bioinformatics/phylogenetic for teach me about phylogenetic analysis.

I thank Dr. Filipe Castro, Dr. Miguel Santos for their helpful suggestions and I acknowledge all team’s of the laboratories of CIIMAR where I was allowed to use their tools: ECOTOX, LETOX, LANUCE and specially the LEGE (Lab Ecotoxicology, Genomics and Evolution) where I developed all my work.

I am also grateful to Vincent Laudet, Zdenek Kostrouch and Ann Tarrant, who answered by email to my doubts about their reports.

ABSTRACT

Nuclear receptors (NRs) are intracellular transcription factors which are restricted to metazoan organisms. Reduced data has been reported on the diversity of these receptors, especially in early metazoan lineages. Previous studies suggested lineage-specific diversification of several NR genes, such as the diversification of retinoid X receptor (RXR) gene within the Phylum Cnidaria (Reitzel and Tarrant, 2009). Currently, only a few sequences of NRs and/or RXR genes have been reported in Cnidaria, therefore six cnidarian species were analyzed in this study: *Actinia equina, Actinia fragacea, Anthothoe sphyrodeta, Anemonia sulcata, Bunodactis verrucosa a*nd *Urticina felina.* A combination of molecular approaches was used to assess RXR gene in the most basal cnidarian class, the Class Anthozoa. In addition, the amplification of other NRs from the NR2 family previously reported in this phylum was performed. We supported the hypothesis of RXR absence in anthozoans given the non-amplification of DBD fragments of the RXR in the target species. However, in *Anemonia sulcata* five partial NRs sequences were retrieved, such as the FTZ-F1, TR2/4 and COUPT-TF receptors. In *Anthothoe spyrodeta, Bunodactis verrucosa, and Urticina felina*, one fragment of TR2/4 was also obtained. Moreover, phylogenetic analysis using representative NRs reinforced the evidence that four *A. sulcata* sequences found encode also a COUPT-TF receptor, two TR2/4-like receptors and a NR with an inconclusive position within the NR2 family. A conclusion of this work is the need to increase the datasets and knowledge on the NRs existing in the phylum Cnidaria.

Key words: nuclear receptor, RXR, Cnidaria, molecular analysis.

RESUMO

Os recetores nucleares (NRs) são fatores de transcrição intracelulares que são restritos aos metazoários. Poucos dados têm sido documentados acerca da sua diversidade, especialmente nos metazoários menos evoluídos. Artigos anteriores sugerem a diversificação específica de genes de NRs em determinados taxa, tal como a diversificação do gene RXR no filo Cnidária (Reitzel e Tarrant, 2009). Atualmente, poucas sequências de genes de NRs e/ou do gene RXR têm sido publicados em

metazoários; pelo que neste estudo seis espécies de cnidários foram analisadas: *Actinia equina, Actinia fragacea, Anthothoe sphyrodeta, Anemonia sulcata, Bunodactis verrucosa e Urticina felina*. Uma combinação de metodologias moleculares foi usada para estudar o gene RXR na classe mais basal dos cnidários, a classe Anthozoa. Para além disso, foi realizada a amplificação de outros NRs da família NR2 anteriormente publicados em cnidários. Nós apoiamos a hipótese que defende a ausência do gene RXR nos antozoários, atendendo à não amplificação do fragmento do DBD do RXR nas espécies-alvo. Mas, na *Anemonia sulcata* foram obtidas sequências parciais de cinco NRs, nomeadamente dos recetores FTZ-F1, TR2/4 e COUPT-TF. Na espécie *Anthothoe sphyrodeta, Bunodactis verrucosa, Urticina felina* também foi amplificado um fragmento do TR2/4. A análise filogenética com NRs de taxa representativos reforçou o indício de que quatro sequências descobertas na *A. sulcata* codificam um recetor COUP-TF, dois recetores TR2/4 e um NR cuja classificação dentro da família NR2 é inconclusiva. Uma conclusão deste trabalho é a necessidade de aumentar os dados e o conhecimento acerca de NRs existentes no filo Cnidária.

Palavras-chave: recetor nuclear, RXR, Cnidária, análise molecular.

CONTENTS

LIST OF ABBREVIATIONS

A - Absorbance
AF-2 – Ligand Dependent Transactivation Function Domain 2
As_10 - *Anemonia sulcata* sequence similar to NvNR10
As_11 - *Anemonia sulcata* sequence similar to NvNR11
As_14 - *Anemonia sulcata* sequence similar to NvNR14
As_a – *Anemonia sulcata* RACE sequence
BIC - Bayesian Information Criterion
BLAST – Basic Local Alignment Search Tool
C- Cytosine
cDNA – complementary DeoxyriboNucleic Acid
CDS – Coding Sequence
COI - Cytochrome *c*-oxidase
COUP-TF - Chicken Ovalbumin Upstream Promoter Transcription Factor
CTAB - CetylTrimethylAmmonium Bromide
DBD - DNA Binding Domain
ddH_2O – ultra pure water
DHA - DocoHexaneoic Acid
DHR - Diapauses Hormone Receptor
EAR2 - V-erbA-Related receptor
EcR - Ecdysone Receptor
EDTA – DiAminoEthanetetraAcetic Acid
ER - Estrogens Receptor
ERR - Estrogens Related Receptor
EtBr - Ethidium Bromide
FTZ-F1 - Fushi tarazu-Factor1
FXR - Farnesoid X Receptor
G - discrete Gamma distribution
G - Guanine
GCNF - Germ Cell Nuclear Factor
gDNA – genomic DeoxyriboNucleic Acid
GTR - General Time Reversible
HNF4 - Hepatocyte Nuclear Factor-4
HREs - Hormone Response Elements
JTT - Jones-Taylor-Thornton

K2 - Kimura 2-parameter

LBD - Ligand Binding Domain

LEGE - Laboratory of Ecotoxicology, Genomics and Evolution

$MgCl_2$ – Magnesium Chloride

min – minutes

mtDNA – mitochondrial DeoxyriboNucleic Acid

NaCl – Sodium Chloride

NALC - N-Acetyl-L-Cysteine

NCI - Close-Neighbor-Interchange

NGFIB - Nerve Growth factor IB

NR - Nuclear Receptor

NR1B1 – Nuclear Receptor Subfamily 1 Group B Member 1

NR1B2 –Nuclear Receptor Subfamily 1 Group B Member 2

NR1B3 – Nuclear Receptor Subfamily 1 Group B Member 3

NR2B1- Nuclear Receptor Subfamily 2 Group B Member 1

NR2B2 – Nuclear Receptor Subfamily 2 Group B Member 2

NR2B3 – Nuclear Receptor Subfamily 2 Group B Member 3

NRR – Nuclear Receptor Resource

NvNR – Nuclear receptor of *Nematostella vectensis*

PCR - Polymerase Chain Reaction

PPAR - Peroxisome Prolifector-Activated Receptor

PUFAs - Polyunsaturated Fatty Acids

RA - Retinoic Acid

RAR - Retinoic Acid Receptor

RAR ɣ - Retinoid X Receptors Gamma Isoform

RAR α- Retinoid Acid Receptors Alpha Isoform

RAR β - Retinoid Acid Receptors Beta Isoform

RNAse - Ribonuclease

RT – Reverse Transcription

RXR - Retinoid X Receptor

RXR ɣ - Retinoid X Receptors Gama Isoform

RXR α – Retinoid X Receptors Alpha Isoform

RXR β - Retinoid X Receptors Beta Isoform

s - seconds

SF1 - Steroidogenic factor-1

Taq – Thermus aquaticus

TLL – Tailless receptor

TLX - Homologue of the Drosophila tailless gene

TR - Testicular receptor

TR - Thyroid Hormone Receptor

USP - Ultra spiracle

VDR - Vitamin D_3 Receptor

WoRMS - World Register of Marine Species

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. Signalling trough Nuclear Receptors (NRs)

Cellular signalling paths trough molecules are one of the main processes of functional control of an organism. Nuclear receptors (NRs) are critical components of these pathways, as they are a super family of ligand-activated transcriptional factors. Therefore, these receptors bind directly to DNA and have the ability to repress or activate the expression of target genes. Activation of this process occurs upon binding of an agonistic ligand to the receptor, then, in the nucleus of the cell, the complex ligand-receptor bind to specific sequences called hormone response elements (HREs) within the promoter region of the target genes (Philip et al., 2011).

NRs share a common modular structure (Fig. 1); they have a variable N-terminal domain (A/B domain), a DNA binding domain (DBD or C domain), a hinge region (D) and a ligand binding domain (LBD or E domain), which contains the ligand-dependent transactivation function (AF-2). This domain contributes to the dimerization interface of NRs. A few NRs contain also a C-terminal domain (F), which have an unknown function (de Groot et al., 2005).

Fig. 1 - Modular structure of nuclear receptors (Tarrant et al., 2008)

The well-conserved domains are the DBD and LBD. DBD is a highly conserved domain, while LBD is moderately conserved, thus both allow a relatively easy amplification of their sequences and also a reliable phylogenetic reconstruction of the super family (Tocchini-Valentini et al., 2008).

The way how LBD and ligands interact allow to classify three major groups of NRs: in the first class, the LBD can be activated by a specific hydrophobic ligand (classical NRs); in the second class, NRs are ligand-independent transcription factors, so they can activate gene expression in the absence of a ligand; in the third class, NRS can be transcriptional repressors as their LBD can not to interact with transcription coactivators (Bridgham et al., 2010).

A more systematic classification, based on sequence similarity, has shown a clear evolutionary tree of the human NR super family, branching from a single ancestor gene to six subfamilies of NRs (Fig. 2) named as NR1, NR2, NR3, NR4, NR5 and NR6 family (Nuclear Receptor Nomenclature Committee, 1999; Escriva et al., 2000). The

NR1 family corresponds to the thyroid hormone receptors, which is a large group comprising the receptors: TR, RAR, RZR, PPAR, Rev-erb, VDR, FXR and EcR; NR2 family contains the group: RXR, COUP, HNF4, EAR2, TR2, TR4 and TLL; NR3 family clusters the steroid receptors: ER and ERR; NR4 family is a small group of orphan receptors (receptors which do not have a physiologically specific ligand) called NGFIB; NR5 family has SF1, FTZ-F1 and DHR receptors and NR6 family has a single orphan receptor: GCNF. There is also the NR0 family, clustering all the unusual receptors that contain only one of the two conserved domains (Escriva et al., 2000).

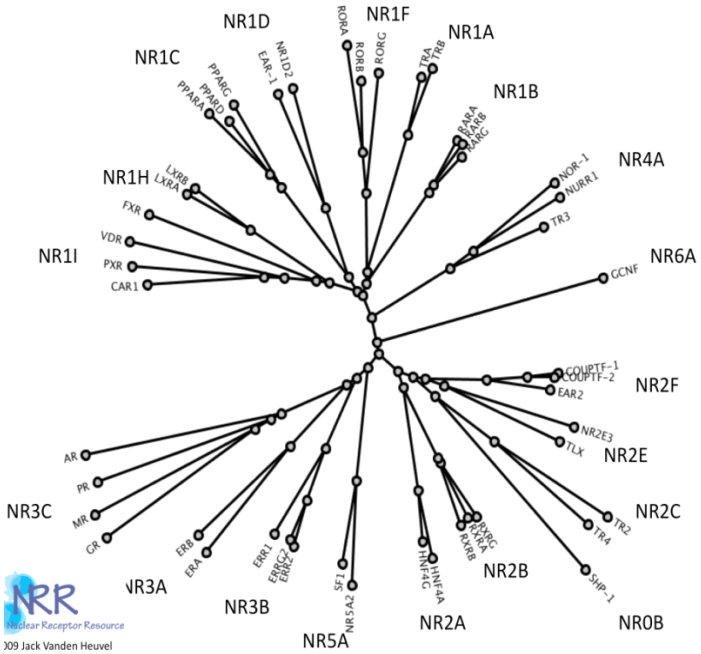


Fig. 2 - Phylogenetic tree of NRs (NRR, http://nrresource.org/).

NR super family appear to be a metazoan innovation, because they are absent from the genomes of choanoflagellates, fungi, plants, and prokaryotes; nevertheless, controversy still exists regarding its evolutionary origin (Bridgham et al., 2010).

The structural diversity of their ligands and the existence of orphan receptors led to different hypothesis: as to whether the ancestral receptor was ligand-dependent receptor or the binding ability of most NRs evolved independently. By one hand, Escriva et al. (2000) suggested that ligand-binding was acquired during NRs evolution. So, the ancestral receptor would have been an orphan ligand-independent receptor and its ligand recognition was gained later. On the other hand, Bridgham et al. (2010) suggested that the ancestral NR was a ligand-dependent receptor, which was activated by its ligand binding and this dependence of ligand activation was lost during evolution. The evidences of diversification of this super family are ambiguous due to limited

sequence sampling and/or use of outdated phylogenetic methods. Therefore, it is relevant to study and to characterize the NRs within groups closer to the root of the metazoan origin (Bridgham et al., 2010).

1.1. Signalling trough Retinoic Acid

Retinoic acid (RA) is the major bioactive metabolite of retinol and it is an essential molecule for several biological functions in metazoans, such as, cell growth and differentiation of embryonic development and reproduction, homeostasis, cell proliferation of adult organism (McKenna, 2012).

In vertebrates retinoid acid signals are transduced by two families of NRs: Retinoic Acid Receptors (RARs; NR1B) and Retinoic X Receptors (RXRs; NR2B), which are both members of the steroid/thyroid hormone receptors (Bouton et al., 2005). RXR has been identified in species ranging from placazoans to mammals (Philip et al., 2011) and RARs have been found in invertebrate chordates, such as amphioxus and tunicates, and in vertebrates (Marlétaz et al. 2006). While in invertebrates there is a single gene, vertebrates have three or more copies of RXR or/and RAR gene; RAR subtypes are RARα (NR1B1), RARβ (NR1B2) and RARγ (NR1B3), in turn, the isoforms of RXR are RXRα (NR2B1), RXRβ (NR2B2) and RXRγ (NR2B3). While RXRγ shows a restricted expression in muscles, pituitary gland and certain regions of the brain; RXRα is mainly expressed in the liver, kidney, epidermis, intestine, skin and RXRβ isoform has ubiquitous distribution (www.nursa.org/10.1621/datasets.02001).

Despite the similarities between these two families, they have different features related to their primary structure and ligand specificity. RARs are known to bind all-trans RA with high affinity. Little is known of the natural activators of RXR, but the 9-cis-RA is pointed as its specific ligand.

Studies related to RA signalling pathway can provide new insights about physiological processes but there are scarce evidences for roles of RA signalling outside chordates and its evolutionary origin itself remains elusive (McKenna, 2012).

1.1.1. Retinoic X Receptor (RXR): gene and protein

RXR binds to repeats targets: a consensus sequence 'AGGTCA' arranged in a direct, inverted or everted manner. These repeats are involve in the formation of dimers, therefore RXR is an heterodimerization partner for many other nuclear receptors (Philip et al.,2011), such as RARs, thyroid hormone receptor (TR), vitamin D_3 receptor (VDR), peroxisome prolifector-activated receptors (PPARs) and numerous orphan receptors. As a heterodimerization partner, RXR have a complex action as it is

involved in the regulation of multiple cellular pathways (Steineger et al., 1998), so it is an interesting therapeutic target for treatment of metabolic syndromes.

There are two classes of RXR dependent heterodimers, permissive heterodimers and nonpermissive heterodimers. RXR heterodimers that contain permissive partners, can be activated by agonists for either RXR or for the partner receptor. When both are activated, they have a synergistic action. In turn, heterodimers with a nonpermissive partner can be activated only by the agonist of the RXR's partner receptor (Schulman et al, 1997).

RXR and some of its dimers are activated *in vitro* by 9-cis-RA, a derivate of RA, but there is very low amount of this molecule *in vivo*, casting doubt of its physiological response as a RXR ligand; 9-cis-RA can be only a pharmacological activator of RXR (Tocchini-Valentini et al., 2008). It has also been shown that docosahexaenoic acid (DHA), docosatetraenoic acid activate RXR, but with low affinity (Steineger et al., 1998).

Structural analysis of RXR has shown that several evolutionary shifts have modified the ligand binding ability. This fact highlights the evolutionary plasticity of LBD's RXR (Tocchini-Valentini et al., 2008).

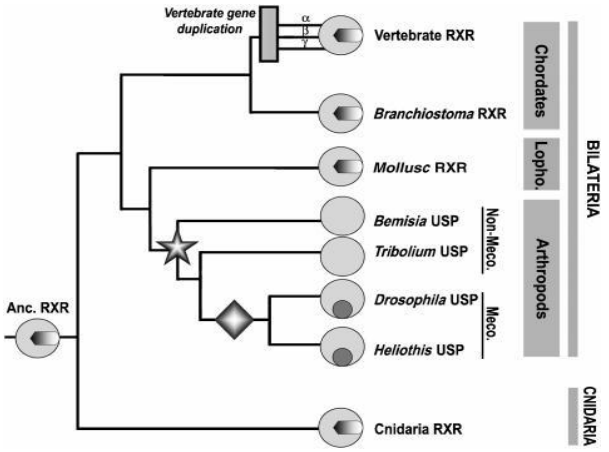


Fig. 3 - RXR cladogram based on LBD ligand-ability (Iwema et al., 2007).

Iwema et al. (2007) showed that the capability of ligand binding 9-cis-RA (indicated by the rectangular pocket, Fig. 3) of Cnidaria RXR apparently was acquired by a common ancestor and it was conserved throughout evolution. However, it is also highlighted the loss of ligand-dependent binding to 9-cis-RA (indicated by the star) in the Arthropod lineage, as they possess a RXR orthologous, USP (Ultraspiracle), which do not bind to this ligand. Furthermore, RXR of *B. glabrata*, a mollusc phylogenetic closer to Arthropods, binds 9-cis-RA essentially in the same way as the human RXR.

This evidence indicates that the potential of binding to 9-cis-RA has been lost in Arthropods after their separation from Molluscs. Moreover, in vertebrate there are typically three genes (RXRα, RXRβ and RXRγ), which arose due to the genome duplication in the vertebrate ancestor. In addition within fishes there were specific whole genome duplications (not shown in Fig. 3), thence fishes have an additional RXRβ gene and only zebrafish have additional copies of the RXRα and RXRβ genes (Philip et al., 2011).

Although the origin of RXR gene has occurred evidently very early during metazoan evolution, the origin of the ancestral RXR gene is not consensual. Wiens et al. (2003) reported a full-length RXR gene from the sponge *Suberites domuncula.* Recent studies used the analysis of whole-genome in early metazoans and they indicated the absence of the RXR gene in *Amphimedon queenslandica* (Phylum Porifera), the presence in *Trichoplax adhaerens* (Phylum Placozoa) and in *Hydra magnipapillata* (Phylum Cnidaria). Thus, within metazoans, Placazoa is the most ancient phylum with RXR (Bridgham et al., 2010; Reitzel et al., 2011).

Crystal structures of either apo-LBD (without ligand) or holo-LBD (ligand-bound) have been determined for several NRs. RXR's structures reveal that LBDs fold into an α-helical sandwich (helix11 and helix12), which envelopes a hydrophobic ligand-binding pocket. Upon binding of an agonistic ligand, helix-12 (also called AF-2 helix) packs against the rest domain, closing the pocket. Thus, the analysis of apo and holo-LBD structures led to predict a mousetrap model to explain the conformational switch activated by the ligand (Gampe et al., 2000).

2. Phylum Cnidaria

The phylum Cnidaria encloses diploblastic animals with a radially or biradially body, which live in aquatic environments, mostly marine. They have a single opening, the mouth, which connects with an internal cavity and they are mostly divided into three anatomic parts: tentacles, body column and base. Cnidarians exist into two different forms, medusa and polyp. This phylum is divided into four classes: Cubozoa (sea wasps or box jellyfish), Scyphozoa (cup animals); Anthozoa (anemones and corals) and Hydrozoa (hydrazoans).

Despite the cnidarians simplicity they are very successful given their diversity and abundance. They are ecologically important as predators, prey and structure-builders; moreover, they occupy a key evolutionary position as basal metazoans (Tarrant, 2007). Even though the bioregulation of cnidarians physiology seems to be complex, cnidarians hormonal signalling pathways are poorly characterized and few

endpoints have been established. Thus, it is urgent to develop more studies about the metazoan bioregulation; new released genomic sequences and better molecular tools will improve the evaluation of the regulation and/or disruption of cnidarian signalling (Tarrant, 2007).

2.1. Class Anthozoa

Bridge et al. (1992) suggested a basal position for class Anthozoa within the phylum Cnidaria, given the presence of a structural alteration in the mitochondrial DNA (mtDNA) of the Hydrozoa, Scyphozoa and Cubozoa classes, which is not evidenced in the class Anthozoa. This fact indicates that these three classes form a monophyletic group relative to the class Anthozoa. A more comprehensive study of the evolutionary relationships among these groups [trough an analysis based on 18S ribosomal DNA (rDNA), mitochondrial 16S rDNA, mitochondrial genome structure and morphological characters] also supports that Anthozoa is the earliest cnidarians class (Fig. 4).

The class Anthozoa is among the most slowly evolving metazoan lineages; because rates of mitochondrial evolution are slower than those of other marine invertebrates, nonetheless, the rate of evolution for nuclear genes has not been well-studied (Daly et al., 2008).

It is supported the split of Anthozoa into the clades Hexacorallia and Octocorallia (Daly et al, 2010). Within the subclass Hexacorallia, the order Actiniaria (sea anemones) is the most diverse and successful. Curiously, studies suggested that an unusually slow molecular evolution is shown within this group and that Actiniaria is monophyletic with respect to other extant hexacorallians (Daly et al., 2008).

The order Actiniaria comes out as an interesting group to study genes, such as nuclear genes, and their rate of evolution. Therefore, the species studied (Table 1) are sea anemones belonging to this order and phylogenetic closer to *Anemonia sulcata*, which has been the first anthozoan reported with the RXR gene (Escriva et al., 1997).

Within the studied species there are some morphological identifying features: *Actinia equina* is an anemone with tentacles arranged radially in six circles around the opening of gastrovascular cavity. Bright blue spots, called acrorhagi, are below the tentacles on the outer margin of the column, which distinguish *A. equina* and *A. fragacea*. These species are likely to be confused, however, *Actinia fragacea* have larger size, have spots all over the column and have less variable coloration (Stachowitsch, 1992). *Actinothoe sphyrodeta* has up to 100 white tentacles irregularly arranged around a yellow/orange disc. The translucent dirty-white column has faint longitudinal greenish stripes and a diameter base up to 2 cm (Picton and Morrow,

2010). *Anemonia sulcata* is a sea anemone with long, sinuous tentacles, which are rarely retracted. The column is smooth, with a row of inconspicuous warts and its colour is brownish or greyish. *Bunodactis verrucosa* is an anemone with a brownish base and a column surrounded by rows of many white, greyish, long and/or small warts. They have four rows of circles of tentacles around the only opening. *Urticina felina* may occur in almost any colour and may have short tentacles are arranged in multiples of 10. They have small size; its diameter of base is around to 120mm (Picton and Morrow, 2010).

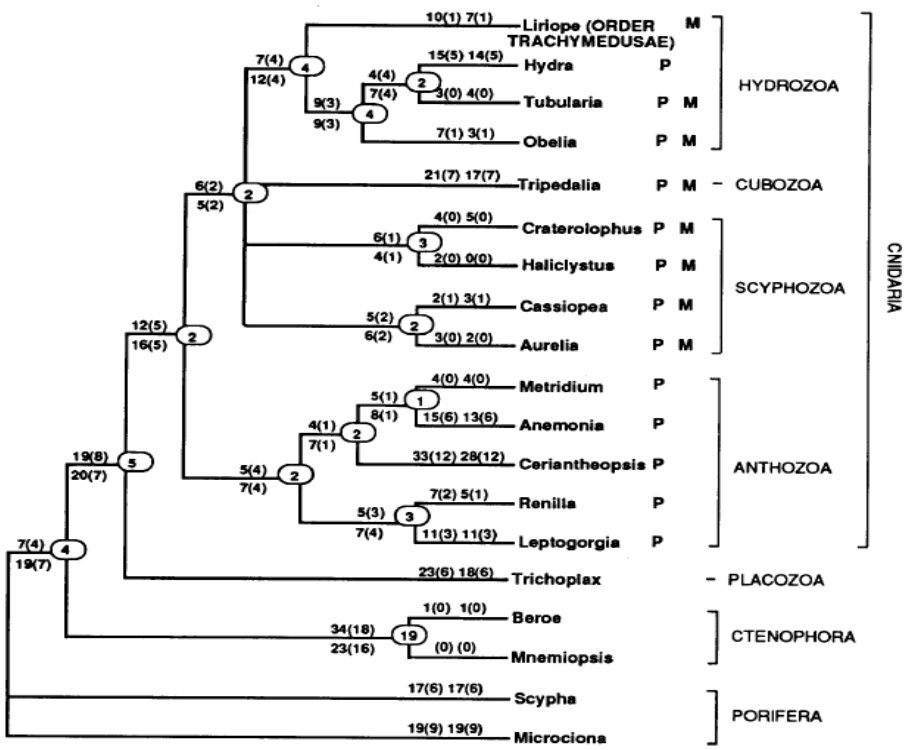


Fig. 4 - Consensus of cladograms of classes from Cnidaria, according to 18S rDNA, mtDNA structure and morphological features analyses (Bridge et al., 1995).

Class Anthozoa					
Subclass	**Order**	**Suborder**	**Family**	**Genus**	**Specie**
Hexacorallia	Actiniaria	Nynantheae	Actiniidae	*Actinia*	*Actinia equina*
Hexacorallia	Actiniaria	Nynantheae	Actiniidae	*Actinia*	*Actinia fragacea*
Hexacorallia	Actiniaria	Nynantheae	Sagartiidae	*Actinothoe*	*Actinothoe sphyrodeta*
Hexacorallia	Actiniaria	Nynantheae	Actiniidae	*Anemonia*	*Anemonia sulcata*
Hexacorallia	Actiniaria	Nynantheae	Actiniidae	*Aulactinia*	*Bunodactis verrucosa*
Hexacorallia	Actiniaria	Nynantheae	Actiniidae	*Urticina*	*Urticina felina*

Table 1 - Taxonomy of anthozoan Actinia equina, *Actinia fragacea, Actinothoe sphyrodeta, Anemonia sulcata and Urticina felina* (WoRMS, www.marinespecies.org).

2.2. NRs in Cnidaria

Researches about NRs from Cnidaria have initially use reverse transcription (RT-PCR) and the polymerase chain reaction (PCR) with degenerate primers. NRs identified through these methods were reported in distinct cnidarians species, such as, the hydrozoan *Hydra vulgaris,* [the orphan COUP-TF receptor (Gauchat et al., 2003)]; the anthozoan *Anemonia sulcata* [the member of RXR and FTZ-F1 group of receptors (Escriva et al., 1997)]; the cubozoan *Tripedalia cystophora* [the RXR gene(Kostrouch et al., 1998)]; the anthozoan *Acropora millepora* [TLX, RXR, COUP-TF, HNF4, FTZ-F1, TR2/TR4 (Grasso et al., 2001)] and the anthozoan *Pocillopora damicornis* [TLL, COUP-TF and a third gene which cannot be unambiguously classified within a currently recognized subfamily (Tarrant et al.,2008)].

In addition, sequencing of the genome of the anthozoan *Nematostella vectensis* (Sullivan *et al.* 2006) and of hydrozoan *Hydra magnipapillata* (Chapman et al., 2010) has provided a recent opportunity to survey cnidarian NRs diversity on a genome-wide scale. Reitzel and Tarrant (2009) identified all NRs from the *N. vectensis*; a set of seventeen NRs (HFN4, TLL, TLX, FAX, PNR, COUTF-TF, TR2/4, GCNF and NRs of NR1 family), moreover in *H. magnipapillata* it was predicted the presence of six NRs, which grouped strongly with COUP-TF, HNF4, RXR subfamily and with NRs members of the NR2E subfamily (Reitzel et al., 2011).

Among the cnidarian NRs identified to date mostly are members of NR family 2, evidence that the rate of diversification for the NR super family is fairly modest in the early diverging animals (Reitzel et al., 2011).

Among these studies only in the jellyfish RXR it was confirmed its ability to bind 9-cis-RA *in vitro,* the others did not analyzed the ability of NRs to bind ligands. Therefore, a hypothesis remains unclear: whether these cnidarian NRs are regulated by endogenous and/or exogenous hormones, which will control cnidarian signalling pathways (Tarrant et al., 2008). Future studies are needed to understand and characterize the mechanisms of bioregulation in cnidarians.

2.2.1. RXR in Cnidaria

There is strong evidence of the presence of RXR in cnidarians because retinoids are reported to influence morphogenetic pattern in the hydrozoan *Hydractinia echinata* (Bouzaiene et al., 2007).

Furthermore, previous studies reported NR sequences displaying substantial similarity with vertebrate RXR genes; Escriva et al. (1997) mentioned that it was identified a probable orthologous of RXR/USP in the anthozoan *Anemonia sulcata.* Grasso et al. (2001) got *Acropora millepora* cDNA sequences, which were similar to the RXR *A. sulcata* sequences*;* both studies did not confirm its ligand-dependence to 9-cis-RA. In a cubozoan, *Tripedalia cystophora,* was got the full-length of the gene encoding RXR and, trough radio-binding analysis, it was evidenced that this NR bound to 9-cis-RA in the same way as the vertebrate´s RXR. Moreover, they showed that this receptor is expressed at all developmental stages of the jellyfish (Kostrouch et al., 1998). These researches identified the gene by amplification of CDS sequences.

New approaches used the complete-genome analysis to study NRs in cnidarians. In 2009, Reitzel and Tarrant reported a complete set of NR from the anthozoan *Nematostella vectensis* and they revealed that occurred a RXR gene loss. In addition, they analyzed the sequences of previous reports and they denied the RXR gene presence in anthozoan *Acropora millepora*. Additionally, Reitzel and Tarrant (2011), analyzed the reported genome of the hydrozoan *Hydra magnipapillata* and they found that most NRs are members of the NR2 family. One of these NRs grouped with strong support to RXRs from bilaterians and from placozoan *Trichoplax*, therefore this hydrozoan has an ortholog of invertebrates RXR genes*.* By one hand, there are no published NRs about the class Scyphozoa and the published results related to class Anthozoa are unclear; on the other hand, it was identified a RXR gene in a cubozoan and hydrozoan specie. Therefore, it is inferred that the cnidarian-bilaterian ancestor had the RXR gene and within Cnidaria occurred a specific diversification.

3. Objectives

The transcriptional regulators inherited capacity to be ligand-activated makes them prime targets of endogenous and/or exogenous hormones. Since cnidarians hormonal signalling pathways are poorly characterized and few appropriate endpoints have been established, research about its NRs has been recognized as urgent. Moreover, the topology of the phylogeny of most these NRs has been uncertain, due to limited sequenced cnidarians samples, thus more data is needed to clear up how NRs evolved. Within Cnidaria RXR gene diversification is controversial, so this nuclear gene is considered a relevant target of the study.

In this research we aimed to assess the presence/absence of RXR gene in new anthozoan species: *Actinia equina, Actinia fragacea, Actinothoe spyrodeta, Anemonia sulcata, Bunodactis verrucosa, Urticina felina*. Furthermore, our goals were the

amplification of the most RXR's conserved domain in order to identify the gene; the sequencing of its full-length CDS and the analysis of the gene evolutionary relationships within class Anthozoa. Throughout the lab work new goals arose in order to support new evidences about other NRs; we intended to amplify sequences of NRs (NR2 family) previous reported in anthozoan species, which allowed complementing the phylogenetic analyses.

MATERIAL AND METHODS

1. Biologic Material

The field work was performed in the intertidal zone; adult *Actinia equina, Actinia fragacea, Actinothoe spyrodeta, Anemonia sulcata, Bunodactis verrucosa, Urticina felina* specimens were collected between January and February at Praia da Memória (Oporto, Portugal) and Praia da Foz (Oporto, Portugal). Fresh tissues of epidermis of the pedal disc were preserved in RNAlater (Sigma). Then anemones were dissected and the tissues were flash frozen in liquid nitrogen to be ground into smaller pieces; using sterilized material, such as mortar and pestle. Then all biological material was stored at −80 ◦C.

2. Genomic DNA Extraction

Different genomic DNA extraction protocols were used: an adapted protocol from phenol: chloroform extraction procedures contributed by Pinto et al. (2000) and the genomic DNA extraction Mini Kit (Bioline).

Firstly a pre-treatment with acetyl cysteine (NALC) of the flash frozen tissues, was performed to decrease the viscosity of the samples. 750 µL NALC solution (1.45% sodium citrate; 2% sodium hydroxide; 0.5% N-acetyl-L-cysteine) was added to each sample, which was incubated at room temperature for 25 min. Then, they were centrifuged at maximum speed for 10 min and the pellet was resuspended in 1ml of 1x PBS solution (137 mM NaCl; 2.7 mM KCl; 4.3 mM Na_2HPO_4; 1.47 mM KH_2PO_4; pH of 7.4).

By one hand, the gDNA extraction with the Mini Kit (Bioline) was performed according to the manufacturer instructions. One the other hand, the phenol: chloroform extraction was performed with initial lyses of the tissues; adding 1000 µL of Lyses Buffer (10 mM Tris-HCl; 100 M EDTA, 0.5% sodium dodecyl sulphate; pH of 8.0) and 10 µL lysozyme, the lysate was incubated for 30 min at 55°C. Then, proteinase K was added to a final concentration of 1 mg/mL and incubation was performed overnight at 37°C. RNAse was added to a final concentration of 0.6 mg/mL and incubation for 30 min was performed at 37°C. It was performed three successive addictions of 60 µL CTAB solution (0.3% CTAB; 0.7M NaCl) plus incubation for 20 min at 65°C plus addictions of 1000 µL (24:1) chloroform: isoamyl plus a centrifugation for 20 min at 12000 x *g*, until the interphase disappears. The supernatant was recovered and it was added 1 volume of 24:1:1 phenol: chloroform: isoamylm, then a centrifugation was

performed for 20min at 16000 x *g*. The precipitation of genomic DNA was achieved with addiction of 0.1 volumes 6M NaCl plus the addiction of 2.5 volumes of cold absolute ethanol. Samples were centrifuged 20min at 4°C and 16000 x *g*. The supernatant was drained off and the pellet was washed with 800 µL 70% ethanol. The gDNA was dried and resuspended in 100 µL of TE buffer (10 mM Tris-HCl; 1 mM EDTA; pH of 8.0).

The integrity of gDNA was evaluated by agarose gel electrophoresis and it was quantified its concentration by Nanodrop 1000. Since the samples of gDNA extracted using protocol of phenol: chloroform extraction had higher concentration and yield values, they were used in PCR amplification.

3. Total RNA Extraction

Two total RNA extraction protocols were evaluated: an adapted single-step method of RNA extraction by acid guanidiniumthiocyanate: phenol: chloroform (Chomczynski and Sacchi, 1987) and the protocol from UltraClean Tissue & Cells RNA Isolation Kit (MO BIO), which suffered some adaptations. Attending on the total RNA quantity and purity, the chosen method was the adapted protocol of UltraClean Tissue & Cells RNA Isolation Kit.

A fragment (20-25 mg) of all tissues kept in RNAlater was dissected and treated with 700 µL of Solution TR1 prepared with β-mercaptoethanol, then 1 volume of Solution TR2 was added to the lysate and it was homogenised by pipetting. A centrifugation of 30 s at 10000 x *g* was performed before the transference of the supernatant to a spin filter column. Then, the supernatant was centrifuged for 2 min at 16000 x *g* and the spin filter was washed with 500 µL of Solution TR3 and with 500 µL of Solution TR4 (a centrifugation for 2 min at 16000 x *g* was performed after both steps). To dry the membrane, an additional centrifugation was performed (3min at 16000 x *g*) and the membrane was air-dry for 2 min. To elute the RNA, it was added 50 µL of Solution TR5 directly onto the membrane. Total RNA was recovered with a centrifugation for 1 min at 13000 x *g* and then it was cleaned up according to the manufacturer instructions of UltraClean Tissue & Cells RNA Isolation Kit (MO BIO).

The material was stored at -80°C. Its integrity was evaluated by agarose gel electrophoresis and to evaluate the quantity and purity of the extracted RNA, Nanodrop 1000 was used.

4. Polymerase Chain Reaction (PCR)

Oligonucleotide primers (see Annex 1.), designed to identify and amplify the RXR gene, were based on the homology of invertebrate sequences deposited on GenBank (www.ncbi.nlm.nih.gov/): *Tripedalia cystophora* (Accession number: AF091121.1); partial CDS gene of *Anemonia sulcata* (Accession number: U93415.1, U93416.1); *Hydra magnipapillata* (Accession number: NW_002152543.1); *Trichoplax adhaerens* (Gene ID: 6751230); *Nucella lapillus* (Accession number: EU024473.1); *Saccoglossus kowalevskii* (Gene ID: 100329058). By one hand, specific primers were designed with Primer3 v.0.4.0 (http://frodo.wi.mit.edu/) using the consensus sequence of the MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) alignment of the conserved domains: DBD and LBD. On the other hand, degenerated primers were designed using amino acid sequences in CODEHOP (http://bioinformatics.weizmann.ac.il/blocks/codehop.html). For each set of primers, PCR reactions were firstly run to determine its optimal annealing temperature and magnesium concentration. Oligonucleotide primers (see Annex 1.) designed by Reitzel and Tarrant (2009) were used to amplify a set of different cnidarians NRs sequences and the PCR parameters were followed according to the author's recommendations.

The different combination of PCR strategies was performed in Biometra T Professional Thermocycler (Alfagene) using 20 µL reactions, according to the next scheme:

Components	Initial concentration	Final concentration	Volume (µL)
ddH_2O	--	--	To a final volume of 20
Buffer Taq (Bioline)	14x NH_4	1.4x NH_4	2.0
dNTPs (Bioline)	10 mM	0.2 – 0.4 mM	0.4 - 0.6
$MgCl_2$ (Bioline)	50 mM	2.0 – 2.5 mM	1.0 - 0.8
Primer Foward	10 µM	0.5 µM	1.0
Primer Reverse	10 µM	0.5 µM	1.0
Biotaq ^{TM} DNA Taq polymerase (Bioline)	5 U/µL	--	0.1
DNA template	≥ 100 ng /µL	≥ 5 ng /µL	2.0 – 1.0

Table 2 – Concentration and volume of the components of the Master Mix per PCR reaction.

PCR cycle parameters included an initial denaturation (5 min, 94°C), then 40 cycles of denaturation (30 s, 94°C), of annealing (50 s, annealing temperature) and of extension (30 s, 72°C); and in addiction it was added a final extension (10 min, 72°C).

For a few set of primers (when many bands appeared) was used a Touchdown PCR program, which increases the amplification specificity, sensitivity and yield. As previously described, it was performed an initial denaturation (5 min, 94°C); followed by

10 cycles of denaturation (30 s, 94°C), of annealing (50 s, higher annealing tempeture) and of extension (30 s, 72°C); plus, 25 cycles of denaturation (30 s, 94°C), annealing (50s, lower annealing tempeture) and extension (30 s, 72°C); the final extension was performed for 7 min at 72°C).

The PCR products were analyzed by agarose gel electrophoresis 1 – 2.5 % p/v (TAE 1x buffer [0.4M Tris-acetate, 0.01M EDTA, pH=8.3]) with 3 µg/mL ethidium bromide (EtBr). After the electrophoresis separation, the PCR products were visualized with transilluminator and photographed with Canon PowerShot G9 system.

The PCR products of interest were purified with Diffinity Genomics Rapid Tip or with PureLink ™ PCR Purification Combo Kit Quick Gel Extraction (Invitrogen), according to the manufacturer instructions. The products obtained with the set of primers designed by Reitzel and Tarrant (2009) were directed sequenced (MACROGEN) and the other PCR products were cloned, sequenced and then the results were analyzed with the FinchTV software.

5. RT (Reverse Transcriptional) PCR

The cDNA used to amplify partial sequence of NRs genes was synthesized with the SuperScript III First-Strand Synthesis System (Invitrogen), according to the manufacturer instructions. Thus, cDNA was synthesized, using total 10 ng–1 µg RNA primed with Oligo(dT)18 (2.5 µM) and the reverse transcription was performed with SuperScript III Reverse Transcriptase (10 U/µl) for 50 min at 50°C.

The template of RACE PCRs was the cDNA obtained using SMARTer™ RACE cDNA Amplification Kit (Clontech). The first-strand cDNA synthesis was performed according to the manufacturer instructions; in a 10 µL reaction, 10 ng–1 µg of total RNA, primed with SMARTer II A oligo (1.2 µM) and 3'/5'-RACE CDS Primer A (1.2 µM), and the target was reverse transcribed at 42 ◦C for 90 min, using SMARTScrib Reverse Transcriptase (10 U/µl).

Then, first-strand reaction products were diluted with 10-20 µL Tricine-EDTA Buffer (10 mM Tricine-KOH; 1.0 mM EDTA; pH of 8.5) and stored at - 20°C. The cDNA synthesis efficiency was evaluated trough the amplification of a housekeeping gene (cytochrome *c*-oxidase, COI). In the reaction, 2 µL of each cDNA sample were added to a reaction mix containing 1x PCR buffer, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 4U of Biotaq ™ DNA *Taq* polymerase (Bioline), 23.8 µL of molecular grade PCR H_2O, 10 µM of both standard universal primers, LCO1490-forward and HC02198-reverse, in a total volume of 40 µL per reaction. The cycling parameters were an initial denaturation (1 min, 94°C), then 35 cycles of denaturation (1 min, 94°C), of annealing (30 s, 45°C)

and of extension (1 min, 72°C); and in addiction it was added a final extension (10 min, 72°C).

6. Rapid Amplification of cDNA Ends (RACE)

Gene specific primers (GSP) of 5'/3' RACE (see Annex 1.) were designed based on partial nucleotide sequences. They were primed with the 10X Universal Primer A Mix (UPM) of SMART RACE kit from Clontech to performed 5'/3' RACE PCRs and, when no bands appeared, the products were amplified using GSPs and Nested Universal Primer A (NUP).

Component	Volume (µL)					
	3' RACE	5' RACE	Nested 3' RACE	Nested 5' RACE	Negative Control	
Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific)	10	10	10	10	10	
RACE - Ready cDNA	1.0	1.0	--	--	--	
RACE PCR - Products	--	--	1.0	1.0	--	
UPM (10X) Clontech	2.0	2.0	--	--	2.0	
Specific Forward Primer (10 µM)	1.0	--	1.0	--	1.0	--
Specific Reverse Primer (10 µM)	--	1.0	--	1.0	--	1.0
NUP Clontech (10 µM)	--	--	0.5	0.5	--	
dd H_2O	To a final volume of 20	To a final volume of 20	To a final volume of 20	To a final volume of 20	To a final volume of 20	

Table 3 - Volume of the components of the 5'/3' RACE PCR or 5'/3' Nested RACE per reaction.

The 20 µL reaction was thermal cycled using the PCR parameters recommended by the Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific) instructions, following a Touchdown program: initial denaturation (10 s, 98°C); followed by 10 cycles of denaturation (1 s, 98°C), annealing (5 s, 68°C) and extension (4 min, 72°C); 25 cycles of denaturation (1 s, 94°C), annealing (5 s, GSP's annealing tempeture) and extension (4 min, 72°C), at least the final extension step was performed for 1 min at 72°C. Then, the products were analyzed on a 1.5 % agarose/EtBr gel and cloned. The clones were sequenced by the MACROGEN sequencing company.

7. TA cloning

Retrieved bands were cloned into pGEM-T Vector System I (Promega). The ligation reaction was performed according the manufacturer instructions; nevertheless the incubation was performed overnight at 4°C. TOP10 competent cells were used to be transformed; transformation was performed by heat shock (samples were put 30 min on ice, subsequently were placed for 30 s on a dry bath at 42ºC and then on ice for 2 min). 50 µL of concentrated and diluted cultures were placed onto ampicillin: X-Gal (100:80µg/mL) plates and they were incubated at 37ºC overnight. The transformed colonies were selected by blue-white screening technique. They were used to perform PCR with pUC/M13 Primers in order to check for products of the expected size. Then, the products were recovered with GenElute Plasmid Miniprep (Sigma-Aldrich).

8. Phylogenetic analysis

The phylogenetic relationships of the *Anemonia sulcata* sequences 5' product of Nested RACE and the products of PCR using primers previously reported by Reitzel and Tarrant (2009), were studied trough the construction of a phylogenetic tree. These cDNA and deduced amino acid sequences were aligned and related with sequences from metazoans representative of the diversity of NRs to the subfamily level (*Homo sapiens, Danio rerio, Xenopus laevis, Drosophila melanogaster and Ciona intestinalis*)[1] [Bertrand et al. (2004), Devine (2002), Thornton (2003), Wu (2006) and Wu (2007)]. To complement the analysis sequences of *N. vectensis* were added; the NvNR4, NvNR5, NvNR10 and NvNR15 (Accession number: XP_001638550; XP_001630386; XP_001629708; XP_001631902) as they had strong hit with Expressed Sequenced Tags (ESTs) reported as containing NR genes (Bertrand et al., 2004). The phylogenetic analysis was based on maximum likelihood trees, which were constructed using MEGA5 (Tamura et al., 2001) and MUSCLE NRs alignment. For each dataset the best-fit model of DNA substitution was calculated trough software tools, the mmodel with the lowest BIC scores (Bayesian Information Criterion) was selected and the data was bootstrapped for 1000 replicates.

[1] Accessions numbers listed in the Annex 4.

RESULTS

1. Nucleic Acid extraction and comparison

DNA extractions were successful (Fig. 5); but the evaluation of both protocols revealed that the adapted phenol-chloroform extraction protocol of Pinto et al. (2000) had better results. While the gDNA extracted with this process showed a mean concentration of 273.9 ± 103.1 ng/µL (n=15) and a mean A_{260}/A_{280} of 1.813 ± 0.2459 (n=15); the values of gDNA extracted with Mini kit of Invitrogen corresponded to a mean concentration of 38.675 ± 28.79 ng/µL (n=15) and a mean A_{260}/A_{280} of 1.477 ± 0.5025 (n=15). Only the gDNA samples with a concentration in a range of 50.0-100 ng/µL and an A_{260}/A_{280} ratio near 1.80 were used to perform all molecular analysis.

The samples of total RNA extracted with the UltraClean Tissue & Cells RNA Isolation Kit (MoBIO) provided better quality of total RNA than the guanidiniumthiocyanate: phenol: chloroform method (Chomczynski and Sacchi, 1987). The RNA integrity (Fig. 6) and quantity (mean concentration of 55.60 ± 29.24 ng/µL) was higher; additionally, the RNA purity was also superior (mean A_{260}/A_{280} of 1.925 ± 0.4534; n=15).

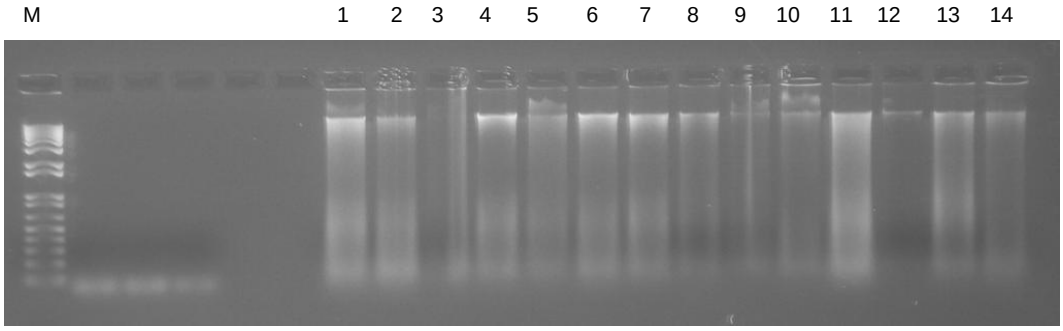


Fig. 5 - 1.5 % Agarose gel electrophoresis; 1- 14: genomic DNA extracted; M - 1 kb plus ladder invitrogen.

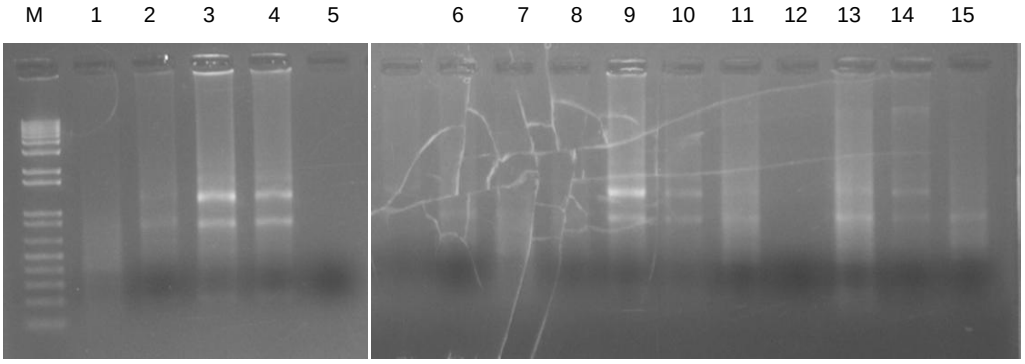


Fig. 6 – 1.5 % Agarose gel electrophoresis; 1 – 16: total RNA extracted; M - 1 kb plus ladder invitrogen.

2. FTZ-F1 and TR2/4 gene identification

The first strategy was the amplification of the well-conserved domain of RXR (DBD), based on gene homology. Since this domain is expected to be in the second exon of the gene (Kostrouch et al., 1998), PCRs were performed with genomic DNA.

Initially, the PCR using the degenerated primers f1 and r1 (see Annex 1.) was performed to determine the optimal annealing temperature (53°C) trough a gradient (see Annex 2.). A PCR product was detected by the formation of a band with the expected size, 133 bp, in *Anemonia sulcata* (Fig. 7). The results of BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi/) for four clones indicated an amplification of a FTZ-F1 (NR5E) gene previously reported by Escriva et al. (1997) (Accession number: U93413.1). According to the results, the sequences shared 92% of identity. This suggests that primers f1 and r1 are too degenerated.

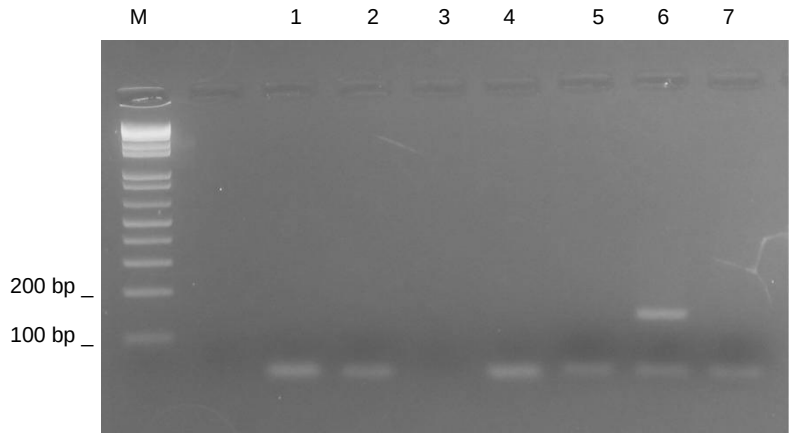


Fig. 7 - 3% Agarose gel electrophoresis - PCR products for primers f1 and r1, 1 - Actinia equina, 2 - Actinia fragacea, 3 - Bunodactis verrucosa, 4 - Actinothoe sphyrodeta, 5 - Urticina felina, 6 - Anemonia sulcata, 7 - Negative Control. M - 1 kb plus ladder invitrogen,

Escriva et al. (1997) also deposited two partial CDS of RXR gene (Accession number: U93415.1, U93416.1) from *A. sulcata* and the sequence U93416.1 was used to design specific primers (f2 and r2, see Annex 1.); therefore *Anemonia sulcata* was used as the positive control throughout the experiment. Initial PCR using these primers resulted in nonspecific bands, so the PCR conditions were optimized *(*see Annex 2.). Then a fragment was amplified (Fig. 8) in four species, the *Urticina felina, Anemonia sulcata, Actinothoe sphyrodeta and Bunodactis verrucosa*. The alignments of the cloned sequences revealed low identity between all species (see Annex 3.). Moreover, their analyses were inconclusive, even querying them against ESTs of NR genes reported by Bertrand et al. (2004) (not shown). Any protein or NRs presented high similarity with these sequences. Therefore, the deposited *A. sulcata* sequence U93416.1 seems to be irreproducible.

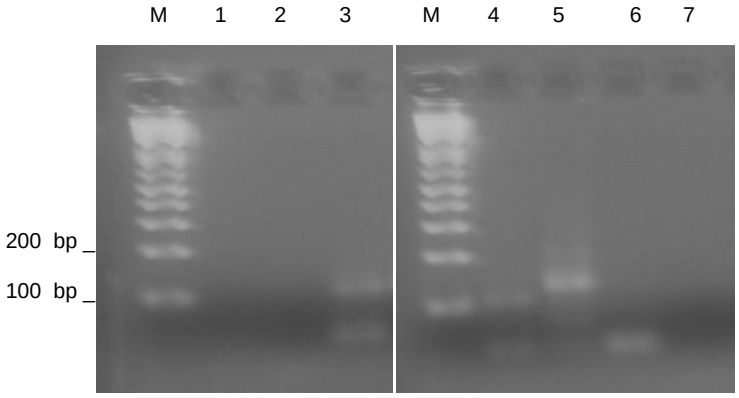


Fig. 8 - 3% Agarose electrophoresis gel - PCR products for primers f2 and r1, 1 - Actinia equine, 2 - Actinia fragacea, 3 - Urticina felina, 4 - Anemonia viridis, 5 - Bunodactis verrucosa, 6 - Actinothoe sphyrodeta, 7 - Negative Control, M -1kb plus ladder Invitrogen.

Since primer R1 was designed based on the conserved FFR/KR/K amino acid signature of the DBD of RXR and primer F2 is specific to a RXR, we tried to combine the primers in order to identify the RXR gene in all species of interest. The amplification was successful using the gDNA of *Urticina felina, Anemonia sulcata, Bunodactis verrucosa* and *Actinothoe sphyrodeta* (Fig. 9). The clones revealed the presence of a sequence with ~ 130 bp. They are potentially cnidarians TR2/4-like receptors because they had a 70-75% similarity with invertebrate receptors (88-98% of coverage).

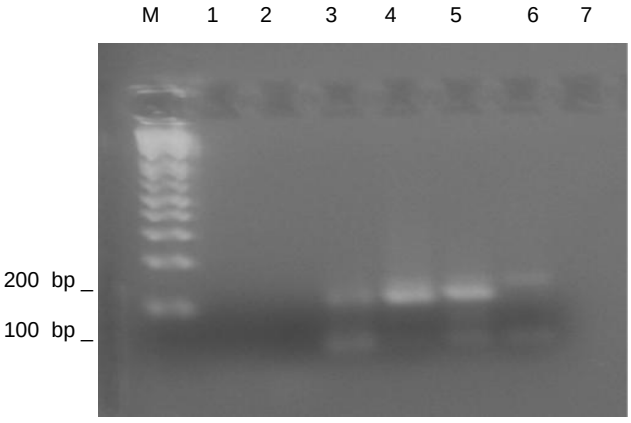


Fig. 9 - 3% Agarose electrophoresis gel - PCR products for primers f2 and r2, 1 - Actinia equina, 2 - Actinia fragacea, 3 - Anemonia sulcata, 4 - Actinothoe sphyrodeta, 5 -Urticina felina, 6 - Bunodactis verrucosa, 7 - Negative Control, M - 1 kb plus ladder Invitrogen.

Another set of degenerated primers was designed, primer f3 and r3 (see Annex 1.), in order to amplify a partial sequence of another conserved domain, the LBD, and they were used to perform PCRs using cDNA. No product was amplified (Fig.10) using these primers even when performing a gradient of temperature and magnesium concentration. The hypothesis of a PCR template problem was discarded since all

cDNA samples were assessed performing the amplification of the gene encoding COI (Fig. 11).

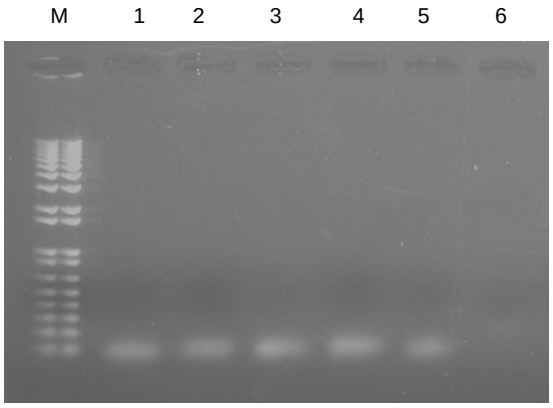


Fig. 8 - 1.3 % Agarose gel electrophoresis - PCR products for primers f4 and r4, 1 - *Actinia equina, 2 - Actinia fragacea, 3 Anemonia sulcata, 4 - Bunodactis verrucosa, 5 - Urticina felina,* 6 – Negative Control, M - 1kb plus ladder invitrogen.

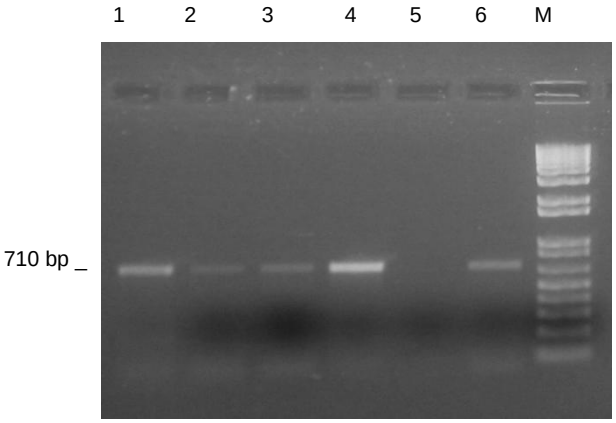


Fig. 9 - Fig. 11 – 1.5 % Agarose gel electrophoresis; 1- 6: PCR products for primers LCO1490 and HC02198 using cDNA as the template; M - 1 kb plus ladder Invitrogen.

3. Partial NR gene isolation in *A. sulcata*

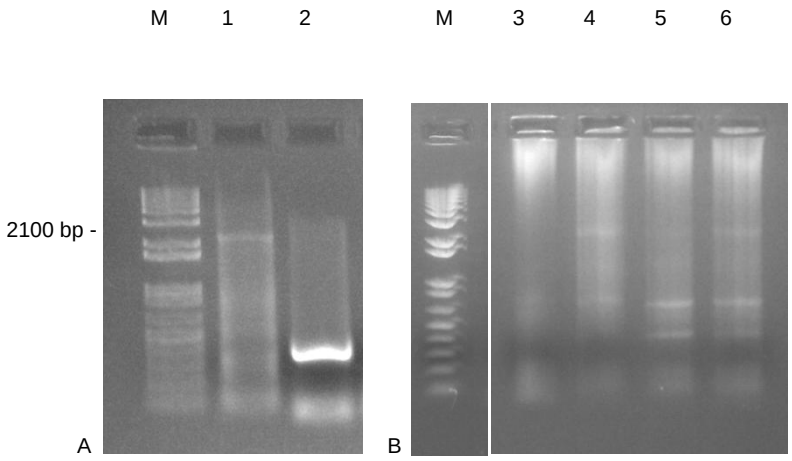
Rapid Amplification of cDNA Ends (RACE) was the chosen technique to provide full-length sequences, starting from a small sequence of the gene. Despite the unsuccessful preliminary results, the next approach was experimented with a reverse primer described by Kostrouch et al. (1998), named in this work as GSP5R (see Annex 1.).

Firstly, the DNA polymerase master mix was optimized for performing a 5’ RACE with the positive control and the 5' RACE TFR Primer (10 µM) from SMARTer RACE Kit and it worked as expected, a 2100 bp 5’ product was amplified (Fig. 12.A). In addition, a Touchdown RACE PCR program was adapted to each GSP (Gene-Specific Primer), according to their annealing temperature.

cDNA template from *A. sulcata* served as a positive control for 5' RACE-PCR with the primer GSP5R. Smears were visible, thus Nested 5´RACEs were performed with varying dilutions (1:2, 1:4 or 1:10) of the 5´products. Different patterns of bands were visualized with each dilution (Fig. 12.B). Then, the Nested 5´RACEs were repeated with the 5'product of 1:4 and 1:10 dilution, implementing a temperature gradient; resulted in two bands (Fig 12.C). However, the sequencing results of the two 5'products were insignificant, indicating that they are nonspecific RACE products; probably because of nonspecific binding of the primer GSP5R in the cDNA.

Due to these results the isolation of other NRs seemed more promising. Since the reverse primers designed in the study are predicted to be closer to the 5´ends of the cDNA. The next step was the development of 5' RACE-PCRs with the positive control using the primer r1 as a GSP. The Nested reactions allowed the appearance of a band near 400 bp (Fig.12.D), and this 5'sequence was named as As_a. The conclusions were ambiguous, because BLAST results showed that it matched with PPARs and RARs which are NRs present just in bilaterian animals (Bridgham et al., 2010), it also matched with a TR4 sequence and moreover it presented 93% identity with a 27 bp portion of the DBD of *Tripedalia cystophora* RXR (Accession number: AF091121.1).

Therefore, a new set of primers was designed within this sequenced product; the primers GSP1f and GSP1r (see Annex 1.). These GSPs were tested with the six target species of the study. Although aall 5' RACE-PCR and 3' RACE-PCR did not revealed formation of bands, the Nested RACEs showed many different products (Fig.12.E and Fig.12.F). However, the sequencing results of the excised bands were not significant because apparently no NR was found.



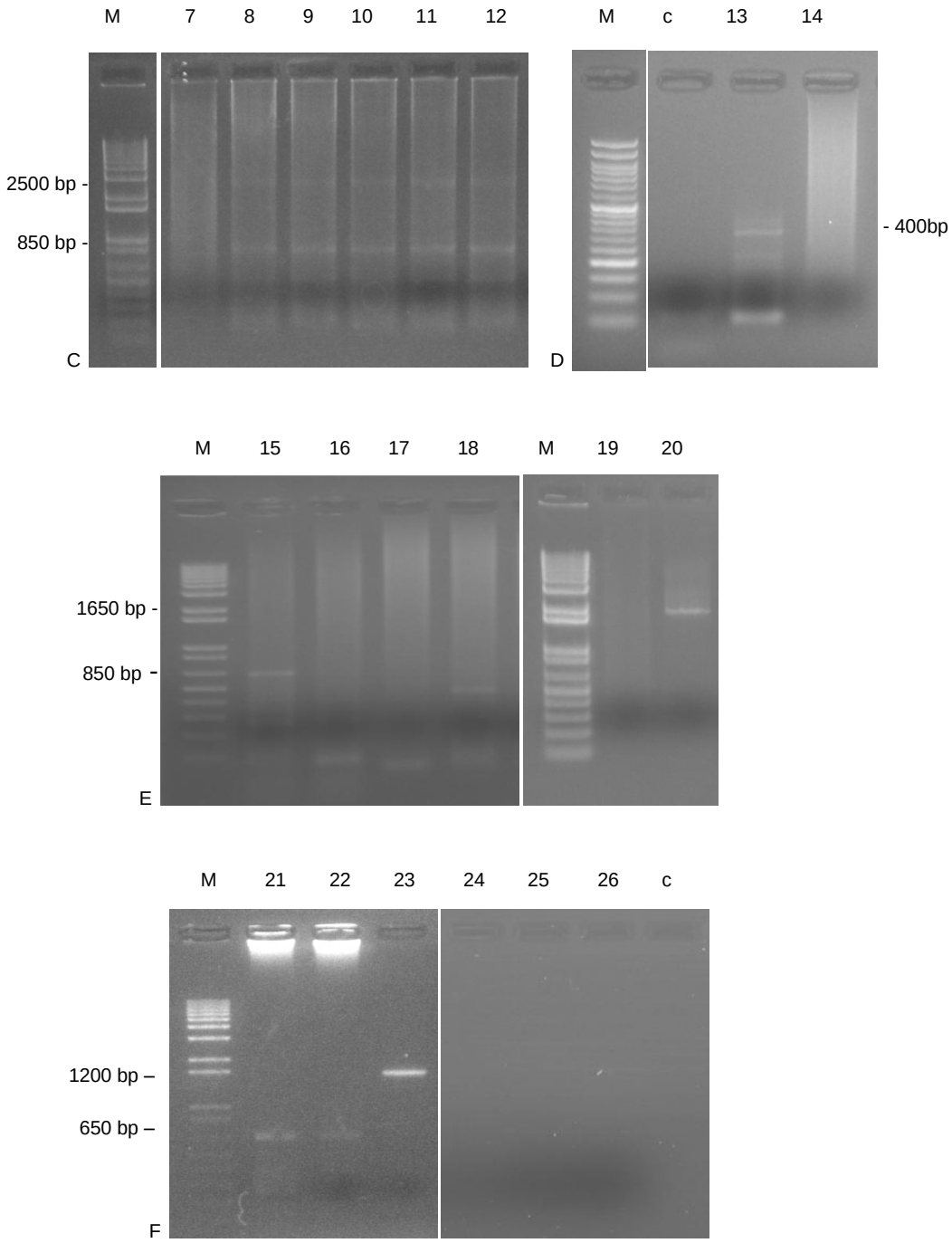


Fig. 10 - 1.5 % Agarose gel electrophoresis; RACE-PCR products. A - Biotaq TM DNA Taq polymerase (Bioline) master mix test with the 5' RACE-cDNA of Control Mouse Heart Total RNA of Clontech, 1 – amplification with Control 5'RACE TFR Primer plus UPM/NUP primer, *2* - amplification with Control 3' RACE and 5'RACE TFR Primers. B and C – Nested 5'RACE products with the GSP5R using Anemonia sulcata cDNA, 3 – without dilution, 4- dilution 1:2, 5- dilution 1:4, 6- dilution 1:10, 7- dilution 1:4 plus 56°C annealing temperature, 8- dilution 1:4 plus 60°C annealing temperature, 9- dilution 1:4 plus 64°C annealing temperature, 10- dilution 1:10 plus 56°C annealing temperature, 11- dilution 1:10 plus 56°C annealing temperature, 12- dilution 1:10 plus 56°C annealing temperature; D - 5'products with primer r1, 13 - 5´RAC product, 14 - Nested RACE product; E – Nested 5'RACE products with GSPr1, 15- *Actinothoe sphyrodeta,* 16- *Anemonia sulcata*, 17- *Actinia fragacea*, 18- *Actinia equina,* 19- *Bunodactis verrucosa*, 20- *Urticina felina*; F – Nested 5'RACE products with GSPr2, 21 - *Actinothoe sphyrodeta*, 22 - *Anemonia sulcata* , 23 - *Urticina felina*, 24 -*Actinia fragacea* , 25 - *Actinia equina*, 26 - *Bunodactis verrucosa*. c – Negative control, M - 1 kb plus ladder Invitrogen, N - NZYDNA Ladder VI Nzytech.

4. Three genes amplification of NR2 family in *A. sulcata*

Most of the *Nematostella vectensis* NRs published (Reitzel and Tarrant, 2009) have strong support to belong to the NR family 2, therefore our interest was the evaluation of relationships between the sequence As_a with other NRs identified in anthozoan species. These NRs were expected to be amplified in *Anemonia sulcata* trough PCR using the primers designed by Reitzel and Tarrant (2009), named in this study as NR3f, NR3r, NR4f, NR4r, NR5f, NR5r, NR6f, NR6r, NR7f, NR7r, NR8f, NR8r, NR9f, NR9r, NR10f, NR10r, NR11f, NR11r, NR12f, NR12r, NR13f, NR13r, NR14f, NR14r, NR15f, NR15r, NR16 and NR16r (see Annex 1.). According to the authors NvNR4 is a homologous of HNF4 (NR2A); NvNR5-NvNR9 seems close to subfamily 2E (TLL/TLX, FAX, PNR) and NvNR10-NvNR14 to COUP-TFs (NR2F family), moreover, NvNR15 and NvNR16 grouped with TR2/4 receptors (NR2C family).

Through this approach, negative and positive results were obtained. On the one hand, the samples with absence of product or multiple products (amplification using the set of primers NR3/4/5/6/8/9)were not sequenced (Fig. 13). On the other, the sequencing data did not hit with the corresponding NvNR sequences (amplification using the set of primers NR7/12/13). Nevertheless, the *Anemonia sulcata* products named as As_10 (~ 1000 bp), As_11 (~ 300 bp) and As_14 (~ 400 bp) matched with its correspondent NvNR, so they are apparent COUP-TF receptors. They should be different isoforms as Reitzel and Tarrant (2009) mentioned “NvNR10 was most closely related to bilaterian COUP-TFs, while NvNR11-NvNR14 are supported as an independent radiation with this subfamily”.

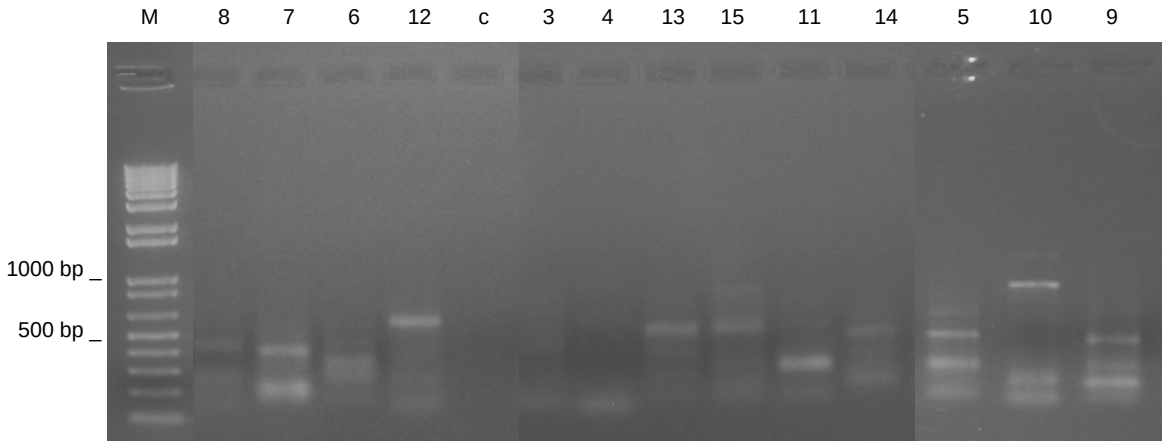


Fig. 11 - 1.5% Agarose gel electrophoresis: products of the set of primers designed by Reitzel and Tarrant (2009) using cDNA of *A. sulcata*. 4 – set of primers NR4f and NR4r; 5 – set of primers NR5f and NR5r; 6 – set of primers NR6f and NR6r; 7 – set of primers NR7f and NR7r; 8 – set of primers NR8f and NR8r; 9 – set of primers NR9f and NR9r; 10 – set of primers NR10f and NR10r; 11 – set of primers NR11f and NR11r; 12 – set of primers

NR12f and NR12r; 13 – set of primers NR13f and NR13r; 14 – set of primers NR14f and NR14r; 15 – set of primers NR15f and NR15r; 16 – set of primers NR16f and NR16r; M - 1 kb plus ladder Invitrogen.

5. Phylogenetic relationships within NR2 family

To understand the evolutionary relationships between the sequenced NRs and the NR2 family, phylogenetic analyses were carried out. In Maximum likelihood (ML) phylogeny the model implemented was Jones-Taylor-Thornton (JTT) for amino acid dataset and Kimura 2-parameter for nucleotide data. Moreover, the evolutionary rate was modeled by the discrete Gamma distribution and the ML heuristic analysis was performed by Close-Neighbor-Interchange (NCI) method.

Both trees were rooted with the HNF4 group. The analyses using nucleotide data had better results; comparing the bootstrap percentage (BP) scores (Fig. 14 and Fig.15). Therefore this tree was considered a better estimation of the arrangement of the NR subfamilies.

Since the interior branch of As_10 and NvNR10 bootstrap percentage is ≥ 80%, it is strongly supported their relationship, so primer NR1f and NR1r should be efficient to the COUP-TF gene amplification in anthozoan species. As_a sequence was predicted as a possible TR2/4 receptor, validating previous results using primer r1 in PCR. As_a and As_11 grouped together so they may represent gene duplication of TR2/4-like receptor. As_14 sequence grouped differently in both analyses; the nucleotide sequences' analysis showed As_14 related to As_a, as NR2C members, but in the amino acid sequences' analysis it is identified as a NR2E member. In both cases the BP are low (BP=32% and BP=25%, respectively). Therefore, these outcomes contradict our predictions of As_14 sequence relationship with COUP-TF receptors.

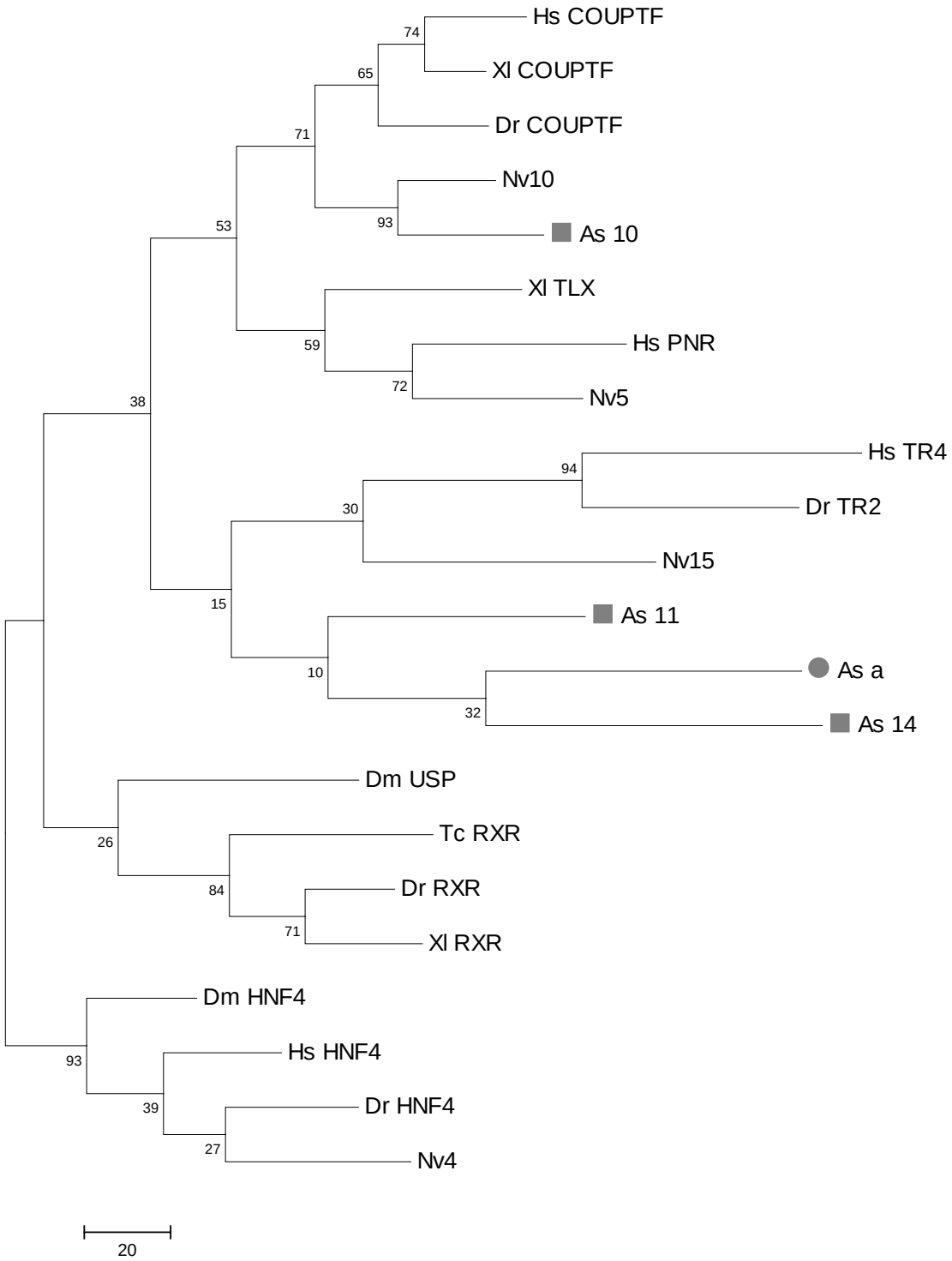


Fig. 12 - ML tree constructed using partial NRs PCR/RACE nucleotide sequences from Anemonia sulcata. Grey square represents the resultant products of the NvNRs primers amplification; grey circle represents the product of 5'RACE. *As, Anemonia sulcata; Hs, Homo sapiens; Tc, Tripedalia cystophora; Ds, Danio rerio; Xl, Xenopus laevis; Dr, Drosophila melanogaster; Ci, Ciona intestinalis; Nv, Nematostella vectensis.*

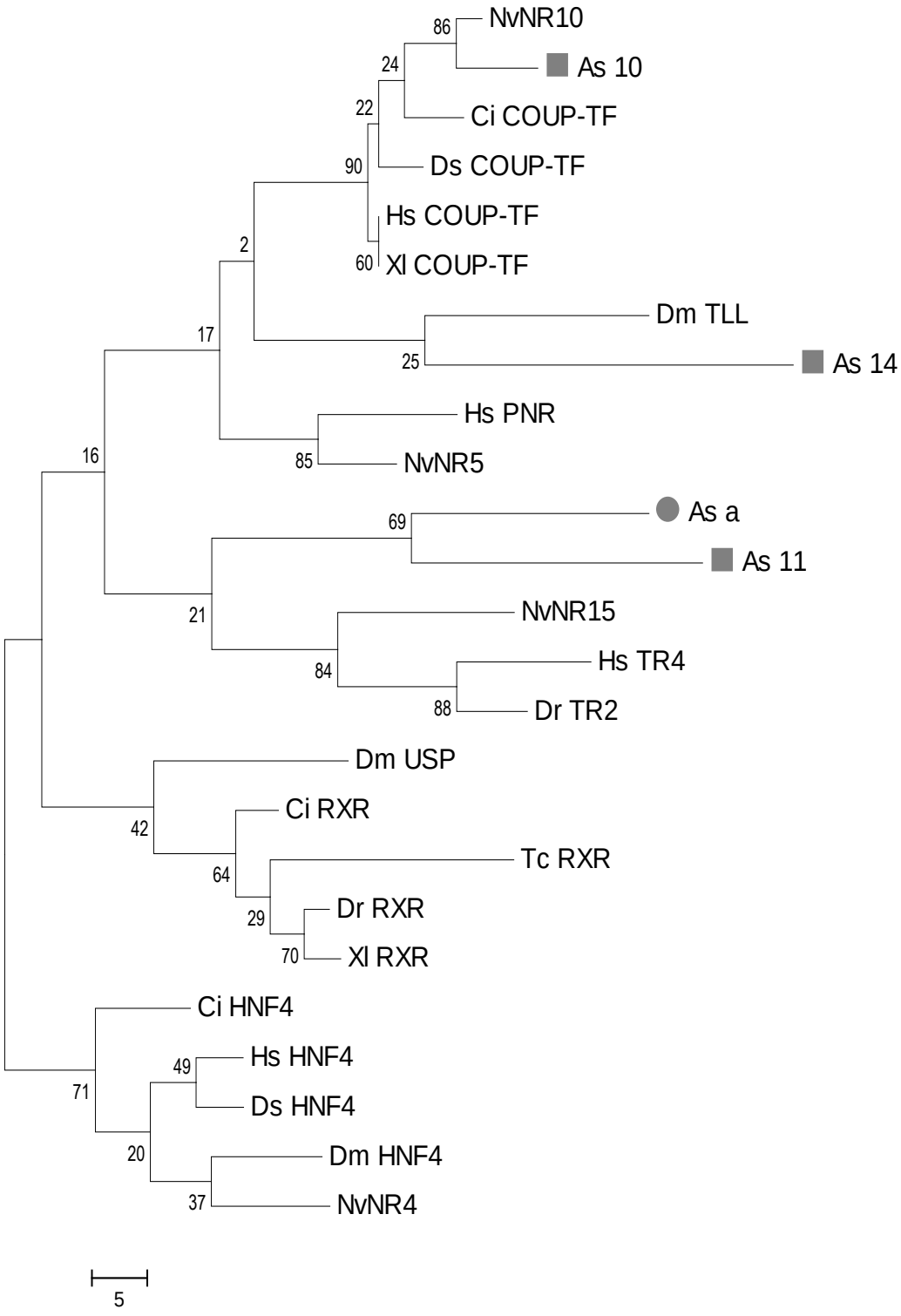


Fig. 13 - ML tree constructed using translated NRs PCR/RACE products from Anemonia sulcata and their relationship to other anthozoan nuclear receptors. Grey square represents the resultant products of the NvNRs primers amplification; grey circle represents the product of 5'RACE. As, *Anemonia sulcata; Hs, Homo sapiens; Tc, Tripedalia cystophora; Ds, Danio rerio; Xl, Xenopus laevis; Dr, Drosophila melanogaster; Ci, Ciona intestinalis; Nv, Nematostella vectensis.*

DISCUSSION

Previous inferences about NRs in the class Anthozoa indicate that the cnidarian-bilaterian ancestor was subjected to RXR gene specific diversification. Grasso et al. (2001) and Escriva et al. (1997) published two probable RXR isoforms of *Acropora millepora* and *Anemonia sulcata*, respectively. But Reitzel and Tarrant (2009) argue that the RXR was lost within the class, as it was not found in *Nematostella vectensis*. To clear up whether RXR gene is present or was lost specifically in the anthozoan-lineage, more species should be analyzed. Therefore, in this study, more five species and one control species were analyzed trough molecular methodologies; we performed PCR and/or RACE-PCRs with degenerated and specific primers in order to get sequences encoding the conserved domains and/or full-length genes. Due to the initial outcomes we inferred that RXR is absent in *Actinia equina, Actinia fragacea, Anthothoe spyrodeta, Anemonia sulcata, Bunodactis verrucosa a*nd *Urticina felina.* Thus, the new objective was to identify other NRs of the most well characterized NR family in phylum Cnidaria, the NR2 family (Reitzel et al., 2001). Since NR2 family contains a diversity of receptors: RXR, COUP, HNF4, EAR2, TR2, TR4 and TLL, we decided to analyse the NR2 members identified in the study of Reitzel and Tarrant (2011).

In this study the *A. sulcata* gDNA and cDNA were amplified and we identified: a previous reported fragment of DBD of FTZ-F1 receptor (NR5E) (Escriva et al., 1997), three partial sequences of a TR2/4-like receptor (NR2C), a gene sequence of a COUP-TF receptor (NR2F) and one sequence of a NR, which was not conclusively assigned to a NR2 family group. Moreover, a fragment of TR2/4-like receptor's DBD was also identified in *Anthothoe spyrodeta, Bunodactis verrucosa a*nd *Urticina felina* specimens.

In fact, the levels of similarity exhibited between sequence As_11 (TR2/4-like receptor) and As_a (TR2/4-like receptor) are only slightly (46%), but both aligned with the fragments of TR2/4-like receptor's DBD amplified in *Anemonia sulcata, Anthothoe spyrodeta, Bunodactis verrucosa a*nd *Urticina feline* (see Annex 3). The alignment of DBD TR2/4 sequences of these four species has a high level of identity (see Annex 3) validating that this TR2/4 domain is well conserved. This similarity can be a clue that evolutionary rates for nuclear genes are slow, congruent with previous studies that showed a slow molecular evolution within the order Actiniaria (Daly et al., 2008). But the data available in this work is not enough to make strong conclusions about the evolution rate of some NR among anthozoan species.

Interestingly, TR2/4 receptor was not identified in *Actinia equine* and *Actinia fragacea.* However, the out-group of the six target species is *Anthothoe spyrodeta*, because it belongs to a different family, the family Sagartiidae. In addition, the FTZ-F1 receptor was just amplified in the *A. sulcata* specie; possibly a mutation occurred in its DNA allowing the proper annealing of primer f1 and r1, which did not occur in the other species.

Nucleotide and amino acid sequences evolutionary relationships were analyzed; the analyses using amino acid data resulted in more poorly supported nodes throughout the tree, comparing the bootstrap percentage (BP). Thus, nucleotide analysis was chosen as the reference analysis, confirming the Simmons (2000) inference that trees from nucleotide sequence data are preferred because the amino acid sequences disregard some information due to the redundancy of the genetic code.

The trees were rooted with the HNF4 group attending to the many references of evolutionary studies; they mentioned that a HNF4-like is potential the original nuclear receptor, which was identified in the sponge *Amphimedon queenslandica* genome according a parsimonious analysis (Bridgham et al., 2010)

The phylogenetic analysis indicates a difficulty placing of As_14 with one of the established groups. This fact could indicate that it represents an ancestral gene or it is a member of other NR family. Although As_a and As_11 were not clustered with strong supported (BS=15%) to the TR2/4 group, these phylogenetic relationship was reinforced by the alignment performed with all probable TR2/4 DBD sequences retrieved (see Annex 3). Therefore, these results contrast with the preliminary inferences, As_11 and As_14 are not COUP-TFs members and As_10 is the only COUP-TF receptor isoform identified.

Moreover, since Reitzel et al. (2009) have reported the first complete set of NRs in the cnidarian *N. vectensis* from suborder Nyantheae, the same order of the studied species, the same PCRs conditions were used to explore the relations between the products and the correspondents NvNRs.

Despite the specificity of the primers to NvNRs sequences, we amplified three different NRs from the NR2 family from the species, which was used as the positive control for the current study. Phylogenetic analysis of the *A. sulcata* sequences was performed with representative *Acropora millepora* NRs, instead of the NvNRs, and the results were not significant (not shown).

With our results we showed the presence of COUP-TFs (Chicken Ovalbumin Upstream Promoter Transcription Factor) in the class Anthozoa. They are considered

as orphan receptors (Gauchat et al., 2004) and there are two widely expressed main of COUP-TFs, isoforms I and II, which have high homology (Ferrer-Martínez et al., 2004). Moreover, these NRs have been identified in anthozoans *Acropora millepora, Pocillopora damicornis* and *Nematostella vectensis.* COUP-TFs have also been identified in hydrozoans, flatworms and lancelets (Reitzel and Tarrant, 2009).

We found two *A. sulcata* TR2/4-like receptors (testicular receptor 2/4) isoforms and many isoforms have been identified in previous studies; numerous TR2 and TR4 variants have been identified from different species, including vertebrates (e.g. human, murine, rabbit, fish, and amphibian) and invertebrates (e.g. *Drosophila, sea urchin* and *nematode)* (Ghang et al., 2002). In addition, TR2/4 receptors have been also identified in the anthozoans *Acropora millepora* and *Nematostella vectensis*. These receptors also do not have specific ligands identified (Young et al., 1998) and, curiously, they suppress the transcriptional activity of other NRs.

The inefficiency of the 5'RACE technique using primer GSP5R, in order to obtain a RXR sequence of an RNA transcript (from an internal sequence to the 5' end of the RNA transcript) was not an evidence of the loss of the gene. This primer was based on the DBD signature, the spacing of Cyst and the conserved FFR/KR/K sequence of this NR (Kostrouch et al., 1998) but it was only proved the amplification of a 5'product in a hydrozoan, thus it can be not able to anneal an anthozoan RXR sequence.

Other problem to gene isolation trough the RACE-PCR can be related to gene expression level; if it is weak, no product is obtained. Little is known about the RXR expression in cnidarians despite a study, using an antibody, confirmed the presence of a RXR-like receptor in the anthozoan *Renilla koellikeri* and determined the localization of this receptor type in epithelial nerve tissue in both *R. koellikeri* and *A. millepora* (Bouzaiene et al., 2007). The antibody was raised against a specific epitope of human RXRα, with significant sequence similarity with a sequence reported by Grasso et al. (2001). However, this coral RXR sequence was indicated as most similar to an orphan receptor by the phylogenetic analyses of Reitzel and Tarrant (2011). This is evidence that expression patterns of this gene are not well characterized.

Nevertheless, the implementation of the 5'/3'RACE-PCR was also difficult in the isolation of a TR2/4 gene. The problems can be related to the complexity of some reactions and some patterns of bands appeared almost as smears, as we verified in the 5'RACE-PCRs. Moreover, in the study it was not confirmed whether this sequence corresponds to a complete or incomplete 5'product, since no method of cDNA synthesis can guarantee a full-length cDNA in the 5' end.

The aim to assess the presence/absence of RXR gene was achieved, and we inferred that its loss occurred within class Anthozoa But, given this fact; it was not possible to neither obtain the complete sequence of the gene nor assess their evolutionary relationships. The last appointed goal was achieved because other NRs were amplified and its identification was supported by phylogenetic analyses. Therefore, with this work we increase a little the dataset related to the NR2 family, specifically in an early metazoan lineage, so these results are relevant. Due to limited sequence sampling, the evidences of diversification of NR super family are ambiguous, especially, about NRs within groups closer to the root of the metazoan origin (Bridgham et al., 2010).

For instance, RA signalling evolution outside chordates remains elusive (McKenna, 2012), despite the increase of analysis of RXR receptor. According to some reports previously mentioned, the presence RXR in Phylum Cnidaria is proposed in *Anemonia sulcata* (Escriva et al., 1997), *Tripedalia cystophora* (Kostrouch et al., 1998), *Acropora millepora* (Grasso et al., 2001) and *Hydra magnipapillata* (Reitzel et al., 2011). However, new sequenced genomes led to controversial results. For instance, Reitzel et al. (2009) proposed that RXR had a secondarily lost within class Anthozoa; since RXR is absent in *Nematostella vectensis* and it was found in *Trichoplax adhaerens* (Phylum Placozoa) and in *Hydra magnipapillata* (Phylum Cnidaria) genomes (Bridgham et al., 2010; Reitzel et al., 2011). Within metazoans, Placazoa is most ancient than phylum Cnidaria and class Anthozoa is the basal class of Cnidaria. This fact means that specific evolutionary events occurred within the class Anthozoa.

Our analyses emphasize the Reitzel et al. (2011) hypothesis as, in the study, our sequences did not match or group with RXRs sequences or ESTs. To confirm the presence/absence of the gene, further it is promising to perform expression pattern analysis within the different cnidarian classes; these studies can be carried out with antibody-linked probes, an approach reported by Anctil et al. (2007). Or alternative techniques like cDNA library screening can be employed. Moreover, the increase of sequencing genomes of evolutionary key positioned groups of organisms will provide more insight about genes and consequently its diversification.

Moreover, expression analyses also play an important role in improving our understanding of the signalling and metabolic pathways. Future studies are also needed to understand and characterize the mechanisms of bio regulation of cnidarians, in order to explore possible consequences of the loss of the RXR gene in the signalling pathway; to evaluate whether cnidarians conserve the ability to be regulated by endogenous and/or exogenous 9-cis-RA. Then, the knowledge about these

mechanisms in early metazoan-lineages can provide new insights about signalling pathways in higher lineages.

REFERENCES

Bouzaiene, M., Angers, a, & Anctil, M. (2007). Immunohistochemical localization of a retinoic acid-like receptor in nerve cells of two colonial anthozoans (Cnidaria). *Tissue & cell*, *39*(2), 123–30. doi:10.1016/j.tice.2007.02.001

Bridgham, J. T., Eick, G. N., Larroux, C., Deshpande, K., Harms, M. J., Gauthier, M. E. a, … Thornton, J. W. (2010). Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS biology*, *8*(10). doi:10.1371/journal.pbio.1000497

Escriva, H., Delaunay, F., & Laudet, V. (2000). Ligand binding and nuclear receptor evolution. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *22*(8), 717–27. doi:10.1002/1521-1878(200008)22:8<717::AID-BIES5>3.0.CO;2-I

Escriva, H., Safi, R., Hänni, C., Langlois, M. C., Saumitou-Laprade, P., Stehelin, D., … Laudet, V. (1997). Ligand binding was acquired during evolution of nuclear receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 6803–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=21239&tool=pmcentrez&rendertype=abstract

Grasso, L. C., Hayward, D. C., Trueman, J. W., Hardie, K. M., Janssens, P. a, & Ball, E. E. (2001). The evolution of nuclear receptors: evidence from the coral Acropora. *Molecular phylogenetics and evolution*, *21*(1), 93–102. doi:10.1006/mpev.2001.0994

Kostrouch, Z., Kostrouchova, M., Love, W., Jannini, E., Piatigorsky, J., & Rall, J. E. (1998). Retinoic acid X receptor in the diploblast, Tripedalia cystophora. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(23), 13442–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24838&tool=pmcentrez&rendertype=abstract

Philip, S., Castro, L. F. C., da Fonseca, R. R., Reis-Henriques, M. a, Vasconcelos, V., Santos, M. M., & Antunes, A. (2011). Adaptive evolution of the Retinoid X receptor in vertebrates. *Genomics*, 1–9. doi:10.1016/j.ygeno.2011.12.001

Reitzel, A. M., Pang, K., Ryan, J. F., Mullikin, J. C., Martindale, M. Q., Baxevanis, A. D., & Tarrant, A. M. (2011). Nuclear receptors from the ctenophore Mnemiopsis leidyi lack a zinc-finger DNA-binding domain: lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily? *EvoDevo*, *2*(1), 3. doi:10.1186/2041-9139-2-3

Reitzel, A. M., & Tarrant, A. M. (2009). Nuclear receptor complement of the cnidarian Nematostella vectensis: phylogenetic relationships and developmental expression patterns. *BMC evolutionary biology*, *9*, 230. doi:10.1186/1471-2148-9-230

Wiens, M. (2003). Retinoid X receptor and retinoic acid response in the marine sponge Suberites domuncula. *Journal of Experimental Biology*, *206*(18), 3261–3271. doi:10.1242/jeb.00541

de Groot, A., de Rosny, E., Juillan-Binard, C., Ferrer, J.-L., Laudet, V., Pierce, R. J., … Borel, F. (2005). Crystal structure of a novel tetrameric complex of agonist-bound ligand-binding domain of Biomphalaria glabrata retinoid X receptor. *Journal of molecular biology*, *354*(4), 841–53. doi:10.1016/j.jmb.2005.09.090

Bertrand S, Brunet F, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi, M. (2004). Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Molecular Biology and Evolution*, 21,1923-1937.

Bouton, D., Escriva, H., de Mendonça, R., Glineur, C., Bertin, B., Noël, C., Pierce, R. (2005). A conserved retinoid X receptor (RXR) from the mollusk *Biomphalaria glabrata* transactivates transcription in the presence of retinoids. *Journal of molecular endocrinology*, *34*(2), 567–582.

Bridge, D., Cunningham, C., Schierwater, B., DeSalle, R., Buss, W. (1992) Class-level relationships in the phylum Cnidaria: Evidence from mitochondrial genome structure. *Proc. Nat. Acad. Sci. USA*, *89,* 8750-8753.

Bridge, D., Cunningham, R., DeSalle, Buss, W. (1995). Class-level relationships in the phylum Cnidaria: Molecular and morphological evidence. *Molec. Biol. Evol., 12,* 679-689.

Bridgham, T. (2010). Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Super family from a Ligand-Dependent Ancestor. *PLoS Biology, 8,* 1-13.

Chang, C., Kokontis, J., Acakpo-Satchivi, L., Liao, S., Takeda, Chang Y. (1989) Molecular cloning of new human TR2 receptors: A class of steroid receptor with multiple ligand-binding domains. *Biochemical and Biophysical Research Communications,* 165(2), 735–741.

Chapman A., Kirkness, F., Simakov, O., Hampson, E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, G., Borman, J., Busam, D., (2010). The dynamic genome of *Hydra*. *Nature*, *464*, 592-596.

Chapman, A., Chapman J., Kirkness E., Simakov O., Hampson S., Mitros T., Weinmaier T., Rattei T., Balasubramanian P., Borman J., Busam D. (2010). The dynamic genome of Hydra, *Nature, 464*, 592 -595.

Chomczynski, P., Sacchi, N., (1987). Single-step method of RNA isolation by acid guanidiniumthiocyanate-phenol-chloroform extraction. *Anal. Biochem., 162*, 156–159.

Daly, M., Chaudhuri, A., Gusmão, L., Rodríguez, E., (2008) Phylogenetic relationships among sea anemones (Cnidaria: Anthozoa: Actiniaria). *Molecular Phylogenetics and Evolution*, *48(1*), 292–301.

Daly, M., Chaudhuri, A., Gusmão, L., Rodríguez, E., (2010) Phylogenetic Signal in Mitochondrial and Nuclear Markers in Sea Anemones (Cnidaria, Actiniaria). *Comp. Biol*., *50 (3)*, 371-388.

Devine, C,.Hinman, V., Degnan, B. (2002). Evolution and developmental expression of nuclear receptor genes in the ascidian Herdmania. *International Journal of Developmental Biology, 46:*687-692.

Gampe, R., Montana, V., Lambert, Wisely, Milburn, (2000) Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix. *Genes & Development., 17*, 2229–2241.

Gauchat, D., Escriva, H., Miljkovic-Licina, M., Chera, S., Langlois, M., Begue, A., Laudet V, Galliot B. (2004). The orphan COUP-TF nuclear receptors are markers for neurogenesis from cnidarians to vertebrates. *Developmental Biology*, *275*, 104-123.

Gittoes, McCabe, Verhaeg, Sheppard, Franklyn (1998) Retinoid X receptor expression in the normal pituitary and clinically 'non-functioning' pituitary tumours. *Clinical Endocrinology, 48*, 425-433.

Iwema, Billas, Beck, Bonneton, Nierengarten, Chaumot, Richards, Laudet, Moras (2007). Structural and functional characterization of a novel type of ligand-independent RXR-USP receptor. *EMBO Journal, 26 (16),* 3770–3782.

Mangelsdorf, D., Borgmeyer, U., Heyman, R., (1992) Characterization of three RXR genes that mediate the action of 9cis retinoic acid. *Genes & Development., 6*, 329-344.Mark, M., Ghyselinck, B., Chambon, P., (2009). Function of retinoic acid receptors during embryonic development. *Nucl Recept Signal,* 7, e002.

Martínez, Marotta, Baldán, Haro, Gómez-Foix (2004). Chicken ovalbumin upstream promoter-transcription factor I represses the transcriptional activity of the human muscle glycogen phosphorylase promoter in C2C12 cells. BBA - *Gene Structure and Expression*, 1678, 157–162.

McKenna, J. (2012). EMBO Retinoids 2011: Mechanisms, biology and pathology of signalling by retinoic acid and retinoic acid receptors. *Nucl Recept Signal*., *10,* e003.

Nuclear Receptor Nomenclature Committee (1999). A unified nomenclature system for the nuclear receptor super family. *Cell, 97,* 161-163.

Picton, B., Morrow, C., (2010) Encyclopedia of Marine Life of Britain and Ireland. *(Online), http://www.habitas.org.uk/marinelife/taxonfulllist.asp.*

Pinto, S., Fernandes-Matioli, F, Schlenz, E. (2000) DNA extraction from sea anemone (Cnidaria: Actiniaria) tissues for molecular analyses. *Genetics and Molecular Biology*, *23 (3)*, 601-604.

Schulman, I., Li, C., Schwabe, W., Evans, M., (1997). The phantom ligand effect: allosteric control of transcription by the retinoid X receptor, *Genes Dev., 11*, 299-308.

Simmons, M. (2000). A fundamental problem with amino-acid-sequence characters for phylogenetic analyses. *Cladistics,* 16, 274-282.

Stachowitsch, M. (1992). The Invertebrates: An Illustrated Glossary, *New York: Wiley - Liss*.

Steineger, H., Arntsen, B., Spydevold, O., Sorensen, H., (1998). Gene transcription of the retinoid X receptor alpha (RXRalpha) is regulated by fatty acids and hormones in rat hepatic cells, *J. Lipid. Res., 39,* 744–754.

Sterrer, W. (1986). Marine fauna and flora of Bermuda: A systematic guide to the identification of marine organisms. New York: *John Wiley & Sons, Inc.,* 742.

Sullivan, C., Ryan, F., Watson. A., Webb, Mullikin, C., Rokhsar, Finnerty, R. (2006). *StellaBase*: The *Nematostella vectensis* Genomics Database. *Nucleic Acids Res., 34*, 495–499.

Tamura, D., Peterson, Stecher, G., Nei, Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis UsingMaximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol*. 28(10), 2731–2739.

Tarrant, A., Cortes, J., Atkinson, M., Atkinson, S., Johanning, K, Chiang, Vargas, J., McLachlan, J. (2008) Three orphan nuclear receptors in the scleractinian coral Pocillopora damicornis from the Pacific coast of Costa Rica. *Rev Biol Trop*, *56*, 39-48.

Thornton, W., Need, E., Crews, D. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signalling. *Science, 30,* 1714-1717.

Tocchini-Valentini, Rochel, Escriva, Germain, Peluso-Iltis, Paris, Sanglier-Cianferani, Dorsselaer, Moras, Laudet, (2008). Structural and functional insights into the ligand binding domain of a non-duplicated RXR from the invertebrate chordate amphioxus *JBC,* Manuscript M805692200, 1-15.

Wu, W., Edward, G., Niles, Hirai, H., LoVerde, T. (2007). Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evolutionary Biology*.

Wu, W., Niles E., El-Sayed, N., Berriman, M, LoVerde T. (2006). Schistosoma mansoni (Platyhelminthes, Trematoda) nuclear receptors: Sixteen new members and , Sixteen new members and a novel subfamily. *Gene*, 366, 303-315.

Young, Y., Lee, S., Smithǁ, Chang, S. (1998) A Bidirectional Regulation between the TR2/TR4 Orphan Receptors (TR2/TR4) and the Ciliary Neurotrophic Factor (CNTF) Signaling Pathway. *The Journal of Biological Chemistry*, 273, 20877-20885.

ANNEXES

Annex 1: List of primers

Primer Name	Sequence (5' to 3')	Length (bp)	[2] Tm (°C)	Method	Template
f1	TGYGAKGGMTGTMRWGGGTT	20	57.6	PCR	gDNA
r1	CATTTCTGGAAWCGACARTAYTGRCA	26	56.4	PCR	gDNA
f2	TGTGAGGGGTGTAAAGGGTT	20	55.0	PCR, RACE	gDNA
r2	CAATATTGACACCTGTTGCG	20	52.2	PCR, RACE	gDNA
f3	CGGAACCGATGCCARTWYTGYMG	23	60.5	PCR	cDNA
r3	AARYGNTTYVACGACGACGCCGAC	24	63.1	PCR	cDNA
GSP5R	ACAGTACTGACACCGGTTACGGCT	24	61.7	RACE	RACE-cDNA
GSP1f	TATACTGCATGCGCTTGACC	20	55.1	RACE	RACE-cDNA
GSP1r	TCGACAGTATTGGCACAAGAA	21	54.1	RACE	RACE-cDNA

1.A - List of the sequence primers of the RXR gene analyses. Degenerated primers: K = G or T, M = A or C, R = A or G, S = C or G, W = A or T, Y = T or C.

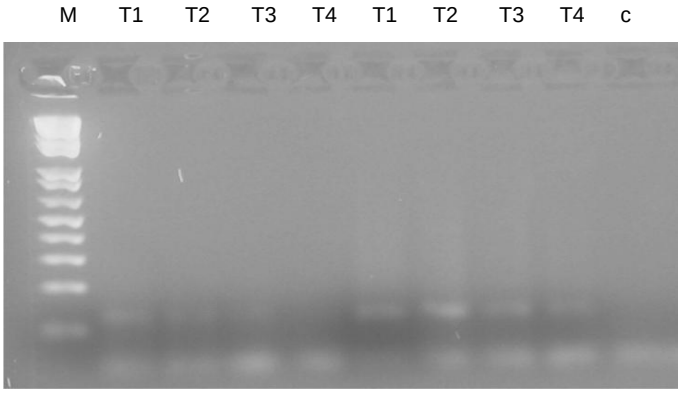
Primer Name	Primer Sequence (5 to 3)	Length (bp)	Tm (°C)	Template
NR3r	GCATCAGGACTTAGTAGCACCAAAC	25	58.0	cDNA
NR4f	CTTCTTATTTTGGTGGAGTGGGC	23	56.1	cDNA
NR4r	GCAGTTCTTTTTGGTTGTGAGCG	23	57.9	cDNA
NR5f	CGACAAACGGAAAGGGGAACTG	22	58.4	cDNA
NR5r	AATCGGGCTACCTGTGTGGGATAG	24	59.9	cDNA
NR6f	TTCATCAGCAGGCAACCGAG	20	57.3	cDNA
NR6r	TGGGTGGGGTGTTTTGTCATCG	22	59.6	cDNA
NR7f	GACGGTTGTAGCGGTTTCTTTATG	24	56.7	cDNA
NR7r	CCCAAATCGGACTTGTTGACC	21	56.0	cDNA
NR8f	TGTCAGCCGATCAAATCAAA	20	52.3	cDNA
NR8r	ACGAAGACCCCGTATGTCAG	23	55.9	cDNA
NR9f	GCAAGTGTGTTGTGGATGTAGCC	22	58.8	cDNA
NR9r	CAGTCGTTTGAACCTCGTGACG	22	58.5	cDNA
NR10f	CACAAACGATGGCTCAAGGAAC	22	56.8	cDNA
NR10r	GGACGATTGCTTTCAGACACGC	21	59.1	cDNA
NR11f	GCGACAAATCTTCGGGGAAAC	24	56.7	cDNA
NR11r	AAAACAGGTTCTCAACTGACACCG	23	57.9	cDNA
NR12f	AGAAATAGCAGAAGTCGTTGCCC	20	57.4	cDNA
NR12r	GGAGAAGTTTGCCGAATCGC	20	56.7	cDNA

[2] Tm (annealing temperature) was evaluated by AmplifiX 1.5.4 (specific primers) and it was evaluated by OligoAnalyzer 3.1 (degenerated primers)

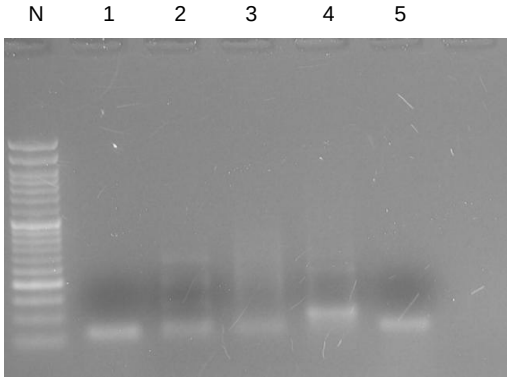
NR13f	CCCAGACAAGACCATCGAGT	20	55.7	cDNA
NR13r	ACCAACGAGAACGCAGTACC	20	57.0	cDNA
NR14f	AAGTGCTTAGCCGTCGGAATG	21	57.6	cDNA
NR14r	CTTGTCCAAATGTTGACCTGGTG	23	56.5	cDNA
NR15f	AATCCATCCCCATTTGCCTG	20	54.8	cDNA
NR15r	CCTTCCCAATCTTATGTCACGC	22	55.8	cDNA
NR16f	CAGGAAAACACTACGGAGTCGTTG	24	58.1	cDNA
NR16r	GGGGAGTCTGAGGAGAATCTTAGC	24	57.9	cDNA

1.B - List of the sequence primers described by (Reitzel and Tarrant, 2009).

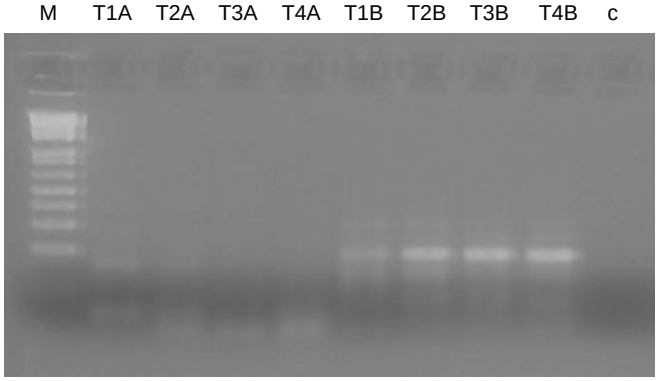
Annex 2: Electrophoresis gel of PCR products - RXR gene



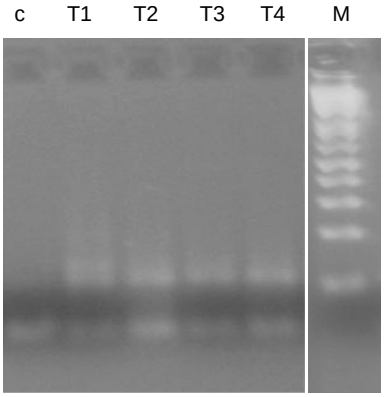
2. A - 3% Agarose gel electrophoresis: PCR products of degenerated primer f1 and r1 using gDNA of *Anemonia sulcata* as template. T1 – 48.0°C, T2 – 49.5°C, T3 – 53.0°C, T4 – 56.0°C, c - Negative control. M - 1 Kb plus ladder invitrogen.



2. B - 2% Agarose gel electrophoresis: PCR products of specific primer f2 and r2 using gDNA. 1 - *Actinia equina; 2 - Bunodactis verrucosa;* 3 – *Actinia fragcea; 4 - Anemonia sulcata;* 5 – *Urticina felina*; N - NZYDNA Ladder VI Nzytech.



2. C - 3% Agarose electrophoresis gel: PCR products with specific primers primer f2 and r2 using gDNA. A - *Anemonia sulcata.* B - *Bunodactis verrucosa.* T1 – 48.0°C, T2 – 49.5°C, T3 – 52.8°C, T4 – 54.4°C, T5 – 56.0°C. c – Negative control. M - 1 Kb plus ladder Invitrogen.



2. D – 2% Agarose gel electrophoresis: PCR products with degenerated primer f2 and r1 using cDNA of *Bunodactis verrucosa*. T1 – 52.2°C, T2 – 54.0°C, T3 – 55.7°C, T4 – 56.8°C, c – Negative control. M -1 Kb plus ladder Invitrogen.

Annex 3: Alignments

```
Saccoglussus/1-228       ----------------------------------GGAAACACTATGGCGT--CTATAGCT
Strongylocentrotus/1-195 ----------------------------------GCAAGCATTATGGAGT--CTACAGTT
Schistosoma/1-231        CCAATTTGTGTGATTTGTGGTGATAAAGCCAGTGGTAAACATTATGGTGTGATTTCA--T
Tricoplax/1-222          ------------------------------------GGCATTATGGCGT--TTACAGTT
Tripedalia/1-225         CAACCATGCTCTGTATGCTCCGATAAGGCCTATGTCAAACATTATGGTGTGTTTGCA--T
Hydra/1-64               ---------------------------------------CATTATGGTGTTTTTACATGT
Anemonia2/1-121          -------------------------------------------GTAAGGGGTTTTT-TAGGC

Saccoglussus/1-228       GCGAAGGATGTAAAGGCTTTTTCAAACGAACAGTTCGAAAAGATTTGCAT-TACGCCTGC
Strongylocentrotus/1-195 GTGAAGGCTGCAAAGGGTTCTTCAAGAGGACAGTACGGAAAGACCT-CACGTATACCTGT
Schistosoma/1-231        GTGAAGGATGTAAGGGATTTTTTAAACGAACAGTGCGTAAACAATT-AGTATATGTTTGT
Tricoplax/1-222          GTGAAGGTTGCAAGGGTTTCTTCAAGCGTACAGTGCGCAAAAATTT-AACTTACACATGC
Tripedalia/1-225         GCGAGGGCTGTAAAGGGTTTTTTAAGAGGTCAGTTCGCAACAACCG-GAAGTATTCTTGC
Hydra/1-64               GATGGATGTCGTGGGTTTTTTAAACGTGCAGTTCGCAGAAACC-----------------
Anemonia2/1-121          GAAGTAT-CAAAAAGCGCCTCTATTATACATGCCGT-GTGGCTGG-----TTGCTGCCC-

saccoglussus/1-228       CGGGATGAGAAGAACTGCATCGTTGACAAAAGACAGCGCAACCGTTGCCAATATTGTCGA
Strongylocentrotus/1-195 CGGGACGATCGTAACTGTATGGTCGACAAGAGACAGAGAAATCGGTGCCAATATTGCCGC
Schistosoma/1-231        CGGGAATCAGGTCAATGTCCTGTAGATAGACGGAAGCGCACACGTTGTCAGCATTGCCGT
Tricoplax/1-222          CGTGATAATCGAAACTGTGATATTGATAAAAAGCAACGAAATCGATGTCAGTATTGTCGA
Tripedalia/1-225         CTTGGAAAGCGGCATTGCGACACTGATAAAAAGAGCCGTAACCGGTGTCAGTACTGTCGA
Hydra/1-64               ---------------------------------------------
Anemonia2/1-121          CC-AAGTGT-------------------------------------

Saccoglussus/1-228       TATCAGAAATGCATTGCTATGGGGATGCGCCGAGAAG--
Strongylocentrotus/1-195 TATCAGAAGTGCCTGGGCATGGGCATGCGCAGAGAAG--
Schistosoma/1-231        TTCGAGCAGTGCTTAGCAAAGGGAATGAAGAAGGAAG--
Tricoplax/1-222          TACCAGAAGTGTCTGCAAGTTGGTATGAAGCAAGAAG--
Tripedalia/1-225         TTCCAGAAATGTGTGCAAGTTGGCATGAAACCAGAAGCTGTCCAGGAT------
```

3. A – Muscle alignment of RXR genes sequences of DBD (design primers).

```
Saccoglossus/1-612       --GACATGCCGGTGGAGAAGATCCTGGAGGCAGAGATACACGTC---GAACCCAAAACAG
Strongylocentrotus/1-612 ATGACATGCCTGTGGAAAAGATCTTGGATGCCGAACTAGCAGTA---GAACCTAACAATG
Tricoplax/1-612          --GAAATGCCAGTTGAAGCTATCCGAGATGCTGAATCTACGCTT---AATATGAATAGTG
Schistosoma/1-636        -----CTGTCGGCTGAATTAAGCATGGATCCTAAATTGGCTGTATCAGAAAGAGGTGAAG
Tripedalia/1-560         ATGA--TTCCCATCGAATCTATCATCGCAGCTGAAACCCTGGTA---GACCCCGGTATAC
Consensus

Saccoglossus/1-612       ATACATACGTAGATAC-ACCGAACGAT-----------------CCAGTCACCAATA-TA
Strongylocentrotus/1-612 GCCCCTACGTAGACAC-CCCGCGTGAC-----------------CCTGTGACAAACA-TT
Tricoplax/1-612          TACCTTACGTCGAAAT-GCAAAGTAAC-----------------CCTGTCTTAAATA-TT
Schistosoma/1-636        CAATTTATGAAGATAT-ACCTGGTGATGATGATACTGGTTTACATCCATTGACCATAATC
Tripedalia/1-560         AGACTTTCGCTAGTGCGAACACG-GAT-----------------CCCATCCGCCACG-TT
Consensus

Saccoglossus/1-612       TGCCAAGCAGCAGATAAACAGTTATTCACT-TTGGTTGAATGGGCAAAGAGAATTCCA-C
Strongylocentrotus/1-612 TGCCAAGCTGCCGACAAACAACTCTTTACC-CTGGTAGAATGGGCCAAGCGAATCCCA-C
Tricoplax/1-612          TGTCAAGCTGCAGATAAACAACTATTTAAT-TTGGTTGAATGGGCCAAGAAAATACCC-C
Schistosoma/1-636        TGTCAGTCTATTGAACAACAATTACCT-CGAATAGTTAATTGGGCTCGTCAGTTGCCAGT
Tripedalia/1-560         TGCCTGGCTGCTGACAAACAGCT-TGCATCGCTTGCAGAGTGGGCTAAACGCCTCCCC-C
Consensus

Saccoglossus/1-612       ACTTTA--CAG---AACTTCCTTTGGATGATCAAGTTATTTTACTACGGGCAGGTTGGAA
Strongylocentrotus/1-612 ACTTCA--CAG---AACTGCCACTGGATGATCAGGTCACGCTCCTCAGGGCAGGGTGGAA
Tricoplax/1-612          ATTTCT--GTG---ATTTATGTGTCGATGACCAAGTAATACTTCTGAGATCTGGTTGGAA
Schistosoma/1-636        GTTTTCATCTGTCTATCTTAGTTTTGATGATCAATTTTGTTTAATAAAAGCGGCTTGGCC
Tripedalia/1-560         ACTTTC--GTG---ATCTGTCAATAGCAGATCAGGTTGTCTTGCTGCAGTGGAGCTGGCC
Consensus

Saccoglossus/1-612       TGAACTGTTAATCGCAGCTTTTTCGCATCGCTCTATCGCTGTTAAAGATGGAATCTTATT
Strongylocentrotus/1-612 CGAGCTGTTGATTGCAGCCTTCTCACATCGCTCCATCCAAGTGAAGGATGGCATCCTCCT
Tricoplax/1-612          TGAATTACTGATAGCTGCGTTTTCGTTTCGATCAATAGCTGTTGAAGATGGCTTGCTCTT
Schistosoma/1-636        TGAATTAGTTTTAATCAGCTCAGCGTATCATTCAACTGTTATTAGAGACGGTTTGCTTTT
Tripedalia/1-560         TGAGCTGCTGATTGGTGGCTTTTGCCACCGTTCCTGTGCTGTCAAGGATGGCATTCTGTT
Consensus

Saccoglossus/1-612       AGCTACTGGTTTACATGTGCACA-GAAACAGCGCACACAGTGC-CGGAGTAGGTACTATC
```

```
Strongylocentrotus/1-612 TGCCACCGGCCTCCACGTCCACC-GTAACAGCGCCCACAGTGC-AGGGGTGGGCACAATC
Tricoplax/1-612          ATCGACTGGGCATTATATTCATC-GCACTAGCGCGCATAATGC-TGGTATTGGGGCTATA
Schistosoma/1-636        ATCGATTGGACGTCATCTTGGTA-GAGA-GGTGGCTAAATCACATGGTCTAGGTCCTCTT
Tripedalia/1-560         GTCGACGGGCTTGCATCT-TACAAGGGACAATCTGAAAAAGGC-CGGCGTGGGAGCTATC
Consensus

Saccoglossus/1-612       TTTGACCGTGTACTCACCGAACTGGTGGCAAAAATGAGAGAAATGAAGATGGATAAGACA
Strongylocentrotus/1-612 TTTGACAGGGTCCTCACCGAGCTGGTGGCTAAGATGAGAGAAATGAAGATGGATAAGACA
Tricoplax/1-612          TTTGATAGGATTTTAACAGAGTTGGTTAATCAAATGAGATATTTAAAAATGGATAAAACT
Schistosoma/1-636        GTTGATAGAATTCTTCATGAACTTGTTGCACGTTTTCGTGATTTATCGTTACAAAGAACT
Tripedalia/1-560         ATTGACAAAATCTTCTCTGAAGTCATAGAAAAGATGCAGGAAATTCAAATGGACCGTGCG
Consensus

Saccoglossus/1-612       GAGTTAGGATGCTTAAGAGCCATTGTTTTGTTCAACCCAGATGCTAAGAATCTTGGAACT
Strongylocentrotus/1-612 GAGCTTGGCTGTCTCAGAGCAATCGTGCTTTTCAACCCAGATGCCAAGAACTT--GACCT
Tricoplax/1-612          GAATTGGGCTGTTTGAGAGCTATTATATTGTTTAATCCAGATGTTCGCGGTTTGACATCA
Schistosoma/1-636        GAATTAGCTTTACTACGTGCTATTATTCTTTTTAATCCTGATGCTAATGGCTTGTCATCA
Tripedalia/1-560         GAATGGGGTTGTTTGCGTGCCATCATGCTATTTTCCCCCGATGCGAAAGGACTGACAGCC
Consensus

Saccoglossus/1-612       --GTTCAGAAGGTTGAAGAGCTAC--GAGAGAAAGTTTATGCTT-CGCTGGAAGAATATT
Strongylocentrotus/1-612 CGGTACAGAAAGTGGAGGAGCTAC--GGGAGAAGGTGTACGCCT-CTCTTGAGGAATACT
Tricoplax/1-612          --GCTGATAGGGTTGAAAAGTATC--GAGAGCTGGTATATGGAG-CTCTAGAGGCTTACG
Schistosoma/1-636        C--GTCATCGCGTGGA--AGCTGTCAGGGAGCAGCTTTATTCAG-CTCTTCATTCGTATT
Tripedalia/1-560         --ATTGACCAAGTAGA-GAACTAC--CGGGAGCTTTATACGTCCACTTTAGAAGATCACG
Consensus

Saccoglossus/1-612       GTAGGAAGACTTACCCAGATGAACCTGGTCGATTTGCCAAACTTCTTCTCCGTCTGCCAG
Strongylocentrotus/1-612 GCCGCAACCAGTACACGGACGAACCCGGCCGCTTCGCCAAGCTGCTCCTCAGACTGCCTG
Tricoplax/1-612          TCAAGAAAAGGTTTCCTGACCAGCTCTGTCGCTTTGCTAAACTACTGCTTCGATTGCCGG
Schistosoma/1-636        GTACGACTAATCAACCTCAAGATACTTCCCGTTTCACAAAATTACTACTAAGATTACCTC
Tripedalia/1-560         TCAAACGAAAGCACCCTGAGCAACCCGATCGGTTTACCAAGGTAATCCTTCGTATACC--
Consensus

Saccoglossus/1-612       CTCTGCGATCAATTGGCCTCAAGTGTCTGGAACATTTATTTTTCTTCAAA
Strongylocentrotus/1-612 CCCTGCGCTCCATCGGTCTCAAATGCCTGGAGCATCTCTTCTTCTTTA--
Tricoplax/1-612          CCTTAAGAGCGATAAGTTTAAAGACCTTGGAGCATCTTTTCTTTTATAAA
Schistosoma/1-636        CTTTACGATCCATCGCATCCAAGTGCCTTGAACATTTAGTATTTGTCAAA
Tripedalia/1-560         ----------------------------------------------------
Consensus
```

3. b – Muscle alignment of RXR genes sequences of LBD (design primers).

```
Tripedalia_DBD/1-69          QPCSVCSDKAYVKHYGVFACEGCKGFFKRSVRRK--YSCLGKRHCDTDKKSRNRCQYCRF
Hydra/1-20                   -------------HYGVFTCDGCRGFFKRAVRR--------------------------
Anemonia/1-42                --------------------CEGCKGFFKRSVQKKKTYTCRDTKDCPMDKRHRNRCQYCSY
Tricoplax_DBD/1-67           ---SICGQRSLRRHYGVYSCEGCKGFFKRTVRKDLTYTCRDNRCDIDKKQRNRCQYCRYQ
Saccoglossus_DBD/1-69        -VCAVCGDRASGKHYGVYSCEGCKGFFKRTVRKDLHYACRDEKCIVDKRQRNRCQYCRYQ
HumanRXRb_DBD/1-65           --CAICGDRSSGKHYGVYSCEGCKGFFKRTIRKDLTYSCRDKDCTVDKRQRNRCQYCRYQ
HumanRXRA_DBD/1-65           --CAICGDRSSGKHYGVYSCEGCKGFFKRTVRKDLTYTCRDKDCLIDKRQRNRCQYCRYQ
HumanRXRg_DBD/1-68           HICAICGDRSSGKHYGVYSCEGCKGFFKRTIRKDLIYTCRDKDCLIDKRQRNRCQYCRYQ
Strongylocentrotus_DBD/1-58  ------------KHYGVYSCEGCKGFFKRTVRKDLTYTCRDDRCMVDKRQRNRCQYCRYQ
Thais retinoid/1-70          HICAICGDRASGKHYRVYSCEGCKGFFKRTVRKDPTYACRDDKCMIDKRQRNRCQFCRYM
Nucella retinoid/1-70        HICAICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTYACRDDKCMIDKRQRNRCQYCRYM
AY048663.1|RXR-like/1-70     HICAICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTYACRDDKCMIDKRQRNRCQYCRYM

Tripedalia_DBD/1-69          QKCVQVGMKPE
Hydra/1-20                   -----------
U93415.1|Anemonia/1-42       Q----------
Tricoplax_DBD/1-67           KCLQVGMKQE-
Saccoglossus_DBD/1-69        KCIAMGMRRE-
HumanRXRb_DBD/1-65           KCLATGM----
HumanRXRA_DBD/1-65           KCLAMGM----
HumanRXRg_DBD/1-68           KCLVMGMK---
Strongylocentrotus_DBD/1-58  KCLGMGMRRE-
Thais retinoid/1-70          KCLAQGMKRE-
Nucella  retinoid/1-70       KCLAQGMKRE-
AY048663.1|RXR-like/1-70     KCLSMGMKRE-
```

3. C – Muscle alignment of RXR DBD amino acid sequences – FFK sequence (yellow) Cyst residues (green) of zinc finger are highlighted.

```
Bv_TR2/4_DBD/1-135 ----------TTCATTTCTGGAATCGACAGTATTGACATCTGTTTCTAGTCATTTGGTTC
Ap_TR2/4_DBD/1-138 ------------CATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTCATTTGGTTC
As_TR2/4_DBD/1-133 -----------TCATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTCATTTGGTTC
Uf_TR2/4_DBD/1-147 GCACTAGTGATTCATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTCATTTGGTTC

Bv_TR2/4_DBD/1-135 ACTTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTTA
Ap_TR2/4_DBD/1-138 ACTTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTTA
As_TR2/4_DBD/1-133 ACTTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTTA
Uf_TR2/4_DBD/1-147 ACTTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTTA

Bv_TR2/4_DBD/1-135 AAGAAACCCTTTACACCCCTCACAA-----
Ap_TR2/4_DBD/1-138 AAGAAACCCTTTACACCCCTCACAAATCCC
As_TR2/4_DBD/1-133 AAGAAACCCTTTACACCCCTCACA------
Uf_TR2/4_DBD/1-147 AAGAAACCCTTTACACCCCTCACAAAT---
```

3. D – Muscle alignment of TR2/4 RXR sequences amplified by PCR. *Bv, Bunodactis verrucosa; Ap, Actinothoe sphyrodeta; As, Anemonia sulcata; Uf, Urticina feline.*

```
As_a/1-229       CACTGTTCCATTCTTGACTTCTATA-CATTGGTTTCACGCCATCTTGATCCATTGGTTGG
TR2/4_DBD/1-135  ------TTCATTTCTGGAATCGACAGTATTGACATC-TGT--TTCTAGTCATTTGGTT--

As_a/1-229       ATGTTCTTGTAGGCTCCCAAAAAAGATTTTACTACTTCAAACCTAAATTTCTTGTCGTAT
TR2/4_DBD/1-135  ---CACTTCACA------------GTTG-CCATCAGCTTTGCAGGTGTAGTCTAACTGTT

As_a/1-229       ATCTGTCGTATACTGCATGCGCTTGACCATGTGGTTCACATGTTCAAACGTCCTCCTCCA
TR2/4_DBD/1-135  CTGTA-------------------CGGTACGTTTAAAGAAACCCTTTACACCCCTCACAA

As_a/1-229       TGGTAT-TTCGGGGCTACAACTTGTGCCAATACTGTCGATTCCAGAAATGA
TR2/4_DBD/1-135  ---------------
```

3. E – Muscle alignment of TR2/4 fragments.

```
As_11/1-288     TACCGATGGGGACGTACGAGGCGTAGTACCTTTATCTGTATCTAAGATACCGGCAAGCTT
c3bf/1-132      ---TGTGAGGGGTGTA-AAGG----GTTTCTTTA---------AACGTACCG-------T

As_11/1-288     GCAGCAAGGCTTCTCTCCAGCGTGGTGGTAGGGGTCTACCACGAATCCTTCTAGCCAGAT
c3bf/1-132      ------------------------------------------------------------

As_11/1-288     GCTGCTACCTCTGTATTAGGGGACTTGACGATAATTTCATTGAGCATGTCGGCTGCCTGT
c3bf/1-132      ---ACAGAAA--CAGTTAGA---CTACACCTGCA-----AAGCTGATGGCAA---

As_11/1-288     CTAGAATAAAGTAGTAAACGTGTGTCAGCTTCTTCGAGGTCGGACTCCAATTCTGGGACT
c3bf/1-132      CTGGAAGTGAAC—CAAATGACTAGAAAC------AGATGCCAGTACTG-----

As_11/1-288     TCTCGCTCTAAGACTCCGTTGAACGGTGTCAGTTGAGAGCCTGTTTTA
c3bf/1-132      --TCGATTCCAGA--------AATG--
```

3. F – Muscle alignment of TR2/4 fragments, amplified by primers f2 and r1.

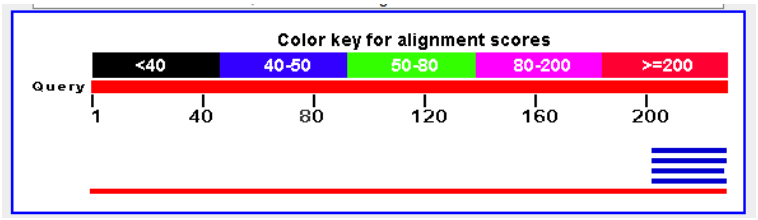
```
c1ar/1-75       ----GGTTTCGAGCATCAGTTTGTCTCCCTTAGTACTTTTGATTTCAGCAGCTTTGTAAC
c1dr/1-62       ----------AAAAGTAAAAT-------------------AGTATAATATCCATCCTGCT
c1br/1-59       GATTTGTGAGGGGTGTAAA---------------------GGGTTCGTAACCATAATAAT
c1cr/1-56       ----TGTGAGGGGTGTAAAG--------------------GGTTGTAAAAGCACAATA-T

c1ar/1-75       CCTTTACACCCCTCACAAA------------
c1dr/1-62       TATTCATTAGCATTAATTAAGGTCTGTCCTC
c1br/1-59       CATTAAA---TTTTACAGAGTCC--------
c1cr/1-56       AATAAAT---CATTACTGAAAACC-------
```

3. G – Muscle alignment of sequences amplified by primers f2 and r2.

```
As_a/1-229      CACTGTT----CCATTCTTGACT-----------TCTATACATTGGTTTCACGCCATCTT
As_11/1-288     TACCGATGGGGACGTACGAGGCGTAGTACCTTTATCTGTATCTAAGATACCGGCAAGCTT

As_a/1-229      GATCCATTGGTT-----GGATGTTCTTGTAGG---CTCCCAAAAA-----------AGAT
```

```
As_11/1-288     GCAGCAAGGCTTCTCTCCAGCGTGGTGGTAGGGGTCTACCACGAATCCTTCTAGCCAGAT

As_a/1-229      TTTACTACTTCAAACCTA--------------AATTTCTTGTCGTATATCTGTCGTATAC
As_11/1-288     GCTGCTACCTCTGTATTAGGGGACTTGACGATAATTTCATTGAGCATGTCGG-------C

As_a/1-229      TGCATGCGCTTGACCATGTGGTTCACATGTTCAAACGTCCTC---------CTCCATGGT
As_11/1-288     TGCCTGTCTAGAATAAAGTAGTAAACGTGTGTCAGCTTCTTCGAGGTCGGACTCCA----

As_a/1-229      ATTTCGGGGCTACAACTTGTGCCAATACT------GTCGATTCCAGAAATGA--------
As_11/1-288     ATTCTGGGACT---TCTCGCTCTAAGACTCCGTTGAACGGTGTCAGTTGAGAGCCTGTTT

As_a/1-229      --
As_11/1-288     TA
```

3. H – Muscle alignment of As_a and As_11 sequences; As, *Anemonia sulcata*.



3. I – Multiple alignment (BLAST) of the sequences: As_TR2/4_DBD, Bv_TR2/4_DBD, Ap_TR2/4_DBD, Uf_TR2/4_DBD and As_a; As, *Anemonia sulcata*.

Annex 4: Accessions number of ESTs and sequences of NRs

Accession number	Database	Specie	Sequence	Gene
AF091121.1	GenBank	*Tripedalia cystophora*	nucleotide	RXR
U93416.1	GenBank	*Anemonia sulcata*	Nucleotide	RXR
NW_002152543.1	GenBank	*Hydra magnipapillata*	Genome	RXR
6751230	GenBank	*Trichoplax adhaerens*	Gene	RXR
EU024473.1	GenBank	*Nucella lapillus*	Nucleotide	RXR
AB594846.1	GenBank	*Thais clavigera*	Nucleotide	RXR
AY048663.1	GenBank	*Biomphalaria glabrata*	Nucleotide	RXR
100329058	GenBank	*Saccoglossus kowalevskii*	Gene	RXR
X52773	GenBank	*Homo sapiens*	Nucleotide	RXR
X63522	GenBank	*Homo sapiens*	Nucleotide	RXR
U38480	GenBank	*Homo sapiens*	Nucleotide	RXR
AF121129.1	GenBank	*Homo sapiens*	Nucleotide	PNR
X16155.1	GenBank	*Homo sapiens*	Nucleotide	COUP-TF
X76930.1	GenBank	*Homo sapiens*	Nucleotide	HNF4
DQ017616.1	GenBank	*Danio rerio*	Nucleotide	HNF4
BC162963.1	GenBank	*Danio rerio*	Nucleotide	COUP-TF
NM_131275.1	GenBank	*Danio rerio*	Nucleotide	RXR
NM_001085811.1	GenBank	Xenopus laevis	Nucleotide	TLX
NM_001094481	GenBank	Xenopus laevis	Nucleotide	COUP-TF
NM_001087467.1	GenBank	Xenopus laevis	Nucleotide	RXR
NM_057433.3	GenBank	Drosophila melanogaster	Nucleotide	USP
NM_079857.3	GenBank	Drosophila melanogaster	Nucleotide	TLL
NM_001103656.1	GenBank	Drosophila melanogaster	Nucleotide	HNF4
XP_001638550	GenBank	*Nematostella vectensis*	Protein	HNF4
XP_001630386	GenBank	*Nematostella vectensis*	Protein	PNR/TLL
XP_001635112	GenBank	*Nematostella vectensis*	Protein	PNR/TLL
XP_001630385	GenBank	*Nematostella vectensis*	Protein	PNR/TLL
XP_001634999	GenBank	*Nematostella vectensis*	Protein	PNR/TLL
XP_001624815	GenBank	*Nematostella vectensis*	Protein	COUP-TF
XP_001629708	GenBank	*Nematostella vectensis*	Protein	COUP-TF
XP_001634340	GenBank	*Nematostella vectensis*	Protein	COUP-TF
XP_001634378	GenBank	*Nematostella vectensis*	Protein	COUP-TF
XP_001636010	GenBank	*Nematostella vectensis*	Protein	COUP-TF
XP_001636637	GenBank	*Nematostella vectensis*	Protein	TR2/4
XP_001631902	GenBank	*Nematostella vectensis*	Protein	TR2/4
XM_001638500.1	GenBank	*Nematostella vectensis*	Nucleotide	HNF4
XM_001630336.1	GenBank	*Nematostella vectensis*	Nucleotide	PNR/TLL
XM_001635062.1	GenBank	*Nematostella vectensis*	Nucleotide	PNR/TLL
XM_001630335.1	GenBank	*Nematostella vectensis*	Nucleotide	PNR/TLL
XM_001634949.1	GenBank	*Nematostella vectensis*	Nucleotide	PNR/TLL
XM_001624765.1	GenBank	*Nematostella vectensis*	Nucleotide	PNR/TLL
XM_001629658.1	GenBank	*Nematostella vectensis*	Nucleotide	COUP-TF
XM_001634290.1	GenBank	*Nematostella vectensis*	Nucleotide	COUP-TF
XM_001634328.1	GenBank	*Nematostella vectensis*	Nucleotide	COUP-TF
XM_001635960.1	GenBank	*Nematostella vectensis*	Nucleotide	COUP-TF
XM_001636587.1	GenBank	*Nematostella vectensis*	Nucleotide	COUP-TF
XM_001631852.1	GenBank	*Nematostella vectensis*	Nucleotide	TR2/4
XM_001631008.1	GenBank	*Nematostella vectensis*	Nucleotide	TR2/4
ci0100142690	JGI	Ciona intestinalis	Protein	COUP-TF
ci0100146308	JGI	Ciona intestinalis	Protein	HNF4

ci0100150561	JGI	Ciona intestinalis	Protein	RXR
ci0100138574	JGI	Ciona intestinalis	Protein	TR2/4
Accession number	**Database**	**Sequence**		**Gene**
BI781831	GenBank	EST		RXR
BI782081	GenBank	EST		RXR
BM280724	GenBank	EST		RXR
BM568658	GenBank	EST		RXR
BQ382934	GenBank	EST		RXR
CA304237	GenBank	EST		RXR
CB014960	GenBank	EST		RXR
BI510638	GenBank	EST		RXR
BI511231	GenBank	EST		RXR
BI513792	GenBank	EST		RXR
BG930255	GenBank	EST		RXR
BG930250	GenBank	EST		RXR
CD124719	GenBank	EST		RXR
CD124789	GenBank	EST		RXR
CD132154	GenBank	EST		RXR
CD065670	GenBank	EST		RXR
CD181869	GenBank	EST		RXR
CD124616	GenBank	EST		RXR
CD181801	GenBank	EST		RXR
CD163651	GenBank	EST		RXR
CD116659	GenBank	EST		RXR
CD125564	GenBank	EST		RXR
CD196231	GenBank	EST		RXR
CD191347	GenBank	EST		RXR
CD114207	GenBank	EST		RXR
CD114191	GenBank	EST		RXR
CD090814	GenBank	EST		RXR
BI781435	GenBank	EST		RXR
CD308971	GenBank	EST		TLL
CD308972	GenBank	EST		TLL
CD334321	GenBank	EST		TLL
CD340492	GenBank	EST		TLL
BI515764	GenBank	EST		COUP-TF
BI516773	GenBank	EST		COUP-TF
BG786614	GenBank	EST		TR2/4
BG787668	GenBank	EST		TR2/4
CD290031	GenBank	EST		TR2/4
CD153024	GenBank	EST		TR2/4
CD168281	GenBank	EST		TR2/4
CD124909	GenBank	EST		TR2/4
CD069500	GenBank	EST		TR2/4
CD159107	GenBank	EST		TR2/4
CD069750	GenBank	EST		TR2/4
BI783431	GenBank	EST		HNF4
BI783711	GenBank	EST		HNF4
BI514984	GenBank	EST		HNF4
BI503347	GenBank	EST		HNF4
BI508133	GenBank	EST		HNF4
BI509848	GenBank	EST		HNF4

4. A – List of the accession numbers of ESTs, nucleotide and amino acid sequences used in the molecular analyses.

Annex 5: Sequenced nucleotide and deduced amino acid sequences

>As_FTZ-F1_DBD
GTTGCCATCAGCTTTGCAGGTGGGTTCTTTAAGCGGGCGGTACAGAAACA
GTTAGACTACACCTGCAAAGCTGATGGCACACTGTGAAGTGAACCAAATGACTAG
ATAACAGATGTCACCCCTGCAGCTTCA

>As_TR2/4_DBD
TCATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTCATTTGGTTCACT
TCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTT
AAAGAAACCCTTTACACCCCTCACA

>Bv_TR2/4_DBD
TTCATTTCTGGAATCGACAGTATTGACATCTGTTTCTAGTCATTTGGTTCAC
TTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTT
TAAAGAAACCCTTTACACCCCTCACAA

>Ap_TR2/4_DBD
CATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTCATTTGGTTCACTT
CACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTGTACGGTACGTTTA
AAGAAACCCTTTACACCCCTCACAAATCCC

>Uf_TR2/4_DBD
GCACTAGTGATTCATTTCTGGAATCGACAGTACTGGCATCTGTTTCTAGTC
ATTTGGTTCACTTCACAGTTGCCATCAGCTTTGCAGGTGTAGTCTAACTGTTTCTG
TACGGTACGTTTAAAGAAACCCTTTACACCCCTCACAAAT

>As_a
CACTGTTCCATTCTTGACTTCTATACATTGGTTTCACGCCATCTTGATCCAT
TGGTTGGATGTTCTTGTAGGCTCCCAAAAAAGATTTTACTACTTCAAACCTAAATTT
CTTGTCGTATATCTGTCGTATACTGCATGCGCTTGACCATGTGGTTCACATGTTCA
AACGTCCTCCTCCATGGTATTTCGGGGCTACAACTTGTGCCAATACTGTCGATTCC
AGAAATGA

>As_a
LTVPFLTSIHWFHAILIHWLDVLVGSQKRFFKPKFLVVYLSYTACAPCGSHVQT
SSSMVFRGYNLCQYCRFQK

>As_10
TATTGCTCTACTGTTACGAGCGAAGCCGTACCATACCACCCCATTTCAACA
ACAATCGGCTCAACATGCCCTGTGGTATCATGGGCATCGAGAATATCTGTGAGCT
ATCTGCTAGGTTGCTGATTTAGTGCTGTTGAAGGGGCACGTAGTATCCCTTTCTTT
CCTGACTTGCCCGTGACTGACCTAATGGCTTTGCTACCTCTTGTATGGAGTGAGC
TGTTTGTGTTAAATGCATCACAGTGTCCCATGCGATTACAAGTTGCGCCGCTTCTT
GCTACACCGGGAATTCACTCTAATCATATGTCTCCTGACACAATGGTCACCTTCAT
GGATAATATGAAAATATTTCAAGAACAAGTGGATAAACTAAAAAATCTTCATGTACA
TGCTGCCGATTTCGCCTGCTTGAAAGCAATTGTTTTGTTTACTTCAGACGCATCAG
GGCTAACCGACCCACACTACATTGAAAGCCTACAAGAAAAGACACAGTGTGCCTT
GGATGAATAGAACAAGAATCAGTATCCTAACCAACCCTCTCGCTTTGCAAAACTAC
TGCTCCTGCTCCCATCACTGAGGAGCATAAGCACCAATGTGGTAGATCAGTTGTA
CTATGTACGTCTAGTTGGGAAACTCCAAAGATACTCTTCTCAACATATGCTACTCT

CTGGTACCCAAGATCTTGGCCATATCTGACCCTGTTCATGATGACTGTGAACTGCT
GATCATTAATACGGAGACTAGGGAGGATAGGAACTGTAAGGGGCTCGTGCCGCG
TGTGTAATGGGTGAATCGAACAATATTTTTAGATTGAGGACCTTCGAAGGTTCCTC
GATACACCATAGTTTGTCAGTTGCTAAGGCGTAAAGTTTCATTTACAGATGTAACT
AGATCTAAAATTTATACTGGAAACAGATTCTTGATTT

>As_10

YCSTVTSEAVPYHPISTTIGSTCPVVSWASRISVSYLLGCFSAVEGARSIPFFP
DLPVTDLMALLPLVWSELFVLNASQCPMRL

QVAPLLATPGIHSNHMSPDTMVTFMDNMKIFQEQVDKLKNLHVHAADFACLK
AIVLFTSDASGLTDPHYIESLQEKTQCALDENKNQYPNQPSRFAKLLLLLPSLRSISTNV
VDQLYYVRLVGKLQRYSSQHMLLSGTQDLGHI

>As_11

TACCGATGGGGACGTACGAGGCGTAGTACCTTTATCTGTATCTAAGATACC
GGCAAGCTTGCAGCAAGGCTTCTCTCCAGCGTGGTGGTAGGGGTCTACCACGAA
TCCTTCTAGCCAGATGCTGCTACCTCTGTATTAGGGGACTTGACGATAATTTCATT
GAGCATGTCGGCTGCCTGTCTAGAATAAAGTAGTAAACGTGTGTCAGCTTCTTCG
AGGTCGGACTCCAATTCTGGGACTTCTCGCTCTAAGACTCCGTTGAACGGTGTCA
GTTGAGAGCCTGTTTTA

>As_11

XTIARYRWGRTRRSTFICIDTGKLAARLLSSVVVGVYHESFPDAATSVLGDLTII
SLSMSAACLESSKRVSASSRSDSNSGTSRSKTPLNGVSEPV

>As_14

CACTACCCAGTTTTTGCGCGAACCAGACAAGCTCAAGAATTGGCTGCGAG
AACCCAAAGTGATATTTGGCCCAATTTATGTCTTCATGGTGTGTGCAGGGCTCCGA
TTTTGCACTGCAATCGACGATTGTCCATCCTTGCTAACAGCACTGTCGAAAAAACA
CCAGGTCAACATTTGGACAAGGTGCTACACACCATGGGAAAGTTGAATGAAGTTC
TGTTCCATTTTGGCTGCAGCGCTCTTTCGTCCCGACAACAACGAGACTCTGAAGC
TTGTTCTTCCAGGATCCTGGAACA

>As_14

TTQFLREPDKLKNWLREPKVIFGPIYVFMVCAGLRFCTAIDDCPSLLTALSKKH
QVNIWTRCYTPWESMKFCSILAAALFRPDNNETLKLVLPGSW

5. A – List sequenced/translated sequences in the study. As*, Anemonia sulcata*; *Bv, Bunodactis verrucosa; Ap, Actinothoe sphyrodeta; Uf, Urticina felina.*

Annex 6: pGEM®-T Vector Sequence

```
   1  GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCCGCGGGA
  51  T*ATCACTAGT GCGGCCGCCT GCAGGTCGAC CATATGGGAG AGCTCCCAAC
 101  GCGTTGGATG CATAGCTTGA GTATTCTATA GTGTCACCTA AATAGCTTGG
 151  CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA
 201  ATTCCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC
 251  CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA CTGCCCGCTT
 301  TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC
 351  GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC
 401  TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT
 451  CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG
 501  AACATGTGAG CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC
 551  GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA
 601  ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 651  CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT
 701  GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC
 751  TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC
 801  TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC
 851  CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT
 901  CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 951  GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG
1001  AAGAACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA
1051  AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT
1101  GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA
1151  AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA
1201  ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
1251  TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA
1301  TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA
1351  TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC
1401  GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC
```

```
1451 AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA
1501 ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC
1551 GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC
1601 GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG
1651 TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA
1701 TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC
1751 CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC
1801 TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA
1851 AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA
1901 GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA
1951 CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT
2001 TCGGGGCGAA AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT
2051 GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA
2101 GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA
2151 ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA
2201 TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG
2251 AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA
2301 AAAGTGCCAC CTGATGCGGT GTGAAATACC GCACAGATGC GTAAGGAGAA
2351 AATACCGCAT CAGGAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT
2401 TAAATTTTTG TTAAATCAGC TCATTTTTTA ACCAATAGGC CGAAATCGGC
2451 AAAATCCCTT ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTTGT
2501 TCCAGTTTGG AACAAGAGTC CACTATTAAA GAACGTGGAC TCCAACGTCA
2551 AAGGGCGAAA AACCGTCTAT CAGGGCGATG GCCCACTACG TGAACCATCA
2601 CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC TAAATCGGAA
2651 CCCTAAAGGG AGCCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAACG
2701 TGGCGAGAAA GGAAGGGAAG AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG
2751 GCAAGTGTAG CGGTCACGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA
2801 TGCGCCGCTA CAGGGCGCGT CCATTCGCCA TTCAGGCTGC GCAACTGTTG
2851 GGAAGGGCGA TCGGTGCGGG CCTCTTCGCT ATTACGCCAG CTGGCGAAAG
2901 GGGGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG GTTTTCCCAG
2951 TCACGACGTT GTAAAACGAC GGCCAGTGAA TTGTAATACG ACTCACTATA
```

pGEM-T flankers:

Fwd GCCGCGGGAT**T**, **Rev A**ATCACTAGTG, **Rev Fwd** CGGCGCCCTA**A** → AATCCCGCGGC, **Rev T**TAGTGATCAC → CACTAGTGAT**T**

6. A – The sequence supplied is that of the circular pGEM(R)-5zf(+) Vector from which the pGEM(R)-T Vector is derived. The pGEM(R)-T Vector has been linearized with EcoR V at base 51 (*) and a T-residue has been added to both 3'-ends. The added T is not included in this sequence. pGEM-T flankers are supplies below.