INTEGRATED STRATEGIES FOR INVESTIGATING ENDOCRINE MECHANISMS IN *BIOMPHALARIA GLABRATA* AS A TEST ORGANISM FOR ANDROGENIC CHEMICAL TESTING

*A thesis submitted for the degree of Doctor of Philosophy*

*By*

*Satwant Kaur*

Institute for the Environment, Health and Societies
Brunel University

March 2015
DECLARATION

The work submitted in this thesis was conducted between 2010 and 2014 at Brunel University (Uxbridge, West London, UK). This work was carried out independently and has not been submitted for any other degree.
Abstract

Endocrine and metabolic disease or dysfunctions are of growing concern in modern societies across the globe, underlining the need for continued focus on the development of pharmaceuticals. Subsequent scientific research has revealed a trend in the increase of such abnormalities and expansion of chemical industries, highlighting concerns that these disorders may, in part, be caused by exposure to environmental pollutants. This has led to changes in legislation concerning chemicals safety testing involving an increasing number of vertebrate animal tests as a part of environmental risk assessment process, at significant financial and ethical costs. A solution that is appropriate and aligned with the three R’s (reduction, refinement and replacement) in relation to animal research is to exploit the use of small invertebrate organisms as possible replacements for mammals. In line with the above approach/solution, this thesis is based on the null hypothesis that common genes, proteins and processes in gastropod molluscs and humans underlie the response of male reproductive organs to androgenic chemicals. Using a freshwater pulmonate snail, *Biomphalaria glabrata*, physiological effects of two steroid androgens on the development of mollusc secondary sexual organs were studied. Furthermore, an exhaustive investigation on the mollusc nuclear receptor repertoire and reproductive type neuropeptides was conducted. This also included the study of the evolutionary degree of conservation of these genes in non-model molluscs. The results obtained suggest that the snails did not respond to, and were not affected by exposure to the androgens. These results were supported by the absence of the members of subfamily 3C of nuclear receptors, which includes some of the “vertebrate” steroid hormone targets, suggesting that this mollusc may be an inappropriate model for steroid hormone mediated mammalian endocrine function. The nuclear receptor (NR) repertoire of *B. glabrata* comprised of 39 nuclear receptors representing all the known subfamilies of the NR superfamily. 21 reproductive type neuropeptide genes were identified encoding precursors that are predicted to release over 124 bioactive cleavage products. The consequence of these findings is significant in the context of the development of alternative model organisms for chemical testing as well as elucidating the taxonomic scope of nuclear receptor mediated endocrine disruption.
Acknowledgement

I take this opportunity to express my sincere gratitude towards my supervisors, Professor Susan Jobling and Dr Edwin Routledge to give me the opportunity to do research in their group, for their advice, support, motivation and encouragement throughout my work, and especially for the amount of time and dedication to discuss my work and beyond.

I would also like to thanks all the students and members of staff of the Institute for the Environment (now Institute of Environment, Health and Societies) for creating such a friendly and supporting environment, helpful discussions, and pieces of advice and for just being great colleagues. Alice Baynes, Tamsin Runnalls, Catherine Harris, Margaret Town, Anne Lockyer, Erin Losty, Luigi Margiotta-Casaluci, Elizabeth Nicol, Elizabeth Lawton, Adam Lynch, Graham Harris, Nicola Beresford, Safeyeh Haghani, Lourdes Lopez-Merino and Alpa Patel. In particular I would like to thank Alice Baynes and Anne Lockyer for their incessant guidance and support, day to day assistance during my practical research help with proof reading of my chapters and keeping me sane with their words of encouragement. In addition to them I would also like to thank Elizabeth Lawton, Margaret Town and Richard Evans for proof reading of my thesis chapters. I would also like to thank Richard for helping me with my sudden computational doubts. Thanks are also due to the technical staff of the animal unit (Julie, Sue and Amy and Steve) and IFE’s admin crew (Anne Smith and Sue Maddix) for making things easier. I owe my gratitude to NC3Rs for their financial support.

I would also like to take this opportunity to thanks my friends and family for all the good times, laughter, seamless support and little words of encouragement. In particular my parents for providing me financial and psychological support without which this dream would have been too difficult to pursue.
# Table of contents

DECLARATION............................................................................................................................................. II

ABSTRACT.................................................................................................................................................. III

ACKNOWLEDGEMENT........................................................................................................................... IV

TABLE OF CONTENTS........................................................................................................................... V

LIST OF FIGURES.................................................................................................................................. X

LIST OF TABLES....................................................................................................................................... XIII

1 CHAPTER 1: GENERAL INTRODUCTION............................................................................................... 1

1.1 ENDOCRINE DISRUPTION AND ANIMAL TESTING ........................................................................... 2

1.2 THE VERTEBRATE ENDOCRINE SYSTEM .......................................................................................... 4

1.2.1 Endocrine glands............................................................................................................................. 6

1.2.2 Hormones........................................................................................................................................ 7

1.2.3 Hormone synthesis........................................................................................................................... 8

1.2.4 Hormone action............................................................................................................................... 9

1.3 ENDOCRINE DISRUPTOR ACTION ....................................................................................................... 11

1.4 ENDOCRINE DISRUPTOR ANIMAL TESTING PROGRAMMES ....................................................... 13

1.5 THE 3RS (REPLACEMENT, REFINEMENT AND REDUCTION OF ANIMALS IN TESTING CHEMICALS) ......................................................................................................................... 18

1.6 INVERTEBRATES AS REPLACEMENTS FOR VERTEBRATES IN REPRODUCTIVE TOXICITY TESTING ............................................................................................................................................. 19

1.7 MOLLUSCS AS PROPOSED TEST SYSTEMS FOR REPRODUCTIVE ENDOCRINE DISRUPTER TESTING .............................................................................................................................................. 23

1.8 THE GASTROPOD NEURO-ENDOCRINE SYSTEM ............................................................................. 26

1.9 CONCLUSION .................................................................................................................................... 29

1.10 AIM OF THE THESIS .......................................................................................................................... 30

2 CHAPTER 2: INTRODUCTION TO THE TEST SPECIES, BIOMPHALARIA GLABRATA AND PRELIMINARY WORK ................................................................................................................................. 31

2.1 INTRODUCTION.................................................................................................................................. 32

2.2 GASTROPODA .................................................................................................................................... 32

2.2.1 Prosobranch .................................................................................................................................. 32

2.2.1.1 Test species 1- Lottia gigantea (for computational purposes) .................................................... 34

2.2.2 Pulmonates .................................................................................................................................... 34

2.2.2.1 Test species 2- Biomphalaria glabrata ....................................................................................... 36

2.2.2.1.1 General Anatomy ...................................................................................................................... 37

2.2.2.1.2 Life cycle and reproduction ...................................................................................................... 38

2.2.2.1.3 Molecular tools to study the biology of B. glabrata ................................................................ 40
3  CHAPTER 3: EXPOSURE OF THE FRESHWATER SNAIL BIOMPHALARIA GLABRATA TO STEROID ANDROGENS DURING DEVELOPMENT DOES NOT AFFECT REPRODUCTIVE PHYSIOLOGY ........................................................................................................... 61

3.1  INTRODUCTION ................................................................................................................. 62
3.2  AIM OF THE STUDY ........................................................................................................... 64
3.3  EXPERIMENTAL APPROACH .......................................................................................... 66
  3.3.1  Test compounds ............................................................................................................. 66
  3.3.2  Stock solution preparation ............................................................................................ 66
  3.3.3  Experimental Design ..................................................................................................... 67
  3.3.4  Egg collection and initial static-renewal exposure of embryos ....................................... 68
  3.3.5  Tank water analyses ...................................................................................................... 69
    3.3.5.1  Extraction and concentration of androgenic chemicals from water samples ............... 69
    3.3.5.2  Yeast Androgen Screen (YAS) Assay ...................................................................... 70
    3.3.5.3  Calculating androgenic equivalents ......................................................................... 72
  3.3.6  Biological Sampling and analyses ................................................................................ 73
    3.3.6.1  Terminal sampling .................................................................................................... 73
    3.3.6.2  B. glabrata dissection and preservation .................................................................... 73
    3.3.6.3  Morphometric analysis of glandular complex .......................................................... 74
    3.3.6.4  Statistical Analysis ................................................................................................... 75
3.4  RESULTS ........................................................................................................................... 76
  3.4.1  Measured water concentration of MT and DHT ............................................................... 76
  3.4.2  Hatchling survival and growth (shell diameter and weight) .......................................... 77
  3.4.3  Quantitative assessment of the size of reproductive glands .......................................... 81
  3.4.4  Histopathological analyses of reproductive tissues ..................................................... 83
3.5  DISCUSSION ...................................................................................................................... 91
  3.5.1  Possible effects on survival and growth ........................................................................ 91
  3.5.2  Possible effects of androgens on reproductive tissue size and histopathology ............. 93
3.6  CONCLUSION ..................................................................................................................... 96

4  CHAPTER 4: INVESTIGATION OF NUCLEAR RECEPTORS IN BIOMPHALARIA GLABRATA AND LOTTIA GIGANTEA ......................................................................................................................... 97

4.1  BACKGROUND ............................................................................................................... 98
5 CHAPTER 5: REPRODUCTIVE TYPE NEUROPEPTIDES IN BIOMPHALARIA GLABRATA - IDENTIFICATION FROM THE B. GLABRATA GENOME .........................142
List of figures

Figure 1.1: Schematic diagram of the vertebrate endocrine system showing feedback mechanism of hypothalamus-pituitary-gonadal axis. ................................................................. 6
Figure 1.2: The steroid hormone synthesis pathway in vertebrates showing the enzymes and cytochrome P450 (CYP) genes participating in this process.................................................. 9
Figure 1.3: (A) Pie chart representing animal usage in different areas of research. (B) Pie chart showing percentage of animals used in toxicological research........................................ 16
Figure 1.4: Comparison of anatomy of male mouse urogenital tract (A) with reproductive tract in gastropod snail (B). ................................................................................................. 24
Figure 1.5: Neuroendocrine system of gastropod snails (Koen et al., 2010) showing involved ganglia of the central nervous system................................................................................. 28
Figure 2.1: Reproductive tract of Dioctocardian (A,B) female and male; Monotocardian (C,D) female and male.................................................................................................................. 33
Figure 2.2: Figure showing anatomical parts of the Biomphalaria glabrata ........................................ 37
Figure 2.3: Figure showing the schematic of reproductive tract in B. glabrata ................................ 39
Figure 2.4: Pictomicrograph of B. glabrata hatchlings showing the gonad at different stages of development. ..................................................................................................................... 48
Figure 2.5: Technique of gonadectomy .......................................................................................... 51
Figure 2.6: Figure showing mortalities in gonadectomised and sham operated snails over a period of six weeks after a recovery period of one week post-gonadectomy. ........................................ 52
Figure 2.7: Change in weight measured in grams (mean ±SD) of gonadectomised and sham-operated (controls) snails ...................................................................................................... 53
Figure 2.8: Change in diameter measured in grams (mean ±SD) of gonadectomised and sham-operated (controls) snails ...................................................................................................... 54
Figure 2.9: Boxplot showing the normalized albumen gland weight (NAW) of the gonadectomised snails compared to the sham operated snails after six weeks of observation .............................. 55
Figure 2.10: Diagram of the layout of the experiment in the exposure room ........................................ 57
Figure 2.11: (A) Mean weight ±SD and (B) mean diameter ±SD of the snails pre- and post DHT exposure preliminary study ................................................................. 59
Figure 2.12: Survival of snails at the termination of DHT exposure preliminary study. ................. 59
Figure 3.1: Gantt chart showing the experimental design for the 30 days exposure study using B. glabrata .......................................................................................................................... 68
Figure 3.2: Standard curves showing the starting molar concentrations of DHT and MT with respect to ethanol .................................................................................................................. 71
Figure 3.3: Figure showing the schematic of reproductive tract in B. glabrata ................................. 74
Figure 3.4: Photomicrographs of part of glandular complex tissue from DWC of 17α-MT exposed snails. .............................................................................................................................. 75
Figure 3.5: Mean tank water concentration of (A) DHT and (B) MT during exposure period. 77
Figure 3.6: Numbers of B. glabrata snails surviving until experiment termination (30 days exposure) in each replicate tank, exposed to (A) Dihydrotestosterone (DHT) and (B) Methyltestosterone (MT). 77

Figure 3.7: Boxplots of shell diameter (A, B) and wet weight (C, D) of B. glabrata developmentally exposed to Dihydrotestosterone (A, C) and Methyltestosterone (B, D). 79
Figure 3.8: Mean shell diameter of snails from replicate tanks exposed to (A) Dihydrotestosterone (DHT) and (B) Methyltestosterone (MT). 79
Figure 3.9: Square root transformation of the total body weight as a function of shell diameter for the snails exposed to (A) Dihydrotestosterone and (B) Methyltestosterone. 80
Figure 3.10: Linear regression on the B. glabrata mean shell diameter (mm) against density (no. of snails per litre) from the (A), DHT exposure and (B), MT exposure. 81
Figure 3.11: Boxplots of normalized surface area of glandular complex (A, B), normalized albumen gland weight (C, D) and normalized ovotestis weight (E, F) of B. glabrata developmentally exposed to dihydrotestosterone (A, C, E) and methyltestosterone (B, D, F). 83

Figure 3.12: Photomicrographs of reproductive tissues (albumen gland (A-B), glandular complex (C-D), and ovotestis (E-F)) preserved in Bouin’s (A, C, E) and in RNA later followed by Bouin’s (B, D, F). 85

Figure 3.13: (A-F) Photomicrographs of part of albumen gland tissue rom DWC, SC, 62.5 ng/l, 125 ng/l, 250 ng/l and 500 ng/l of 17a-MT exposed snails. 86
Figure 3.14: (A-F): Photomicrographs of part of albumen gland tissue from DWC, SC, 62.5 ng/l, 125 ng/l, 250 ng/l and 500 ng/l of 17a-MT exposed snails. 88

Figure 3.15: (A-F): Photomicrographs of part of glandular complex tissue from DWC, SC, 62.5 ng/l, 125 ng/l, 250 ng/l and 500 ng/l of 17a-MT exposed snails. 89

Figure 3.16: (A-E): Photomicrographs of glandular complex tissue from DWC, SC, 62.5 ng/l, 125 ng/l, 250 ng/l and 500 ng/l of 17a-MT exposed snails. 90

Figure 4.0.1: Schematic representation of an nuclear receptor (NR). 99
Figure 4.0.2: Molecular mechanism of class I and II NRs. 102
Figure 4.0.3: Phylogenetic tree showing nuclear receptors identified in humans and fly and the structure of identified ligand in the nuclear receptors as mentioned in Table 4.0.2. 106

Figure 4.0.4: The UV image of the PCR products representing NRs in B. glabrata. 115
Figure 4.0.5: Example of a chromatogram file indicating the nucleotide sequence of a fragment of a NR expressed in B. glabrata analysed after sequencing. 118
Figure 4.0.6: Example of the domain analysis using InterProScan. 119

Figure 4.0.7: Phylogenetic relationships of NRs in molluscs, humans, fly and worm. 129
Figure 4.0.8: Phylogenetic relationships of THR in molluscs, humans, fly and worm using Bayesian inference showing posterior probability values. 130

Figure 5.1: Schematic of the post-translational processing of the prepro-hormone. 145
Figure 5.2: Figure showing Standard gene structure. 151

Figure 5.3: Screen shot of the Artemis Gene Builder for APGWamide showing the hierarchical order of
List of Tables

TABLE 1.1: SNAPSHOT OF THE EFFECTS OF ENDOCRINE DISRUPTORS OBSERVED IN WILD LIFE ........................................ 2
TABLE 1.2: VERTEBRATE ENDOCRINE GLANDS, TARGET GLANDS, THEIR HORMONE AND CHEMICAL NATURE OF THE HORMONES ........................................................................................................ 7
TABLE 1.3: EDSAT TIER 1 BATTERY TESTS TO DETECT HORMONE-MEDIATED EFFECTS ON RODENTS .................. 14
TABLE 1.4: OECD (2012) CONCEPTUAL FRAMEWORK FOR TESTING AND ASSESSMENT OF ENDOCRINE DISRUPTORS. 15
TABLE 1.5: THE ESTIMATED COSTS PER CHEMICAL FOR TIER 1 SCREENING, ADAPTED FROM EDSAT ....................... 17
TABLE 1.6: THE ESTIMATED COSTS PER CHEMICAL FOR TIER 2 SCREENING, ADAPTED FROM EDSAT ....................... 17
TABLE 2.1: SUMMARY OF THE TISSUE PROCESSING METHOD OF THE LEICA TISSUE PROCESSOR, INDICATING THE VARIOUS SOLUTIONS USED AND TIME TAKEN FOR EACH STEP .......................................................................................................................... 47
TABLE 2.2: THE STAINING PROTOCOL FOR HISTOLOGICAL SLIDES OF VARIOUS TISSUES FROM B. GLABRATA ........ 47
TABLE 2.3: VOLUME OF DHT ADDED TO PREPARE DOSING STOCKS MADE UP TO 1L AND NOMINAL CONCENTRATION IN THE TANKS .................................................................................................................. 56
TABLE 3.1: VOLUME OF CHEMICAL (DHT/MT) ADDED TO PREPARE DOSING STOCKS MADE UP TO 1L AND NOMINAL CONCENTRATION IN THE TANKS .................................................................................................................. 56
TABLE 3.2: MEAN VALUES (±SD) OF NORMALIZED GLANDULAR COMPLEX SURFACE AREA (MM²), NORMALIZED ALBUMEN GLAND WEIGHT (G) AND NORMALIZED OVOTESTIS WEIGHT (G) FOR B. GLABRATA DEVELOPMENTALLY EXPOSED TO DHT ........................................................................................................ 67
TABLE 3.3: MEAN VALUES (±SD) OF NORMALIZED GLANDULAR COMPLEX SURFACE AREA (MM²), NORMALIZED ALBUMEN GLAND WEIGHT (G) AND NORMALIZED OVOTESTIS WEIGHT (G) FOR B. GLABRATA DEVELOPMENTALLY EXPOSED TO MT .................................................................................................................................................................................. 81
TABLE 4.0.1: NOMENCLATURE OF NUCLEAR RECEPTORS AND THEIR ASSOCIATED LIGANDS ........................................ 104
TABLE 4.0.2: RNA EXTRACTION REPORT USING NANO DROP SHOWING THE YIELD FROM ADULT, 96 HOURS AND 20 HOURS EMBRYOS (3 EGG MASSES (EMS) IN EACH TIME POINT X TWO REPLICATES) .................................................................................................................. 112
TABLE 4.0.3: OLIGONUCLEOTIDE PRIMERS USED IN RT-PCR ANALYSES OF NUCLEAR RECEPTOR GENE PRODUCTS WITH THEIR PRODUCT SIZE (AS SUBMITTED TO GENBANK) AND THEIR CORRESPONDING GENBANK ACCESSION NUMBERS ........................................................................................................................................ 114
TABLE 4.0.4: DISTRIBUTION OF PUTATIVE NEUROPEPTIDE PRECURSORS IN THE 12 TISSUES OF ADULT BB02 SNAIL 123
TABLE 4.0.5: SEQUENCE IDENTIFICATION OF ALL NRs FOR SPECIES USED IN THIS STUDY ........................................ 124
TABLE 4.0.6: SUMMARY OF ALL THE TOP BLAST EXPRESSED SEQUENCE TAG (EST) HITS FOR L. GIANTEA NRs, INCLUDING THE EST DESCRIPTION, GENBANK ACCESSION NUMBER AND E-VALUE .................................................................................. 126
TABLE 4.0.7: RESULTS FROM BLAST2GO, INTERPROSCAN AND GO ANNOTATIONS SHOWING THE BEST HITS, THEIR CORRESPONDING E-VALUE AND PROTEIN SIGNATURES IDENTIFIED USING DIFFERENT INTERPROSCAN DATABASE .................................................................................................................. 138
TABLE 5.1: EXPLANATION OF THE GFF3 FILE FORMAT EDITED DURING ANNOTATION PROCESS ................................. 153
TABLE 5.2: DISTRIBUTION OF PUTATIVE NEUROPEPTIDE PRECURSORS IN THE 12 TISSUES OF ADULT BB02 SNAIL ... 156
TABLE 5.3: IDENTIFIED REPRODUCTIVE TYPE NEUROPEPTIDES FROM B. GLABRATA SHOWING PUTATIVE BIOACTIVE CLEAVAGE PRODUCTS AND THE NUMBER OF PEPTIDES ENCODED IN THE PROHORMONE .......................................................................................................................... 159
TABLE 5.4: RESULTS FROM BLAST2GO, INTERPROSCAN AND GO ANNOTATIONS SHOWING THE BEST HITS, THEIR
CORRESPONDING E-VALUE AND PROTEIN SIGNATURES IDENTIFIED USING DIFFERENT INTERProScan DATABASE.
1 Chapter 1: General Introduction
1.1 Endocrine disruption and animal testing

Endocrine and metabolic diseases are among the most contemporary human afflictions, the prevalence of which has been well defined in large population-based studies (Golden et al., 2009; Ford et al., 2008). To name a few these, neurodevelopmental disorders, diabetes, obesity and birth defects have been reported in pediatric population and cancer, obesity, diabetes in adult population (Bloom, Cohen and Freeman, 2010; Pleis JR, Ward BW, Lucas JW, 2010). A snapshot from different studies has been presented in the Table 1.1. The prevalence and incidence of certain disorder has been well defined in large population based studies (Golden et al., 2009). The inability to effectively reduce the incidence of these diseases and disorders stems from the failure to completely understand the mechanisms underpinning these diseases. Scientific research on the relationship between the environment and health as well as a recently released World Health Organisation (WHO) 2012 report has revealed that a larger portion of these disorders in humans may be caused by exposure to environmental pollutants than previously thought possible (Prins, 2008; Grandjean et al., 2008; Heindel, 2006). There is the suspicion that exposure to endocrine disrupting chemicals (EDCs) in the environment may be partially to blame (WHO report, 2012) (http://www.unep.org/).

Table 1.1: Snapshot of the effects of endocrine disruptors observed in wild life.

<table>
<thead>
<tr>
<th>Contaminant</th>
<th>Gender</th>
<th>Observation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diethylbesterol</td>
<td>Male</td>
<td>Decreased testosterone level</td>
<td>(Yamamoto et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Advanced development of primary and secondary follicles in ovary</td>
<td></td>
</tr>
<tr>
<td>Tributyltin</td>
<td>Male</td>
<td>Reduced no of sertoli cells in fetal testis</td>
<td>(Kishta et al., 2007)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Reduced number of germ cells in fetal ovary</td>
<td></td>
</tr>
<tr>
<td>Vinclozolin</td>
<td>Male</td>
<td>Apoptosis in germ cells in testis</td>
<td>(Cowin et al., 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Delay in puberty</td>
<td>(Wu, Buck and Mendola, 2003)</td>
</tr>
</tbody>
</table>

According to WHO (2012) the definition of an endocrine disruptor is as follows:

‘An exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism or its progeny or (sub) population’.

Typically, EDCs are natural or synthetic compounds present in food we eat, water we drink and air we breathe (Diamanti-Kandarakis et al., 2009). Modern societies produce thousands of chemicals that have the potential to interfere with the endocrine system.
These chemicals come from different sources such as energy production, agriculture, manufacturing, mining, transportation and health care (http://deainfo.nci.nih.gov/). The widespread presence of EDCs in the environment has led to proposed changes in legislation concerning the safety testing of chemicals. As a consequence, more rigorous testing of potential chemicals and drugs is required which will inevitably lead to significantly more vertebrate animals being sacrificed. Available figures for usage of rodents in reproductive toxicity tests in the UK and Europe state that around 210,000 rodents/year are used in chemical and drug testing for reproductive toxicity and that this number is rising (http://ec.europa.eu/environment/). Thus it is timely to find more ethically acceptable alternative methods for testing endocrine disrupting chemicals and this is the focus of this thesis.

The use of animals in experiments has been a long standing concern within society dating back to 1865 when a government policy related to Prevention of Cruelty to Animals was framed. Nonetheless it was not until the second half of the 20th century, with increasing scientific knowledge, that it was established that at least all vertebrate animals, are sentient beings (capable of experiencing pain and distress). This led to a significant shift in community attitudes towards greater compassion for the welfare of animals. This led to the Prevention of Cruelty to Animals Act (1986) followed by the Prevention of Cruelty of Animals Regulations (1996) and the current Prevention of Cruelty of Animals Regulations (2008). These proposed changes are in response to people's increasing concern that animals should be treated humanely and the need to avoid causing unnecessary or unjustified suffering. The European Directive 2010/63/EU specifically mentions the need to use alternatives to living animals for testing chemicals and promotes the development of alternative methods. Moreover, the Organisation of Economic Co-operation and Development (OECD) has also committed itself to the implementation of the 3Rs-principle that stands for (Replacement, Refinement and Reduction). In 2005 all European Union member countries adopted a guidance document ‘The Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment’ that is central to animal welfare and emphasises a harmonised approach towards validation of new alternative test methods for regulatory purposes. In addition to the European and UK specific regulations, the US have a number of institutions who have guidelines on to how to work towards animal welfare and implementation of alternative to animal tests. These are namely the Institute for Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals 2010, and the ICLAS guiding principles (2010) are some of the many institutions that are working towards animal welfare and implementation of
alternatives to animal tests.
In order to design any test system for evaluating potential endocrine disruptors, it is important to understand how these chemicals alter the function of the endocrine system causing adverse health effects. A brief description of the endocrine system therefore follows:

1.2 The Vertebrate endocrine system

The endocrine system in vertebrates is a system of ductless glands and the chemical messengers (hormones) that they produce. This system functions through a series of cascade mechanisms, triggered by external or internal stimuli causing secretion of hormones into the blood that then travel to target organs binding to specific receptors leading to a coordinated physiological response (Hadley, 1992). These coordinated physiological responses are regulated by negative and positive feedback mechanisms to maintain homeostasis and prevent alterations in hormone secretions that could cause adverse effects.

The endocrine system cascade acts through a number of axes or sub-systems, e.g. the hypothalamus-pituitary-gonadal (HPG) axis consisting of the hypothalamus, pituitary and gonads (ovaries and testes), regulating reproductive and immune function; the hypothalamus-pituitary-thyroid (HPT) axis and hypothalamus-pituitary-adrenal (HPA) axis for maintaining energy homeostasis and energy maintenance respectively. The different glands within (and sometimes between) each axis, signal each other in sequence.

During male sexual maturation, for example, (Figure 1.1) after genetic sex determination, genes located on the Y chromosome send a signal for sexual differentiation through neurotransmitters to the hypothalamus in the brain. In response the hypothalamus secretes the gonadotropin-releasing hormone (GnRH), which travels to the anterior pituitary (adenohypophysis), binds to the GnRH receptor and stimulates the secretion of gonadotropins such as follicle stimulating hormone (FSH) and Luteinizing hormone (LH). FSH and LH travel to the testis, where they stimulate the production and secretion of testosterone from the Leydig cells in the testis (Dorner, 1978). In the testis LH binds to the receptors on the Leydig cells and stimulates the synthesis of testosterone whilst FSH binds to the receptors on the Sertoli cells and release and metabolise factors necessary for spermatogenesis (Campbell and Reece, 2005). Testosterone is responsible for differentiation and development of the male urinogenital system. Testosterone then circulates back to the pituitary and hypothalamus and this leads to a reduction in the
secretion of GnRH (negative feedback). The testis then undergoes quiescence until puberty when they are once again activated by GnRH secretion from the pituitary. The hypothalamus is the neural control center for most of the endocrine systems of glands and hormones involved in physiological and behavioural responses that are necessary to survive, through the anterior pituitary gland. The hypothalamic neurosecretory cells produce hormones (some releasing and some inhibitory) that reach the anterior pituitary gland via hypothalamo-hypophyseal system to regulate its hormone (tropic hormone) production. These tropic hormones regulate the endocrine activities of the thyroid gland (secreting thyroid hormone), the adrenal cortex (secreting corticosteroids), gonads (secreting reproductive steroids) and the liver (secreting growth factors). This control system acting through hypothalamus does not directly control a number of endocrine organs like the adrenal medulla, the parathyroid glands, the pancreas, the thymus, the gastrointestinal tract, pineal gland and the kidney. The feedback mechanism in these glands is controlled by signals of the regulated variable they control for example the blood glucose levels control the secretion and inhibition of insulin and glucagon from the pancreas.
Figure 1.1: Schematic diagram of the vertebrate endocrine system showing feedback mechanism of hypothalamus-pituitary-gonadal axis.

The positive feedback mechanism stimulates the secretion of Gonadotropic-releasing hormone (GnRH) from hypothalamus to anterior lobe of pituitary gland that produces and secretes gonadotropins (LH and FSH). These gonadotropins travel to testes and enhance the production of testosterone. The negative feedback mechanism is inhibitory to reduce the secretion of GnRH and thus maintain homeostasis (Taken from http://pubs.niaaa.nih.gov/publications/arh25-4/282-287.htm)

1.2.1 Endocrine glands

As mentioned above endocrine glands are ductless glands, they are highly vascularised and as secretory cells they are located close to the blood vessels to allow rapid transmission of their secretions. These glands have specific biological organisation to suit their function like the peptide secreting cells have well-developed rough endoplasmic reticulum rich in protein filled storage granules and the mitochondria has flat cristae while steroid secreting cells have more of smooth endoplasmic reticulum rich in lipid droplets containing cholesterol (precursor of steroids) and the mitochondria has tubular cristae (Norris and Carr, 2013). The following Table 1.2 shows the various vertebrate endocrine glands and their hormones.
Table 1.2: Vertebrate endocrine glands, target glands, their hormone and chemical nature of the hormones.
(Adapted from Adkins-Regan, 2007).

### 1.2.2 Hormones

A hormone is a biochemical molecule produced by an endocrine gland that travels via blood to distant cells and tissues where it produces irreversible effects (Melmed et al., 2011). Hormones regulate a variety of physiological processes in the body from cell differentiation during embryonic development to the control of organ functions in adulthood and behavioural activities such as responding in appropriate ways to other individuals (social behaviour), that varies with age, between and with the sexes (Adkins-Regan, 2007). A single hormone can affect more than one function and several hormones
Vertebrate hormones fall into three chemical classes:

1. Amino acid derivatives are the derivatives of tyrosine such as catecholamines (epinephrine or norepinephrine) secreted by adrenal medulla and thyroxine secreted by the thyroid gland.

2. Peptide hormones are glycoproteins secreted from anterior pituitary gland such as thyroid stimulating hormone and luteinizing hormone. These also include gastrointestinal tract hormones and pancreatic hormones such as insulin.

3. Lipid derivatives are cholesterol derived steroid hormones like androgen, estrogen, and those secreted by adrenal cortex. These are synthesised on demand and have the same basic ring structure but their side chain and spatial orientation renders them specific. Another class of lipid derivatives includes the major classes of eicosanoids including prostaglandins, prostacyclins, thromboxanes and leukotrienes (Nussey and Whitehead, 2001).

1.2.3 Hormone synthesis

Amino acid derivatives like catecholamines are hormone regulators and are synthesised from amino acid phenylalanine or tyrosine by enzymatic reaction taking place in the cytosol. These are then stored in vesicles and are exocytosed as and when required. Being water soluble (hydrophilic) they do not require any transport proteins and can be easily transported through plasma.

The synthesis of peptide hormones starts with transcription of a single gene in the presence of RNA polymerase II and various transcription factors followed by modification of 3’ and 5’ ends of the mRNA (messenger RNA). This mRNA is used as a template that docks with the signal recognition particle on the surface of rough endoplasmic reticulum where protein synthesis takes place. The peptides in their inactive state (pro-hormone) are packed in secretory granules within the endoplasmic reticulum from where they pass into the Golgi apparatus and in the presence of peptidases undergo cleavage forming active hormones. Like catecholamines, these are also transported through plasma (Norris and Carr, 2013; Nussey and Whitehead, 2001).

Steroid hormones are synthesised through a biosynthetic pathway called steroidogenesis in the mitochondria and endoplasmic reticulum. The complete process of steroidogenesis involves a number of steps starting with signal transduction pathway, which causes the increased transport of cholesterol to mitochondria in Leydig cells. Cholesterol is the starting precursor molecule for all steroid hormones and all steroid hormones have similar basic structure. Enzymatic conversion of cholesterol into pregnenolone in the presence of
cytochrome P450 cholesterol side-chain cleavage enzyme (P450SCC), initiates this biochemical pathway that culminates with the formation of hormones (Figure 1.2). For the purpose of illustration the pathway for the formation of androgens will be described. Pregnenolone is converted into progesterone by 3β-hydroxysteroid dehydrogenase. Progesterone is converted into 17α-hydroxyprogesterone (by 17α-hydroxylase), which is then converted into androstenedione (by 17, 20 lyase). 17β-hydroxysteroid dehydrogenase converts androstenedione into testosterone, which is then converted into dihydrotestosterone (DHT) by 5α-reductase (Hinson et al., 2007). DHT is significantly more potent androgen compared to testosterone and is an end-product hormone. The last enzyme of steroidogenic pathway is aromatase that converts testosterone into estradiol, an estrogen.

![Figure 1.2: The steroid hormone synthesis pathway in vertebrates showing the enzymes and cytochrome P450 (CYP) genes participating in this process. The enzymes are 3β-HSD: 3beta-Hydroxysteroid dehydrogenase; 17β-HSD: 17beta-Hydroxysteroid dehydrogenase. (Adapted from Fretter, 1984).](image)

1.2.4 Hormone action

Most hormones initiate a cellular response by combining with an intracellular or cell membrane associated receptor. Hormone receptors provide specificity to hormone actions, both in terms of the time and the place of hormone action. Hormones cannot act on all the cells but only those where the respective receptors are present and also many hormones do
not act throughout the life span of an organism but at specific time periods of life.

- Peptides are large water soluble molecules that do not readily cross the hydrophobic lipid cell membrane. For this reasons their receptors are membrane receptors lying extracellularly and they couple with the intracellular signal transducing molecules by transversing the cell membrane (Nussey and Whitehead, 2001). Most peptide receptors are members of G-protein coupled receptor family and these may either have a short extracellular amino terminal domain (For example, GnRH) or a long extracellular domain (For example thyroid stimulating hormone) (Darlison and Richter, 1999). Hormone-receptor interaction causes dissociation of associated G-protein, activation of membrane bound enzymes that stimulate the synthesis of cyclic AMP (cAMP) (second messenger), which triggers protein kinases (Cooper and Hausman, 2000). The transcription factor targets for kinases include c-jun and c-fos that form heterodimeric complex, which may be targeted by mitogen activated protein kinase (MAPK) or cAMP response element binding protein (CREB) that initiate transcription (Nussey and Whitehead, 2001). A GTP-binding membrane protein mediates between the peptide receptor protein and the enzyme for the synthesis of AMP, adenylate cyclase. However, some peptides like vasopressin have more than one receptor subtype (Goodson and Bass, 2001). There is evidence that steroids, acting through intracellular receptors stimulate the transcription of membrane receptor genes thus upregulating peptide action (Norris, Madsen and Freestone, 1996). In addition, peptides and their receptors are key players in the downstream pathway for steroid action on brain and behaviour (Ferguson, Young and Insel, 2002; Albers and Bamshad, 1999).

- Unlike peptides, steroid and thyroid hormones are lipophilic and can easily diffuse across the cell membrane. Their receptors are intracellular proteins acting as transcription factors regulating the expression of target genes in target tissues. These receptors belong to the superfamily of nuclear receptors (NRs) that are classified based on their ability to form dimers and the sequence of DNA they bind (Mangelsdorf et al., 1995). Not all the NRs are activated by hormones (also called ligands), for some NRs regulatory ligands are not known or might not exist and these could be regulated by dietary components. All these receptors have a similar structure with the highly conserved DNA binding domain and less conserved the ligand binding domain regions (discussed in detail in chapter 4) (Germain et al., 2006). The hormone binds the NR in the ligand binding domain causing structural changes in the receptor that allows the DNA binding domain to bind with the
hormone response elements, localized in the 5’ flanking region of the target genes and bring about transcription. In addition to the nuclear receptors, there are also membrane receptors for some steroid hormones like estrogens and progestins. The membrane receptors for these receptors are coupled with fast acting pathways that result in rapid effects, unlike NRs, which take several hours to stimulate production of protein and exert action.

1.3 Endocrine disruptor action

Endocrine disrupting chemicals (EDCs) mimicking the endogenous hormones can act to disrupt physiology by interfering with the synthesis, secretion, transport, binding, action or elimination of hormones in the body so changing homeostasis, reproduction or development and/or behaviour (Vos et al., 2000; Tyler, Jobling and Sumpter, 1998). These chemicals have been shown to affect the hormone delivery to the receptor, have effect on hormone receptors and alter their functions (Diamanti-Kandarakis et al., 2009). Some EDCs may have mixed properties or they may metabolise to yield substance having different properties. For example estrogen agonist dichlorodiphenyltrichloroethane (DDT) metabolises into dichlorodiphenyldichloroethylene (DDE), which is androgen antagonist (Diamanti-Kandarakis et al., 2009; Sohoni and Sumpter, 1998). These conversions can have significant biological effects as human reproduction (male and female) is an interplay of both androgens and estrogens. Along with direct effects on androgens and estrogens they can also interfere in the negative and positive feedback mechanisms of the different axes in vertebrates (HPG, HPT and HPA) (Kester et al., 2002).

Among the first compounds to be identified as an EDC is diethylstilbestrol (DES). This was specifically designed drug to help prevent miscarriage and restore hormonal balance in pregnant women (Bamigboye and Morris, 2003; Smith, 1949; Smith and Smith, 1949). However in 1971, Herbst and colleagues found that in utero exposure to DES is associated with vaginal adenocarcinoma in female offspring (Herbst, Ulfelder and Poskanzer, 1971). This study was followed by reports on DES induced cryptorchidism in male offspring (Newbold, 2004; Lasserguere et al., 2003; Skakkebaek, Rajpert-De Meyts and Main, 2001; Giusti, Iwamoto and Hatch, 1995). Transgenerational studies using DES suggest that in utero exposure to DES alters the normal programming of genes that play an important role in development and differentiation of reproductive tract, which increased risk of reproductive tract anomalies and dysfunction in their sons and daughters (Titus-Ernstoff et al., 2010; Palmer et al., 2005; Schrager and Potter, 2004; Kim et al., 2002) and
have been associated with 80% increased risk of endometriosis (Missmer et al., 2004). Subsequent research confirmed the transplacental mechanism of DES transmission (Kerlin, 2005; Maydl et al., 1983) and classified DES as a mutagen (Stopper, Schmitt and Kobras, 2005), teratogen and carcinogen (Mittendorf, 1995). Animal studies using prenatally DES exposed outbred mice showed numerous transgenerational effects including uterine fibroids (Hoover et al., 2011); ovarian cancer (Blatt et al., 2003); increased menstrual irregularities (Titus-Ernstoff et al., 2006) and others. One study by Viglietti-Panzica et al. (2005) investigating sexual behaviour of Japanese quail concluded that embryonic exposure to DES caused sex reversal of behavioural phenotype and a significant decrease in vasotocin in the males (Viglietti-Panzica et al., 2005). In another recent study using Balb/c pregnant mice it was demonstrated that 67μg/kg body weight of DES injected on gestation day 13 induced symptoms of cervical carcinogenesis in the F1 generation (Zulfahmi et al., 2013). Even though this medicine was banned for use during pregnancy in the 1980s, animal studies are being performed to understand the mechanism of action. In a review Newbold (2012) suggested that DES exposed population should be continuously monitored as they age and the mechanism of DES induced toxicities would be useful in predicting effects of other environmental estrogens (Newbold, 2012).

Another example of cause and effect of EDCs is the case of the skewed sex ratio of fish in the rivers receiving effluent from paper mills effluents and sewage treatment plants which has been reported consistently across geographical regions and habitats and different species (Lepomis megalotis, Catostomus catostomus, Coregonus clupeaformis, Perca fluviatilis, Gambusia affinis; Pimephales promelas) (Hinck et al., 2009; Blazer et al., 2007; Fentress et al., 2006; Orlando et al., 2004; Jobling et al., 1998; Munkittrick et al., 1998; Gagnon et al., 1994; Munkittrick et al., 1991). Masculinisation of fish was reported among the embryos of viviparous eelpout Zoarces viviparous in the vicinity of a paper mill in Swedish Baltic coast (Larsson and Forlin, 2002). During the period of mill shutdown for 17 days, which coincided with the duration of embryonic gonad differentiation, sex ratios returned to normal, and then subsequently shifted towards males again when the mill was reactivated the following year, providing evidence in favour of chemicals in the effluent causing the shift in sex ratio. Compounds present in pulp/Kraft mill effluents have been shown to bind to androgen receptor in both humans and fish (Hewitt et al., 2002; Parks et al., 2001) and also bind to the retinol receptors, RXR and RAR (Alsop et al., 2003).

In contrast, Lange et al. (2011) found sex ratios of roach deviating from 1:1 ratio in uncontaminated environment to 98% feminised phenotype, when exposed to wastewater
effluent from one month post hatch to 3.5 years of age (Geraudie et al., 2011; Lange et al., 2011). Long-term studies in a lake in northwestern Ontario, Canada, indicated that exposure to the potent estrogen, ethinylestradiol, at 4–6 ng/L over a period of 3 years resulted in the collapse of the population of the fathead minnow (*Pimephales promelas*) (Harris et al., 2011; Kidd et al., 2007). A strong correlation between predicted steroidal estrogens in the UK rivers and feminisation in the wild fish was demonstrated by Jobling et al. (2006) (Jobling et al., 2006). Statistical modeling of different steroids in the UK wastewater also provided a strong argument of multicausal etiology of the widespread feminisation of fish (Jobling et al., 2009).

There are many more instances of endocrine disruption having huge effects like the population-level consequences of Tributyltin (TBT) on molluscs (will be discussed in later chapters) and the story goes on but the important issue to be tackled is the increasing number of chemicals that are being introduced in the environment through varied sources and the increased number of pre-clinical tests to meet safety requirements.

### 1.4 Endocrine Disruptor Animal Testing Programmes

The increased awareness that EDCs might play a role in the increasing prevalence of reproductive endocrine diseases has led to stringent laws for chemical testing being devised both by the US Environmental Protection Agency (EPA) and by the OECD. Currently these test protocols focus mainly on interactions of chemicals with the estrogen, androgen and thyroid systems. The Endocrine Disruptor Screening and Testing Advisory Committee (EDSAT, 1998) stated in their final report, that as a part of the protocol, every chemical has to undergo a two-tiered screening process firstly consisting of *in vitro* and *in vivo* assays to identify and classify substances relative to their potential interaction with the endocrine hormone (androgen, estrogen and thyroid) system (Tier 1) Table 1.3. Tier 2 tests focused on developing a dose-response relationship in animal models (http://www.epa.gov/). The screening process is a battery of tests designed to ensure estrogen and androgen mediated effects through various modes of action (receptor binding and steroidogenesis pathway) and also feedback mechanisms of the HPG- and HPT-axis are detected. The battery is comprised of a combination of *in vitro* and *in vivo* assays that complement each other and help identify specific endocrine mechanisms of chemicals and relate them to physiological effects observed at higher levels of biological organisation. If a chemical is identified as a potential endocrine disruptor, it is further screened using Tier 2 battery of processes. The main purpose of Tier 2 is to provide a comprehensive profile
of the effects the chemical could induce via different modes/mechanisms of action.

Table 1.3: EDSAT Tier 1 Battery tests to detect hormone-mediated effects on rodents.

<table>
<thead>
<tr>
<th>Screening Assays</th>
<th>Modes of action</th>
<th>Steroidogenesis</th>
<th>HPG- HPT-Axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Vitro</td>
<td>Receptor Binding</td>
<td>Anti-Androgen</td>
<td>Anti-Androgen</td>
</tr>
<tr>
<td>ER Binding</td>
<td>Estrogen</td>
<td>Androgen</td>
<td>Steroidogenesis</td>
</tr>
<tr>
<td>ERx Transcriptional activation</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>AR Binding</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Steroidogenesis</td>
<td>H29SR</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Aromatase</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Recombinant In Vivo</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Uterotrophic Hershberger</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Pubertal Male</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Pubertal Female</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Amphibian Metamorphosis</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fish short-term Reproduction (Male and Female)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

In short Tier1 should inform Tier2 in terms of dose range, life stage of exposure and most appropriate organism in which adverse effects are likely to occur. Tier 2 tests are long-term animal studies that could even be transgenerational with a broad range of taxa (Hecker and Hollert, 2011). In addition to the US EPA activities in the same year, the OECD also initiated an Endocrine Disrupter Testing and Assessment (EDTA) process. The EDTA has five levels of organisation that are not to be used as a tiered testing strategy but rather as a conceptual framework with which to provide information regarding the hazards of a chemical with regards to its potential endocrine disrupting properties Table 1.4. In 2007 the European community released a summary scientific report for prioritising chemicals for their endocrine disrupting effects. Of the 575 chemicals investigated, 320 substances were identified as potential EDCs (European Commission, 2007). The European Commission has also co-operated with the WHO (WHO, 2012). Besides these, REACH (Registration, Evaluation, Authorisation, and restriction of CHemical substances) is another European community regulation on chemicals and their safe use and came into force on 7th June 2007. REACH allows identification of intrinsic properties of chemicals such as carcinogenicity, mutagenicity, reproductive toxicity and their persistence and potential for bioaccumulation. The testing requirements for chemicals are according to their supply in tonnage. The tests are those included in OECD conceptual framework (Table 1.4). Taken together, the available information suggests that national and international governments are in the process of establishing testing programs and strategies to assess the risk associated with endocrine
disrupting chemicals.

Table 1.4: OECD (2012) conceptual framework for testing and assessment of endocrine disruptors.
Adapted from OECD (2012).

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Existing data/literature</td>
</tr>
<tr>
<td>2</td>
<td>In vitro assays providing data about selected endocrine mechanisms/pathways</td>
</tr>
<tr>
<td>3</td>
<td>In vivo assays providing data about selected endocrine mechanisms/pathways</td>
</tr>
<tr>
<td>4</td>
<td>In vivo assays providing data on adverse effects on endocrine relevant endpoints</td>
</tr>
<tr>
<td>5</td>
<td>In vivo assays providing more comprehensive data on adverse effects on endocrine relevant endpoints encompassing different time-points in life-cycle of organisms</td>
</tr>
</tbody>
</table>

One of the key issues remaining in context with the increasing number of testing programs across the globe to evaluate the likelihood of a chemical to disrupt endocrine functions is the reliance on a large number of in vivo tests using live animals. With the increasing testing demands the number of rodents, amphibians and fish that will be required will increase proportionally. It is difficult to estimate the number of animals being used globally for animal testing. In 1995, researchers at Tufts University, US Center for Animals and Public Policy had an estimate of 14–21 million animals being used in American laboratories in 1992, which was a reduction from 50 million used in 1970 (Andrew NR, 1995). According to the US Department of Agriculture (USDA) the total number of animals used in 2005 was almost 1.2 million (USDA report on the Enforcement of the Animal Welfare Act, 2005), excluding rodents, which make up about 90% of research animals. The British Union for the Abolition of Vivisection (BUAV) has estimated that 115 million vertebrates are experimented on around the world every year of which approximately 27 million of them are in European Union laboratories (http://www.buav.org). The US Department of Agriculture (USDA) has estimated over 950,000 animals were used in research in 2012 as compared to 1.1 million in 2010 (http://speakingofresearch.com/) of which 98% of the research was done using rodents, birds and fish. According to the UK Home Office report HC317 on animal testing (2011), just over 2.18 million animals were used in scientific experimentation in 2011. Sector-wise, the report shows an increase of 2% in the number of scientific studies for toxicological safety testing as compared to 2010; 78% of these procedures were for pharmaceutical procedures and 73% involved rodent species in testing. Of the total species used, mice contributed to 71% in these studies Figure 1.3 (https://www.gov.uk/).
According to the Annual Statistics of Scientific Procedure on living animals, Great Britain (2012) this number went down to 2.13 million animals used in scientific experimentation in 2012 (excluding breeding of genetically modified and harmful mutants) (https://www.gov.uk/). The US EPA’s EDSP (Endocrine Disruptor Screening Program) report highlights that subjecting a single chemical to the complete Tier 1 testing battery would require approximately 130 rats, 30 tadpoles/froglets and 60 fish (Hecker and Hollert, 2011). These estimates give us an overview of the large number of animals being used. None of these figures include animals bred in surplus for research studies that are then killed as surplus and or animals used for breeding purposes. Invertebrates like fruit flies, nematodes or molluscs are not part of these calculations, as they are not yet regulated. It is not only the huge number of animals used that should be the reason for concern but also the cost involved in these studies. Table 1.5 and Table 1.6 below shows the estimated total cost for conducting EDSAT Tier1 and Tier2 screening for a chemical which totals $922300. This does not include the feeding, housing, and other maintenance costs applied.

With the changes in the legislation concerning the safety testing of chemicals to incorporate endocrine disruption, increased numbers of vertebrate animal tests involving rats, mice, fish and frogs will grow proportionally (http://www.epa.gov/endo/pubs/edsat/) and almost 80,000 chemicals need testing (Krisberg, 2010).
Table 1.5: The estimated costs per chemical for Tier 1 screening, adapted from EDSAT final report.
The values listed are the mean of estimates for each assay obtained from a cost estimate survey conducted by Applied Pharmacology and Toxicology, Inc. in May 1998.

<table>
<thead>
<tr>
<th>Name of Assay</th>
<th>Cost Required endpoints in mammalian assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat estrogen receptor equilibrium exchange assay</td>
<td>$7,200.00</td>
</tr>
<tr>
<td>Rat androgen receptor equilibrium exchange assay</td>
<td>$7,300.00</td>
</tr>
<tr>
<td>Steroidogenesis assay in minced testes</td>
<td>$11,600.00</td>
</tr>
<tr>
<td>Uterotrophic Assay in Ovariectomized Rats</td>
<td>$29,600.00</td>
</tr>
<tr>
<td>The Hershberger Assay in Male Rats</td>
<td>$37,200.00</td>
</tr>
<tr>
<td>Pubertal Assay in Female Rat</td>
<td>$38,900.00</td>
</tr>
<tr>
<td>Fish Gonadal Recrudescence Assay</td>
<td>$37,800.00</td>
</tr>
<tr>
<td>Frog Metamorphosis Assay</td>
<td>$22,200.00</td>
</tr>
<tr>
<td>Subtotal (Assays only)</td>
<td>$191,800.00</td>
</tr>
<tr>
<td>Analytical costs</td>
<td>$30,000.00</td>
</tr>
<tr>
<td>Total for Tier 1 screening</td>
<td>$221,800.00</td>
</tr>
</tbody>
</table>

Table 1.6: The estimated costs per chemical for Tier 2 screening, adapted from EDSAT final report.
The values listed are the mean of estimates for each assay obtained from a cost estimate survey conducted by Applied Pharmacology and Toxicology, Inc. in May 1998.

<table>
<thead>
<tr>
<th>Name of Test</th>
<th>Cost of test</th>
<th>Analytical cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avian Reproductive Toxicity Test</td>
<td>$172,100.00</td>
<td>$55,000.00</td>
</tr>
<tr>
<td>Fish Life Cycle Toxicity Test</td>
<td>$221,400.00</td>
<td>$80,000.00</td>
</tr>
<tr>
<td>Mysid Toxicity Test</td>
<td>$38,600.00</td>
<td>$23,000.00</td>
</tr>
<tr>
<td>Amphibian Reproductive / Developmental Toxicity Test</td>
<td>$75,400.00</td>
<td>$35,000.00</td>
</tr>
<tr>
<td>Subtotal</td>
<td>$507,500.00</td>
<td>$193,000.00</td>
</tr>
<tr>
<td>Total for Tier 2 screening</td>
<td>$700,500.00</td>
<td></td>
</tr>
</tbody>
</table>

The use of alternative methods for testing endocrine disruptors could bring substantial economic savings in terms of time and resource. In vitro assays, high throughput screening assays and working with invertebrates with short life cycles is increasing the speed with which results can be obtained, and there is decreased cost of maintaining microorganisms/invertebrates compared with maintaining vertebrate animals in laboratories. A good example is the use of in vitro tests in high-throughput screening (HTS) for drug discovery in the pharmaceutical industry. The search for compounds with the ability to interact with particular molecular targets among millions of chemical
structures would be practically impossible without this technology. HTS technologies have found application in a wide range of important topics in biology and medicine, now allowing researchers to address biological questions using innovative approaches such as discovery and development of new chemical probes as research tools for use by the research community to enhance the understanding of biological functions and disease mechanisms (Soon, Hariharan and Snyder, 2013). One of the candidate drugs, FT1050 (16,16-dimethyl Prostaglandin E2), that influences adult stem cells was discovered through HTS on zebra fish. This drug has entered clinical trials for optimising transplantation of human hematopoietic stem cells from umbilical cord blood (North et al., 2007). Newly developed gene sequencing methods have provided new perceptions into the global organisation of eukaryotic genomes that were previously unattainable. Analyses of individual regions in the genome have revealed the role of some regulatory factors in controlling the expression of genes such as transcription factors and non-coding RNAs (Eskiw et al., 2010; Branco and Pombo, 2007; Fraser and Bickmore, 2007). This information has helped researchers to map genetic variants to gene regulatory regions and to analyse multiple pathways linked to disease (Boyle and Silver, 2012).

Both the governmental and non-governmental funding bodies do realise that animal research must continue for the well being of all communities (animal and plants) on the earth and that immediate abolition of animal testing is not possible. Nevertheless, more research is being focused on finding alternatives to vertebrate animal testing for endocrine disruptors in different areas of research such as biomedical science, endocrinology and drug development. My thesis describes one of the examples of many such projects being funded by the NC3Rs (National Centre for Replacement, Refinement and Reduction of animals in research).

1.5 The 3Rs (Replacement, Refinement and Reduction of animals in testing chemicals)

The 3Rs principle was first set out by William Russell and Rex Burch in 1959 as ‘The Principles of Humane Experimental Technique’. Since then, the 3Rs have become established as an internationally accepted approach to research involving animals, and have been written into legislation in many countries (Guhad, 2005; Russell, Burch and Hume, 1959). Besides this the European Union Reference Laboratory for Alternatives to Animal Testing (ECVAM), Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) and EU funded, European Phased Adaptive Approach (EPAA) are
also conducting research into alternative methods and their validation.

Reduction – includes methods that minimise the number of animals used per experiment or study including maximising the information obtained from a single animal. Examples include improved experimental design and statistical analysis, sharing data and resources within the scientific community and organisations by depositing experimental design and results in public repositories and the use of technologies, such as imaging to enable longitudinal studies in the same animal.

Refinement includes methods that minimise pain, suffering and distress or lasting harm experienced by the animals like surgical and other painful procedure. Pain, distress causes physiological changes in the animal and could lead to variability in results. Refinement reduces variability in experimental results and thus reduces the number of animal tests that need to be done. Experiments encompassing painful procedures or surgeries should not be carried out without proper sedation or anesthesia and if any procedure could lead to chronic pain then the animals should be humanely killed. Refinement also means using techniques and choice of animal that is best suited for the projects in terms of their behaviour, genetic constitution and general health.

Replacement includes the use of:

- Human volunteers, tissues and cells
- Non-testing approaches like mathematical and computer models
- Established animal cell lines, or cells and tissues taken from animals killed solely for this purpose (i.e. not having been subject to a regulated procedure)
- Immature forms of vertebrates, or invertebrates, such as *Drosophila* and nematode worms.
- Use of ‘omics’ technologies like transcriptomics, proteomics and metabolomics.

The research described in this thesis focuses on invertebrates as substitutes for vertebrate testing. One of the dysfunctions for which there seems at present to be no current invertebrate alternative to vertebrate animal testing is reproductive disorders and this is also a particular focus of this thesis.

### 1.6 Invertebrates as replacements for vertebrates in reproductive toxicity testing

Historically, invertebrates have made a major contribution to biomedical research (Wilber, 1964) because certain aspects of their physiology are very similar to that of mammals. There has always been a strong belief that organisms closest to humans in evolutionary
terms will provide the best ‘model’ for experiments on human biology and medicine. For example, a monkey is a better model for the human than is a rat, and a rat is better than a worm. However recent research has shown this is not necessarily the case as the very basic biological functions amongst all animals are quite conserved. For example, Pourquie (2003) reported that genetic machinery responsible for somitogenesis, a multistep patterning process forming mesoderm of the embryo, is conserved throughout the animal kingdom from rodents to protostomes through deuterostomes (Pourquié, 2003). With the growing wealth of sequenced genomes, computational comparative genomic tools and system biology approaches there is a potential to utilise the divergence of DNA sequences and physiology between species to find a suitable invertebrate model. Invertebrates raise less societal concern as compared to vertebrates and are not included in the list of protected animals under the UK Home Office Animals (Scientific procedures) Act 1986 (except Octopus vulgaris) (https://www.gov.uk). They generally have a short life span so developmental studies can yield faster results and they are easy and less expensive to culture and maintain in laboratories. Invertebrates have a simpler body plan as compared to vertebrates and this has been pivotal in understanding basic biological systems in vertebrates (Andrews, 2011). The ability of invertebrates to regenerate, as seen in echinoderms (Mashanov et al., 2010; Carneveli, 2006) can be quite useful in optimising techniques such as tissue engineering. The primary advantage of designing assays using lower vertebrates/invertebrates is that the embryos develop rapidly and we can get the complete repertoire of their gene expression and molecular signaling from fertilization to organogenesis in short span of time. During this critical window of development, embryos can be monitored for a series of physiological or behavioural endpoints and there is a high probability of detecting an adverse interaction between the chemical and its molecular targets. In addition to the rapid developmental profile, invertebrate embryos are often transparent and develop externally (in contrast to mammalian in utero development), allowing for non-invasive microscopy techniques to observe them across different developmental stages. Moreover, using whole animals is quite useful when integrative endpoints are assessed, keeping in mind the complexity of an endocrine system functioning in a living organism, and especially that of a developing organism. For example in the area of reproductive toxicity, one of the problems is the lack of understanding of many integrated mechanisms of action leading to toxicity (EURL-ECVAM report, 2013) (Zuang et al., 2013) and real time exposure studies in invertebrates could be an answer. Nevertheless, working with whole animals/embryos would also allow observing physiological processes without the knowledge of mechanisms, including
currently unknown mechanisms.

**Drosophila melanogaster** (fruit fly) is the best understood animal model for detecting mutagenicity (Brusick, 1980), teratogenicity (Bournias-Vardiabasis et al., 1983) and female reproductive toxicity (Avanesian, Semnani and Jafari, 2009; Goss and Sabourin, 1985). The importance of *D. melanogaster* to research can also be estimated by the Nobel prizes awarded to Thomas H Morgan in 1933, for his discovery related to the role of chromosomes in heredity using *D. melanogaster* as model organism.

*D. melanogaster* is a poikilothermic, diploid outbreeding organism and adult flies have body parts divided into three regions (head, thorax with attached wings and abdomen) with distinct tissues and homologues of most human organs (Bier and McGinnis, 2004). However, *Drosophila*, like all insects, does not have a calcareous endoskeleton like mammals, but a chitinous exoskeleton. Further distinct features are the lack of an adaptive immune system, its circulatory system comparable to a mix of lymph and blood, the tracheal breathing system, and the structure of the nervous system, to name a few. *D. melanogaster* is bisexual with male and female flies and has a high rate of reproduction. It undergoes indirect development passing through distinct phases from egg to larva and pupa, reaching adulthood in a short time span of about 10 days under common laboratory conditions. The reproductive organs of female fly include a pair of ovaries, accessory glands, spermatheca, spermathecal gland, gonophore and vagina making it structurally similar in many regards to mammalian reproductive organs (King, Aggarwal and Aggarwal, 1968). The male reproductive anatomy comprises of testes, vas deferens, seminal vesicle, accessory gland and ejaculatory duct leading into the gonophore from where the sperm exits. There is a pair of testes comprised of testicular follicles and encased in epithelial tissue (nutritive layer) and all the follicles are encased in connective tissue. The accessory glands are responsible for secreting chemical infusion for stimulating spermatozoa. There is increasing evidence of conserved sexual development genes between *D. melanogaster* and mammals (Nanda et al., 2009; Lalli and Sassone-Corsi, 2003; Carney and Bender, 2000; De Loof and Huybrechts, 1998) and there are studies suggesting *D. melanogaster* as an alternate animal model for quick screening of chemicals for their reproductive toxicity. Indeed, the *Drosophila* female reproductive system has been studied to get a better understanding of human ovarian cancer (Naora and Montell, 2005) and some studies have demonstrated ovarian impairment in female *Drosophila* when exposed to 0.12 and 0.24mg/ml of methotrexate (causes folic acid deficiency) (Avanesian, Semnani and Jafari, 2009; Affleck and Walker, 2007; Affleck et al., 2006). Xenobiotics like dichlorvos and chlorpyrifos caused tissue damage in the
accessory gland of adult male flies in a dose dependent manner (Gupta et al., 2007). However, the major limitation for using Drosophila as a model organism for male reproductive toxicity is the absence of mammalian molecular machinery, such as the estrogen receptor and androgen receptor as well as mammalian steroid hormones. A few studies have suggested ecdysone as a potential steroid hormone-mimicking mammalian steroid hormone (Rauschenbach et al., 2000; De Loof and Huybrechts, 1998) but clear evidence is still lacking (Tiwari et al., 2011).

**Caenorhabditis elegans** - The round worm (Caenorhabditis elegans) is a model organism popularised by Sydney Brenner in the field of developmental biology (Brenner, 1974). Since then C. elegans have been used to study neurodegeneration, signal transduction, apoptosis, ageing and RNA interference (Teschendorf and Link, 2009; Leung et al., 2008; Antoshechkin and Sternberg, 2007; Timmons et al., 2003). Moreover Nobel Prizes have been awarded to three geneticists working with C. elegans. In 2002, the Nobel Prize in Medicine was awarded to Sydney Brenner, H. Robert Horvitz and John Sulston for their work on the genetics of organ development and programmed cell death in C. elegans. In 2006, Andrew Fire and Craig C. Mello were awarded the Nobel Prize for their discovery of RNA interference in C. elegans. The 2008 Nobel Prize in Chemistry was shared between Martin Chafie, Osamu Shimomura and Roger Y. Tsien for their work on green fluorescent protein in C. elegans. Furthermore, the intensively studied genome, complete cell lineage map, knockout mutant libraries and established genetic methodologies such as mutagenesis and RNA interference to name a few, provide a variety of methods to manipulate and study C. elegans at the molecular level (Leung et al., 2008).

Like Drosophila, C. elegans is also poikilothermic, can be easily cultured in the laboratory and has an even shorter life span of 3 weeks. Unlike Drosophila it is small in size (adult worm measuring 1mm) and has a transparent body. It is naturally a highly inbred organism of mostly self-fertilising hermaphrodites and occasional males with high reproductive output and a short time span of 3 days from egg to sexual maturation (Kaletta and Hengartner, 2006). The hermaphrodite germ line produces both oocytes and sperms. Sperms are produced by maturing germ cells only during L4 larval stage (approximately 37 hours post oviposition) while in the adult stage germ cells mature to oocytes (Zanetti et al., 2012). Development is indirect passing through four distinct larval stages L1-L4 in three days at 20°C. C. elegans as a model organism has its own advantages and disadvantages. The small size (adult is approximately 1mm long) of C. elegans is useful as in vivo assays can be done in 96-welled plate but is experimentally challenging for biochemical assay due to the amount of material in single worm. The transparent body of
the worm is useful for non-invasive microscopic examination of the internal anatomy and development but the thick outer cuticle forms a barrier for certain hydrophilic chemicals limiting its use to monitor chemical toxicity (Pandey and Nichols, 2011; Avanesian, Semnani and Jafari, 2009). A big disadvantage of *C. elegans* is its evolutionary distance from humans with less conserved genes, absence of steroid hormones as well as steroid hormone nuclear receptors and the lack of metabolising enzymes like cytochrome P450s. Furthermore, almost one third of the organism is comprised of neurons with less defined organs and tissues.

**Other invertebrate models** – As recently highlighted by Bolker (2012), very few invertebrate models have been investigated, and in order to find the best model organism for endocrine function we may need to look beyond the traditional non-vertebrate model organisms (flies, worms) (Bolker, 2012). Apart from *Drosophila* and *C. elegans* other invertebrates like the slime mould, *Dictyostelium* have been used in research on epilepsy and bipolar disorder (Williams *et al.*, 2006). In toxicology, effects such as mutation can be conveniently detected using several strains of the bacterium *Salmonella typhimurium* in the Ames test. This test is used to identify almost all types of chemicals for mutagenicity (Mortelmans and Zeiger, 2000). The sea urchin has long been the test organism for basic developmental biology (McClay, 2011; Semenova, Kiselyov and Semenov, 2006; Anderson, Hose and Knezovich, 1994). Apart from all the research done on these organisms, there are no satisfactory invertebrate models for male reproductive endocrinology, but as the basic concept of sexual reproduction is similar in most animal phyla, it could be hypothesised that regulation of reproduction must be highly conserved between invertebrates and vertebrates. Lower vertebrates like zebra fish were not used in this study even though it is well-established model organism to study drug-induced developmental toxicity. Like *C. elegans*, the embryos of this fish are transparent allowing non-invasive observation of developmental changes but the difference in reproductive morphology such as lack of prostate glands makes them unsuitable for preliminary testing for androgenic compounds.

### 1.7 Molluscs as proposed test systems for reproductive endocrine disrupter testing

Recent research has shown that the effects of androgenic and anti-androgenic chemicals on reproduction and development in crustaceans, gastropod molluscs and echinoderms seem to be comparable to those in mammals. Of the three phyla, molluscs are the most
promising group because the sex steroids in molluscs seem to be chemically similar to vertebrates (Matthiessen, 2000). Gastropod molluscs have long been used as models of neuronal mechanisms of learning and memory (Glanzman, 2006), including the work of Eric Kandel who won a Nobel Prize in 2000 for his studies using Aplysia californica (Californian sea slug). Molluscs have also been suggested as models of autism (Hatfield, 2005). Although gastropods have been used as models of neuronal mechanisms, its use as a surrogate for testing reproductive toxicity can be suggested owing to the similarities between mollusc and mammalian reproductive system. Gastropod molluscs, in particular have a reproductive anatomy that, at least superficially, is very similar to that of vertebrates. Like mammals, the snail male reproductive tract extends from the testis (or ovotestis in hermaphrodite snails), has seminal vesicles, and a prostate gland (shown to be involved in stimulating sexual drive) (Adiyodi and Adiyodi, 1988) Figure 1.4 and terminates in a penile glandular complex (around the preputium) containing sebaceous glands and a penis. The eversion of the penis is controlled by a retractor muscle (analogous to the levator ani and bulbocavernous muscle in mammals) (discussed in more detail in Chapter 2).

There is evidence that the gonads of molluscs are able to synthesise estrogen and testosterone like molecules in the reproductive cycle (Lafont and Mathieu, 2007), known
to regulate vertebrate reproductive function (Bannister et al., 2007). Preliminary studies have shown that a molluscan testosterone may have a role in male reproductive tract development, but there is no evidence of whether this activity is dependent on an androgen receptor (Sternberg, 2007). The presence of androgen receptors has been suggested but not yet reported (Koehler et al., 2007). Biosynthetic pathway of the vertebrate steroids (involved in reproduction) uses 3β-hydroxy-Δ5-steroid dehydrogenase (3β-HSD) enzyme to convert pregnenolone to progesterone or dehydroepiandrosterone (DHEA) to androstenedione (Ad). 3β-HSD enzyme activity has been reported in different mollusc species (including, Crassostrea gigas (oyster), Octopus vulgaris (octopus) and Pecten yessoensis (scallop)) (Matsumoto et al., 1997; D’aniello et al., 1996; Krusch et al., 1979; Mori, Muramatsu and Nakamura, 1969). The evidence for the existence of 17β-hydroxysteroid dehydrogenase that catalyse the interconversion of Ad and testosterone (T), estrone and 17β-estradiol in vertebrates is very strong in molluscs (Lyssimachou, Ramón and Porte, 2009; Lyssimachou, Bachmann and Porte, 2008; Janer et al., 2005). Another enzyme important in metabolising T into 5α-dihyrdrotestosterone (DHT) in vertebrates has been identified in molluscs tissues (Lyssimachou, Bachmann and Porte, 2008; Janer et al., 2005; Ronis and Mason, 1996). Developmental exposure of gastropods to the exogenous androgen, Testosterone (T) or to the environmental androgen tributyltin (TBT) caused hypertrophy of the prostate gland and inhibition of spermatogenesis in males (Wang and Croll, 2004). In the female these exposures led to development of a penis and transformation of the albumen gland into a prostate gland (Watermann et al., 2008). This raises the possibility that the molluscan male reproductive tract is a functional analogue of the mammalian male reproductive tract. When exposed to mixture of androgens and anti-androgens the above mentioned effects seemed to be suppressed, indicating the effects to be hormonal dependent (Hagger et al., 2009; Hagger et al., 2006). Spooner et al. (1991) and Bettin et al. (1996) suggested that TBT could prevent the conversion of testosterone to E2 by inhibiting aromatase (Bettin, Oehlmann and Stroben, 1996; Spooner et al., 1991). Later on Gooding et al. (2003) and Le Blanc et al. (2005) suggested that TBT prevented esterification of testosterone, thus leading to free testosterone and the above effects (Leblanc, Gooding and Sternberg, 2005; Gooding et al., 2003).

Recently, retinoid X receptor has since been identified as being affected by TBT and leading to masculinisation in different species of mollusc (Thais clavigera (sea snail), Nucella lapillus (dog whelk) and Nassarius reticulatus (netted dog whelk)) (Pascoal et al., 2012; Sousa et al., 2010; Horiguchi et al., 2007). Targeted deletion of RXR-alpha in
mice led to prostatic growth, but not invasive carcinoma and metastasis (Tanaka and De Luca, 2009; Huang et al., 2002). There may be some similarity in the mode of action of development of prostate gland in molluscs and mammals. To understand the likelihood of the effects of chemicals on vertebrates and invertebrates, a basic understanding of their endocrine system is required.

1.8 The Gastropod neuro-endocrine system

The neuro-endocrine system in Gastropods is by far one of the best investigated and consists of neurohemal organs made up of neuro-secretory cells, nerve cells and endocrine glands (cerebral, pleural, pedal and abdominal) that produce neuro-peptides. These neuro-peptides are secreted to elicit appropriate response to the stimuli (external and physiological) and regulate processes such as development, growth and reproduction (Ketata et al., 2008). The nervous system with cerebroid ganglia and a ventral nerve chain has a central role in coordination along with endocrine glands as in vertebrates. Many hormonally active regions have been identified as integral parts of their neuro-endocrine system. The dorsal body present in the dorsal part of the cerebral ganglia produces dorsal body hormone (DBH), which is a female gonadotropic hormone stimulating vitellogenesis, oocyte maturation and development of female accessory organs. The caudo-dorsal cell hormone (CDCH) produced by caudo-dorsal cells (CDC) in the cerebral ganglia are involved in control of ovulation egg mass formation and egg laying behaviour in freshwater pulmonates (Fig. 1.5) (Geraerts, 1976a; Geraerts and Algera, 1976; Geraerts, WPM and Joosse, J, 1975). Moreover, an egg laying hormone (ELH) has been reported in the sea hare Aplysia californica (Garden et al., 1998; Mahon et al., 1985; Arch and Smock, 1977; Arch, 1976; Arch, Earley and Smock, 1976). The pleural ganglia have the dark green cell system, which generates a hormone structurally analogous to vertebrate thyroid-stimulating releasing hormone (TRH) having a diuretic effect (Pinder et al., 1999). The dark green cells are supported by yellow and yellow-green cells that produce peptides stimulating sodium influx in the body and thus control osmosis in the body (Pinder et al., 1999). Male copulatory behaviour in pulmonates is controlled by Ala-Pro-Gly-Trp-amide-related peptide (APGWamide) synthesised in the right arterial lobe (AL) of the cerebral ganglion that is connected to the penial complex in both Lymnaea and Aplysia (Koene, 2010; Fan et al., 1997; Croll and Van Minnen, 1992) (Figure 1.5). Although some peptide hormones are clearly specific to molluscs, there are few that seem to be functionally, or structurally, similar between vertebrates and molluscs. The lateral lobes in
molluscs generate gonadotropin-releasing hormone (GnRH) gene, a peptide hormone that plays a role in reproduction and on gamete differentiation in gonads in invertebrates (Gorbman and Sower, 2003; Rastogi et al., 2002). In vertebrates GnRH (via other peptide hormones) acts on steroidogenic cells in the gonads, and it is the steroid hormones that affect gametogenesis and reproductive behaviour (Hartenstein, 2006). Using anti-peptide antibodies, several studies have substantiated the presence of GnRH-like peptides in a wide range of molluscs including gastropods (*Helisoma trivolvis* (freshwater snail), *Lymnaea stagnalis* (great pond snail), * Loligo edulis* (European squid) and *A. californica* (Onitsuka et al., 2009; Zhang et al., 2008; Young, Chang and Goldberg, 1999; Goldberg et al., 1993), cephalopods (*O. vulgaris*) (Iwakoshi-Ukena et al., 2004; Di Cristo et al., 2002), bivalves (*Crassostrea gigas*, *Mytilus edulis* (blue mussel), *Patinopecten yessoensis* (Yesso scallop)) (Treen et al., 2012; Rodet et al., 2005) and the squid (*Loligo bleekeri*) (Amano et al., 2008). The light green cell system with the cerebral ganglia secretes molluscan insulin-like peptides (MIP). Five members (I, II, III, V and VII) of the MIP family have been characterised from the pulmonate snail *L. stagnalis* (Lardans et al., 2001; Smit et al., 1996), land snail *Otala lacteal* (Abdabra and Saleuddin, 2000) and in *C. gigas* (Hamano, Awaji and Usuki, 2005). Studies show that MIPs control body growth, development and metabolism (Gricourt et al., 2003; Lebel et al., 1996; Smit et al., 1996) and are excellent candidates for insulin signaling pathways in schistosoma-host relationships (Dissous et al., 2006). These peptides stimulate adenylcyclase and guanosine 5’ triphosphate binding activity of G-proteins similar to insulin and insulin related factor1 in vertebrates (McClellan-Green, 2013). Another vertebrate peptide hormone analogue found in molluscs is Conopressin belonging to vasopressin/oxytocin family (neurohypophyseal hormones) (Van Kesteren et al., 1995a). This family of peptides is of very ancient origin, found in organisms as diverse as worms and man (Beets et al., 2013; Hoyle, 1998) and is believed to play a role in modulation of male sexual organs and contribute to the control of complex behaviour such as copulation (Garrison et al., 2012; Van Soest and Kits, 2002).
In addition to the peptide hormones, recent research has been focused on the possibility that vertebrate type steroid hormones are functionally active and that they are endogenously produced in molluscs, although there is some controversy as to which steroids are naturally produced (Scott, 2012a). Vertebrate-like steroids (17β-estradiol, androgens and progesterone), have however, been measured in several species of molluscs and in numerous publications the potential for biosynthesis of vertebrate-type steroids has been demonstrated for molluscs (For example (Fernandes, Loi and Porte, 2011; Lafont and Mathieu, 2007; Janer and Porte, 2007; Osada, Tawarayama and Mori, 2004; Siah et al., 2002; Le Curieux-Belfond et al., 2001; Bose, Majumdar and Bhattacharya, 1997; D'aniello et al., 1996; Jong-Brink et al., 1981; De Longcamp, Lubet and Drosdowsky, 1974). Moreover, enzymes involved in steroidogenesis have also been found in gonad tissue (Osada, Tawarayama and Mori, 2004; Di Cosmo, Di Cristo and Paolucci, 2001; D'aniello et al., 1996). Although the functional role of sex steroids in molluscs is questionable, there is evidence suggesting that steroids have marked effects when administered to molluscs. In 1980, aquaculture operators used immersion of oyster spats in either methyltestosterone (MT) or 17β-estradiol to produce unisex clutches of oysters (Koehler et al., 2007). Research suggested that testosterone administration to castrated male slugs (*Euhadra prelionphala*) could stimulate the production of male secondary sex
characteristics (Koehler et al., 2007). Similarly, testosterone exposure of female gastropods (N. lapillus, N. reticulatus, Ilyanassa obsolete (Eastern mud snail) and Marisa cornuarietis (Giant rams horn snail) caused pseudohermaphroditism or imposex measured by vas deferens sequence index (an index to measure the development of vas deferens and blockage of oviduct) (Oehlmann et al., 1996) and penis length (Bettin, Oehlmann and Stroben, 1996b; Spooner et al., 1991). Developmental exposure of snails to the potent steroidal androgen, MT or to the environmental androgen tributyltin (TBT) caused hypertrophy of the prostate gland and inhibition of spermatogenesis in males and transformation of the female albumen and capsule gland to prostate tissue and penis development in females (Watermann et al., 2008; Oehlmann et al., 1996). The growth induced by TBT were blocked by co-administration of the anti-androgenic drug CPA (Hagger et al., 2009; Bettin, Oehlmann and Stroben, 1996) suggesting that the growths caused by TBT may be androgen-dependent as are many prostate tumours in humans and rodents. Overall there are many findings that suggest the existence of steroids similar to vertebrates but the primary issue is the precise mechanism in which these steroids exert endocrine control in molluscs, which is still not clear and require more research.

1.9 Conclusion

Modern societies produce thousands of chemicals some of which have the potential to interfere with the endocrine system of the organisms causing diseases. Declining health is of particular current concern, underlining the need for continued focus on the development of pharmaceuticals with which to treat these disorders and diseases. Moreover, scientific research on the relationship between the environment and health has highlighted concerns that these disorders may, in part, be caused by exposure to environmental pollutants. This has led to more stringent laws for screening of chemicals, devised by government bodies like EPA in the US and OECD in Europe. The screening process consists of a battery of tests designed to ensure steroid hormones such as estrogens and androgens do not mediate the mode of action of the chemicals. These tests included animal studies with a broad range of taxa such as fish, amphibians, rodents and a few other species of mammals. More tests meant more animals being used in research and all this led to finding alternatives to vertebrate animal testing for endocrine disruptors. These alternatives include in vitro assays, high throughput screening assays, use of invertebrate animals with short life cycle. My thesis discusses the potential of Biomphalaria glabrata, a mollusc as test organism for
androgenic chemical testing. Additionally, this new knowledge provides a broad evolutionary perspective on the molecular conservation of metabolic ad developmental pathways.

My research was carried out to test the following null hypothesis:

**Androgenic chemicals affect the reproductive tract development in Biomphalaria glabrata and the process is mediated through homologues of sex steroid receptors as reported in vertebrates.**

### 1.10 Aim of the thesis

The aim of my Ph.D. project was to test the hypothesis that common genes, proteins and physiological processes in both molluscs and humans underlie the response of the male reproductive system to androgens. In order to fulfill this aim the following objective are being addressed:

- Examine the physiological responses of androgenic chemicals (17α-Methyl testosterone and Dihydrotestosterone) on the development of reproductive tract in the snail (*Biomphalaria glabrata*).

- Investigate the fully sequenced genomes from *Biomphalaria glabrata* and *Lottia gigantea* to identify putative nuclear receptors from both the species of snails and cross-species comparison, looking at the similarities and differences that might exist.

- Identification and annotation of reproductive type neuropeptides in *Biomphalaria glabrata*. 
2 Chapter 2: Introduction to the test species, *Biomphalaria glabrata* and preliminary work
2.1 Introduction

The phylum Mollusca is one of the most diverse group of triploblastic protostomes (Brusca et al., 2005) ranking second to arthropoda with approximately 200000 extant marine, freshwater and land species (Ponder and Lindberg, 2008) that trace their origin to the earliest period of Cambrian at the commencement of the radiation of coelomate animals (animals with a true body cavity) (Brusca et al., 2005). Based on morpho-anatomical evidence and molecular data, the phylum is divided into seven or eight classes, Aplacophora (Solenogastres, Caudofoveata), Monoplacophora, Polyplacophora, Scaphopoda, Cephalopoda, Bivalve and Gastropoda (Stöger et al., 2013; Ponder and Lindberg, 2008). Of these, the focus of my thesis is on gastropods.

2.2 Gastropoda

The class Gastropoda emerged during the Ordovician period of important evolutionary differentiation within the molluscan phylum (Passamaneck, Schander and Halanych, 2004). If we consider the origin of life on the earth on a 24 hour clock with each second equivalent to 52,000 years then humans that diverged 2.6 million years ago, appear 54 seconds before the clock complete its 24 hours and the gastropods appear 500 million years ago, which is approximately 3 hours. Since their origin they have undergone dramatic radiations with over 100,000 extant species comprising about 80% of all mollusc species, forming one of the truly diverse classes of molluscs (Haszprunar and Wanninger, 2012). Based on the anatomical layout of the respiratory system of the gastropods, it has been classified into Opisthobranchia (gills behind the heart), Prosobranchia (gills in front of the heart) and Pulmonata (lungs instead of gills) (Brusca et al., 2005). The main focus of this thesis is a freshwater pulmonate (*Biomphalaria glabrata*), although a marine prosobranch (*Lottia gigantea*) was studied for comparative cross-species computational analyses. What follows is an overview of the physiology and anatomy of these mollusc groups followed by detailed description of the test species.

2.2.1 Prosobranch

Prosobranch are the most primitive of all gastropods and are mainly gonochoristic with two different organisations, the diotocardian and the more advanced monotocardian. In diotocardian, fertilization is external and it discharges by the opening of the functional right kidney, whereas the monotocardian practice internal fertilization and they discharge through a separate glandular pallial genital duct (Fretter, 1984). In addition to the normal
sexual reproduction, some species of prosobranch exhibit parthenogenesis for example *Potamopyrgus antipodarum* (Collin, 1995; Wallace, 1992) and some species show protandric sex reversal (e.g. *Crepidula fornicata*, (Collin, 1995); *Busycon carica*, (Castagna and Kraeuter, 1994); *Lottia gigantea*, (Lindberg and Wright, 1985). A male phase precedes the female phase and is linked by a hermaphrodite phase (Geraerts, 1976b).

The male reproductive tract in the diotocardian is simple, having a single gonad and the tubules of the gonad join a short testicular duct that opens into the right kidney and discharge sperm directly into the water Figure 2.1. In the more advanced monotocardian, it is comprised of a long and highly coiled testicular duct that runs along the visceral mass leading into a ciliated vas deferens, which opens into the pallial duct and the prostate (Fretter, 1984). Usually sperm is produced in the tubules and passes along the vas deferens and the prostate where nutritive secretions are added. The sperm is then discharged through the penis, situated on the head behind the tentacle (Fretter, 1984) (Figure 2.1). The female reproductive tract in diotocardian is similar to males and the gonad forms the protective covering for the eggs, which are shed singly into the water bodies. In monotocardian, there is a more elaborate female reproductive tract starting with an ovary that runs into the ovarian duct and the fertilization pouch. The fertilized embryos move from the fertilization pouch along the oviduct into the proximal albumen gland, where a nutritive secretion is molded around the eggs, and then into the distal capsule gland or the jelly gland that produces a viscous fluid in which the egg capsules are embedded. This provides an extra protection to the egg capsules from osmotic pressure. The egg capsules are then harbored in the pallial duct (modified into a brood pouch) until hatched and then discharged through the vaginal opening (Fretter, 1984).

*Figure 2.1: Reproductive tract of Diotocardian (A,B) female and male; Monotocardian (C,D) female and male.*
Thickened lines indicate boundary of mantle cavity. ag: albumen gland; b: bursa copulatrix; cp: capsule gland; od: ovarian duct; om: opening to mantle cavity; ov: ovary; rc: receptaculum seminis; rk: right kidney; rpc: renopericardial canal; te: testis; ug: urinogenital papilla (Adapted from (Fretter, 1984)).

2.2.1 Test species I- Lottia gigantea (for computational purposes)

The prosobranch snail *L. gigantea* was chosen as the first lophotrochozoan for whole genome sequencing to provide a foundation for elucidating the evolutionary and ecological success of the species and its lineage for conservation purposes (Simakov et al., 2012). The genome data was significant to define the core animal gene set, enabling the identification of many ancient animal genes that have apparently been lost in flies and nematodes but retained in molluscs (Chapman, 2007). The genome is small relative to other molluscs, consisting of about 500 mega base pairs ([http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html](http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html)). The genome organisation, gene structures and functional content of these species are more similar to invertebrate deuterostomes (sea urchin and amphioxus) than to the sequenced protostomes (flies, nematodes and planarians). These similarities reveal features that might have been lost or diverged in the protostomes studied so far (Simakov et al., 2012). *L. gigantea* has been used for computational analysis of the relevant genes in my thesis ([http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html](http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html)).

2.2.2 Pulmonates

The pulmonates include the large majority of land snails and slugs; freshwater snails and a few families of estuarine and shallow water marine snails and slugs (Ponder and Lindberg, 2008) and are comprised of two groups; the Stylommatophora (one pair of tentacles with eyes on the end of the tentacle) and the Basommatophora (two pairs of tentacles and eyes at the base of the posterior tentacle). The Stylommatophora are entirely terrestrial (land slugs and snails) and the Basommatophora are mainly freshwater snails inhabiting ponds, ditches, rivers and lakes (Geraerts et al., 1984). Basommatophora are hermaphrodite and show internal fertilization. Their reproductive system consists of a bisexual gonad (ovotestis), male and female ducts and accessory glands (albumen gland in female and prostate gland in male). Male and female gametes are produced in the ovotestis, which is located in the apical part of the visceral hump. When mature, both male and female gametes pass anteriorly into the hermaphrodite duct. With an exception of a few species, most of them have a short seminal vesicle along this duct (Geraerts et al., 1984) and then male and female gametes are transported via separate paths. The junction of male and female ducts and hermaphrodite duct is called carrefour. Both albumen gland and the
fertilization pocket also diverge at this point. At copulation the sperm passes through the carrefour and are transported to the male duct. This region also receives sperm from cross-fertilization, which fertilize in the fertilization pouch and secretions from the albumen gland cover the oocytes preventing further sperm entrance (Egonmwan, 2007). The female duct is mainly glandular and secretes nutritive and packaging materials to support and protect the developing embryos (Heller, 1993). The female duct receives and often stores the sperms from a copulative partner to fertilize the eggs until the stores are exhausted (Geraerts, 1976b). The fertilized egg receives yolk from the albumen gland and passes into the uterus where tertiary membranes are added to the zygote before it moves to the oviduct and is extruded through the muscular vagina (Egonmwan, 2007). When the albumen gland stops secreting fluid, the ovulated oocytes that remain in the hermaphrodite duct and carrefour are transported to the bursa copulatrix where they are hydrolyzed and reabsorbed (Boyle and Yoshino, 2000).

The male duct is muscular and lined by ciliated epithelium helping to transport the sperm and to transfer it to its partner during mating. A prostate gland is usually present but there is a wide range of species-specific variation in structure. The prostate gland in higher Basommatophora is an elaborate and glandular with a number of diverticula (De Jong-Brink, 1969) associated with vas deferens, but in lower Basommatophora the prostate gland is just an elongated part of the male duct (Geraerts et al., 1984). The male copulatory organ consists of penis transversed by the vas deferens and enclosed in the penis sheath, that has a terminal opening through which the penis can be everted. This is the general anatomical plan of the reproductive tract of all pulmonates, although a wide range of species-specific variation in these structures occurs (Geraerts et al., 1984).

As evident from the above discussion the pulmonates differ in an important respect from prosobranch snails. Of the three extant subclasses, Prosobranchia diverged about 400 million years ago and underwent strong radiation, occupying most diverse habitats and different reproductive mechanisms (bisexual, hermaphrodite, parthenogenic) (Ram et al., 1998) compared to pulmonates that are all hermaphrodite (Adiyodi and Adiyodi, 1988).

Development in prosobranch is planktotrophic where they spend days or weeks in the plankton stage depending on the species. As they age, the larvae spend more time at greater depths and under favourable conditions undergo metamorphosis (Fretter and Shale, 1973). Studies have shown that chemotactic and thigmotactic responses initiate settlement, which may be induced by organic properties of the substratum beneficial to the adult (Struhsaker and Costlow, 1968). Conversely, pulmonates undergo direct development in eggs passing through morula, blastula, gastrula, trochophore and veliger stages (Little et
hatching out as miniature versions of adult snails. The rate of development is dependent on abiotic factors like temperature and salinity (Geraerts et al., 1984). Unlike prosobranch, there is no evidence of pulmonates showing chemotactic and thermotactic responses but the Basommatophora are known to have a neuroendocrine center that responds to external and internal stimuli through various hormones and neuropeptides (Lever and Boer, 1983). Endocrine control of reproduction in prosobranch is achieved by first order neurosecretory mechanisms through cerebral ganglia, pedal ganglia, cerebropleural connective areas controlling non-endocrine target organs, but in pulmonates, both first and second order neurosecretory mechanisms occur; that is neurosecretory cells control endocrine and non-endocrine target organs (discussed in detail in Chapter 1).

2.2.2.1 Test species 2-Biomphalaria glabrata

The most extensively studied order of pulmonates is Basommatophora (basal eye-bearing stalk). More specifically, Biomphalaria in the genus Planorbidae has become a focus for research due to the integral role it plays in the transmission of human intestinal schistosomiasis. The tropical freshwater snail Biomphalaria glabrata is an intermediate host for several digenean trematode parasitic worms, including Schistosoma mansoni, the causative agent of schistosomiasis. This is the most widespread trematode infection affecting around 200 million people, leading to a chronic debilitating disease and up to 200,000 deaths per year, across 75 developing countries (WHO report, 2011). Schistosomiasis is a disease acquired by people when they come in contact with freshwater infested with larval form (cercariae) of Schistosoma mansoni. B. glabrata is the intermediate host of S. mansoni that is infected by free-swimming larvae (miracidium) of S. mansoni (Bayne, 2009a). For all the above reasons B. glabrata has been a focus of research in the field of taxonomy, morphology, cytology, physiology, ecology, genetics, and reproduction (Raghavan and Knight, 2006; Camey T and Verdok NH, 1970).

My test species, Biomphalaria glabrata, strain BB02, has been widely studied as a model intermediate host for Schistosoma mansoni since the mid-1950s when it was found that susceptibility in B. glabrata to S. mansoni is a heritable trait (Ittiprasert and Knight, 2012; Richards and Shade, 1987; Newton WL, ). Laboratories across the world are actively studying the molecular biology of B. glabrata trying to identify genes and their products that interfere with parasite survival in the snail and may reduce transmission of schistosomiasis to humans e.g. (Ittiprasert and Knight, 2012; Bayne, 2009a; Raghavan et al., 2003). From a biological perspective, B. glabrata possesses some unique properties. It
is a tropical, aquatic, hermaphrodite organism and has been used as a research organism in comparative immunology (Rollinson, 2011; Loker, 2010), evolutionary systematics, biogeography (DeJong et al., 2001) and environmental toxicological studies (Ansaldo et al., 2009; Salice, Miller and Roesijadi, 2009; Ansaldo et al., 2006). It can be easily maintained under laboratory conditions and its natural diet consists of aquatic plants and the cyanobacterium, nostoc (found in soil and mud) (Toledo and Fried, 2011; Eveland and Haseeb, 2011) and in the laboratory can be fed dried lettuce or dried fish food.

2.2.2.1 General Anatomy

The shell of *B. glabrata* is composed of calcium carbonate embedded in small amount of organic matrix secreted by the mantle. The mantle is a line of skin, which transcends the shell and is divided into three chambers (pulmonary, water inflow and water outflow) that help in flotation, oxygen reservoir, excreta circulation and elimination (Toledo and Fried, 2011; Jurberg, Cunha and Rodrigues, 1997). Inner layers of the outer mantle fold secrete the outermost organic layer of the shell, the periostracum, during embryogenesis (Hasse et al., 2000). The *Biomphalaria* genus (as well as all gastropods) undergoes a process called torsion and the shell undergoes counter clockwise spiral (known as sinistral coiling). As the snail ages, the shell slowly increases in size from the apex (specifically the pruntoconch) towards the aperture (Marxen et al., 2003). As a consequence, the mantle cavity (as well as the anus) is prostrated over the head-foot Figure 2.2. As the name suggests, the head-foot functions both, as a sensory organ (which has protruding tentacles, mouth and feeding structures) and provides locomotion. The heart is located in a pericardial chamber and contains a distinctive atria and ventricle.

![Figure 2.2: Figure showing anatomical parts of the Biomphalaria glabrata.](image-url)
Although the anatomical features involved in food ingestion by *B. glabrata* are basically similar to other pulmonates, there are some unique features. The adult snail has a total of about 4425 radula teeth arranged in 75 rows well adapted to feed upon diatoms from the surface of freshwater macrophytes and the area grazed can be regulated by the change in direction and velocity of head (Thomas, Nwanko and Sterry, 1985).

### 2.2.2.1.2 Life cycle and reproduction

*B. glabrata* snails start their reproductive cycle as protandric males and then change to simultaneous hermaphrodites (Trigwell and Dussart, 1998; Vernon and Taylor, 1996) with the ability to both self-fertilize (when isolated) and cross fertilize, although they prefer to cross fertilize (Vianey-Liaud, Dominique and Dussart G, 1996; Jarne, Vianey-Liaud and Delay, 1993). They can exchange bundles of sperm to avoid self-fertilization.

The reproductive tract of *B. glabrata* is subdivided into three parts (hermaphrodite, male and female), each including several structures (accessory gland, ducts, etc. Figure 2.3). Both male and female gametes are produced within the hermaphrodite gonad called the ovotestis, situated in the inner whorls of the shell and dorso-laterally covered by the orangish digestive gland (hepatopancreas). The collecting canal of the ovotestis carries male and female gametes into a hermaphrodite duct, which leads into seminal vesicles. The spermatozoa are stored in seminal vesicles until maturation, whereas oocytes proceed anteriorly to the receptaculum seminis, where fertilization occurs. The carrefour is present at the junction of the hermaphrodite duct and the receptaculum seminis and its main role is to receive sperms from copulating partner and transfer it to fertilization pouch.
Figure 2.3: Figure showing the schematic of reproductive tract in B. glabrata.
Starting with ovotestis leading into the hermaphrodite duct to the carrefour where fertilization occurs and the male and female ducts get separated. The female accessory gland is shown that leads into the glandular complex comprising of both male (prostate diverticulum and spermduct) and female accessory glands (mucous gland and membrane gland) closely intertwined into each other.

(Egonmwan, 2007). Albumen gland is one of the female accessory reproductive glands, is a large tubulous gland that pours its secretions into the Carrefour. These secretions (galactogen – a polysaccharide) form a part of the perivitelline fluid that is used as a nutrient by the developing embryos in their eggs and after hatching (Mukai et al., 2004; Ramasubramaniam, 1979). The first part of the female duct, the oviduct, has a tube like appearance, continuing into a swollen convoluted part, the muciparous gland. This gland is not distinctly separate but is closely associated to the oothecal gland, the uterus, and the bursa copulatrix, which are connected with the vagina by a small duct. The external orifice of vagina opens a short distance behind the male genital outlet. The male part starts with sperm duct and runs parallel to the oviduct, curving along the muciparous gland. Between the transitions of the muciparous gland, oothecal gland and bursa copulatrix, the sperm duct has numerous branched and unbranched diverticula forming the prostate gland (De Jong-Brink, 1969). The foremost diverticulum is usually inserted between the oothecal gland and the muciparous gland, but in some instances it partially covers the latter part of the prostate diverticula. This runs into the vas deferens that continues into the penial
complex, consisting of penis sheath and prepuce. The penis sheath is uniformly cylindrical enclosing the penis consisting of well developed erectile tissue surrounding the penis canal and enveloped by a double muscular coat. The vas deferens and prostate help in conveying the semen and also aids in propelling the penis through the prepuce during copulation (De Jong-Brink, 1969). The main function of the male duct is to transport sperm to the copulative partner.

Reproductive behaviour in this species has been divided into four phases by (Trigwell, Dussart and Vianey-Liaud, 1997) and starts with precourtship during which an active searching of mate takes place and the most mobile snail acts as a male. This is slightly different from the pond snail, Limnaea stagnalis in which the period of isolation is the determining factor for the male actor (Van Duivenboden and Maat, 1985). This phase is followed by courtship, copulation and post copulation stages. Copulation involves penetration of female genital pore by the penis of the partner followed by insemination, when the sperm is introduced into the female genital tract and fertilizes the egg internally producing embryos. This could be followed by a second copulation, which is not necessarily with the same individual except in isolated pairs (Vernon and Taylor, 1996), or copulation could end with both the snails moving apart. Ovulation and oviposition follows thereafter, laying capsules with an average of 40-60 eggs and this continues until the allosperms stored in the female tract are exhausted (Vianey-Liaud et al., 1989). There is a typical 4-6 week egg-egg generation time. Since the generation time is approximately 5 weeks, the snail can undergo several generations in a year and under ideal conditions one snail can produce 14,000 eggs in its life span (Toledo and Fried, 2011).

2.2.2.1.3 Molecular tools to study the biology of B. glabrata

Biomphalaria is probably the most widely studied mollusc as a model intermediate host for the human parasite Schistosoma mansoni with more than 50 years of research invested in its genetics and physiology. There are several genomic tools available for B. glabrata that make this species a potential model for various purposes.

2.2.2.1.3.1 Bacterial artificial chromosome (BAC) Library

A BAC is an engineered DNA molecule used to clone large DNA segments (100,000-300,000bps) from the organism, say B. glabrata, by inserting these segments into the bacterial cells (for example, E. coli) (Shizuya et al., 1992; O'Connor, Peifer and Bender, 1989). As the bacterial cells grow and divide, they amplify the BAC DNA, which can then be isolated and used in sequencing DNA. The sequenced parts are rearranged in silico, resulting in the genomic sequence of the organism. BAC libraries provide a platform for
the creation of physical maps and complete genome sequencing, elucidation of gene function and regulation, map-based cloning of quantitative trait loci used in modeling genetic diseases, and comparisons of specific regions between different species or ecotypes, including establishment of synteny and functional single nucleotide polymorphism (SNP) analyses in coding and regulatory sequences (Farrar and Donnison, 2007; Shizuya and Kouros-Mehr, 2001). Within molluscs, BAC libraries have been constructed from bivalves, Eastern oyster and Pacific oyster (Cunningham et al., 2006); Scallop (Chlamys farreri) (Scheuring, 2008) and Aplysia californica as a prerequisite for sequencing. *B. glabrata* snails were collected from Brazil, South America and the genomic DNA from this strain (BB02 (*Biomphalaria* from Barreiro caught in 2002)) was used for the production of BAC library (Adema et al., 2006a). A BAC library from BB02 provided collectively a total of 61824 clones with an average insert size of 136.3kb and provides 9.05X coverage of the 931Mb genome. Odoemelam and colleagues (2009) developed protocol to perform physical mapping of BAC onto chromosome by Fluorescence in-situ hybridization (FISH) for *B. glabrata* (Odoemelam et al., 2009).

### 2.2.2.1.3.2 Microarrays

DNA microarray technology is used to examine the expression of thousands of genes at once (Schena et al., 1995) in a hybridisation-based assay. In microarray experiments, mRNA from one target is made into cDNA labelled with red fluorescent dye, and the other target is labelled with green fluorescent dye. The two cDNA samples are combined into a single solution and hybridized to the microarray slide, which is a collection of single stranded DNA spots, each representing a gene sequence. The cDNA binds to the DNA spots and the colour of the fluorescence indicates the gene expression. When both red and green fluorescence exist on the same dot it appears to be yellow (Hoopes, 2008; Heller, 2002). The application of cDNA microarray technology has brought about a significant change in the breadth of hypotheses that can be explored. For example, transcriptional response patterns can generate a better understanding of mechanisms of disease, identifying disease sub phenotypes or predicting disease progression in different organisms. The first application of microarray technology in *B. glabrata* identified 98 differentially expressed genes between snail strains resistant or susceptible to *S. mansoni*, 94 resistance-associated and 4 susceptible-associated (Lockyer et al., 2008). A 70-mer oligonucleotide-based microarray approach has also been used in *B. glabrata* to investigate defense response of the snails depending on the nature of stimulus (wounding, exposure to gram +ve bacteria (*Micrococcus luteus*), gram –ve bacteria (*Escherichia coli*).
or infection with trematode parasites (*Schistosoma mansoni* and *Echinostoma paraensei*). Wounding yielded a modest response in the differential expression and it was different from the other infections. Distinct differential expression was obtained of stress/defense related features by bacterial infection but significantly altered expression of immune related genes was obtained from the trematodes and the immunosuppressive interference of *E. paraensei* is stronger than *S. mansoni* (Adema and Loker, 2009). The application of these types of approaches are a step forward in understanding the molecular interplay between parasite and host and the availability of these resources make *B glabrata* a potential model organism for other purposes.

### 2.2.2.1.3.3 RNA interference

RNA interference (RNAi) is a process of introducing an exogenous single-stranded (siRNA) or double-stranded (dsRNA) RNA corresponding to a mRNA sequence and through specific binding, the RNA sequence coding for the target gene is depleted resulting in significant reduction in the levels of target mRNA (Campbell and Choy, 2005; Scherer and Rossi, 2003; Sharp, 2001; Fire et al., 1998). The effects of this reduction can be assessed by its phenotype change, which could indicate the biological role of the gene. Since RNAi may not totally abolish expression of the gene, this technique is sometimes referred as a “knockdown”, to distinguish it from “knockout” procedures where the gene is completely erased from the genome through homologous recombination (exchange of nucleotide sequences between non-sister chromatids). The first reported RNAi knockdown for a pulmonate snail was in *L. stagnalis*, which was injected with dsRNA corresponding to the transcript for nitric oxide synthase that led to the reduction in expression of this gene, affecting the feeding behaviour of treated snails (Korneev et al., 2002). In *B glabrata* RNAi mediated knockdown of the fibrinogen-related protein (FREP2) gene involved in molluscan internal defense was successful (Jiang, Loker and Zhang, 2006). Recently a non-invasive technique of post-transcription gene silencing using cationic polymer polyethyleneimine (PEI) mediated delivery of long double-stranded and small interfering RNA (soaking the snails in dsRNA/PEI or siRNA/PEI) was tried in juvenile snails. The technique was successful and led to knockdown in expression for the gene peroxiredoxin at nucleotide and protein level (Knight et al., 2011).

### 2.2.2.1.3.4 Biomphalaria glabrata genome

The *B. glabrata* genome initiative is an effort by an international consortium to identify comparable sequence information from the snail host to that has already been achieved for *S. mansoni* (parasite) (Berriman et al., 2009) and human (definitive host) genomes.
(Lander et al., 2001; Venter et al., 2001), which may provide predictors for epidemiological studies on parasite transmission. With the advent of modern genome biology, genomic information plays a significant role in the development of treatment and control measures for such widespread diseases, as emphasised by the recent publication of the complete genome sequences of the malaria parasite (*Plasmodium falciparum*) and its mosquito vector (*Anopheles gambiae*) (Gardner et al., 2002; Hoffman et al., 2002; Holt et al., 2002). The genome sequencing project for *B. glabrata* was initiated in 2005 at the Washington University Genome Sequencing Centre (WUGS) (Raghavan and Knight, 2006). The first large scale set of genomic sequence data employing whole genome shotgun sequencing was contributed by WUGS and the expressed sequence tags (ESTs) derived either from whole snail tissue or different tissues. The genome of *B. glabrata*, using BB02 (susceptible) strain from Brazil, is composed of 2n = 36 small chromosomes (Toledo and Fried, 2011; Burch, 1960) with a genome size of ~ 950Mb with high heterozygosity and is AT rich (74.6%) (Ittiprasert and Knight, 2012). All the background information and the availability of the draft genome make *B. glabrata* an appropriate test species for my project. The fact that it can be easily bred in the laboratory with a generation time of approximately two months is helpful in terms of undertaking developmental studies in a short time span that is one of the main objectives of the project. *L. gigantea* genome will be used for cross-species comparison of the relevant genes of interest.

2.2.2.2 Preliminary work done on test species (B. glabrata) and optimization of techniques

Before commencing the main experimental study, a number of preliminary experiments and procedures were conducted with *B. glabrata* to inform the main experimental design. These were primarily divided into the following subsections,

- Determination of the likely critical window for androgen/anti-androgen exposure in immature/developing snails (the time period between hatching and onset of puberty). Identification and description of the reproductive tract/accessory sex organs and optimization of dissection and histopathological techniques.
- Optimisation of the protocol for gonadectomy, including the success rate; survival following gonadectomy and the utility of the gonadectomised snail model as a part of this project.
- Pilot study exposing gonadectomised snails to one of the androgenic chemical, 5α-dihydrotestosterone to be used in the exposure study.
2.2.2.2.1 Sexual development in *B. glabrata* from day of hatch to sexual maturity

In *B. glabrata* there is literature related to the embryogeny of development (Vasta, LeSage and Fried, 2011; De Jong-Brink, 1969; Boyce, Tieze-Dagevos and Larman, 1967) but not much information is available on time-specific development, which would be helpful to identify the time of gonadal differentiation and the onset of puberty. This information was needed before the androgen exposure study could be initiated.

**Materials and methods**

**Biomphalaria glabrata** acquisition

An initial stock of *B. glabrata* (BB02 strain) was kindly supplied by the Natural History Museum, London, UK in 2010. Thereafter a captive-bred stock has been maintained at our facility at Brunel University, London (UK). For the preliminary developmental study, one group of 10 adult snails (diameter>12mm) from same breeding stock was selected.

**General laboratory culture conditions**

During non-experimental periods, adult and juvenile snails were housed in a recirculation system in glass aquaria supplied by de-chlorinated tap water. Snails were fed Tetramin fish food flakes (roughly ~10% of their body mass) *ad libitum* three times a week. Any uneaten flakes or rotting vegetation and dead snails were siphoned out prior to feeding. Ammonia (0mg/l), nitrate (0mg/l), nitrite (<40mg/l) (API aquarium test strips) and dissolved oxygen levels (7 ± 0.5 mg/L) using Haan Lange probe were monitored in the water system 3 times a week, as was pH (7 ± 0.21) and general water hardness (120mg/l). Low pH and carbonate levels were adjusted by the addition of commercial aquaculture re-mineralizer (Remin) to sustain snail growth and prevent shell thinning (White *et al.*, 2007).

**Temperature and photoperiod**

The snail room temperature was maintained at 27⁰C by the use of an internal air-conditioning unit and similar temperature was maintained in the snail tanks. The photoperiod for all work conducted in the snail room was 16:8 Light:Dark incorporating 20 min dawn/dusk transition periods and this was controlled externally by a digital timer switch.

**Routine collection of egg masses**

Egg masses were easily observed sticking to the glass or pipe-work in the aquaria. Egg
masses laid on flat surfaces were removed using a sterile scalpel pressed against the edge of the egg mass taking care not to damage the eggs inside the egg mass. If the egg mass was laid on pipe-work or in a corner, it could be easily taken off using thumbnail in similar manner. Once removed, egg masses were placed, 2-3 egg masses per basket, in metal mesh baskets, which were suspended using glass rods in the tanks to keep the egg masses at optimal temperature and provide sufficient mobility space for the juveniles.

Biomphalaria glabrata rearing protocol

Under these optimum conditions *B. glabrata* eggs start hatching 5 days post oviposition and hatched snails were fed finely crushed Tetramin fish food flakes (approximately 0.05gms). Most of the snails hatched 6-7 days post-oviposition and hatching completed by day 11 post-oviposition. The hatchlings were left in the baskets, fed *ad libitum* three times a week until big enough not to escape the tanks. Once released into the tanks at the density of 100-150 hatchlings per tank with working volume of 3.7L, they were fed @ 0.5 gm/tank/alternate-day and this was increased to 3.0 gm/tank/alternate-day as the juvenile snail’s age. Under these conditions they attain reproductive age by 30 days post oviposition and they lay, on average, 1-2 egg masses a day (containing 20-50 eggs per mass) until six months of age, after which their reproductive fitness steadily declines.

For experimental purposes, egg masses were taken from adult snails of 2-3 months of age to maintain ideal hatching and developmental conditions. The snails were monitored for two weeks to get an estimate of egg masses laid per day. The snail tanks were cleaned one day before the beginning of this study and any egg masses present or uneaten flakes or rotting vegetation and dead snails were siphoned out.

Husbandry of hatchlings and juvenile snails

Four egg masses laid on the same day were collected and one egg mass was placed in each of the four wells of a six-well plate (nunc, sterile non-treated), filled with tank water, and floated in the snail flow through recirculation system maintained at optimum conditions mentioned above. A 50% water change was conducted every 48 hours and the wells were refilled with the tank water. They were closely monitored for any hatching. Hatching started on day 5 post-oviposition with maximum hatching in the 6-7 days post-oviposition. When hatching started egg masses containing unhatched embryos were moved to a new well and the egg mass were washed off using a small jet of water from a pipette to remove any newly hatched snails not yet left the surface of the egg mass. The well containing the newly hatched snails was re-labeled specifying their date of hatch. This process was repeated until all the snails had hatched. Hatchlings were fed on finely crushed Tetramin
fish food flakes twice a week. At 10 days post hatch, the hatchlings were taken out of the
six well plates using a small 1ml plastic pipette with the nozzle diagonally cut and
transferred into a basket, suspended in the tanks using glass rods. The hatchlings were
kept at a density of ~30-40 snails per basket but due to natural mortality or continual
sacrificing for histology, the density was not constant. Cleaning and feeding continued
every 48 hours with the feed increasing @ 0.5gm/tank/day, after 15 days post-hatch (dph).
These juvenile snails were kept in the baskets in the tanks till 60 days post hatch.

**Sampling of the hatchlings and juvenile snails**

Hatchlings and juvenile snails were sacrificed and fixed for histology at 3dph and then on
5, 7, 9, 11, 13, 15, 19, 21, 23, 27, 30, 34, 42, 49, 56 and 60dph. Snails were narcotized in
5% w/v Magnesium Chloride hexahydrate (MgCl₂ 6H₂O) (≥99%, Sigma-Aldrich) solution
(using purite water) for 30 minutes, or until the snails showed no mobility, which ever was
the longer. Once the snails were fully narcotized, the hatchlings from 3dph – 15dph were
fixed whole (with shells) in Kahel’s fixative (Frase, 1986) and kept at room temperature in
7ml glass vials until further histological processing. The bottles were labeled with the
species name, age in dph and the sampling date.

**Histopathology processing and analyses**

Prior to processing, the hatchlings (3dph -9dph) were placed in biopsy cassettes
(Histosette II) and to prevent any loss of sample, the tissue was sandwiched between two
Whatman filter papers dipped in 70% IMS. Each specimen was designated its own biopsy
cassette and labelled with the individual snail reference number. The fixed tissues were
serially dehydrated in an automatic tissue processor (TP 1020, Leica Inc.) passing the
samples through a series of increasing alcohol concentrations and cleared with Histoclear
II (National diagnostics) before being vacuum impregnated with wax and subsequently
embedded in wax blocks taking care that the tissue was not too much folded, especially
the ovotestis Table 2.1. Tissue samples were then sectioned on a rotary microtome (Lieca
RM2235) using disposable microtome blades (Shanndon MB35 premier 35°/ 80 mm), at a
thickness of 5µm. All sections were mounted on glass polysine treated microscopic slides
(VWR) and each slide was labelled with the individual snail reference number and slide
number on the opaque end. They were dried overnight (minimum 12 hours) on warming
racks (@45°C photoxdishwarmer) before staining with Mayer’s haematoxylin (VWR) and
1% aqueous eosin (VWR) according to the protocol in Table 2.2. After staining the slides
were left to dry for a few minutes in the fume cupboard to allow Histoclear II (RA Lamb)
to evaporate. A droplet of histomount was placed on the sections of tissue and a coverslip
(22 x50 mm, No. 1 thickness, RA Lamb) was mounted onto the slide. Care was taken to avoid any bubbles formation on or around the tissue sections and any excess of histomount on the slide was removed using tissue paper dipped in histoclear. The slides were dried in the fume cupboard for at least 24 hours and then were examined under an Olympus (BX51) light microscope using x2, x10 and x40 objectives (total magnification x20, x100 and x400 respectively). For each tissue, the cell types and structure and any abnormalities were assessed. Photomicrographs were taken using a digital camera (Q Imaging Micropublisher 5.0RTV) linked to a personal computer. Q Capture Pro 5.1 software was used to capture and view images from the digital camera.

<table>
<thead>
<tr>
<th>CYCLE NUMBER</th>
<th>TREATMENT</th>
<th>HOURS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70% IMS</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>90% IMS</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>95% IMS</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>100% IMS</td>
<td>1.5</td>
</tr>
<tr>
<td>5</td>
<td>100% IMS</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>100% IMS</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>100% IMS</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>HISTOCLEAR</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>HISTOCLEAR</td>
<td>1.5</td>
</tr>
<tr>
<td>10</td>
<td>HISTOCLEAR</td>
<td>1.5</td>
</tr>
<tr>
<td>11</td>
<td>WAX</td>
<td>1.25</td>
</tr>
<tr>
<td>12</td>
<td>WAX</td>
<td>1.25</td>
</tr>
</tbody>
</table>

**Table 2.1: Summary of the tissue processing method of the Leica Tissue Processor, indicating the various solutions used and time taken for each step.**

**Table 2.2: The staining protocol for histological slides of various tissues from B. glabrata.**

<table>
<thead>
<tr>
<th>Step</th>
<th>Bath</th>
<th>Chemical</th>
<th>In bath time (secs)</th>
<th>In bath agitation</th>
<th>Out bath time (secs)</th>
<th>Out bath agitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Histoclear</td>
<td>900</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>100% IMS</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>90% IMS</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>70% IMS</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Water</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Haematoxylin</td>
<td>600</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>Water</td>
<td>600</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>Acid alcohol</td>
<td>20</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Water</td>
<td>20</td>
<td>Y</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>Sat. Li2CO3</td>
<td>20</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>Water</td>
<td>20</td>
<td>Y</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>Eosin</td>
<td>40</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>Water</td>
<td>300</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>70% IMS</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>90% IMS</td>
<td>120</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>100% IMS</td>
<td>300</td>
<td>N</td>
<td>8</td>
<td>Y</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>Histoclear</td>
<td>300</td>
<td>N</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>Histoclear</td>
<td>0</td>
<td>N</td>
<td>0</td>
<td>N</td>
</tr>
</tbody>
</table>

**Results and Discussion**

*B. glabrata* undergoes direct development and when hatched the snails look like
miniatures of an adult snail with a transparent shell that is separated from the head region by the mantle fold, which is depressed to form the mantle cavity. The relative position of the digestive gland and gut were established at 3dph. The first evidence of the hepatopancreas could be seen at 6dph while the kidney appeared on the 9th day after hatch. The glandular complex comprising of prostate, membrane gland and mucous gland could be seen in primary stages of development at 9dph. No gonad was observed in *B. glabrata* hatchlings from the day of hatch until 12dph Figure 2.4(A). There is evidence that juvenile snails (less than 8mm in diameter) invest first in the development of ovotestis and male organs while development of the albumen gland and female organs starts later in the sub adult stage (when they are greater than 8mm in diameter) (Koene and Ter Maat, 2007; Trigwell and Dussart, 1998). The genital rudiment had extended its full length posteriorly showing early stages of gonadal acini formation from 14dph to 16dph Figure 2.4(B). Mature gonads with large numbers of sertoli cells were seen at 20dph Figure 2.4(C) and full spermatogenesis with spermatozoa and oocytes in the lumen was seen at 28dph Figure 2.4(D). The albumen gland and glandular complex were fully developed on the 22dph.

![Figure 2.4: Pictomicrograph of B. glabrata hatchlings showing the gonad at different stages of development.](image_url)

(A) 9dph with no gonad visible (x100); (B) 14dph: developing acini first become apparent (x100); (C) 20 dph showing a number of acini with sertoli cells, spermatogonium, spermatocytes and early stage oocytes (x100); (D) 24dph acini have full spermatogenesis and oogenesis to stage 3 (x100). 5μm sections stained with H & E. digestive tissue – dig; foot – f; gut – g; kidney – k; oocyte – O; shell – S; Sertoli cells – ser; spermatocytes – spc; spermatids – spd; spermatogonium – spg; spermatozoa – spz; tentacle – t.
Overall the preliminary experimental results indicate that the time period between hatching and onset of puberty is very short, between 12dph to 24 dph, which only provides a short opportunity to evaluate effects of the androgens on the developing reproductive system before natural puberty begins. In order to avoid any confounding factors that might affect the response or missing this critical period, it was decided to conduct androgen exposures from the egg mass stage through to sexual maturity (approximately 30 days post hatch (dph)).

2.2.2.2 Optimization of gonadectomy technique

In the rodent Hershberger assay, castration (Moon et al., 2009) is usually practiced in the adult version of the assay as it removes any endogenous sources of androgen and improves the response rate. Procedures for castration of adult snails have been described in the older scientific literature (Vianey-Liaud and Dussart, 2002; Miksys and Saleuddin, 1987) where castration of hermaphroditic pulmonate snails has been reported to cause acceleration in the growth of female accessory sex organs (albumen gland) (Miksys and Saleuddin, 1987; Geraerts and Mohamed, 1981; De Jong-Brink et al., 1979) and the size of the male accessory sex organ (prostate organ) remained unaffected (Harry, 1965). In contrast, Harry (1965) reported that castration of juvenile B. glabrata prevented the development of accessory sex organs and concluded that the growth of the accessory sex organs is dependent on the presence of the gonad and thus on hormones (Harry, 1965). Geraerts and Josse (1975) also reported that two blood-borne hormones were responsible for the development of prostate gland and female glands respectively (Geraerts, WPM and Joosse, J, 1975).

The hypothesis was that the gonadecotomised adult snail could serve as a useful model with which to explore whether steroidal substances produced by the gonads are necessary for development and maintenance of the snail prostate gland as is the case in mammals. If so, removing the source of endogenous hormones could produce a snail equivalent of the castrated rodent, which could be used to look at the response of the snail to androgens.

Material and Method

Procedure for gonadectomy

The snails used in this study were sexually mature snails from the same breeding stock maintained at our facility at Brunel University, London (UK). Snails having shell diameter
within the range of between 10-12mm (measured across the widest portion of the shell) were selected for the procedure (Dussart G, personal communication). Snails were removed from their tanks, blotted dry, weighed (g) using an electronic balance and their shell diameter measured to the nearest 0.1mm using hand-held digital calipers. Castration was carried out on ten snails at one time, by gently breaking open the apex of the central whorl of the shell on the right side of the posterior body wall with fine forceps, thereby exposing the hepatopancreas and the ovotestis Figure 2.5. The ovotestis can be observed as a narrowing pink or orange spotty structure at the very posterior of the animal; it is often curled (much like the shell) while the digestive tissue is much larger and generally brown in colour, dorso-laterally covering the ovotestis. The membrane overlying the ovotestis was cut using a pair of dissecting scissors with fine point open shanks curved (Fishers Scientific, UK) and with fine forceps along the length of the curvature and stripped away on either side laying bare the ovotestis. The hermaphrodite duct was ligatured using synthetic absorbable sterile surgical sutures (Vicryl), and the ovotestis was gently aspirated using plastic pipette not exerting too much force, taking care not to damage the hepatopancreas. The shell opening was then sealed with autoclaved shell pieces of dead snails with sterile veterinary tissue adhesive (Vetloc). After surgery, animals were allowed to recover in standing water containing an antibiotic gentamycin (Sigma-Aldrich) at a final concentration of 40µg/ml and 0.25µg/ml Amphotericin B (fungizone) (Sigma-Aldrich) for 1-2 hours. The gonadectomised snails were transferred in newly assigned tanks in the flow through system. Ten sham-operated controls were treated the similar manner as the gonadectomised snails, except that no gonadal tissue was removed. The process was repeated until >30% survival rate could be obtained (in line with others performing a similar procedure, Dussart G, personal communication). A batch of 50 snails of similar age was then gonadectomised and 50 snails were sham-operated. The gonadectomised and sham-operated snails were kept in separate tanks, and monitored for mortality and growth for six weeks.
Figure 2.5: Technique of gonadectomy
A: Making an incision in the innermost whorl of the shell using a fine forceps. B: showing the incision exposing the ovotestis and hepatopancreas, C: the hermaphrodite duct being ligatured to prevent excessive blood loss, D: the light orangish region towards the tip is the ovotestis and the dark orangish part is the hepatopancreas, E: The incision sealed using autoclaved shell pieces from dead snail, F: the recovering snail.

The snails were weighed and their diameter was measured every week and any mortality noted. At the end of six weeks, the snails were euthanized by quick decapitation with a scalpel. Once euthanized, the shell was carefully crushed and removed and the wet weight of each snail was recorded. Dissections took place under a stereomicroscope (Leica); the visceral cavity was opened up and the albumen gland was removed, weighed and fixed in a minimum of three times the tissue volume of Bouin’s solution (Sigma-Aldrich) for 1 hour, rinsed twice in Industrial Methylated Spirits (IMS, Charles Tenant) (70% IMS: 30 % distilled water) and then stored in 70% IMS at room temperature before processing. The histopathological processing and analysis was done on the albumen gland following the procedure described earlier in this chapter.

Statistical Analyses
All statistical analyses were carried out using Sigma-stat (v3.5). Shell diameter and total weight pre- and post-experiment were each compared by analysis of variance (ANOVA) followed by Dunnett’s test to compare the treatment means with the respective controls.
Data sets found to lack variance homogeneity were alternatively subjected to non-parametric analyses (Kruskal-Wallis), followed by Dunn’s post hoc test (Zar, 1999). The normalized albumen gland weight of the sham operated snails and the castrated snails at the time of termination of the experiment was compared by Student’s T-test or Mann-Whitney U test depending on whether the data were parametric or non-parametric.

**Results and Discussion**

Maximum mortality was observed in the first week after gonadectomy. This could have been due to excessive hemolymph loss and fungal growth and prior to death the snails were much less mobile and did not consume food. The survivors behaved normally in terms of mobility and food consumption. After one week there were 30 survivors from the gonadectomised snails and 45 survivors from the 50 sham-operated snails. To keep the numbers similar, 30 from each gonadectomised and sham-operated snails were monitored and the remaining sham-operated snails were killed. In the gonadectomised snail tank, there were 5 mortalities in week 3 followed by 2 in week 4 and 1 in week 5 Figure 2.6.

![Figure 2.6: figure showing mortalities in gonadectomised and sham operated snails over a period of six weeks after a recovery period of one week post-gonadectomization.](image)

In the sham-operated snail tank, only 1 snail died in week 5. At the end of the six week period there were 22 gonadectomised snails as compared to 29 sham-operated snails. In the gonadectomised snails the increase in mean weight was from 0.363g ±0.05 in the first week to 0.844g±0.19 in the sixth week while the sham-operated snail had mean weight of
0.441±0.064 in the first week that increased to 0.753±0.24 in the sixth week Figure 2.7. In the first week the gonadectomised snails had a mean diameter of 11.39mm± 0.08 that increased to a mean value of 15.79mm±1.84 in week six Figure 2.8. The sham-operated group had mean diameter of 11.38mm±0.19 in the first week that increased to a mean value of 15.62mm±1.78 in the sixth week.

![Graph showing weight change over weeks](image)

*Figure 2.7: Change in weight measured in grams (mean ±SD) of gonadectomised and sham-operated (controls) snails.*

During the first two weeks the gonadectomised snails weighed less than sham-operated snails Figure 2.7 that was likely due to loss of the gonadal tissue. A maximum increase in weight in the gonadectomised snails was observed between weeks three and four (24% increase) and then from week 4 onwards, their weight has risen over the sham-operated animals and this difference persisted until the end of the study period. Overall the gonadectomised snails showed an average increase of 132% in weight when compared to sham-operated snails with only a 71% increase. The overall average increase in diameter for the gonadectomised snails was 39% and a similar increase of 37% in diameter was measured in sham-operated snails. The increase in shell diameter and total wet weight of the gonadectomised snails was significant compared to the sham-operated snails (P=<0.001). These results are similar to earlier studies looking at the effects of castration on somatic growth in *Bulinus truncatus* (De Jong-Brink *et al.*, 1979), *Heloisoma duryi* (Miksys and Saleuddin, 1987) and *B. glabrata* (Vianey-Liaud and Dussart, 2002).
Castration in hermaphrodites seems to favor somatic growth due to reduced reproductive activity (Miksys and Saleuddin, 1987).

The weight of the albumen gland was normalized against the total body wet weight \( \frac{\text{weight of albumen gland (g)}}{\text{Total wet weight of the snail body (g)}} \) in both gonadectomised and sham-operated snails. The normalized albumen gland weight (NAW) for the sham-operated snails had a mean of 0.0419 but the gonadectomised snails had a larger albumen gland and the mean normalized albumen gland weight was 0.0739 Figure 2.9. There was a significant difference in the normalized albumen gland weight between the gonadectomised snails and the sham operated snails (controls) (P=<0.001). The increased body weight cannot be the only reason for the increased albumen gland weight in the gonadectomised snails. One of the reasons could be accumulation of secretory products (perivitelline fluid) due to reduced reproductive activity (Miksys and Saleuddin, 1987). A similar increase in the size of albumen gland post-gonadectomy has been reported earlier in *H. duryi* and *B. truncatus* (Miksys and Saleuddin, 1987; De Jong-Brink *et al.*, 1979). The smaller overall size as well as the smaller albumen glands is due to egg laying in the control snails compared to gonadectomised snails where all energy gained through food was invested in growth (Koene and Ter Maat, 2007).
Figure 2.9: Boxplot showing the normalized albumen gland weight (NAW) of the
gonadectomised snails compared to the sham operated snails after six weeks of observation.
Boxes represent median with 25th and 75th percentiles. The bars extend to the 10th and 90th
percentiles with outliers represented as dots. Sample size is indicated above each treatment label on
the x-axis. * p< 0.05 compared to the Control group.

Thus it could be concluded that the technique for gonadectomy with a survival rate of
~51% was successful. Post-gonadectomy increase in body size (both body weight and
shell diameter) and the increased albumen gland weight agree with previous studies on
other species of snail (Miksys and Saleuddin, 1987; De Jong-Brink et al., 1979).

2.2.2.2.3 Exposure of gonadectomised snails to 5α-dihydrotestosterone
The next step was to conduct a preliminary study exposing gonadectomised snails to
androgen, DHT with the aim of investigating effects of DHT on the overall condition
(mortality and growth) of the gonadectomised snails.

Materials and Methods

Test compounds
All chemicals were purchased in pure form (≥97%). The androgen 5α-dihydrotestosterone
(DHT) (CAS No. 521-18-6) was purchased from Sigma Aldrich Co. Ltd (UK) and HPLC
grade N, N-dimethylformamide (DMF) >99% (CAS No. 68-12-2) was purchased from
Fisher Scientific, UK.

Stock solution preparation
Chemical super stock solution was prepared by dissolving 0.025g (weighed on an
electronic balance to four decimal places) DHT in 50ml of DMF making a concentration of 0.5 g/L and were stored in glass screw top bottles at 4°C until required to make dosing stock solutions. Dosing stocks were then prepared in 1litre amber glass bottles by spiking double distilled water with the desired volume of concentrated stock in DMF (Table 2.3). 1 litre of each stock solution was prepared with DMF as the carrier solvent in all stock solutions. Total DMF solvent concentration in all dosing stock bottles was maintained at 6%. Fresh stock solutions were prepared twice a week. Stock bottles (1 litre) were washed and were double solvent rinsed (using methanol and ethanol) prior to new stock storage.

Table 2.3: Volume of DHT added to prepare dosing stocks made up to 1L and nominal concentration in the tanks.

<table>
<thead>
<tr>
<th>Dosing stock (DHT/MT)</th>
<th>0.5g/l DHT in DMF (ml)</th>
<th>100% DMF (ml)</th>
<th>ddH₂O (ml)</th>
<th>Total (ml)</th>
<th>Nominal tank concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 μg/l</td>
<td>1</td>
<td>59</td>
<td>940</td>
<td>1000</td>
<td>500 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>250 μg/l</td>
<td>0.5</td>
<td>59.5</td>
<td>940</td>
<td>1000</td>
<td>250 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>125 μg/l</td>
<td>0.25</td>
<td>59.75</td>
<td>940</td>
<td>1000</td>
<td>125 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>62.5 μg/l</td>
<td>0.125</td>
<td>59.875</td>
<td>940</td>
<td>1000</td>
<td>62.5 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>60</td>
<td>940</td>
<td>1000</td>
<td>Solvent control</td>
</tr>
</tbody>
</table>

The continuous flow through dosing system
Dechlorinated carbon filtered (5 and 10 μm) tap water (27°C ±1) was delivered via flow control-meters into individual glass mixing chambers via medical grade silicon tubing (VWR, UK) at a rate of 200 mL/min (Figure 2.10). From each mixing chamber, water was evenly distributed into two replicate tanks at the rate of 100Ml/min. Test compound dosing was controlled via a peristaltic pump (Watson Marlow, UK). Dosing stock solutions were pumped from 1 L amber glass bottles through pumpsil platinum-cured silicone tubing (Watson Marlow), to the mixing chambers at a rate of 0.2 mL/min, in order to achieve 1000 fold dilution and nominal concentrations in the exposure tanks of 62.5, 125, 250 and 500 ng/L (DHT). The solvent control (SC) tank received DMF at the same rate; such that the water in all tanks (except DWC) contained DMF at 0.006% v/v (below the suggested maximum solvent concentration of 0.01% v/v OECD limit for aquatic vertebrates OECD (2000) (http://www.epa.gov.uk/endo/pubs/ref-2_oecd) and gastropods OECD (2010) (http://www.oecd.org/). The glass exposure tanks (two per treatment) each had a working volume of 3.5 L (16x 15 x 23.5cm) and received approximately 41 tank volume renewals per day. Tank water was gently aerated via 2 pumps and 12 air stones to maintain optimum levels of dissolved oxygen. Dilution water and dosing flow rates were
monitored and if necessary corrected daily and dosing stock solutions were freshly prepared and replaced every 3 days to avoid any potential chemical degradation.

![Diagram of the layout of the experiment in the exposure room.](image)

*Figure 2.10: Diagram of the layout of the experiment in the exposure room.*

The diagram shows the main water supply from the header tank flowed through 6 flow meters into 6 mixing chambers at the rate of 200ml/min. The chemical from the dosing bottle flowed into the mixing chamber at the rate of 0.2ml/min and then from the mixing chamber the diluted chemical flows into the tanks @100ml/min.

**Experimental procedure**

This preliminary study was designed to chemically expose 5 adult gonadectomised snails in each replicate tank (working volume of 3.5L) for 10 days following the “rodent castrated adult male assay” (Moon et al., 2009), to observe the effects of the chemical on accessory reproductive glands. The study consisted of six treatments (tested in duplicate): Dilution water control (DWC), solvent control 0.006% DMF (SC) and DHT at nominal concentrations of 62.5ng/l, 125ng/l, 250ng/l and 500ng/l. Test concentrations were chosen to be in the range of androgen concentrations previously reported to affect gastropod molluscs and fish (Margiotta-Casaluci and Sumpter, 2011; Lyssimachou, Ramón and Porte, 2009; Watermann et al., 2008; Schulte-Oehlmann et al., 2006; Pawlowski et al., 2004; Schulte-Oehlmann et al., 2004; Tillmann et al., 2001). The study was designed to chemically expose 5 adult gonadectomised snails in each replicate tank (working volume of 3.5L) for 10 days. Based on the previous gonadectomy survival rate, 140 snails were gonadectomised over a period of 2 days, to allow for expected mortalities (i.e. aprox. 50% mortality expected) following the same gonadectomy protocol mentioned above in section 2.2.2.2.2. In total 60 gonadectomised snails were used in the DHT study.
(6 treatments X 2 replicate tanks X 5 gonadectomised snails in each of the 12 tanks). At the end of 10 days of DHT exposure, the snails were removed from their exposure tanks, blotted dry, weighed (g) using an electronic balance and shell diameter measured to the nearest 0.1mm using hand-held digital callipers. Snails were then euthanized by quick decapitation with a scalpel. Once euthanized the shell was crushed, shell pieces carefully removed and each unshelled snail weighed.

**Statistical Analysis**

All statistical analyses were carried out using Sigma-Stat (v3.5). Shell diameter and total weight pre- and post-experiment were each compared by analysis of variance (ANOVA) followed by the Dunnett’s test to compare the treatment means with the respective controls. Data sets found to lack variance homogeneity were alternatively subjected to non-parametric analyses (Kruskal–Wallis), followed by Dunn’s post hoc test (Zar, 1999). A Spearman's Rank Order correlation between shell diameter and total weight of the snails was performed to determine overall condition of the snails.

**Results and discussion**

The weight of the gonadectomised snails in the different treatments at the beginning of the experiment varied between 0.304-0.336±0.018 (g) and their diameter varied between 10.51-10.93±0.22 (mm). The average weight post-exposure varied from 0.283±0.008 (g) in solvent control to 0.351±0.045 (g) in DWC. In the DWC, 62.5ng/l DHT and 250ng/l DHT treatments the average weight increased but a decrease in average weight was observed in the SC, 125ng/l DHT and 500ng/l DHT Figure 2.11(A). A similar trend was observed for diameter except in SC, where even though the average weight post exposure decreased but diameter did show a slight increase (1.62%) Figure 2.11(B). Both the weight and diameter of the snails post-exposure was significantly positively correlated ($R^2=0.69, P<0.001$). Overall there were no significant differences observed in the average weight and average diameter of the gonadectomised snails in the different treatments post-exposure (ANOVA; P<0.05).
Figure 2.11: (A). Mean weight±SD and (B). mean diameter±SD of the snails pre- and post DHT exposure preliminary study. A group of 60 sexually mature B. glabrata was exposed to 62.5, 125, 250 and 500ng/l nominal concentration of DHT.

Figure 2.12: Survival of snails at the termination of DHT exposure preliminary study. A group of 60 sexually mature gonadectomised B. glabrata exposed to 62.5, 125, 250 and 500ng/L of DHT for 10 days.

High mortality was recorded over the duration of the exposure, with only 5 survivors (2 in tank1+3 in tank 2) in DWC, 6 survivors in 62.5ng/l DHT (3 in tank1+3 in tank 2) and 125ng/l DHT (4 in tank1+2 in tank 2), 4 survivors in 500ng/l (2 in tank1+2 in tank 2) (Figure 2.12). The surviving snails were not very healthy, which was evident by the fact that they were not consuming food and were lethargic (not moving actively in the tanks). These surviving snails had pale hemolymph, as observed at the time of dissection.

The initial aim of this study was to quantify any changes in the accessory reproductive glands post-exposure. Due to low survival there was not enough data to provide statistically meaningful results. The survivors were sick and so could not be truly representative. The results obtained during this preliminary study indicated the likely unreliability of using the gonadectomised snail model. Another factor was that the number of snails needed to be gonadectomised for the main study would be approximately 500, which would be very labour intensive taking approximately seven days (approximately 70 snails per day). Taking together the work involved and the mortality rate during the
exposure study it was not considered feasible to have a gonadectomised snail version of the mammalian Hershberger test.

2.2.3 Conclusion

All the information gained about the molluscs and the test species through the available literature was vital for my PhD project as it would set the ground for the main exposure studies and the bioinformatics investigation in *B. glabrata* and *L. gigantea*, that will be discussed in the later chapters. The preliminary experiments conducted in *B. glabrata* provided baseline information to compare the reproductive tract development of the experimental snails. The protocol for gonadectomising snails was well optimized but the pilot study provided vital evidence that led to the decision of excluding gonadectomy as a part of my exposure study protocol.
Chapter 3: Exposure of the freshwater snail *Biomphalaria glabrata* to steroid androgens during development does not affect reproductive physiology
3.1 Introduction

Androgens are steroid hormones that regulate a wide range of physiological processes in male vertebrates, often mediated through the androgen receptor (AR). Various synthetic androgens and androgen antagonists can therefore act via the androgen receptor and are used in pharmaceutical medicines for treatment of some common diseases, such as androgen deficiency and prostate cancer, respectively (Tabb and Blumberg, 2006). In pre-clinical and safety testing of these pharmaceutical drugs, the Hershberger assay, originally developed to identify the drivers of male sexual development, is the gold standard test (Hershberger, Shipley and Meyer, 1953). This assay employs either immature or castrated peripubertal male rats, which are dosed with the compound to be tested over a number of days (3-10 days depending on assay). At the end of the exposure the reproductive organs and tissues that are known targets for androgens are assessed (e.g. weight, morphology, histopathology etc.). Chemicals, which increase organ weights are considered stimulatory, i.e. a compound, which increases the weight of the prostate in the Hershberger assay, would be considered androgenic. Similarly a reduction in target organ weight would suggest anti-androgenic or inhibitory properties of the chemical. These types of mammalian assays, including pubertal assays (to detect developmental effects) have been validated by the Organisation for Economic Co-operation and Development (OECD) and the United States Environment Protection Agency (US EPA) for screening suspected endocrine disrupting chemicals for androgenic and anti-androgenic activity (Moon et al., 2009; Owens et al., 2007; Shin et al., 2007; Owens et al., 2006).

New legislation concerning chemical safety testing will undoubtedly involve increased numbers of vertebrate animal tests and will bring with it significant financial and ethical costs. As a consequence fish tests are also on the increase, as they are part of the new environmental risk assessment process involved in the safety testing of pharmaceuticals and industrial chemicals both for toxicity and endocrine disrupting properties. As discussed earlier in chapter 1, the Annual Statistics of Scientific Procedure on living animals, Great Britain (2012) stated that over 2.13 million animals were used in scientific experimentation in 2012 and the estimated total cost for conducting EDSTAC Tier 1 and Tier 2 screening for a chemical sums up to approximately $900000 (approximately over £540,000). Considering the vast amount of chemicals ultimately to be tested will involve a huge expenditure. Now more than ever, there is increased pressure to implement the three R’s (reduction, refinement and replacement) in relation to the testing of chemicals and medicines and in general biomedical research. By exploiting the use of small invertebrates
as surrogate mammalian and fish test organisms the number of vertebrates can be reduced or in some cases replaced. This has been successfully achieved with *C. elegans* and *D. melanogaster* in, for example, the fields of ageing, development and physiology (Harrington *et al.*, 2010; Ghosh-Roy and Chisholm, 2010; Beckingham *et al.*, 2007; Roberts, 2006). Neither of these organisms, however, have male reproductive anatomy remotely similar to that of the mammals. Even in the non-mammalian vertebrate classes (e.g. fish), suitable models for mammals have not been found, with respect to the existence of a prostate and other accessory sexual glands used as endpoints in the Hershberger assay. Notwithstanding this, very few invertebrates have been investigated as potential surrogates of male or female reproductive function, for either biomedical or chemical safety testing.

At the time this research was initiated, homologues of androgens and steroidogenic androgen and estrogen producing enzymes had, been reported in the gonads and digestive tube of several invertebrate groups (Markov *et al.*, 2009; Lafont and Mathieu, 2007; Lavado, Janer and Porte, 2006a; Lehoux and Sandor, 1970), including molluscs and annelids. In addition molluscs have a reproductive anatomy that is, at least superficially, similar to that of vertebrates. Like mammals, the snail’s male reproductive tract extends from the testis (or ovotestis in hermaphrodite snails), has seminal vesicles, and a prostate gland and terminates in a penile glandular complex (around the preputium) containing sebaceous glands and a penis (Adiyodi and Adiyodi, 1988) (discussed in detail in Chapter 2).

There are also various examples in the literature that suggest developmental exposure of gastropod snails (*Marisa cornuarietis*) to the synthetic androgen, 17α-methyltestosterone (MT) or to the environmental pollutant “androgen” tributyltin (TBT) causes hyperplasia in the epithelial tissue of reproductive glands and inhibition of spermatogenesis in males (Schulte-Oehlmann *et al.*, 2006; Schulte-Oehlmann *et al.*, 2004). Indeed TBT and triphenyltin (TPT) have been documented to alter female reproductive tissue (the albumen gland) to histologically resemble male reproductive tissue (the prostate gland) in some gastropod snails (Watermann *et al.*, 2008; Albanis *et al.*, 2006). Female adult *M. cornuarietis* (a freshwater prosobranch gastropod mollusc) developed imposex (imposition of penis growth in females) after 150 days of MT exposure (Janer *et al.*, 2006a; Janer *et al.*, 2005b) and adult pond snails, *Lymnaea stagnalis* (a pulmonate gastropod), demonstrated weak histological damage to the albumen and prostate glands after 8 weeks of exposure to 100ng/L of MT (Czech, Weber and Dietrich, 2001). Reduction in fertility and embryo production was observed in *Potamopyrgus*
antipodarum, M. cornuarietis and Nassarius reticulates when exposed to environmentally relevant concentrations of MT (Duft et al., 2007). Similar “androgenic” effects were reported to be significantly reduced by co-exposure to the anti-androgen cyproterone acetate and to an aromatase inhibitor (1-methy-1, 4-androstadiene-3, 17-dione) (Ketata et al., 2008; Oberdörster and McClellan-Green, 2002; Bettin, Oehlmann and Stroben, 1996), suggesting the existence of vertebrate-like mechanisms of male reproductive function in molluscs (Hagger et al., 2009). Moreover, exposure of males from a number of gastropod molluscan species (Marias cornuarietis, Nucella lapillus and Hinia reticulata) to 17α-MT caused a reduction in the length of the penis and penis sheath (Watermann et al., 2008; Oehlmann et al., 2007; Tillmann et al., 2001), adding further strength to the suggestion that mammalian-like mechanisms through which androgens and anti-androgens could act are present in molluscs. These documented impacts (mentioned in detail in Chapter 1) raise the possibility that the effects of androgenic and anti-androgenic chemicals on reproduction and development in molluscs might be comparable to those in mammals (Oehlmann and Schulte-Oehlmann, 2003) and that snails may hold promise as a model for male reproductive diseases and for testing medicines and chemicals for interference with male reproductive system.

As the basis of the mammalian Hershberger assay is the absolute requirement of the male reproductive tract for testosterone (T) and/or dihydrotestosterone (DHT), it seemed sensible to test the sensitivity of B. glabrata reproductive tract to androgens. In rodents, the Hershberger assay has been tested at three different life stages, the prepubertal intact male (Ashby and Lefevre, 2000), the castrated adult male and the peripubertal castrated male (Gray, Furr and Ostby, 2005). The choice of time points is based on the fact that accessory sex glands and tissues (seminal vesicles and prostate, coagulating glands and Cowper’s glands, preputial glands and glans penis) are dependent upon androgen stimulation to gain and maintain weight during and after puberty (Owens et al., 2006).

### 3.2 Aim of the study

The aim of this study was to investigate if an invertebrate “Hershberger-type” test could be developed. These tests included intact unhatched snails exposed until sexual maturity and adult castrated (gonadectomised) snails to replicate the “rodent prepubertal intact assay” and the “rodent castrated adult male assay” respectively. These tests employed 5α-Dihydrotestosterone (DHT) and 17α-methyltestosterone (MT) as model androgenic compounds, to see if similar stimulatory effects could be produced in developing and adult snails as would be expected in rodents.
5α-Dihydrotestosterone (DHT) is one of the most physiologically important and active androgens in many male vertebrates (George, 1997), and is formed by the 5α-reduction of testosterone by the enzyme 5α-reductase. Both 5α-reductase enzyme activity and endogenous levels of testosterone have been previously reported in molluscs (Fernandes, Loi and Porte, 2011; Lavado, Janer and Porte, 2006a; Janer, LeBlanc and Porte, 2005a), even though their functions have not been clearly demonstrated, as the knowledge of steroid metabolism in invertebrates is still limited and fragmentary. The drug, Finasteride has been shown to inhibit 5-alpha reductase activity in the ovaries and testis of the sea urchin *Lytechinus variegatus* with an IC50 of 2.7 μM (Wasson et al., 2000; Wasson and Watts, 1998b). Moreover, 5-alpha reduced metabolites were found in the gonads and digestive tube of another sea urchin, *Paracentrotus lividus* when exposed to 10 μM Fenarimol imidazole-like fungicide (Lavado et al., 2006b; Janer, LeBlanc and Porte, 2005a). Subcellular fractions isolated from the visceral coil of the gastropod snail *M. cornuarietis* reported 5-alpha reduced metabolites such as 5α-androstenedione and 5α-DHT (Janer, LeBlanc and Porte, 2005a) indicating that steroid 5α reductase activity appears to be involved in androgen metabolism in these invertebrate species. 17α-methyltestosterone (MT) was tested in addition to DHT, as it is known to have similar binding affinity as T and the addition of the 17 α-alkyl group has been shown to slow metabolism at the 17-position and prolong its in vivo half-life (Gao, Bohl and Dalton, 2005). This synthetic androgen is widely applied in aquaculture to control sex determination and sex-reversal of genetic female fish to phenotypic males (Vick and Hayton, 2001; Kitano et al., 2000; Papoulias, Noltie and Tillitt, 2000; Hunter and Donaldson, 1983) and is therefore considered to be an endocrine disruptor in fish (Rivero-Wendt et al., 2013; Kang et al., 2008; Pawlowski et al., 2004; Hornung et al., 2004). The specific mechanism through which MT acts in invertebrates remains to be resolved, but dose, timing and duration of treatment can influence the effects, such as inducing masculinization in prosobranch snails and crustaceans (Fujioka, 2002; Czech, Weber and Dietrich, 2001). In another study conducted by Schulte-Oehlmann and colleagues (2004), 0.1μg/L MT caused virilization in female *M. cornuarietis* and also affected the formation of germ cells in male and female gonads (Schulte-Oehlmann et al., 2004).

The aim of this study, therefore, was to test the hypothesis that common physiological processes might underlie the response of the male reproductive system to androgens in both molluscs and mammals.
3.3 Experimental Approach

The overall approach was to expose *B. glabrata* to two potent androgens; the natural vertebrate steroid, 5α-dihydrotestosterone (DHT) and the anabolic steroid 17α-Methyltestosterone (MT), after which reproductive tissues and organs were assessed for stimulatory (e.g. male tissue reproductive growth or advanced spermatogenesis) or possible disruptive (e.g. to female reproductive tissues or oocyte development) effects. Due to the unreliability of the gonadectomised model (high rate of mortality and unhealthy snails), it was not considered feasible to have a gonadectomised snail version of the mammalian Hershberger test. In vertebrates, early life and pubertal stages are more susceptible to endocrine interference by natural or exogenous steroid androgens (Mensink et al., 1996) than in adulthood. Therefore, to investigate possible androgenic effects in *B. glabrata*, the snails were developmentally exposed with the expectation that a similar ‘sensitive window’ during sexual development would also present in our test species.

3.3.1 Test compounds

All chemicals were purchased in pure form (≥97%). The androgen 5α-dihydrotestosterone (DHT) (CAS No. 521-18-6) was purchased from Sigma Aldrich Co. Ltd (UK) and HPLC grade N, N-dimethylformamide (DMF) >99% (CAS No. 68-12-2) was purchased from Fisher Scientific, UK.

3.3.2 Stock solution preparation

Chemical super stock solution was prepared by dissolving 0.025g (weighed on an electronic balance to four decimal places) either MT or DHT in 50ml of DMF making a concentration of 0.5 g/L and were stored in glass screw top bottles at 4°C until required to make dosing stock solutions. Dosing stocks were then prepared in 1litre amber glass bottles by spiking double distilled water with the desired volume of concentrated stock in DMF (Table 3.1). 1 litre of each stock solution was prepared with DMF as the carrier solvent in all stock solutions. Total DMF solvent concentration in all dosing stock bottles was maintained at 6%. Fresh stock solutions were prepared twice a week (Figure 3.1). Stock bottles (1 litre) were washed and were double solvent rinsed (using methanol and ethanol) prior to new stock storage. Personal protective equipment was worn while preparing and handling stock solutions.
Table 3.1: Volume of Chemical (DHT/MT) added to prepare dosing stocks made up to 1L and nominal concentration in the tanks.

<table>
<thead>
<tr>
<th>Dosing stock (DHT/MT)</th>
<th>0.5g/l DHT in DMF (ml)</th>
<th>100% DMF (ml)</th>
<th>ddH₂O (ml)</th>
<th>Total (ml)</th>
<th>Nominal tank concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 µg/l</td>
<td>1</td>
<td>59</td>
<td>940</td>
<td>1000</td>
<td>500 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>250 µg/l</td>
<td>0.5</td>
<td>59.5</td>
<td>940</td>
<td>1000</td>
<td>250 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>125 µg/l</td>
<td>0.25</td>
<td>59.75</td>
<td>940</td>
<td>1000</td>
<td>125 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>62.5 µg/l</td>
<td>0.125</td>
<td>59.875</td>
<td>940</td>
<td>1000</td>
<td>62.5 ng/l (DHT/MT)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>60</td>
<td>940</td>
<td>1000</td>
<td>Solvent control</td>
</tr>
</tbody>
</table>

3.3.3 Experimental Design

Two consecutive 30 days exposure studies (one for DHT and one for MT) were conducted in April-May, 2011. Each experiment was designed to have 20 hatchlings in each tank, in six treatments, tested in duplicate (Figure 2.10) up until sexual maturity (240 animals in total per chemical). The range of doses for exposure was informed by the doses used in the preliminary adult castrated snail study which itself had been informed by literature on fish developmental studies (Margiotta-Casaluci and Sumpter, 2011) and other mollusc research (Lyssimachou, Ramón and Porte, 2009; Watermann et al., 2008; Schulte-Oehlmann et al., 2006; Schulte-Oehlmann et al., 2004; Pawlowski et al., 2004; Tillmann et al., 2001). Based on this, the six treatments for the experiment (tested in duplicate) were dilution water control (DWC), solvent control 0.006% DMF (SC), and MT or DHT at nominal concentrations of 62.5 ng/L, 125 ng/L, 250 ng/L, and 500 ng/L. Tanks were pre-dosed for 3 days before introducing egg masses to allow for the dosing to equilibrate and water samples were collected for analysis prior to and during the snail exposure (see Figure 3.1).
3.3.4 Egg collection and initial static-renewal exposure of embryos

Snails were exposed to each of the test chemicals, MT or DHT, throughout embryonic and early development until sexual maturation (age at which egg laying begins) total of 30 days. 24 egg masses, laid on the same day, were collected from a breeding group of 30 adult *B. glabrata* (BB02). Egg masses were carefully removed from the glass walls of the parental tank using a scalpel. Each egg mass was inspected under dissecting microscope for any damage prior to being included in the study. *B. glabrata* naturally hatch out of the egg mass over a period of days (i.e. within a single egg mass some hatchlings will emerge before others), however for this experiment design, to remove variations in age and development at the point of sampling, hatchlings were required to be monitored (and selected) over the hatching period. Therefore, egg masses were initially exposed to the test solutions individually within wells of six-well plates. The multi-well plates (nunc, sterile non-treated) were floated on top of the flow through exposure tanks to maintain...
temperature and light regime consistency. Four egg masses were exposed per treatment. Each egg mass was inspected daily for development and hatching and 50% of the dosed water (3ml total volume) was renewed every 24 hours. All newly hatched snails were carefully separated from egg masses (and continued to be dosed in a new well) and the date that they hatched was recorded. The majority of snails hatched on days 6-7, at which point 20 hatchlings were collected and placed into the appropriate tank (40 snails per treatment), enabling all hatchlings to be the same age at the point of sampling and for the density of snails to be fixed at the start of the experiment. After hatchling snails were placed into the tanks (day 7 of the exposure) and were exposed under continuous dosing (flow-through) conditions for 23 days until sexual maturation.

### 3.3.5 Tank water analyses

Water samples (250 mL) were collected, in amber glass bottles, from each tank on day 0 (day before egg-masses were exposed), and subsequently every 7 days until the end of the experiment (Figure 3.1). The target chemical was extracted from the water sample by solid phase extraction (SPE) using C18 SPE cartridges (Sep-pak Plus, Waters Ltd., Watford, UK) and a Supelco SPE manifold (Sigma Aldrich) as per the manufacturers instructions. Additional “blank” and “spiked” samples were also taken to check SPE efficiency, these were double distilled water ‘blanks’, MT/DHT ‘spiked’ using 10µl of 0.5g/l chemical in 250ml double distilled water. These were extracted at the same time and using the same SPE method as the 250 ml tank water samples.

#### 3.3.5.1 Extraction and concentration of androgenic chemicals from water samples

Extraction and concentration of androgenic chemicals from water samples was accomplished using a ‘visiprep’ vacuum manifold (Supelco) and reverse-phase C18 cartridges (Sep-Pak C18, Waters Associates, inc). A silicone liner was inserted into each port of the manifold and C18 cartridge was positioned on the top of the liner. A tube adapter and plastic reservoir were placed above the cartridge. The vacuum unit was then attached to the manifold and turned on. 5ml of double distilled water was added to each sample well, followed by 5ml of methanol and then a further 5ml of double distilled water to condition the C18 cartridge (these were sucked thought the cartridge by the vacuum). While conditioning, it was important that the C18 cartridge did not dry out so another 5ml of double distilled water was filled in the reservoir while the well bung was being applied to the reservoir. This bung had an attached plastic straw, which was suspended into the sample bottles, and weighed down by a stainless steel weight at the other end. Once the vacuum was started and taps opened, water was sucked through the plastic straw (at the
rate of \( \leq 5\text{ml/min} \), into the reservoir, and through the C18 cartridge (where the test chemicals were bound). The water sample dripped through the disposable liner (attached to the C18 cartridge) and into the base of the manifold. The water collected at the bottom of the manifold was removed through a rubber tube to a collecting vessel for disposal. Once the entire water sample had passed through the C18 cartridge, the bung and attached straw were removed. The vacuum was left on for at least 30 minutes to remove any water from C18 cartridge. 7ml glass vial double solvent rinsed were placed inside the manifold beneath the C18 cartridge and disposable liners taking care that the cartridge/liner/vials were properly aligned. The C18 cartridges were eluted using 5ml of analytical grade methanol (Fisher scientific, UK) under a vacuum @ approximately 2ml/min. Then the methanol-elutant was dried (evaporated) under nitrogen gas using the supelco ‘visidy’ attachment for the extraction manifold. The extract was then re-suspended in 1 ml of analytical grade ethanol (Hayman, UK). The concentration factor calculated as 
\[
\left( \frac{\text{Volume of water extracted}}{\text{Volume of ethanol in which the sample was resuspended}} \right)
\]
was recorded. Vials containing extract in ethanol were stored with tight fitting lids at 4\(^\circ\)C to prevent evaporation until further analysis using a recombinant Yeast androgen screen (YAS) (Sohoni and Sumpter, 1998).

### 3.3.5.2 Yeast Androgen Screen (YAS) Assay

The yeast assay is a receptor/reporter assay using recombinant yeast plasmid for detecting estrogenic or androgenic activity of the chemical compounds. The activity can be monitored by eye and measured by spectrophotometer. The assay was carried out in a type II laminar flow cabinet (to minimize aerosol formation) following the standard protocol (Sohoni and Sumpter, 1998).

A day before the YAS assay was conducted, yeast growing medium was prepared (Appendix 3.1). 125 \( \mu l \) of 10X concentrated yeast stock, from a cryogenic vial, was added to prepared sterile yeast growth medium and then incubated at 28\(^\circ\)C for approximately 24 hours on an orbital shaker (250 r.p.m). Chemicals in the extract (prepared from water extraction in the earlier section) were serially diluted in absolute ethanol in a 96-well microtitre plate. To start with, 100\( \mu l \) aliquots of absolute ethanol were pipetted into each row of wells, except for the first well. In the first and the second well 100\( \mu l \) of sample was pipetted. The 200\( \mu l \) mixture of extract and ethanol were mixed using the pipette and then 100\( \mu l \) of the 50:50 mixtures of extract and ethanol was removed from the second well and transferred to the third well. This was mixed carefully using a pipette and then 100\( \mu l \) was taken from this well and transferred into the fourth well and this process was repeated until all 12 wells had been completed (the final well contained 100\( \mu l \) of absolute ethanol
and 100μl of the mixture). The dilution series for the Dihydrotestosterone (DHT) and 17α-methyltestosterone (MT) standard curves were prepared in similar manner (DHT starting concentration 1x10^{-6} M and MT 6.6x10^{-7} M) (Figure 3.2). MT (6.6x10^{-7} M) showed similar activity in the YAS screen with DHT standard solution and so this concentration was selected as the starting concentration for MT. 10μl aliquot of each concentration was then transferred from the dilution plate to a new 96-well microtitre assay plate, starting from the most dilute concentration from the right side of the dilution plate. Each sample dilution was replicated so two identical rows were produced and to avoid cross contamination, between each double row of extract one row was left completely empty (blank). DHT and MT were used as positive controls in their respective screens and a negative control (absolute ethanol) was used in both the experiments. The purpose of the negative control was to account for any kind of activity due to possible chemical contamination in the ethanol. Once all assay plate dilutions were complete the ethanol was evaporated to dryness.

![Figure 3.2: Standard curves showing the starting molar concentrations of DHT and MT with respect to ethanol.](image)

While the assay plates dried the final assay medium was prepared as per the primary
culture (appendix 3.1) with the addition of 0.5ml of Chlorophenol red-β-D-galactopyranoside (CPRG) and 4x10^7 yeast cells from the 24-h culture prepared the previous day, per 50ml of growth medium. The volume containing 4x10^7 yeast cells was calculated by taking two 100μl aliquots of the 24-hour culture, in wells of a 96-well microtitre plate and the absorbance at 640nm were measured using a spectrophotometer plate reader. The mean of the two readings obtained was used to extrapolate the number of cells per ml using the absorbance graph. The volume required to give 4 x 10^7 cells per 50ml of growth medium was calculated using the following equation.

\[
\frac{4 \times 10^7}{\text{Number of cells calculated from the absorbance graph}} = \text{Vol of 24 hour culture}
\]

200μl of the yeast inoculated final assay medium was then added to each well of the assay plate using a multichannel pipette. To prevent evaporation of the medium (which might change the concentration of CPRG) and therefore the colour of the background, each plate was sealed round its edges using autoclave tape. The plates were then agitated briskly for 2 minutes on a titre plate shaker before being incubated at 32°C in a naturally ventilated heating cabinet. The plates were removed after 24 hours and incubated overnight at 28°C to avoid a significant increase in background expression of the enzyme β-galactosidase. Colour development of the medium was checked at regular intervals using a Titertek Multiscan MCC/340 spectrophotometer plate reader. The plates were read at an absorbance of 540nm (for colour) and 620nm (for turbidity). Following incubation, a deep red colour indicates positive activity and due to slight background production of β-galactosidase, the negative control wells appear light orange in colour. Growth of the yeast is accompanied by a turbid appearance. Therefore the corrected value = chemical absorbance at 540nm-(chemical absorbance at 620nm-blank absorbance at 620nm) to correct for the turbidity.

3.3.5.3 Calculating androgenic equivalents

The androgenic equivalents were calculated by first plotting the corrected absorbance values, the DHT standard curve and ethanol blank (to check for contamination), against the relative concentration of each extract. When plotted the DHT standard and extract dilutions produced sigmoidal curves. Only the linear part of the graph was used for calculations. The midpoint of this linear part was noted and then two corrected absorbance values on the either side of the selected midpoint were obtained. Based on the well numbers of these two values, the corrected absorbance was extrapolated using the standard
curve and the concentration of the first well was calculated (concentration before dilution). The concentration obtained were then multiplied by the molecular weight of DHT (290.44)/ MT (288.424), to provide a concentration in weight per volume, rather than a Molar concentration. The concentration (weight per volume) was then multiplied by 20 to take into account the volume of assay medium. To obtain the concentration of the chemical in initial water sample in g/l, these values were divided by the original concentration factor. An average value of the two concentrations was used as the final androgenic equivalent concentration of the chemical in the sample.

3.3.6 Biological Sampling and analyses

3.3.6.1 Terminal sampling
At the culmination of the experiment (after 30 days of exposure), snails were removed from their exposure tanks, blotted dry, weighed (g) using electronic balance and their shell diameter measured to the nearest 0.1mm using hand-held digital callipers. Snails were then euthanized by quick decapitation with a sterile scalpel. Once euthanized the shell was carefully crushed and removed. Once the snail was removed from the shell it was washed in a petri dish containing distilled water to remove any remains of shell pieces. When handling the de-shelled snail with forceps, mantle was used to hold the tissue, to avoid damaging any reproductive organs. The soft body weight of the whole snail was recorded and were quickly fixed in RNAlater and stored at 4°C in 2ml cryovial tubes (Nalgene).

3.3.6.2 B. glabrata dissection and preservation
Each snail fixed in RNAlater was carefully dissected under stereomicroscope (Leica), the visceral cavity was opened up and the albumen gland, glandular complex and the ovotestis were removed Figure 3.3. The ovotestis is situated in the innermost whorls of the shell and is dorso-laterally covered by the digestive gland (hepatopancreas). Using forceps and a sterile scalpel the ovotestis was carefully teased away from the digestive tissue, weighed on an electronic balance up to 4 decimal places, weight recorded on the data sheet and cut transversally into two equal sized pieces, placing one half into the new RNAlater cryovial and the other into Bouin’s fixative. The albumen gland or the female accessory reproductive gland is a large tubulous gland, that can be distinctly seen above the hepatopancreas as a ball like or folded structure, was removed using forceps (Figure 3.3), weighed and divided into two equal sized pieces for molecular analysis and for histopathological analysis. The glandular complex (GC) is the term used to describe the prostate diverticulum, membrane gland (oothecal gland) and mucous glands. This term has
been previously used in the literature to describe the albumen gland and the mucous gland (Pal, 2007). Due to its complex nature (i.e. 3 gland types intertwining into each other) GC were only dissected and preserved for histopathology and not split between molecular and histopathology analysis. Total tissue weights or sizes (gland complex, surface area cm², see section below) were normalized by dividing them by the individuals total body weight to produce, normalized glandular complex, normalized ovotestis or normalized albumen gland;

\[ \frac{\text{Albumen gland weight (g)}}{\text{Total body weight (g)}}. \]

To prevent cross treatment contamination, dissection equipment (scalpel, scissors, and forceps) were washed in 100% ethanol between different snails and a new plastic petri dish and weigh boat used for each snail. Tissues for histology were fixed in a minimum of three times the tissue volume of Bouin’s solution (Sigma-Aldrich) for 1 hour, rinsed twice in Industrial Methylated Spirits (IMS, Charles Tenant) (70% IMS: 30 % distilled water) and then stored in 70% IMS at room temperature before processing.

3.3.6.3 Morphometric analysis of glandular complex

Prior to fixation for histopathological analysis, the whole GC were dissected out and placed on a grid (cm²) to assess their cross-sectional area through image analysis (Figure 3.3).
3.4. To determine the area of the GC, photos were taken using a Leica MZ FLIII binocular dissecting microscope and captured with a Leica DC300F digital camera. The images were captured using 8bit/channel colour depth. The resolution of the objective was varied between 0.8x to 1.25x. The images were then objectively analysed using a computer-assisted image analysis program, ImageJ (http://rsb.info.nih.gov). The program allows the user to analyze, edit, enhance and print the images. Image intensity was measured using grey values with an arbitrary scale of 0 (dark, black) to 255 (clear, white) units. The scale was calibrated depending on the resolution of the image. Images were cropped and using freehand selection tool, an outline was drawn for every image before it could be converted to RGB stacks consisting of three primary colour channels. All the images in the red channel of the RGB stack were selected and were threshold at a constant value and their cross-sectional area measured.

![Image of glandular complex tissue](image)

**Figure 3.4:** Photomicrographs of part of glandular complex tissue from DWC of 17α-MT exposed snails.

Tissue section dissected out from the visceral cavity and placed on a cm² grid to measure the cross-sectional area through image analysis. Glandular complex comprises of Mg: mucous gland, Memgl: membrane gland and Pd: prostate diverticulum. Scale bar 100μm is used.

### 3.3.6.4 Statistical Analysis

All statistical analyses were carried out using Sigma-Stat (v3.5). Possible intra replicate differences in shell diameter, total weight, CG surface area, albumen gland weight and ovotestis weight were analysed with either Student's T-test or Mann-Whitney U test depending on whether the data were
parametric or non-parametric. A Spearman’s Rank Order correlation between shell
diameter and total weight of the snails was performed to determine overall condition of the
snails. Correlations between snail density, in terms either of total number of snails (n)
present at the end of the exposure were conducted using either Pearson’s correlation
coefficient or Spearman’s rank order to assess the effects of different survival/density
profiles on overall condition. When no significant intra replicate differences were found
using T-test/Mann Whitney tests then shell diameter, total weight, GC area, ovotestis and
albumen gland weights were compared by analysis of variance (ANOVA) followed by the
Dunnett’s test to compare the treatment means with the respective controls. Data sets
found to lack variance homogeneity were alternatively subjected to non-parametric
analyses (Kruskal–Wallis), followed by Dunn’s post hoc test (Zar, 1999). Statistical
significance was considered at \( \alpha \leq 0.05 \) for all comparisons.

3.4 Results

3.4.1 Measured water concentration of MT and DHT
Analysis of biological activity in tank water extracts dosed with MT and DHT using the
YAS assay showed that snails were exposed to the test substances in the expected range of
concentrations. No androgenic activity was detected in either the dilution water control
(DWC), solvent control (SC) or extraction ‘blanks’ at any time point measured. Solid
phase extraction efficiencies, calculated from spiked samples, were 93±37.7% for DHT
and 71.13±21.1% for MT. Overall the actual tank concentrations (as measured by the YAS
assay) were between 80 and 110% of nominal concentrations for MT and 65% to 70% of
nominal concentration for DHT and increased in a dose-dependent manner (Figure 3.5).
The lower level of DHT relative to the nominal is likely due to lower stability of DHT in
water as compared to MT (Katsiadaki et al., 2006). No significant intra-replicate
differences were found in the concentrations of either chemical dosed (P<0.05).
Figure 3.5: Mean tank water concentration of (A) DHT and (B) MT during exposure period. Water samples analyzed using in vitro recombinant yeast androgen screen (YAS) (Sohini and Sumpter, 1998). Nominal tank concentration for MT and DHT were Dilution water control, solvent control, 62.5, 125, 250 and 500ng/l.

3.4.2 Hatchling survival and growth (shell diameter and weight)

Hatchling survival and growth of *B. glabrata* exposed to either Methyltestosterone (MT) or Dihydrotestosterone (DHT) was monitored over the exposure period to assess the effects of these chemicals on snails. For both DHT and MT exposures, 20 hatchling snails were placed into each replicate tank at the start of the experiment. At the end of the 30 day exposure hatchling survival varied between treatments, and in some cases between replicate tanks within a treatment (Figure 3.6A). Overall survival in the DHT experiment was lower (61.7%) than in the MT experiment (80.4%). In the DHT treatment most snails survived in the SC (39 total) and 62.5ng/l DHT (38 total) nominal treatments, fewer snails survived in the 125ng/l DHT (31 total) and DWC (25 total), with much lower survival found in the 500ng/l (13 total) and 250ng/l DHT (2 total) treatments.

Figure 3.6: Numbers of *B. glabrata* snails surviving until experiment termination (30 days exposure) in each replicate tank, exposed to A) Dihydrotestosterone (DHT) and B) Methyltestosterone (MT).
In the MT treatment maximum survival was seen in the DWC (39 total) and SC (38 total) treatments Figure 3.6(B). Survival was slightly lower in the 125ng/l nominal MT (33 total) and 62.5ng/l MT (30 total); as with the DHT exposure the lowest survival in MT treatment was seen in the 500ng/l (28 total) and 250ng/l MT (25 total). Although overall survival was higher in the MT exposure compared to that of DHT, more variation between replicate tanks was observed with MT, with more than double the number of snails surviving in one 250ng/l MT replicate compared to the other (Figure 3.6B).

Shell diameters in the DHT exposure ranged from 6.78 to 15.76mm with an average of 11.52mm over the whole experiment, wet weight ranged from 0.108g to 0.701g, with an average of 0.377g (Figure 3.7 A, C). The mean diameter ranged from dilution water control (DWC) (12.04 ± 1.52mm) to 62.5ng/l (10.39 ± 1.32mm), 125ng/l DHT had the largest mean diameter (12.83 ± 1.49mm) and 500ng/l DHT (11.93 ± 1.66mm) had a smaller mean diameter. The mean wet weight also followed similar pattern decreasing from dilution water control (DWC) (0.383 ± 0.117 g) to 62.5ng/l (0.300 ± 0.072 g), 125ng/l had the largest mean weight (0.473 ± 0.124 g) and then 500ng/l DHT (0.428 ± 0.113 g). No intra replicate differences were found in the shell diameter and weight of the DHT treated snails (Figure 3.8A). Although significant differences (P <0.05) were found between the shell diameter and weight of the snails from the DWC and DHT dosed snails (62.5, 125 and 500ng/l). However when compared to the solvent control (SC) treatment, no dose dependent trend was seen for mean shell diameter or wet weight, with the largest snails occurring in the 125ng/l DHT treatment and the smallest from the 62.5ng/l DHT treatment (Figure 3.7A, C). There was also significant variability in snail size within the 62.5ng/l DHT treatment (Figure 3.8A).

Snails from the MT exposure were generally smaller than those in the DHT experiment, ranging from 2.97 to 15.03mm shell diameter, with an average of 9.42mm (Figure 3.7 B, D). However, the mean total weight of the snails from the dilution water control (DWC) was the largest (10.86 ±1.26) and smallest in 250ng/l MT (7.57 ±1.9) and 500ng/l MT (8.93 ±2.65) compared to other treatments (Figure 3.7D). Individual wet weight of snails from the MT experiment ranged from 0.023g to 0.685g with an average of 0.260g. The mean wet weight was the largest in DWC snails (0.332 ± 0.071) and followed a decreasing trend in the exposed snails and 250ng/l was the smallest (0.153 ± 0.096) (Figure 3.7B). Significant intra replicate differences both, in the shell diameter and total wet weight were found within a number of treatments (DWC (P = 0.013), SC (P <0.001), 125ng/l MT (P <0.001) and 500ng/l MT (P = 0.006)) (Figure 3.8B). Due to lack of uniformity between
replicates further ANOVA analysis was not conducted for diameter and wet weight in the MT study.

Figure 3.7: Boxplots of shell diameter (A, B) and wet weight (C, D) of B. glabrata developmentally exposed to Dihydrotestosterone (A, C) and Methytestosterone (B, D).

Boxes represent median (full line) with 25th and 75th percentiles. The bars extend to the 10th and 90th percentiles, with outliers represented as dots. * p< 0.05 compared to the Solvent Control group. Sample size is indicated above each treatment label on the x-axis.

Figure 3.8: Mean shell diameter of snails from replicate tanks exposed to (A) Dihydrotestosterone (DHT) and (B) Methytestosterone (MT).

Figure showing intra replicate differences in shell diameter of snails exposed to dilution water control (DWC), solvent control (SC), 62.5ng/l, 125ng/l, 250ng/l and 500ng/l of DHT (A) and MT (B) for 23 days post hatch.
To assess possible impacts of androgens (MT/DHT) on somatic development of the snails, a linear regression of snails shell diameter and total body weight (square root transformation of weight) was conducted. Significant linear correlation was observed for both the DHT ($R^2 = 0.884$, $P = 0.003$) and MT ($R^2 = 0.976$, $P = 0.001$) studies (Figure 3.9A and B). For both studies no obvious deviation from the expected pattern of growth was found for any of the treatments i.e. all treatments followed the similar growth pattern as the controls.

Significant relationships were found between density (number of surviving snails per litre) and diameter as well as density and wet weight in both the MT and DHT exposures ($P<0.05$) (ANCOVA) (Figure 3.10 A, B). However, the two studies showed opposite trends, in the DHT study, a significant negative trend was observed between mean shell diameter and snail density ($R^2 = -0.402$, $P = 0.026$) (Figure 3.10 A) and also between mean wet weight and snail density ($R^2 = -0.502$, $P = 0.009$) meaning tanks with lower survival had larger snail size (both weight and shell diameter). Whereas in the MT study, both the mean diameter ($R^2 = .384$, $P = 0.031$) (Figure 3.10 B) and the mean wet weight ($R^2 = .424$, $P = 0.021$) of the snails was significantly positively correlated to snail density, therefore, in this case tanks with higher survival (and therefore higher density) was associated with larger snail size.

Figure 3.9: Square root transformation of the total body weight as a function of shell diameter for the snails exposed to (A) Dihydrotestosterone and (B) Methyltestosterone.

The figure shows coefficient of linear regression ($R^2$) and statistical significance at ($P<0.05$) of B. glabrata exposed to dilution water control (DWC), solvent control (SC), 62.5ng/l, 125ng/l, 250ng/l and 500ng/l.
Figure 3.10: Linear regression of B. glabrata mean shell diameter (mm) against density (no. of snails per litre) from the (A). DHT exposure and (B). MT exposure. Figure showing weak but significant correlation in both experiments. Each treatment tank is represented by a dot (DHT: violet and MT: pink). Linear trend line, Regression coefficient ($R^2$) and significance ($P$) are highlighted on each graph.

3.4.3 Quantitative assessment of the size of reproductive glands

To assess possible androgenic effects on growth and development of reproductive glands in developmentally exposed B. glabrata, the size or weight of reproductive tissues were compared. In the DHT study, no significant intra-replicate variation was found for normalized glandular complex (NGC) size, although no comparisons could be made for the 250ng/l dose due to very small sample size ($n = 2$). No significant treatment related effects on NGC size was observed ($P>0.05$). Overall the mean NGC size was the largest in 62.5ng/l ($0.236±0.052$), similar size in DWC, 500ng/l DHT and SC ($0.219±0.077$, $0.214±0.042$ and $0.20±0.067$) and smallest in 125ng/l DHT ($0.191±0.065$) (Figure 3.11 B and Table 3.2).

Table 3.2: Mean values (±SD) of normalized glandular complex surface area (mm$^2$), normalized albumen gland weight (g) and normalized ovotestis weight (g) for B. glabrata developmentally exposed to DHT. Asterisks represent a significant difference compared to the solvent control (SC).

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Treatment</th>
<th>DWC</th>
<th>SC</th>
<th>62.5</th>
<th>125</th>
<th>250</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Glandular complex</td>
<td></td>
<td>0.219±0.077</td>
<td>0.20±0.067</td>
<td>0.236±0.052</td>
<td>0.191±0.065</td>
<td>xxxxxx</td>
<td>0.214±0.042</td>
</tr>
<tr>
<td>Normalized Albumen gland weight</td>
<td></td>
<td>0.019±0.007</td>
<td>0.032±0.009</td>
<td>0.029±0.009</td>
<td>0.023±0.008</td>
<td>xxxxxx</td>
<td>0.023±0.006</td>
</tr>
<tr>
<td>Normalized Ovotestis weight</td>
<td></td>
<td>0.023±0.008</td>
<td>0.021±0.009</td>
<td>0.21±0.006</td>
<td>0.019±0.005</td>
<td>xxxxxx</td>
<td>0.018±0.003</td>
</tr>
</tbody>
</table>

Table 3.3: Mean values (±SD) of normalized glandular complex surface area (mm$^2$), normalized albumen gland weight (g) and normalized ovotestis weight (g) for B. glabrata developmentally exposed to MT. Asterisks represent a significant difference compared to the solvent control (SC).

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Treatment</th>
<th>DWC</th>
<th>SC</th>
<th>62.5</th>
<th>125</th>
<th>250</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Glandular complex</td>
<td></td>
<td>0.243±0.132</td>
<td>0.21±0.069</td>
<td>0.287±0.098</td>
<td>0.209±0.047</td>
<td>0.253±0.099</td>
<td>0.23±0.136</td>
</tr>
</tbody>
</table>
In the MT study no dose dependent trend was observed in the size of the normalized glandular complex (NGC). The gland size decreased from DWC to solvent control. The gland size was largest in 62.5ng/l MT and smallest in 125ng/l MT (Table 3.3). However significant intra replicate differences were found for NGC size in 62.5ng/l (P = 0.027) nominal treatment (Figure 3.11A) in this experiment.

In DHT exposed snails no intra replicate differences were found for normalized albumen gland weight (NAW). Snails from the dilution water control (DWC) had smallest normalized albumen gland weight (0.019±0.007 g) compared to all the other treatments (Figure 3.11D and Table 3.2). No comparisons could be made for the 250ng/l DHT dose due to very small sample size. The mean normalized albumen gland weight (NAW) in the MT treatment was the largest in 62.5ng/l (0.036±0.018 g) compared to the controls and different treatments (Figure 3.11C and Table 3.3) but significant intra replicate differences were found for NAW size in 62.5ng/l (P = 0.034) and 125ng/l (P=0.042) nominal treatment.

The mean normalized ovotestis weight (NOW) seemed to follow a decreasing trend in the DHT treatment with the largest NOW in dilution water control (DWC) and smallest in the 500ng/l nominal treatment (Table 3.2 and Figure 3.11F). However no significant treatment related effects were found for NOW in the DHT study. No comparisons could be made for the 250ng/l DHT dose due to very small sample size. In MT exposure, the mean normalized ovotestis weight (NOW) was largest in DWC (0.027±0.027 g) and smallest in SC (0.017±0.007 g) (Table 3.3 and Figure 3.11E). However, there was significant intra-replicate variation in ovotestis weight in the solvent control (SC) treatment (P = 0.014) in the MT treatment.
3.4.4 Histopathological analyses of reproductive tissues

To assess possible androgenic effects on the reproductive tissue structure or cell type, histopathology was conducted on the reproductive tissues (albumen gland, glandular complex and ovotestis) from developmentally exposed B. glabrata from both MT and DHT studies. For these experiments RNAlater was used to preserve whole de-shelled
snails, enabling quick preservation followed by later dissection of tissues for both molecular and histopathological analysis. RNAlater did sufficiently preserve the tissue and aid dissection (firmed up snail tissue) and had previously been reported to have similar histopathology fixation results to formalin (Florell et al., 2001). However, compared to Bouin’s solution (more often used in our laboratory for histopathology of reproductive tissues) RNAlater was not as fine a histology fixative with some tissue integrity and fine structural detail lost (Figure 3.12(A-F)) but it was good enough for my purpose to determine that chemicals did not affect these tissues. For example fine details of oocyte development stage could not be determined due to disruption to the shape of the oocytes, supporting cells (follicles) and follicle cleft. Much of the shape, size and staining of the acini, including the structure of the acini walls were distorted. However, some spermatogenic stages, spermatids and spermatogonia as well various stages of oogenesis as described by De Jong-Brink et al. (1976, 1977) could be identified (De Jong-Brink et al., 1977; De Jong-Brink et al., 1976). In RNAlater fixed GC staining was muted and differential staining between different tissues types (membrane gland, mucous gland and prostate gland) was less pronounced. Sectioning artifacts also increased in the tissues due to slight hardening up of the tissue. Therefore RNAlater is not recommended as a fixative, especially for gonadal or ovotestis tissues. Histopathology was conducted on the ovotestis; albumen gland and glandular complex of all exposed and control snails from both the MT and DHT exposures.
Figure 3.12: Photomicrographs of reproductive tissues (Albumen gland (A-B), Glandular complex (C-D), and Ovotestis (E-F)) preserved in Bovin’s (A, C, E) and in RNAlater followed by Bouin’s (B, D, F).

Tissue section stained with H & E. Scale bar 100μm in each case. Staining has been affected in albumen gland and prostate diverticulum and some wear and tear can be seen, but ovotestis structure looks degraded. C: connective tissue, E: epithelium, L: lumen of a gland tubule, MG: mucous gland, Memgl: Membrane gland and Pd: prostate diverticulum, Ooc: Oocyte, Oog: Oogonia, Sc: Sertoli cells, Spd: Spermatid, Spg: Spermatogonia.
Figure 3.13: (A-F) Pictomicrographs of part of albumen gland tissue rom DWC, SC, 62.5ng/l, 125ng/l, 250ng/l and 500ng/l of 17α-MT exposed snails. Tissue sections stained with H&E. Scale bar 100μm in each case. C: connective tissue, E: epithelium, L: lumen of a gland tubule.

The histopathological results in both the MT (Figure 3.13(A-F)) and the DHT (Figure 3.14(A-E)) studies revealed no significant compound/dose related morphological changes in the albumen gland. Structurally, the albumen gland in *B. glabrata* consists of densely packed tubules consisting of columnar glandular epithelial cells with large and conspicuous nuclei towards their base (that stained pink) and apically located cuboidal secretory vesicles (which stained dark pink to violet) (Klussmann-Kolb, 2001). Some of the glandular epithelial cells were heterogeneously stained but this was observed in all
treatments (including the DWC) irrespective of the dose. This could be due to variation in the granules and cytoplasm present in the secretory cells, depending on the egg laying process (Klussmann-Kolb, 2001; De Jong-Brink, 1973). The tissues become slightly grainier when RNa later was used and this has been also reported by Guo and Catchpoole (2003) but they reported no adverse effects on tissue integrity (Guo D, 2003). There was no evidence of the albumen gland changing in structure or becoming more like a prostate gland, as has been reported in other gastropod species exposed to environmental ‘androgens’ (Watermann et al., 2008) (COMPRENDO, Final Report 3/2006), (http://ec.europa.eu/). The GC comprised of membrane gland, mucous gland and prostate gland (Figure 3.15(A-F) and Figure 3.16(A-E)). The histological sections of GC from both the MT study and the DHT study did not show any chemical-related or dose-related morphological changes. The membrane gland (Klussmann-Kolb, 2001) or the oothecal gland (De Jong-Brink et al., 1979) consists of glandular cells (stained light pinkish) filled with secretory fluid in the form of small mucous droplets that stained violet. The membrane gland showed a continuous transition to the mucous gland that was the largest part of the GC. The mucous gland consisted of a glandular mass made up of glandular cells that were columnar in shape, widely coiled and stained light pinkish. Interspersed were cells containing small densely packed secretory droplets that stain dark pinkish. The prostate gland consists of the prostate duct and its diverticula contains long glandular cells with distinct vesicles of homogeneous or amorphous texture with conspicuous nuclei that normally stain blue (Klussmann-Kolb, 2001) but due to the RNa later stained dark pink in these experimental samples. The supporting cells bear short cilia and stained pink to violet.

All ovotestis from the MT and DHT exposures were initially assessed for chemical dose related effect on oogenesis and spermatogenesis. As far as could be determined there were no treatment related effects, all stages of oogenesis and spermatogenesis, as described by De Jong-Brink et al. (1976, 1977), were present (De Jong-Brink et al., 1977; De Jong-Brink et al., 1976). However, due to preservation in RNa later detailed analysis (e.g. scoring) of possible alterations to spermatogenesis or oogenesis was not possible.
Figure 3.14: (A-F): Photomicrographs of part of albumen gland tissue from DWC, SC, 62.5ng/l, 125ng/l and 500ng/l of DHT exposed snails.

Tissue section stained with H & E. Scale bar 100μm in each case. Due to very high mortality observed in 250ng/L, the tissue from two survivor snails has not been included. Histopathological analysis did not show any significantly obvious effects on the albumen gland of the exposed groups as compared to controls (DWC and SC). C: connective tissue, E: epithelium and L: lumen of a gland tubule.
Figure 3.15: (A-F): Photomicrographs of part of glandular complex tissue from DWC, SC, 62.5ng/l, 125ng/l, 250ng/l and 500ng/l of 17α-MT exposed snails. Tissue section stained with H & E. Scale bar 100μm in each case. Glandular complex comprises of mucous gland, membrane gland and prostate diverticula. Histopathological analysis did not show any significantly obvious effects on the albumen gland of the exposed groups as compared to controls (DWC and SC). MG: mucous gland, Memgl: Membrane gland and Pd: prostate diverticulum.
Figure 3.16: (A-E): Photomicrographs of glandular complex tissue from DWC, SC, 62.5ng/l, 125ng/l and 500ng/l of DHT exposed snails. Tissue section stained with H & E. Scale bar 100μm in each case. Due to very high mortality observed in 250ng/l, the tissue from two survivor snails has not been included. Glandular complex comprises of mucous gland, membrane gland and prostate diverticula. Histopathological analysis did not show any significantly obvious effects on the albumen gland of the exposed groups as compared to controls (DWC and SC). MG: mucous gland, Memgl: Membrane gland and Pd: prostate diverticulum.
3.5 Discussion

The first part of my hypothesis was tested by *in vivo* developmental exposure of the freshwater snail *B. glabrata* to 5α-Dihydrotestosterone and 17α-Methyltestosterone, compounds known to be potent androgens in vertebrates. The primary aim of this study was to investigate whether developmental exposure to these androgens would affect the differentiation, development and growth of the reproductive organs and glands of the freshwater snail *B. glabrata* thus providing a non-vertebrate alternative model for male reproductive toxicity testing. The main result of the two chemical exposure studies is that no chemical related dose-dependent effects on sexual development were found, in terms of reproductive organ development, using any of the morphological and histopathology endpoints.

3.5.1 Possible effects on survival and growth

The general pattern in hatchling survival in both the DHT and MT studies was that mortality was higher in the 250ng/l and 500ng/l (MT or DHT) than the dilution water control (DWC) or solvent control (SC) treatments. This suggests there may have been some level of toxicity to the developing snails at these concentrations (although mortality in the MT study was generally lower than in the DHT study). However, toxicity has not been previously reported in other gastropod species exposed to similar nominal concentrations of androgens (e.g. (Rivero-Wendt *et al.*, 2014; Giusti *et al.*, 2014; Duft *et al.*, 2007; Czech, Weber and Dietrich, 2001; Tillmann *et al.*, 2001)) and the dosing series for these experiments were designed to mimic concentrations documented to have endocrine effects in molluscs (e.g. (Duft *et al.*, 2007; Czech, Weber and Dietrich, 2001; Tillmann *et al.*, 2001) and aquatic vertebrates (e.g. (Margiotta-Casaluci and Sumpter, 2011; Seki *et al.*, 2004) but not acute toxic effects. The experimental design of the other gastropod studies differs from this study, in that many were static-renewal rather than continuous dosing using flow-through water system. In addition those studies which measured exposure concentrations in the water often found much lower (sometimes <20%) than nominal exposure concentrations (Giusti *et al.*, 2014), whereas in this study we measured exposure concentrations of 80-110% nominal MT and 65-70% nominal DHT. Therefore snails in these experiments may have been exposed to higher concentrations than similar species in other studies.

Growth is an important factor to consider when assessing the reproductive organ size
endpoints. In vertebrates, exposure to androgens has been reported to increase somatic growth, for example in developmentally exposed fathead minnows (*Pimephales promelas*), DHT significantly increased weight and length in males (20ng/l DHT) and females (200ng/l DHT) (Margiotta-Casaluci and Sumpter, 2011). Not much is known about effects of androgens on somatic growth in molluscs but somatic growth was enhanced in *Radix balthica* (a pulmonate snail) when exposed to 5 µg/l 17α-Ethylestradiol (Hallgren et al., 2012). Mimicking the rodent prepubertal intact male Hershberger assay protocol, *B. glabrata* was exposed to chemicals from egg mass to sexual maturity. Although some significant differences in size were found between the solvent control (SC) and DHT exposed *B. glabrata*, no dose dependent trend could be identified, with significantly larger snails in the 125ng/l DHT treatment and significantly smaller snails in the 62.5ng/l DHT treatment, indeed significant differences were also found between the SC and the dilution water control (DWC). Solvents are used as carriers in aquatic toxicology testing because at high concentrations (such as stock solutions) some test chemicals have low water solubility. The possibility of ‘solvent effects’ have arisen in a number of publications, and the use of solvents is a topic of concern in endocrine toxicology testing (Turner et al., 2012; Hutchinson et al., 2006; Zhang and Baer, 2000). The OECD guidelines suggested maximum solvent concentration of 0.01% v/v (100µl/l) for aquatic vertebrates (OECD 2000) (http://www.epa.gov/endo/pubs/ref-2_oecd) and gastropods (OECD 2010) (http://www.oecd.org/). In this study the nominal concentration of dimethylformamide (DMF) in the SC, MT and DHT tanks was 0.006% v/v (60ul/l), and therefore below the OECD recommendations. However, in a review of the use of solvents in aquatic toxicology testing, Hutchinson et al. (2006) suggest using a maximum concentration of 0.002% solvent, and more recently, Turner et al. (2012) and Lecomte et al. (2013) have reported molecular and reproductive effects at solvent concentrations of 0.0025 and 0.002% v/v respectively (Lecomte et al., 2013; Turner et al., 2012; Hutchinson et al., 2006). Notwithstanding this, studies in our laboratory have used similar concentrations of DMF (0.0067% v/v) to those used in the *B. glabrata* tests and found no significant solvents effects on growth or condition index in fish (e.g. (Margiotta-Casaluci and Sumpter, 2011). All these studies highlight the importance of testing solvent controls alongside dilution water controls and comparing chemically treated subject to the appropriate controls when analyzing outcomes. Interestingly, in the MT study a slightly different pattern emerged, in this study the largest snails were found in the DWC treatment, whereas MT exposed snails were on average smaller than the SC or DWC snails, suggesting that there may have been a negative effect on growth of MT in *B.
*glabrata* in the higher concentrations, however due to significant intra dose variation (i.e. between replicate tanks) no statistically significant effects could be seen. In general the variation in shell size between the replicate tanks, in the DHT exposure was much lower than in the MT exposure and much of this variability seems to be caused not by the higher levels of mortality in the DHT exposure but by the higher survival in the MT exposure consisting of smaller, possibly not well grown/developed snails at the end of the 30 exposure. In the DHT study only one snail was less than 7mm in diameter at the termination of exposure compared to 36 snails in the MT experiment, in the MT study these small snails were found in all treatments bar the dilution water control (DWC).

### 3.5.2 Possible effects of androgens on reproductive tissue size and histopathology

One of the primary markers of androgen exposure in vertebrate models, such as the Hershberger test, is increase size/weight of male specific organs and a reduction in those that are female specific. For example, in a pubertal rat assay (enhanced OECD test guideline 407) MT exposure resulted in increased prostate weight in male rats treated with 200 mg/kg/day (Wason *et al.*, 2003). In contrast no chemical related effect on normalized glandular complex (NGC) (including prostate gland) size was found for *B. glabrata* snails exposed to any of the DHT concentrations tested. Although the normalized glandular complex (NGC) of the 62.5ng/l MT exposed snails were much larger than those from the SC treated snails, there were also significant differences in NGC between the two 62.5ng/l MT replicate tanks, suggesting this was not an effect of MT exposure.

There were no consistent significant dose-related effects on albumen gland weight in the MT or DHT exposure studies. In the MT study due to intra replicate differences in the normalized albumen gland size (NAW) no statistically significant effects could be assessed. In the DHT study, the snails from the dilution water control (DWC) had significantly smaller albumen glands compared to all the other treatments, suggesting the addition of the carrier solvent (DMF), rather than the test compound, had positive effects on the snails albumen gland, perhaps by providing an additional source of nutrition for the production and accumulation of secretory substances. In general, gastropod albumen gland size is thought to be related to accumulation of secretory products in the tissue, due to different stages of egg-laying cycle (Egonmwan, 2007; Miksys and Saleuddin, 1987; De Jong-Brink *et al.*, 1979) but this variability factor will hold true for all the snails irrespective of the chemical exposure.

In vertebrate models, histopathological changes to gland structure or development are also reported in androgen exposed individuals, for example, in the pubertal male rats histology...
of the seminal vesicle and prostate gland show enhanced glandular development (1 mg/kg/day Testosterone Propionate by subcutaneous injection) (Tinwell et al., 2007). In comparison, no histological impact of MT or DHT exposure was observed in the albumen gland or the Glandular Complex (GC) of androgen-exposed B. glabrata. Differences were observed between individual snails in the staining of the epithelium of the tubules in the albumen gland, but these variations were found across all treatments, and were due to the presence of more vacuolated cytoplasm and less secretory granules in the epithelium (De Jong-Brink, 1973), possibly due to variation in the egg-laying cycle of the snails at the time of sampling. Varying structural appearances of the albumen gland in conjunction with different physiological states due to metabolic processes and sexual activity has also been reported by Duncan (Duncan, 1960). The lack of histological effects of MT or DHT exposure on secondary sexual glands in B. glabrata is in contrast to previously published result in another freshwater pulmonate species, L. stagnalis, which showed weak histological damage to the albumen after 8 weeks exposure to 100ng/L MT (Czech, Weber and Dietrich, 2001). The timing of exposure (adult versus juvenile) and duration of exposure (8 weeks versus 30 days) could be an important factor deciding the range of effects seen in these organisms.

Although a detailed histopathology comparison of androgen exposure on the ovotestis could not be conducted due to sub-optimal tissue preservation in RNAlater, there were no indications of effects at 10X resolution and no gross disruption to oogenesis or spermatogenesis were observed in any of the MT or DHT exposed snails. No significant effect of androgen exposure (MT or DHT) was found on the ovotestis weight (although some differences were found between the ovotestis size between the dilution water control (DWC) and solvent control (SC) in the MT study). The lack of gross histopathological changes to the ovotestis in B. glabrata exposed to MT is in contrast to the recent work of Rivero-Wendt et al. (2014), in which adult B. glabrata were exposed semi-statically to MT (0.01mg/l to 1mg/l MT). In that study, Rivero-Wendt et al. (2014) reported that nominal MT concentrations of 0.1 mg/l and 1.0 mg/l significantly increased the area of ovotestis acini occupied by mature spermatozoa. However, their statistical comparisons were between MT dosed and water control snails, rather than between MT dosed and solvent control (0.001% ethanol) exposed snails. Indeed, in Rivero-Wendt et al. (2014) the solvent control snails showed the largest average area occupied by sperm (almost double that of the water control) (Rivero-Wendt et al., 2014). This suggests MT may not have been the overriding factor in increasing sperm production in these snails and again highlights the importance of assessing, and factoring in, possible solvent mediated affects
in toxicology studies. A recent study from Bandow and Weltje (2012), reported enhanced embryonic development in *Lymnaea stagnalis* when exposed to 89.3μl/l of triethylene glycol. More recently accelerated embryonic development has also been reported in *P. antipodarum* at 20μl/l of acetone and inhibition of juvenile growth at similar concentration of ethanol (Lecomte et al., 2013).

No changes in gene expression were found in ovotestis from MT snails compared to the controls were seen using suppressive subtractive hybridization methods is conducted in our laboratory by Dr Anne Lockyer. Overall this suggests it is unlikely that there were androgenic effects on spermatogenesis or oogenesis in these animals as has been documented in other molluscs. Similar experimental procedures to those conducted with the snails i.e. developmental exposure to MT via the water, have been conducted on a number of fish species producing androgen-induced effects. For example, Japanese Medaka (*Oryzias latipes*), developmentally exposed to androgens via the water in a similar way to *B. glabrata* in this study, exhibited male skewed sex ratios and testis-ova at average concentrations of 27.75ng/l MT (Seki et al., 2004). Another fish species, the Fathead Minnow (*Pimephales promelas*), when exposed to MT as adults, displayed atretic follicles in their ovary (≥ 100ng/l MT) and inhibited ovarian growth (≥ 50μg/l MT) (Pawlowski et al., 2004).

The absence of significant dose-dependent physiological or molecular responses to MT or DHT on *B. glabrata* was surprising considering the number of publications that suggest gastropod molluscs have physiological levels of vertebrate-type androgens naturally occurring in their tissues; use vertebrate-type androgens in reproduction or are reproductively disrupted by exogenous androgen or anti-androgen exposure (Giusti et al., 2013; Rivero-Wendt et al., 2013; Goto et al., 2012; Schulte-Oehlmann et al., 2006; Janer et al., 2005b; Tillmann et al., 2001). For example, in the freshwater prosobranch gastropod, *Marisa cornuarietis*, MT showed imposex promoting effects from 100ng/l nominal to complete virilization of females at 1μg/l (Schulte-Oehlmann et al., 2006)and extreme alternation of albumen gland histopathology have been reported in gonochoristic prosobranch gastropods environmentally or experimentally exposed to TBT, TPT or MT (Watermann et al., 2008; Schulte-Oehlmann et al., 2006; Schulte-Oehlmann et al., 2004). However, it is worth highlighting that the majority of publications documenting effects of androgens are in prosobranch (generally gonochoristic) gastropods rather than pulmonate (hermaphrodite) gastropods as tested here. Considering the evolutionary radiation of gastropods it might not be unexpected that they might show different responses. Therefore, in terms of basic molluscan endocrinology research, the question that arises
from this work is that if the classic vertebrate steroids, such as androgens, are not controlling sexual development, spermatogenesis and oogenesis then what are the relevant genes and molecules involved? There is the well-known example of masculinizing (imposex) effects of the marine pollutant tributyltin (TBT) on prosobranch gastropods (Matthiessen and Gibbs, 1998) and study of this chemical in molluscs came up with several hypothesis. One of the groups suggested that TBT might be acting through androgen-mediated pathway (Oehlmann et al., 2007; Matthiessen and Gibbs, 1998) and another group in Japan suggested that TBTs virilizing effects were acting via the retinoid X receptor (RXR) (Horiguchi et al., 2007; Nishikawa et al., 2004). More recently Pascoal et al. (2012) demonstrated the involvement of both retinoids and androgen signaling in inducing imposex in molluscs (Pascoal et al., 2012).

3.6 Conclusion

The data presented provides convincing evidence that growth and development of the reproductive tract in B. glabrata are unaffected by developmental exposure (embryo through to sexual maturity) to MT and DHT, both potent androgens in vertebrates. The consequence of these findings are significant as it suggests that B. glabrata and possibly other molluscs species would not be appropriate models for testing androgenic or anti-androgenic compounds. The reported effects of androgens seen in other mollusc species could be due to alternative NRs interacting with androgens or alternative, perhaps non-genomic, pathways for androgen signalling. The next most feasible step was to investigate and identify nuclear receptors in the publically available genome for B. glabrata.

Nevertheless, it must be borne in mind that molluscs consist of a group with considerable diversity that might be reflected in its endocrine system. This disparity among different species, even within the gastropods, makes it difficult to extrapolate steroid exposure effects seen in one species to another and warrants further research to examine and understand mechanisms for gonadal development and other important phenomena in mollusces, which are second only to arthropods in the number of species occupying the widest variety of habitats in the world (Gruner H, 1993). All this should not limit the investigation and use of new invertebrate models in the field of toxicology (thus supporting important 3Rs strategies) but it does highlight the importance of having supporting molecular programs and mechanistically relevant endpoints for the proposed test species.
Chapter 4: Investigation of nuclear receptors in Biomphalaria glabrata and Lottia gigantea
4.1 Background

One of the main stimuli for this thesis was that the effects of androgenic and anti-androgenic chemicals on reproduction and development in crustaceans, molluscs and echinoderms seem to be comparable to those in mammals (For example (Oehlmann and Schulte-Oehlmann, 2003), the consequent hypothesis that androgens in snails must act through signal transduction pathways similar to those found in vertebrates and that androgen receptors must be present in molluscs was inferred (Omran, 2012; Koehler et al., 2007). The results of Chapter 3 however show clearly that vertebrate androgens do not support this hypothesis in my model organism, B. glabrata. In this chapter I therefore sought to investigate the nuclear receptor repertoire in Biomphalaria glabrata and compare these to known vertebrate nuclear receptors (NRs), with a view to assessing the snail’s potential as a new invertebrate model organism, both as a prospective new test organism and to elucidate fundamental genetic and mechanistic causes of endocrine diseases. For comparative purposes, the genome of second gastropod, the owl limpet, Lottia gigantea was also investigated for nuclear receptors.

4.2 Introduction to Hormone signalling

In vertebrates, the classical genomic model of hormone action is directly related to the ability of the steroid hormone to regulate the expression of genes containing DNA response elements in their promoter region. When the steroid hormone (ligand) binds to a specific intracellular steroid receptor protein (also called nuclear receptor), the bound ligand-activated steroid receptor complex acts as a transcription factor and binds as a homodimer/heterodimer to DNA response elements in the target gene promoter causing activation or repression of transcription and thus protein synthesis (Nikolenko and Krasnov, 2007; Shang, Myers and Brown, 2002). In addition, steroids may use a short-cut, referred to as the ‘rapid’ or non-genomic’ pathway in the absence of the nuclear receptor and operating through a membrane receptor that is possibly associated with the plasma membrane (Heinlein and Chang, 2002). This involves initiation of second messenger signal transduction cascades which leads to an increase in free intracellular calcium, activation of protein kinase A (PKA) and protein kinase C (PKC), resulting in activation of MAP kinase cascade and phosphorylation of cyclic AMP (cAMP) responsive element binding (CREB) protein transcription factor (Walker, 2010a). Steroid hormones can act through either of these mechanisms in order to regulate biological processes.
4.2.1 General structure of nuclear receptors

NRs are members of a large superfamily of evolutionary related multi-domain transcription factors that are believed to have emerged early in animal evolution, as reflected by their biologically essential roles (Bridgham et al., 2010; Bertrand et al., 2004; Owen and Zelent, 2000). Distinct NRs regulate and coordinate multiple physiological processes such as embryonic development, sex determination (Ganter et al., 2011; Finley et al., 1998; Shen et al., 1994), fatty acid metabolism and toxin metabolism (Chiang, 2002), detoxification of foreign substances by integrating internal and external signals, thereby maintaining homeostasis (Markov and Laudet, 2011a; Sladek, 2011). Outside of normal physiology, NRs have been identified to play a role in many endocrine related pathological processes, such as cancer (breast and prostate) (Peehl and Feldman, 2004), hormone-resistance syndromes, diabetes, and rheumatoid arthritis (Ottow and Weinmann, 2008; Moller, 2001).

Figure 4.0.1: Schematic representation of an nuclear receptor (NR).

The variable N-terminal region (A/B) contains the ligand independent AF-1 transactivation domain. The C region contains conserved DNA-binding domain (DBD). A hinge region D connects the DBD to the less conserved E region that contains the ligand binding domain (LBD) as well as the dimerization surface and the ligand-dependent AF-2 core activation domain.

Despite their diversity in function, NRs exhibit strong similarities due to their conserved molecular structure, which consists of single polypeptide chains that can be divided into five to six modular domains, from the N-terminal to C-terminal end (Mangelsdorf and Evans, 1995; Giguère et al., 1986) Figure 4.0.1. Although there is undoubtedly some interaction between the domains, each domain is responsible for encoding a motif that exerts a specified molecular action. The activation function 1 (AF-1), which is located on the N-terminal A/B domain, is important for the ligand-independent constitutive transcriptional activity of the NR (Wärnmark et al., 2003). The length and sequence of the A/B domain is hyper variable between receptors and receptor subtypes (RXRα and β) revealing a weak evolutionary conservation. This region contains a variety of kinase recognition sequences with specific attributed functions and is the most frequent site of alternative splicing. For these reasons, it is thought that the variable N-terminal sequences may be responsible for the species-, receptor-, promoter-context- and cell-specific
properties of NR transactivation (Germain et al., 2006; Renaud and Moras, 2000). The C domain or the DNA binding domain (DBD) is based on two highly conserved cysteine-rich zinc finger motifs and a carboxy-terminal extension (CTE). This is the most conserved region of the nuclear receptors. The first zinc finger contains the proximal-or P-box region, an alpha helix that is responsible for the sequence specificity of the receptor and DNA binding response element (Laudet et al., 2002). Located within the second zinc finger is the distal or D-box, which lies vertical to P-box helix, and is a site responsible for receptor dimerization (Aranda and Pascual, 2001). The CTE provides the interface for protein-DNA or protein-protein monomerization (Khorasanizadeh and Rastinejad, 2001). The hinge region is a mobile linker between the DBD and the C-terminal LBD facilitating the DBD and the LBD to adopt different conformations. Moreover, the hinge domain contains the C-terminal end of the DBD, which may be involved in identifying the 5’ end of the hormone response element (HRE). This region may also contain nuclear localization signals and serve as binding site for corepressor proteins (Aranda and Pascual, 2001). The LBD (E domain) region of NRs is a multifunctional domain that mediates homo- and heterodimerisation, ligand binding, interaction with coregulators, nuclear localization and transcriptional activity (Aranda and Pascual, 2001). The sequence of the LBD varies substantially between NRs although the overall architecture of the LBD is comprised of 11-13 anti-parallel α-helices. Helices 1-11 form a structure comprising a ligand-binding pocket that diverges sufficiently among the NRs to guarantee selective ligand recognition. The pocket is guarded by flexible helix 12 (H12), which is commonly referred to as the ligand-induced activation function (AF-2) and contains residues crucially involved in transcriptional coregulator interactions (Gronemeyer, Gustafsson and Laudet, 2004). A further structurally distinct region within the LBD is a dimerization surface, which mediates the interaction with the partner LBDs (Bourguet, Germain and Gronemeyer, 2000; Gampe Jr et al., 2000). The C-terminal F domain, which is contiguous with the E domain, is not present in all nuclear receptors and its function is poorly understood (Germain et al., 2006).

There are several NRs that do not follow this basic structure function paradigm. For example, the constitutive androstane receptor (CAR) undergoes constitutive transcription (in the absence of the ligand) of the target gene (Dussault et al., 2002). A subfamily (NR0) of nuclear receptor lacks either the DBD or the LBD. The dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX-1) and small heterodimer partner (SHP) are the two NRs in humans that lack the DBDs completely and knirps and embryonic gonad (EGON) in Drosophila lack the LBD completely (Germain
et al., 2006). A novel subfamily of NRs with two DBDs have been identified in flatworm species (Schmidtea mediterranea and Dugesia japonica), from molluscs (L. gigantea) and from arthropods (Daphnia pulex) (Wu et al., 2007) but not yet reported in mammals.

4.2.2 Mechanism of action

Extensive studies over the last three decades have provided a detailed understanding of the molecular mechanisms through which nuclear receptors convert a hormone signal into a transcriptional response. Nuclear receptors bind to DNA as a heterodimer, homodimer, or monomer, depending on the class of NR. The steroid hormone receptors glucocorticoid receptor (GR), progesterone receptor (PR), estrogen receptor (ER), androgen receptor (AR) and Mineralocorticoid receptor (MR) function as homodimer, binding to DNA inverted repeat with a gap of 3 nucleotide (AGAACAnnnTGTTCT). The thyroid hormone receptor (THR) and retinoid acid receptor (RAR) as well as many of the other orphan receptors, heterodimerize with retinoid X receptor (RXR) (Lefebvre, Benomar and Staels, 2010) and activate direct repeats of (GGTCA) and the gap between the repeat motifs depends on the heterodimer. For example RXR-RAR heterodimer has a gap of 5 nucleotides and RXR-THR has a gap of 4 nucleotides (Laudet, 1997). The ER activates a similar repeat to the THR but requires a 3bp gap like other steroid receptors (Forman et al., 1992). However, NR family 5 NRs (steroidogenic factor-1) bind as monomers to an extended core site (Mangelsdorf et al., 1995). Based on sequence homology, NRs have been further classified into subfamilies and owing to their mechanism of action NRs have been categorized into classes.

Class I NRs (includes members of NR subfamily 3) are located in the cytoplasm in the absence of ligands. Ligand (hormone) binding to Class I NRs lead to conformational change resulting in dissociation of heat shock proteins, see Figure 4.0.2(A). Translocation to the nucleus is followed by steroid receptor homodimerisation and binding to specific hormone response elements (HREs). The NR-DNA complex in turn recruits other proteins that are responsible for the transcription of downstream DNA into mRNA. Subsequent translation of this mRNA into protein results in changes in cellular functions.

Class II NRs (NR subfamily 1) are located in the nucleus regardless of their ligand binding status and form obligate heterodimers with RXR to form a functional transcription factor, see Figure 4.0.2(B). In the absence of the ligand, the receptor is bound to corepressor proteins, which is dissociated in the presence of the ligand followed by association with coactivator proteins. These coactivator proteins recruit additional proteins such as RNA polymerase II forming a complex that bring about the translation of the downstream DNA
into mRNA and eventually protein synthesis takes place (Rosenfeld, Lunyak and Glass, 2006; Glass and Rosenfeld, 2000).

Both Class III NRs (NR family 2) and Class IV NRs (NR family 5) mostly comprise of orphan receptors (Mangelsdorf et al., 1995). Class III NRs bind to direct repeats of the HRE as homodimers while Class IV NRs bind as monomers or dimers but only a single half site of the inverted repeat of the HRE, that is preceded by a 5’- flanking A/T rich sequence, is bound to the DNA binding domain of the receptor (Aranda and Pascual, 2001). Monomeric NRs use C-terminal extension (CTE) to recognize their HRE (Glass, 1994).

4.2.3 Species distribution of nuclear receptors and their nomenclature

The NR superfamily genes are specific to metazoans and are not found in algae, fungi (Saccharomyces cerevisiae), yeast, choanoflagellates and plants (Arabidopsis thaliana) (King et al., 2008; Escriva, Bertrand and Laudet, 2004; Launhardt and Munder, 2000). However, there are some suggestions that fungi can use ligand regulated transcription factors that might share structural and functional features with metazoan NRs (Naar and Thakur, 2009; Phelps et al., 2006). With recent progress in genome sequencing projects from the early-branching animal lineages, 2 NRs have been reported from the sponge Amphimedon queenslandica (Bridgham et al., 2010; Srivastava et al., 2010), 2 from the
ctenophore *Mnemiopsis leidy* (Reitzel *et al.*, 2011), 4 from the placozoa *Trichoplax adhaerens* (Bridgham *et al.*, 2010; Srivastava *et al.*, 2008) and 17 from the cnidarian *Nematostella vectensis* (Reitzel and Tarrant, 2009). Over 284 NRs have been reported in *Caenorhabditis elegans* (Sluder and Maina, 2001). Among arthropods, 21 NR genes are found in *Drosophila melanogaster* (Adams *et al.*, 2000), 22 in *Apis mellifera* (Velarde, Robinson and Fahrbach, 2006), and 21 in *Tribolium castaneum* (Bonneton, Chaumot and Laudet, 2008). 21 NRs have also been identified in the trematode *Schistosoma mansoni* (Wu *et al.*, 2006) and recently 43 NRs were identified from the Pacific oyster, *C. gigas*, a marine bivalve, (Vogeler *et al.*, 2014a). Moving on to vertebrates, 48 NRs have been identified in the human genome (Robinson-Rechavi, Garcia and Laudet, 2003), 49 in mice and 47 in rats (Zhang *et al.*, 2004). The unusually large number of NRs in *C. elegans* is due to extensive proliferation of one gene (HNF4) within the nematode phylum (Antebi *et al.*, 2000).

Nuclear receptors are strong phylogenetic markers showing diversification of metazoans from the simplest sponges to vertebrate-specific paralogues at different evolutionary distances (García, Laudet and Robinson-Rechavi, 2003) through the structural conservation of their various domains. Comparative studies examining the evolution of various protein families and their nomenclature in major groups of organisms have been done using various approaches to phylogeny. Based on sequence homology using phylogenetic approaches the Nuclear Receptors Nomenclature committee (1999) organized NRs into six distinct subfamilies (Escriva, Delaunay and Laudet, 2000; Thornton and DeSalle, 2000). Each of these subfamilies is a robust monophyletic group, in which all receptors clustered in a subfamily originate from a single ancestor (Markov and Laudet, 2011b) (Figure 4.0.3). The committee defined a gene nomenclature containing the nuclear receptor abbreviation NR followed by subfamily, group and receptor number. Subfamily 1 includes thyroid hormone receptors (THR)-like receptors; Subfamily 2 has RXR-like receptors; Subfamily 3, estrogen receptors (ER)-like receptors; Subfamily 4, nerve growth factor IB (NGFIB)-like receptors; Subfamily 5, steroidogenic factor like receptors; Subfamily 6, germ cell nuclear factor (GCNF)-like receptors and the last Subfamily 0 includes all the nuclear receptors lacking either of the DBD or LBD regions (Zhang *et al.*, 2004; Maglich *et al.*, 2001) (Figure 4.0.3). In Table 4.0.1 all the members of NR superfamily are listed with their subfamily and groups, nomenclature, gene name, accession numbers and their respective ligands (Chawla *et al.*, 2001).
Table 4.0.1: Nomenclature of nuclear receptors and their associated ligands.
(Germain et al., 2006; King-Jones and Thummel, 2005; Smirnov, 2002)

<table>
<thead>
<tr>
<th>Subfamilies and groups</th>
<th>Name</th>
<th>Nomenclature</th>
<th>Accession number</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>THRα</td>
<td>NR1A1</td>
<td>AAH08851</td>
<td>Thyroid hormone</td>
</tr>
<tr>
<td></td>
<td>THRβ</td>
<td>NR1A2</td>
<td>AAJ06931</td>
<td>Thyroid hormone</td>
</tr>
<tr>
<td>1B</td>
<td>RARα</td>
<td>NR1B1</td>
<td>AAH08727</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td></td>
<td>RARβ</td>
<td>NR1B2</td>
<td>AAH60794</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td></td>
<td>RARγ</td>
<td>NR1B4</td>
<td>AF378827</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td>1C</td>
<td>PPARα</td>
<td>NR1C1</td>
<td>AAB32649</td>
<td>Fatty acids, fibrates</td>
</tr>
<tr>
<td></td>
<td>PPARβ</td>
<td>NR1C2</td>
<td>AAA36468</td>
<td>Fatty acids</td>
</tr>
<tr>
<td></td>
<td>PPARγ</td>
<td>NR1C3</td>
<td>AAH06811</td>
<td>Fatty acids, prostaglandins</td>
</tr>
<tr>
<td>1D</td>
<td>Rev-erb-a</td>
<td>NR1D1</td>
<td>AAH56148</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Rev-erb-b</td>
<td>NR1D2</td>
<td>AAH45613</td>
<td>Orphan</td>
</tr>
<tr>
<td>1E</td>
<td>RORα</td>
<td>NR1E1</td>
<td>AAH08931</td>
<td>Cholesterol</td>
</tr>
<tr>
<td></td>
<td>RORβ</td>
<td>NR1E2</td>
<td>AAH3774</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td></td>
<td>RORγ</td>
<td>NR1E3</td>
<td>AAH3774</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td>1F</td>
<td>VDR</td>
<td>NR1I1</td>
<td>AAB95155</td>
<td>Vitamin D</td>
</tr>
<tr>
<td></td>
<td>PXR</td>
<td>NR1I2</td>
<td>AAD05436</td>
<td>Xenobiotics, steroids</td>
</tr>
<tr>
<td></td>
<td>CAR</td>
<td>NR1I3</td>
<td>AAY56401</td>
<td>Xenobiotics, Phenobarbital</td>
</tr>
<tr>
<td>1J</td>
<td>HR96</td>
<td>NR1J1</td>
<td>AAC46928</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Daf-12</td>
<td>NR1J2</td>
<td>AAD34462</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-8</td>
<td>NR1J3</td>
<td>AAP11437</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-48</td>
<td>NR1J4</td>
<td>CAD6502</td>
<td>Orphan</td>
</tr>
<tr>
<td>1K</td>
<td>HR-1</td>
<td>NR1K1</td>
<td>AAC48174</td>
<td>Orphan</td>
</tr>
<tr>
<td>2A</td>
<td>HNF4</td>
<td>NR2A1</td>
<td>NP_849180</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HNF4α</td>
<td>NR2A2</td>
<td>NP_004124</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HNF4B</td>
<td>NR2A3</td>
<td>Z49827</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>DHHNF4</td>
<td>NR2A4</td>
<td>AAF52702</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-49</td>
<td>NR2A5</td>
<td>CAD5702</td>
<td>Orphan</td>
</tr>
<tr>
<td>2B</td>
<td>RXRa</td>
<td>NR2B1</td>
<td>AAII0999</td>
<td>9-cis retinoic acid</td>
</tr>
<tr>
<td></td>
<td>RXRb</td>
<td>NR2B2</td>
<td>AAC18599</td>
<td>9-cis retinoic acid</td>
</tr>
<tr>
<td></td>
<td>RXRγ</td>
<td>NR2B3</td>
<td>AAG0681</td>
<td>9-cis retinoic acid</td>
</tr>
<tr>
<td></td>
<td>USP</td>
<td>NR2B4</td>
<td>AAF5707</td>
<td>20-Hydroxyecdysone</td>
</tr>
<tr>
<td>2C</td>
<td>TR2</td>
<td>NR2C1</td>
<td>AAJ036761</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>TR4</td>
<td>NR2C2</td>
<td>NP_003289</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>TR2-4</td>
<td>NR2C3</td>
<td>AF378828</td>
<td>Orphan</td>
</tr>
<tr>
<td>2D</td>
<td>HR78</td>
<td>NR2D1</td>
<td>Q24142</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-41</td>
<td>NR2D2</td>
<td>NP_500073</td>
<td>Orphan</td>
</tr>
<tr>
<td>2E</td>
<td>TLL/TLX</td>
<td>NR2E1</td>
<td>AALDS871</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>PNR</td>
<td>NR2E2</td>
<td>AAD28301</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>DSRF</td>
<td>NR2E3</td>
<td>AAF52303</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>FAX-1</td>
<td>NR2E4</td>
<td>AAF54133</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-67</td>
<td>NR2E5</td>
<td>CAH97482</td>
<td>Orphan</td>
</tr>
<tr>
<td>2F</td>
<td>COUP-TFα</td>
<td>NR2F1</td>
<td>AAO04154</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>COUP-TFβ</td>
<td>NR2F2</td>
<td>AAO2897</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>SVP</td>
<td>NR2F3</td>
<td>MZ8863</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>EAR2</td>
<td>NR2F4</td>
<td>AAH02699</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>UNC55</td>
<td>NR2F5</td>
<td>CAP16277</td>
<td>Orphan</td>
</tr>
<tr>
<td>2H</td>
<td>AmNR4</td>
<td>NR2H1</td>
<td>AF323683</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Eκα</td>
<td>NR3A1</td>
<td>AAJ28575</td>
<td>Estradiol-17β</td>
</tr>
<tr>
<td></td>
<td>Eκβ</td>
<td>NR3A2</td>
<td>AAY31799</td>
<td>Estradiol-17β</td>
</tr>
<tr>
<td>3A</td>
<td>Eκα</td>
<td>NR3B1</td>
<td>AAY63795</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Eκβ</td>
<td>NR3B2</td>
<td>AAO94609</td>
<td>Diethylstilbestrol</td>
</tr>
<tr>
<td></td>
<td>Eκγ</td>
<td>NR3B3</td>
<td>EAW93335</td>
<td>Diethylstilbestrol</td>
</tr>
<tr>
<td></td>
<td>EκR</td>
<td>NR3B4</td>
<td>AAL07554</td>
<td>Diethylstilbestrol</td>
</tr>
<tr>
<td>3C</td>
<td>GR</td>
<td>NR3C1</td>
<td>AAO13510</td>
<td>Cortisol</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>NR3C2</td>
<td>AAAY39571</td>
<td>Aldosterone</td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td>NR3C3</td>
<td>AAO6081</td>
<td>Progesterone</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>NR3C4</td>
<td>EAX05380</td>
<td>Testosterone</td>
</tr>
<tr>
<td>4A</td>
<td>NGFI-B</td>
<td>NR4A1</td>
<td>CAG32985</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>NURR1</td>
<td>NR4A2</td>
<td>AAH92388</td>
<td>Orphan</td>
<td></td>
</tr>
<tr>
<td>NOR1</td>
<td>NR4A3</td>
<td>CAB95138</td>
<td>Orphan</td>
<td></td>
</tr>
<tr>
<td>HR38</td>
<td>NR4A4</td>
<td>AAF53914</td>
<td>Orphan</td>
<td></td>
</tr>
<tr>
<td>HR-6</td>
<td>NR4A5</td>
<td>CAA85271</td>
<td>orphan</td>
<td></td>
</tr>
<tr>
<td>5A</td>
<td>SF1</td>
<td>NR5A1</td>
<td>AAH32501</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>LRH1</td>
<td>NR5A2</td>
<td>AAI18572</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>FTZ-F1</td>
<td>NR5A3</td>
<td>AAN11667</td>
<td>Orphan</td>
</tr>
<tr>
<td>5B</td>
<td>HR-25</td>
<td>NR5A4</td>
<td>CA901028</td>
<td>orphan</td>
</tr>
<tr>
<td></td>
<td>HR39</td>
<td>NR5B1</td>
<td>AAN11107</td>
<td>Orphan</td>
</tr>
<tr>
<td>6A</td>
<td>GCNF</td>
<td>NR6A1</td>
<td>AAB96828</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR4</td>
<td>NR6A2</td>
<td>AAX73355</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>HR-91</td>
<td>NR5B2</td>
<td>CAB60329</td>
<td>orphan</td>
</tr>
<tr>
<td>0A</td>
<td>Knirps</td>
<td>NR0A1</td>
<td>CAA31709</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Knirps-like</td>
<td>NR0A2</td>
<td>AAP51627</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>EGON</td>
<td>NR0A3</td>
<td>CA34626</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>ODR7</td>
<td>NR0A4</td>
<td>AAC46497</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>Trithorax</td>
<td>NR0A5</td>
<td>M31617</td>
<td>Orphan</td>
</tr>
<tr>
<td>0B</td>
<td>DAX1</td>
<td>NR0B1</td>
<td>AAH11564</td>
<td>Orphan</td>
</tr>
<tr>
<td></td>
<td>SHP</td>
<td>NR0B2</td>
<td>AAH30207</td>
<td>Orphan</td>
</tr>
</tbody>
</table>
Figure 4.0.3: Phylogenetic tree showing nuclear receptors identified in humans and fly and the structure of identified ligand in the nuclear receptors as mentioned in table 4.0.2.

The nuclear receptors in green belong to D. melanogaster and those in red are found in humans (Robinson-Rechavi, Garcia and Laudet, 2003).

4.2.4 Nuclear receptor repertoire in molluscs: evolutionary and endocrine disruption insights

Nuclear receptors became a focus of research in molluscs after the discovery that an anti-fouling compound, Tributyltin (TBT), caused penis growth in female molluscs living near harbours and estuaries (Oehlmann et al., 2007; Matthiessen and Gibbs, 1998). This is one of the most cited examples for the effect of endocrine disrupting compounds in the environment on masculinization of female snails and has been reported in over 150
species of gastropods worldwide (Janer et al., 2006a; Oehlmann and Schulte-Oehlmann, 2003; Matthiessen and Gibbs, 1998). This led to several studies in molluscs trying to differentiate the mode of action of sex steroids, specifically estrogen receptor (Stange et al., 2012; Benstead et al., 2011a; Bannister et al., 2007; Castro et al., 2007; Kajiwara et al., 2006; Keay, Bridgham and Thornton, 2006). Several in vivo studies in molluscs indicated a possible role for “vertebrate-like” endocrine (e.g. androgen and estrogen) hormones in the reproductive cycle (Lafont and Mathieu, 2007; Bannister et al., 2007; Thornton, Need and Crews, 2003). Sex steroids in molluscs may be chemically vertebrate-like (Matthiessen, 2008), although vertebrate antibodies can cross-react with other steroids or steroid metabolites, including non-vertebrate steroids, leading to false positive results. Experiments focussing on blocking of sex steroid receptors with chemical antagonists suggested receptor-mediated sex steroid signalling in the scallop Placopecten magellanicus and in the gastropod N. lapillus (Croll and Wang, 2007; Santos et al., 2006; Bettin, Oehlmann and Stroben, 1996) but (Koehler et al., 2007) pointed out the fact that the presence of binding sites does not necessarily indicate the presence of a receptor. Recently, Thornton et al. (2003) found that molluscan neural and reproductive tissues express a nuclear receptor gene that is unambiguously orthologous to the vertebrate-like estrogen receptor (ER) (Thornton, Need and Crews, 2003). This was followed by vertebrate-like ER being reported in other molluscan species, including the bivalves, Crassostrea gigas (Matsumoto et al., 2007), Mytilus edulis (Puinean et al., 2006) and Chlamys farreri (Zhang, Pan and Zhang, 2012), and the gastropod, Aplysia californica (Thornton, Need and Crews, 2003), Lymnaea ollula, Marisa cornuarietis (Castro et al., 2007), Thais clavigera (Kajiwara et al., 2006), Nucella lapillus (Bannister et al., 2007), and in the cephalopod, Octopus vulgaris (Keay, Bridgham and Thornton, 2006). The vertebrate-like ER in molluscs activates transcription constitutively in the absence of ligand, different from vertebrate ER (Bannister et al., 2013; Eick and Thornton, 2011; Matsumoto et al., 2007; Castro et al., 2007; Keay, Bridgham and Thornton, 2006; Kajiwara et al., 2006). The presence of an vertebrate-like ER in molluscs and its similarity with vertebrate ER is also confounded by the absence of aromatase (CYP19) (Markov et al., 2009), the enzyme required for the conversion of testosterone to estradiol, but it is possible that another enzyme catalyzes the aromatase reaction (Thornton, Need and Crews, 2003). Another NR that was likely to be involved in the induction of imposex was RXR, found in the mollusc T. clavigera, which could bind 9-cis retinoic acid and organotins (Horiguchi et al., 2007). Recent work of Pascoal et al. (2012) using N. lapillus demonstrated that retinoid signalling is involved in inducing imposex and the group also
indicated the involvement of androgen signalling (Pascoal et al., 2012). Moreover, at the time my experiments were conducted, no AR gene had yet been reported in any mollusc species and this remains an area of controversy (Scott, 2012b). During the course of my studies, the draft genome for *B. glabrata* became available. This presented the ideal opportunity both to search for AR-like receptors in this species and to characterize the full complement of NRs in this species. In tandem with this, NRs were further investigated in the mollusc *Lottia gigantea*, for which 26 nuclear receptors had already been identified (Bridgham et al., 2010; Wu, Niles and LoVerde, 2007; Wu et al., 2007), to compare with the snail NR repertoire. These two species are from different families and, thus represent distant lineages within the class gastropoda. I have also assessed the evolutionary conservation, divergence, and uniqueness of molluscan NRs in comparison with the receptors of a vertebrate (*Homo sapiens*), insect (*Drosophila melanogaster*), and nematode worm (*Caenorhabditis elegans*).

### 4.3 Materials and Methods

The following methodological steps were performed to identify the NRs in *B. glabrata* and *L. gigantea*:

1) Screening of *B. glabrata* and *L. gigantea* sequence databases using the Basic Local Alignment Search Tool (BLAST) looking for expressed sequence tags (ESTs), contigs, transcripts and proteins showing low, medium and high degrees of similarity to known NRs from human, fly and worm;

2) 12 ‘local’ databases were made from the 12 tissue transcript data FASTA files using Seqtools ([http://www.seqtools.dk/](http://www.seqtools.dk/)) to reaffirm the prediction of NRs.

3) *B. glabrata*12 tissue transcriptome database in Seqtools were screened (BlastN) using GenBank ESTs (E-value: 1e-98; Descriptions: 10; Alignments: 2) and the same was done against the genome (Vectorbase E-value: 1e-80).

4) Identification of 40 putative NRs from *B. glabrata* and 33 NRs from *L. gigantea*. For *L. gigantea*, the available expressed sequence tags (EST) were screened against assembled *L. gigantea* transcripts from the filtered gene models ([http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html](http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html)); experimental assessment of NR gene expression in *B. glabrata* by RT-PCR to generate products for sequencing;

5) Editing of *B. glabrata* gene sequence chromatograms and computational analyses to identify domains and motifs followed by reverse blast to reaffirm the sequence characteristics.
6) Multiple sequence alignment followed by comparative cross-species phylogenetic analysis of molluse NRs with the human, fly and worm.

4.3.1 Identification of putative NR genes from B. glabrata and L. gigantea

Human AR (GenBank: EAX05380) and ER (GenBank: AAI28575) gene sequences were used to search for ESTs in B. glabrata and L. gigantea on National Centre for Biotechnology Information (NCBI) using TBlastX with default parameters. ESTs are short fragment of complementary DNA (cDNA) that represents the expressed portion of the genome (Parkinson and Blaxter, 2009). These ESTs were then used to search for homologues at the preliminary B. glabrata genome interface (version 4.01) (1e-04) (http://129.24.144.93/blast_bg/2index.html) and assembled L. gigantea transcripts from the filtered gene models in the JGI Genome Portal (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) using BlastN (1e-99) with scoring matrix as BLOSUM62. This was followed by a wider search using nuclear receptors spanning all the groups in humans and D. melanogaster, Only 15 NRs from C. elegans have been included in this study for comparative analysis as they are orthologues of NRs that are broadly conserved among animal phyla (Robinson-Rechavi, Garcia and Laudet, 2003; Sluder and Maina, 2001). A secondary search using the identified Lottia and Biomphalaria NR genes against both the genomes using a lower threshold value came up with additional family members, some of which were distinct from human and fly NRs. The query sequences were filtered for low complexity regions. All of the identified contigs from B. glabrata were downloaded and processed through GENSCAN (1e-10) (http://genes.mit.edu/GENSCAN.html) to identify predicted coding regions, intron-exon boundaries, and peptide sequences. The putative peptide sequences were then assessed using the BLASTP (1e-10) interface against the non-redundant GenBank database to check for the presence of the NR domains (LBD/DBD). All the nucleotide sequences from both molluscs containing NR domains were aligned separately using Clustal X user interface in Seqtools (8.4ver) and analysed for redundancy using clustering analysis with sequence overlap cut-off 0.5 and segment coverage cut-off of 0.25.

A systematic nuclear receptor nomenclature (functional annotation) for the identified genes was based on the sequence similarity and phylogenetic analysis of the full-length sequence to that of previously characterised human and Drosophila NRs. Where there was no vertebrate or invertebrate homologue, nuclear receptor abbreviation NR followed by subfamily, a capital letter for the group and receptor number for the individual gene was used (Nuclear Receptors Nomenclature committee, 1999) (Germain et al., 2006).
4.3.2 Comprehensive search for NRs in RNA-Seq 12 tissue transcript data and Vectorbase

Towards the end of my PhD, I got an opportunity to be part of an international effort on *Biomphalaria* genome annotation and had access to the final *B. glabrata* genome assembly and all raw data discussed in detail in Chapter 5. This gave me another chance to do an exhaustive search and look for any NRs that did not come up in the previous search on the draft genome. Local libraries were created in Seqtools (8.4ver) corresponding to 12 tissue RNA-Seq transcript data provided by the consortium. ESTs for NRs submitted to GenBank were used to search for identical transcripts in the 12 local libraries in Seqtools using BlastN (E-value: 1e-98; Descriptions: 10; Alignments: 2). These ESTs were also used to search for predicted transcripts from the genome by performing BlastN searches in Vectorbase (https://www.vectorbase.org/) using default parameters. The best hit from transcripts and scaffold was then scanned (Genome Browse button on the top right end of the Blast result page in Vectorbase) to get the entire protein summary and all the domains and conserved features in the identified transcript. The nucleotide sequence of the identified transcripts were downloaded and aligned using MUSCLE in Seqtools and Clustal Ω (http://www.clustal.org/) and analysed for splice variants and if not then got a consensus sequence. An exhaustive search was performed in the genome (Vectorbase) using PFAM domain analysis (PF00104 and PF00105) (Bateman *et al.*, 2004) and PANTHER, a hidden Markov model-based method (PTHR24082) (Thomas *et al.*, 2003) and INTERPRO (IPR001628; IPR000536 and IPR001723) (http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=iprscan-I20131216-231529-0380-30756832-oy) to identify all the predicted transcripts with NR domains. TBlastX with default parameters was done using these sequences on National Centre for Biotechnology Information (NCBI) to perform a cross species check for the domains. A local alignment was performed between the RNA-Seq transcript and the *B. glabrata* scaffold to identify splice sites to signal exon/intron boundaries. The BLAST results for the scaffold with the lowest E-value for the query were selected and downloaded and its corresponding GFF3 (Generic feature format v3, described in details in Chapter 6) file containing the genomic features such as exon-intron boundaries was also downloaded.¹

¹The detailed methods section describing the annotation process can be found in Chapter 6.*
The rationale for using the above mentioned bioinformatics methods was to perform an exhaustive search to identify homologues for androgen receptor and other NRs in the *B. glabrata* draft genome that was released at the start of the project and in *L. gigantea* genome. The steps were repeated when the RNA sequencing data was made available to reaffirm the results with the final assembly released towards the end of my PhD.

### 4.3.3 PCR amplification of *B. glabrata* NRs

#### 4.3.3.1 Snail material

Snail material was obtained from different stages of the life cycle (Adult, 96 hours and 120 hours embryo, 14 dph juvenile). Two adult snails were swiftly killed by chopping off the head/foot region through the shell and their shell removed starting from the outermost whorl to inner whorls. Once the snail was removed from its shell, the wet weight of the body was recorded and the whole snail was transferred into individual clean micro-centrifuge tubes containing 4ml RNAlater™ (Ambion) to prevent RNA (ribonucleic acid) degradation. Six egg masses (2 X 3) each from embryos 96 hours post hatch and 120 hours post hatch embryos were collected and transferred into clean micro-centrifuge tubes containing 3ml of RNAlater™ (Ambion). All micro-centrifuge tubes were subsequently stored in the fridge (4°C).

#### 4.3.3.2 RNA extraction

Total RNA was isolated from whole adult *B. glabrata* snails using TRI-reagent (Sigma-Aldrich). The whole snails were homogenized using motor and pestle adding 1ml of TRI-reagent per 100mg of snail tissue until no lumps could be seen. The tissue homogenate was centrifuged (12,000 x g, for 10 minutes at 4°C) and the supernatant was mixed with 0.1ml of 1-bromo-3-chloropropane per ml of TRI reagent used. The sample was shaken vigorously for 15 seconds and then allowed to stand at room temperature for 15 minutes. The mixture was then centrifuged (13,000 x g for 15 minutes at 4°C). The upper aqueous phase containing RNA was collected in a fresh tube and 0.5ml of isopropanol per 1ml of TRI reagent added. The mixture was allowed to stand for 10 minutes at room temperature before centrifugation (13,000 x g for 15 minutes at 4°C), which led to precipitation of RNA as a pellet. The precipitated RNA pellet was washed with 1ml of 75% ethanol per 1ml of TRI reagent, vortexed followed by centrifugation (13,000 x g for 15 minutes at 4°C) (Based on manufacturer’s instructions). The extracted RNA was solubilized in 30μL of RNase-free H₂O. Each RNA sample was quantified using spectrophotometry (Nanodrop, Fisher Scientific Ltd.), and was visually analysed by agarose gel
electrophoresis. All samples were then stored at – 80 °C.

For the egg masses, the RNeasy Fibrous tissue mini kit (Qiagen kit) was used. The egg masses were homogenized using motor and pestle adding 300µl of lysis buffer per 30mg of tissue for 15-20 seconds. The lysate was mixed with 590µl of RNase free water and 10µl of proteinase K per 30mg of tissues followed by incubation at 55°C for 10 minutes to aid in tissue digestion. The homogenate was centrifuged, and the supernatant was collected and mixed with 96-100% ethanol to provide appropriate binding conditions. The samples were subsequently applied to RNeasy Mini Spin Columns, washed and eluted following manufacturer’s instructions. The kit has a DNase step to remove contaminating genomic DNA. The extracted RNA was solubilized in 30µL of RNase-free H₂O. Each RNA sample was quantified using spectrophotometry (Nanodrop, Fisher Scientific Ltd.), and was visually analysed by agarose gel electrophoresis. All samples were then stored at – 80 °C. (See Table 4.0.2 for the yield).

Table 4.0.2: RNA extraction report using Nanodrop showing the yield from adult, 96 hours and 20 hours embryos (3 egg masses (Ems) in each time point X two replicates).

<table>
<thead>
<tr>
<th>Time point</th>
<th>Egg mass no.</th>
<th>ng/µl</th>
<th>Yield (µg)</th>
<th>260/280</th>
<th>260/230</th>
<th>Total weight (g)</th>
<th>Wet weight (g)</th>
<th>Diameter (cms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>1a</td>
<td>2321.53</td>
<td>116.07</td>
<td>1.71</td>
<td>1.09</td>
<td>15.24</td>
<td>0.6239</td>
<td>0.1549</td>
</tr>
<tr>
<td>Adult</td>
<td>1b</td>
<td>2271.84</td>
<td>113.592</td>
<td>1.88</td>
<td>1.01</td>
<td>15.34</td>
<td>0.644</td>
<td>0.1808</td>
</tr>
<tr>
<td>120hrs</td>
<td>2a</td>
<td>191.97</td>
<td>5.7591</td>
<td>2.13</td>
<td>0.75</td>
<td>0.1273</td>
<td>0.1273</td>
<td></td>
</tr>
<tr>
<td>120hrs</td>
<td>2b</td>
<td>212.12</td>
<td>6.3636</td>
<td>1.88</td>
<td>1.31</td>
<td>0.1708</td>
<td>0.1708</td>
<td></td>
</tr>
<tr>
<td>96hrs</td>
<td>3a</td>
<td>72.33</td>
<td>2.1699</td>
<td>2.12</td>
<td>2.15</td>
<td>0.1462</td>
<td>0.1462</td>
<td></td>
</tr>
<tr>
<td>96hrs</td>
<td>3b</td>
<td>64.05</td>
<td>1.9215</td>
<td>2.07</td>
<td>1.70</td>
<td>0.1128</td>
<td>0.1128</td>
<td></td>
</tr>
</tbody>
</table>

4.3.3.3 Generation of complementary DNA using reverse transcription

Superscript III First-Strand synthesis system for reverse transcription-PCR kit (Invitrogen) was used to synthesise complementary DNA. For each adult sample, 4µg of total RNA and for embryos 1.5µg of total RNA (Table 4.0.2) was reverse transcribed in a 20µl reaction containing 1µl of 100µM oligo(dTAP) primer (TGACTCGAGTCGACATCGAT20) and 1.25mM dNTPs (Bioline). RNA and primer mix was briefly centrifuged, incubated at 65 °C for 5 minutes and then placed on ice for 1 minute. This step allowed the annealing of the oligo (dT) primer to the mRNA poly-A tail. Successively, 2µL of 10X RT buffer, 4µL of 25mM MgCl₂, 2 µL of 0.1M DTT, 1 µL of RNase (40 U/µL) and 1 µL of Superscript III RT (200 U/µL) were added, mixed and briefly centrifuged. The mixture was then incubated at 50 °C for 50 minutes to allow the synthesis of the cDNA. The reaction was terminated by incubation at 85 °C for 5 minutes. Any residual RNA was removed by adding 1 µL of RNase H (2 U/µL) to the reaction and
incubating it at 37 °C for 20 minutes. cDNA samples were stored at -20 °C until further analysis. Reverse transcriptase-PCR (RT-PCR) with 18S primers (18S-F: CGCCCGTCGCTACTATCG and 18S-R: ACGCCAGACCGAGACCAA) (Hertel, Adema and Loker, 2005) verified successful cDNA synthesis.

4.3.3.4 Polymerase Chain Reaction

Each of the identified NR sequences from the B. glabrata draft genome were used to design specific PCR primers (Table 4.0.3) using PRIMER3 web software (www.biotools.umassmed.edu/bioapps/primer3_www.cgi), for amplifying fragments from DBD and/or LBD to confirm transcription and sequence. The primers were designed keeping in mind the following parameters: guanine and cytosine (G and C) content of 40-60%, G and/or C nucleotide at the 3” end, and melting point between 55-63 °C. The primer pairs had similar melting point (ΔT < 4 °C). NetPrimer and the java tool OligoAnalyzer 3.1 (www.eu.idtdna.com/analyzer/Applications/OligoAnalyzer) were used to assess the likelihood to form secondary structures such as hetero-dimers, self-dimers and hairpin. All primers were between 18 and 23 nucleotides long and were purchased from Sigma Aldrich. Stock primer solutions were made in molecular grade water (100μM).

For conducting PCRs, a 25µl mix was prepared containing 2µl cDNA (diluted 1 in 20 for adults and 1 in 10 for embryos), 1 X PCR Buffer (Bioline), 2.5mM MgCl₂, 0.2mM dNTPs, 0.5μM forward and reverse primers and 1.25U Taq DNA polymerase (Bioline). The optimal annealing temperature for each primer pair was determined by performing RT-PCR reactions with a gradient of temperatures (Δ = 0.8 °C). After gel electrophoresis of the PCR products, the temperature at which the amplification of the target fragment had an apparent better yield and specificity was chosen for the subsequent reactions.

Generally, PCR amplifications were performed as follows:

- Hot Start: 95 °C for 2 minutes 1 cycle
- Denaturation: 95 °C for 30 seconds
- Annealing: 55-63 °C for 30 seconds 34 cycles
- Extension: 72 °C for 1.30 minute 1 cycle
- Final elongation: 72 °C for 5 minutes 1 cycle
- Final hold: 4 °C
Table 4.0.3: Oligonucleotide primers used in RT-PCR analyses of nuclear receptor gene products with their product size (as submitted to GenBank) and their corresponding GenBank accession numbers.

<table>
<thead>
<tr>
<th>Putative NRs</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>GenBank Acc.</th>
<th>Fragment size (bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUPTFA</td>
<td>GTTGGCAAACACCAACACAC</td>
<td>CTGGAGAAGAAAGAGGAGTG</td>
<td>JZ390916</td>
<td>773</td>
</tr>
<tr>
<td>NRU1</td>
<td>TGAGTTTTTGAGGCGGTTG</td>
<td>GCAGCAAGAGAACACACAC</td>
<td>JZ390900</td>
<td>867</td>
</tr>
<tr>
<td>HR38</td>
<td>CACCGCAATTAGCTCTCC</td>
<td>CTGTAGTGGCAGTGGCTCTG</td>
<td>JZ390932</td>
<td>816</td>
</tr>
<tr>
<td>HR3</td>
<td>AACGGCTCTTACTTCCAC</td>
<td>GCACGGGATTGAGATGTTG</td>
<td>JZ390936</td>
<td>913</td>
</tr>
<tr>
<td>PPAR2DBD</td>
<td>GGTGCAAAAGCGCTCCATTTGGG</td>
<td>CTTTGGACAGGCTGGATGTA</td>
<td>JZ390908</td>
<td>492</td>
</tr>
<tr>
<td>PPAR2BD</td>
<td>AAGGAGTCCACACCATTTCC</td>
<td>CATTGGCAAGATGCTGTCAC</td>
<td>JZ390909</td>
<td>697</td>
</tr>
<tr>
<td>HNF4</td>
<td>CAAGTGGGCACTCTGTGAGA</td>
<td>GCGTGGAGAAGCTGGTTTGA</td>
<td>JZ390914</td>
<td>914</td>
</tr>
<tr>
<td>ER</td>
<td>CACACCACCTGATCTCTGCT</td>
<td>ACCTGCGGCAATCTTGTGAT</td>
<td>JZ390913</td>
<td>1097</td>
</tr>
<tr>
<td>ERF</td>
<td>GATTCCTTCCACAGACTA</td>
<td>TCAGTCATCAACATTCTCAA</td>
<td>JZ390912</td>
<td>1033</td>
</tr>
<tr>
<td>ROR</td>
<td>CACGCCCTCCCTCACAACCTC</td>
<td>CTATGGCAAGAGTCTGAAG</td>
<td>JZ390933</td>
<td>879</td>
</tr>
<tr>
<td>EcR</td>
<td>AAGCAGTGTCTACACCTTCTTGGGA</td>
<td>AGTCGAGAGCTCTGAAG</td>
<td>JZ390899</td>
<td>999</td>
</tr>
<tr>
<td>FAX1</td>
<td>GCAAACAGCTACGGGCTCCTAC</td>
<td>TCAACAGGTCTGCTGCTCA</td>
<td>JZ390933</td>
<td>731</td>
</tr>
<tr>
<td>HR4BD</td>
<td>CCTGTTCCTTGCGGGTGAAT</td>
<td>CTTGGAGATCTCTTGGCATI</td>
<td>JZ390920</td>
<td>779</td>
</tr>
<tr>
<td>HR4</td>
<td>GGAAGTGTCTGTCAGCT</td>
<td>GCTGGAGCTGCTACACTC</td>
<td>JZ390921</td>
<td>833</td>
</tr>
<tr>
<td>HR4LD</td>
<td>CCCCTCAACTCATAAGCAT</td>
<td>CTGAAGAAGTGGGCTCTGTA</td>
<td>JZ390922</td>
<td>372</td>
</tr>
<tr>
<td>NR1D1</td>
<td>GTGCAAACACTAATTTCTTGAGT</td>
<td>TCAAGTAATCAGCAATGCTG</td>
<td>JZ390906</td>
<td>938</td>
</tr>
<tr>
<td>NR1D2</td>
<td>GCTGTAGGCTGTGGCTTTC</td>
<td>GACCTGAGTGCTCTGTCTG</td>
<td>JZ390907</td>
<td>1070</td>
</tr>
<tr>
<td>NR1D3</td>
<td>AAACCCACTGACACAGACAC</td>
<td>GGCCTCCTACCTTCTTGATA</td>
<td>JZ390902</td>
<td>814</td>
</tr>
<tr>
<td>NR1J2DBD</td>
<td>GAGACTGAGAGACACACACA</td>
<td>GATGCGGACACCTCTCTTATG</td>
<td>JZ390897</td>
<td>885</td>
</tr>
<tr>
<td>NR1J2LD</td>
<td>CAGGTGGAGAGAAAGAAAGGA</td>
<td>GACCAACATTTGGAGTGA</td>
<td>JZ390898</td>
<td>661</td>
</tr>
<tr>
<td>1RBDK</td>
<td>GCTCTTGGAAACAAACAAA</td>
<td>CCCCGAGAACATACCAATT</td>
<td>JZ390926</td>
<td>376</td>
</tr>
<tr>
<td>TRK</td>
<td>GAACCTGCTCAAGTGCAGT</td>
<td>GCTTGAGGTGTCATGTTG</td>
<td>JZ390927</td>
<td>701</td>
</tr>
<tr>
<td>TLKLD</td>
<td>CTTCCTAGCACTACGCTTCA</td>
<td>ACACGATTATCCACCGCAGC</td>
<td>JZ390928</td>
<td>608</td>
</tr>
<tr>
<td>NRAA</td>
<td>GACAGCTACACCCAGATGTTT</td>
<td>AAGGCGATGTGCAATAAGTAC</td>
<td>JZ390905</td>
<td>928</td>
</tr>
<tr>
<td>NR2BD</td>
<td>AATGGGCGAAAAAAAGGAAAC</td>
<td>AGAVAGGATACAGGCTGAAT</td>
<td>JZ390903</td>
<td>805</td>
</tr>
<tr>
<td>NR2LD</td>
<td>ATTCACCCGCTCCCTCT</td>
<td>CCTGACTGCTGACACACTG</td>
<td>JZ390904</td>
<td>894</td>
</tr>
<tr>
<td>2BDKNK</td>
<td>CTCCTCACTCATCTCACCAC</td>
<td>CTCCAGTGGGGCTCTTATTC</td>
<td>JZ390896</td>
<td>737</td>
</tr>
<tr>
<td>TLX</td>
<td>CGGAGATGGGCGGATTTTA</td>
<td>CGAGGGGAGTGGTGAAGA</td>
<td>JZ390910</td>
<td>699</td>
</tr>
<tr>
<td>XRJR</td>
<td>ATCTGTCGCAATATGGGCTG</td>
<td>TAAATGTGCAACTGCTGC GGCTG</td>
<td>JZ390931</td>
<td>1059</td>
</tr>
<tr>
<td>RAR</td>
<td>TCTCTGTTGAGGCTGGTAGA</td>
<td>GCTGGCAGAATTGCTGATG</td>
<td>JZ390929</td>
<td>1030</td>
</tr>
<tr>
<td>E73C</td>
<td>CCTTCCTGACTACAGATACA</td>
<td>CGGATGGTACGCGGCTTCTC</td>
<td>JZ390894</td>
<td>905</td>
</tr>
<tr>
<td>HR39</td>
<td>CCCCTGTGATTGCGGTAAG</td>
<td>CTGGAAGCTCAAGACAGA</td>
<td>JZ390901</td>
<td>772</td>
</tr>
<tr>
<td>NR3DBD</td>
<td>GCCGACAAACCCAAACAGACT</td>
<td>TCTCCAGGAGCAGACCACA</td>
<td>JZ390904</td>
<td>379</td>
</tr>
<tr>
<td>NR3U1</td>
<td>CTGCCGAGATCTATGAG</td>
<td>CTGTCGTGCATTGCTTG</td>
<td>JZ390924</td>
<td>633</td>
</tr>
<tr>
<td>NR3U3LD</td>
<td>TGACCACAAAGACACAGAG</td>
<td>TGTTGAGTGAAGACAGCAG</td>
<td>JZ390925</td>
<td>908</td>
</tr>
<tr>
<td>PPAR1</td>
<td>CGTCCGCAAGTTTGCAACA</td>
<td>TGCCATGCACTGCTGCAAAT</td>
<td>JZ390930</td>
<td>1074</td>
</tr>
<tr>
<td>NR1J2</td>
<td>AACTTCTTGAGCAGCCTACCT</td>
<td>CACTCAATGCTCCTACCTTC</td>
<td>JZ390911</td>
<td>582</td>
</tr>
<tr>
<td>NR1J4</td>
<td>AACAATTGTGGAAGCCCAACT</td>
<td>CTGATGGATGCACTACCTCA</td>
<td>JZ390915</td>
<td>862</td>
</tr>
<tr>
<td>NR1J1</td>
<td>CAATCCCAATGCCGAACAAAGGCTGC</td>
<td>GGGGAGACGTCGTTGAAATGGTA</td>
<td>JZ390895</td>
<td>509</td>
</tr>
<tr>
<td>FTZ-F1BD</td>
<td>CTTCACCCGAGGCATTTGGC</td>
<td>CCTACCCACACAAACACCA</td>
<td>JZ390917</td>
<td>573</td>
</tr>
<tr>
<td>FTZ-F1</td>
<td>GCCAAGCAGCTCCTCTCAG</td>
<td>CAGCACTTGGAGACATC</td>
<td>JZ390918</td>
<td>907</td>
</tr>
<tr>
<td>FTZ-F1LBD</td>
<td>GTTTTGTAGTGATGGGAATGGG</td>
<td>ACCTTGCGCTATCCGTGGTG</td>
<td>JZ390919</td>
<td>415</td>
</tr>
<tr>
<td>DSF</td>
<td>CCTCAAGTCCACGGGAAGAA</td>
<td>AAGAAGATGATGGTGC</td>
<td>JZ390934</td>
<td>518</td>
</tr>
<tr>
<td>E75</td>
<td>GTCTTATGCGGGTGTGGTGGTG</td>
<td>CCTGTGTTGATGGTGATGATG</td>
<td>JZ390939</td>
<td>945</td>
</tr>
<tr>
<td>Rev-erb BD</td>
<td>GTAGCAAGTCTCCGCGTTTTT</td>
<td>TGAAGCGCTGTGTTGGTTTG</td>
<td>JZ390937</td>
<td>367</td>
</tr>
<tr>
<td>Rev-erb LBD</td>
<td>CCTCTTCCAAACACAGACACT</td>
<td>GCACAAAGGACGACCACCT</td>
<td>JZ390888</td>
<td>560</td>
</tr>
</tbody>
</table>

Some of the NRs had separate primers for DBD, intermediate (I) and LBD regions.

4.3.3.5 Gel electrophoresis

The DNA fragments amplified by PCR were visualized by gel electrophoresis using 1-1.2% agarose gel with 5 μL of Gel red (Sigma) added and mixed (0.005%). The gel was decanted into a plastic mould with a comb inserted, in order to create empty wells. 7 μL of gel-loading buffer (Sigma Aldrich) were added to 3μL of each sample and loaded onto the gel. The gel-loading buffer increased the density of the sample, preventing the settlement of the PCR product in the gel’s well. The dye (bromophenol blue) is negatively charged in neutral buffers and migrates in the gel in the same direction as the DNA during
electrophoresis. This allows the monitoring of the progress of the electrophoresis. The wells on both the extreme ends were loaded with 4 μL of 1 Kb DNA ladder (Bioline) to detect the size of the PCR fragment. The gels were typically run at 85 Volts for 40-60 minutes, and successively examined under ultraviolet (UV) illumination. Each sample was analyzed for the presence of amplified fragments having the expected size as calculated during primer design, see (Figure 4.0.4). All the PCR products with expected fragment size and a clear single band were sent for sequencing to the Sequencing facility, Wolfson Welcome Biomedical Laboratory, Natural History Museum.

![Image of DNA electrophoresis](image)

**Figure 4.0.4:** The UV image of the PCR products representing NRs in *B. glabrata.*

The product bands obtained are clean, clear and at approximately the correct size in base pairs (bps).

### 4.3.3.5 Extraction of PCR product from the gel

For PCR products that showed double banding when run on an agarose gel, or which suggested the presence of more than one product when the sequencing electropherogram was examined, cloning was employed to obtain usable sequences for further analysis. The 25μL PCR product was run on the agarose gel and a fragment of the correct size was excised from the agarose gel under UV light using a clean, sharp scalpel (Swan-Morton, Sheffield). DNA extraction was done from the gel using the MinElute gel extraction kit (Qiagen), following the protocol indicated by the manufacturer (not described here). The eluted DNA was collected, quantified by spectrophotometry, and stored at -20 °C until further analysis.
4.3.3.6 Cloning DNA fragments

DNA cloning is a method to assemble recombinant DNA molecules for producing large amount of identical DNA fragments. The fragment of DNA to be amplified is inserted in a vector usually a highly modified phage or plasmid that can replicate in a host cell. Below is the brief description of the various steps involved.

**STEP 1: Ligation step to create recombinant DNA:** The PCR reaction mix was prepared using 1μl of 10X buffer, 1μl of 2mM dNTPs, 0.5μl of 50mM Mgcl₂, 0.1μl of Taq and made to 10μl by adding clean extracted DNA. The mix was incubated for 10 minutes at 72⁰C to replace A overhangs. 3μl of this PCR product was added to 5μl of Rapid Ligation buffer (manufacturer/concentration), 1μl pGEM® easy vector (50ng) and 1μl of T4 DNA ligase in 5μl DNase free micro-centrifuge tube. The contents were spun and left overnight at 4⁰C for maximum number of transformants.

**STEP 2: Preparation for transformation step:** Lactobacillus (LB)/ampicillin/isopropyl thiogalactoside (IPTG)/5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) plates were prepared to grow the colonies. LB medium was prepared following laboratory protocol. The medium was divided into two parts and 3.75g of agar added to one part before autoclaving. The autoclaved mixture (LB+agar) was then allowed to cool to 50⁰C before adding 175μl of 100mg/ml ampicillin, as excessive heat would degrade the antibiotic. 25ml of LB+agar was poured into each petri dish (measurement) and stored at 4⁰C. Before use, these plates were supplemented with 20μl of 500mM IPTG and 20ul of 50mg/ml X-Gal spread over the surface of an LB-ampicillin plate and allowed to absorb at 37⁰C.

**STEP 3: Transformation:** The tubes containing ligation reaction were centrifuged and 2μl of each ligation reaction was added to a sterile 1.5ml micro-centrifuge tube on ice. JM109 high efficiency competent cells, (which are cloning strain of *Escherichia coli* used to generate high quality plasmid DNA) were thawed on ice and then mixed by gently flicking it to avoid disrupting the cells. 50μl of cells were carefully transferred to the prepared ligation reaction tubes and incubated on ice for 20 minutes. The cells were heat-shocked for 45-50 seconds in a water bath at exactly 42⁰C and without shaking returned to the ice bath for 2 minutes. This was followed by adding 950μl of SOC medium (Bioline) at room temperature to the ligation reaction transformations, incubating it for 1.5 hours at 37⁰C on shaking incubator. 50-100μl of each transformation reaction was spread onto duplicate LB/ampicillin/IPTG/X-Gal plates and incubated overnight at 37⁰C to grow colonies. A plasmid that does not have DNA insert in the α-peptide coding region of β-galactosidase will produce blue colour and if the α-peptide is inactivated by DNA insert, it will have
white/beige colour. Single white colonies were picked using a sterile pipette tip and grown up overnight in 10ml of LB and 5μl of ampicillin (100mg/ml) in a 50ml of centrifuge tube shaking at 37°C. The next day, the mix is quite cloudy and ready to be used as a template to run PCR.

**STEP 4: PCR using cloned cDNA:** The culture was diluted 1:10 and the cells denatured at 95°C for 10 minutes. It was then placed on ice, spun down to collect all the content at the bottom and then 5μl of this template used to run a 50μl PCR reaction as described above.

Amplified products were analysed by gel electrophoresis using 1% agarose gel and 5μl of Gel Red for the presence of amplified fragments and the insert size calculated using the formula:

M13 reverse sequence primer binding site.........................197 base pairs (bp)
Total vector size..............................................................3015 bp
T7 RNA polymerase promotor........................................2993bp
M13 forward sequence binding site.................................22bp
Total estimated product size =197+22+estimated insert size. This is the distance from the primer on the gene to the end of the gene and the distance from the primer on the vector to the end of the vector (assuming that one primer sits on the cloned gene and the other sits on the vector).

### 4.3.3.7 Sequencing

The PCR products and 10mM of the forward and reverse primers were prepared and were sent to the Sequencing facility, Wolfson Wellcome Biomedical Laboratory, Natural History Museum for Sanger sequencing using Applied Biosystems 3730xl DNA analyser. The results from the sequencing were delivered as chromas plot showing the nucleotide sequence of the DNA product, as shown in Figure 4.0.5. The quality of the sequencing at the beginning of the process and towards the end is usually low, as indicated by the height and definition of the peaks (indicated in grey Figure 4.0.5). When the nucleotide peaks become well defined and higher, the base calls are considered reliable. Most of the sequences obtained were less than 1000bp. The sequences obtained were analyzed by computational methods, including reverse Blast, ClustalX alignment and Seqtools with the predicted sequences to get the consensus sequence. Sequences were submitted to GenBank accession: JZ390894-JZ390939.
Figure 4.0.5: Example of a chromatogram file indicating the nucleotide sequence of a fragment of a NR expressed B. glabrata analysed after sequencing. Note the increasing quality of the peaks, indicated by the quality bar in grey at the top of the graph going above the green line threshold. Typically the first 20-40 sequenced bases show low quality peaks and are not shown in the figure above.

4.3.4 Computational analysis

NR nucleotide sequences were translated using proteomics tools available on the EXPASy website (http://ca.expasy.org/sprot/). The translation process normally generated six possible reading frames, three in forward frame and three in reverse direction. Translation for frame starts from the first nucleotide, frame 2 from the second and frame 3 from the third. Frames 4 to 6 correspond to 1 to 3 but in the reverse direction. The frame that produced the longest reading frame with no stop codons was selected. The translated peptide sequences from B. glabrata and L. gigantea were analysed using different signatures on InterProScan (Thomas et al., 2003) Figure 4.0.6. A further analysis was done with the Simple Modular Architecture Research Tool (SMART) for the identification of DNA-binding domain (DBD) and ligand binding domain (LBD) regions (Letunic et al., 2006). Multiple sequence alignment of the concatenated domain regions (DBD and LBD) from mollusc NR sequences with the human (H. sapiens), fruit fly (D. melanogaster) and nematode (C. elegans) NRs domains (obtained using conserved domain database (CDD)) (Marchler-Bauer et al., 2007) were performed. Altogether 158 sequences from all the five species (H. sapiens, D. melanogaster, C. elegans, B. glabrata and L. gigantea) were compiled into a data matrix and aligned using ClustalX software (Thompson et al., 1997). The aligned file (Appendix 4.1) was then converted into Nexus format, for phylogenetic analysis using Mesquite (ver2.75).
4.3.5 Phylogenetic analysis

Phylogenetics is an approach of reconstructing evolutionary history and understanding the biological classification of genes, organisms and species (Thain, M. & Hickman, M., 1995). Molecular phylogenetic techniques use sequenced fragments of nucleotides/amino acids to build phylogenetic trees and the analysis is carried out using various methods, including maximum parsimony (MP), Neighbour joining (NJ), maximum likelihood (ML) and Bayesian inference (BI).

BI is a probabilistic model-based method using the software MrBayes version 3.1.2 (Ronquist 1 and Huelsenbeck2, 2003), freely available computing cluster Bioportal (http://www.bioportal.uio.no) run by the University of Oslo. This method of phylogenetic analyses is similar to maximum likelihood. BI methods use a set of prior assumptions about the data matrix to infer the probability that a hypothesis may be true. The posterior probability for a hypothesis is proportional to the likelihood multiplied by the prior probability of the hypothesis. BI analyses allow complex models of molecular evolution to be included (Lewis, Holder and Holsinger, 2005; Holder and Lewis, 2003). The best model for a particular set of alignment can be calculated using ModelTest.
(http://www.molecularevolution.org/software/phylogenetics/modeltest) for nucleotide sequences and Prottest (Abascal, Zardoya and Posada, 2005) for protein sequences. BI uses the Markov chain Monte Carlo (MCMC) algorithm that is used to sample probability distribution of the sample data. A Markov chain is a sequence of random variable and the distribution of each variable depends on the value of previous random variable. The algorithm works by taking a series of steps, each step forming a new link in the chain that moves through different trees and models of evolution. BI was conducted for the identified NRs with MrBayes version 3.1.2 using the WAG model that was selected as the best fitting substitution model with both Akaike information criterion and Bayesian information criterion frameworks (determined using default parameters by Prottest (v1.4)).

There were two independent runs of the MCMC analysis, with four chains (one cold and three hot chains) run for five million generations with temperature 0.10. Priors having default values and sampled every 1000 generations. The first 1.25 million generations were discarded (burn-in fraction) to avoid bias in posterior probability calculations. For this the Log likelihood values were plotted against the generations and it was found to be asymptotic well before the burn-in fraction. Convergence between the independent MCMC runs was examined by the average deviation of the split frequencies (0.006) and the potential scale-reduction factor (PSRF), which was 1.00. Clades with posterior probabilities > 95% were considered well supported. An additional BI tree was constructed using subfamily 1 of the NR superfamily to explore the position of newly identified BgTHR homologue, with respect to other selected species. There were two independent runs of the MCMC analysis, with four chains (one cold and three hot chains) run for one million generations, priors having default values and sampled every 500 generations. The first 600 samples (300000 generations) were discarded as burn-in because the Log likelihood values were plotted and found to be asymptotic well before the burn-in fraction. Convergence between the independent MCMC runs was examined by the average deviation of the split frequencies (0.01) and the potential scale-reduction factor (PSRF), which was 1.00. The tree was midpoint rooted in dendroscope (ver 3.2.1).

The maximum Likelihood method is similar to BI and uses probabilities of change from one character to another to calculate the likelihood that a given phylogenetic tree would lead to the dataset observed (Felsenstein, 1981). This is computationally very intensive and consequently demands a lot of computer time when large data matrices are involved. ML was carried out using PHYML v2.4.4 (Guindon and Gascuel, 2003), under Jones-Taylor-Thorton (JTT) substitution model (Jones, Taylor and Thornton, 1992) with a
gamma distribution of rates between sites (eight categories, parameter alpha, estimated by the program).

Maximum parsimony is a character-based method and looks for trees requiring the minimum number of changes to explain the data or in other words by minimizing the total tree length. This method uses PAUP version 4.0b10 (Swofford, 2003) to run the simulations. This process is only valid if evolution itself always proceeds along the most parsimonious route, a situation that seems very unlikely to be the case (Felsenstein, 1985). Furthermore, parsimony also suffers from the problem of long branch attraction. This means that the sequences at the end of long evolutionary branches are similar and this is not because of the close evolutionary relationship but due to convergent evolution (Holder and Lewis, 2003; Felsenstein, 1978). For MP, the settings used were:

- Heuristic searches
- Branch swapping set to tree-bisection-reconnection
- Topological constraints not enforced and
- Initial maximum tree setting at 100.

4.3.5.1 Statistical support in phylogenetic analyses

In MP analyses, statistical support for clades was calculated using bootstrapping, in which the original data matrix was randomly re-sampled and replaced with duplications of other characters (amino acids) chosen randomly from the same matrix (Felsenstein, 1985). The tree-building algorithm was then run using the modified dataset. This procedure was repeated 1000 times, providing a percentage value (bootstrap value) for each clade that represents the proportion of retrieved trees when the given clade is seen at that position. For example, if the procedures is repeated for 1000 times and a clade occurs in the same position 750 times, the bootstrap would be 75%. The higher the bootstrap value, the more robust is the position of that clade in the tree (Swofford, 2003). Bootstrap was also used as statistical method for ML trees and branch support was evaluated by 1000 bootstrap replicates.

Statistical support for BI analyses is given as posterior probability for each clade (Holder and Lewis, 2003), which is dependent on prior knowledge (prior probability) and new data, weighed by their relative precision (Likelihood function) (http://www.fil.ion.ucl.ac.uk/spm/course/slides10-vancouver/08_Bayes.pdf). As the run proceeds, the posterior probability of the new tree location is calculated and if higher than the previous location, the new value is accepted as the next link in the chain. If the new location has a lower posterior probability, then it can be only accepted as the new position.
if this location appears more number of times in proportion to the present location. Effectively, small moves downward in probability are usually accepted, whereas large falls in probability are rejected. If the new position is rejected, the present position is used again as the new position (Holder and Lewis, 2003; Ronquist 1 and Huelsenbeck2, 2003). The simulation is repeated millions of times and a chain of tree locations is created preferring areas of high posterior probability. The number of generations that a chain reiterates in a particular position is an estimate of the posterior probability of the region. At the end of the simulation the output is an estimated value of posterior probabilities of the given tree being accurate (Holder and Lewis, 2003).

In my analyses of NRs an appropriate out-group to root the sequences was difficult for such a diverse and ancient nuclear receptor family and so the phylogenies were mid-point rooted using PAUP.

4.4 Results

4.4.1 Nuclear receptor genes

40 putative NR sequences were identified from the genome of B. glabrata and 33 from L. gigantea including the previously identified 26 NRs (Bridgham et al., 2010; Wu, Niles and LoVerde, 2007; Wu et al., 2007). This compares to the 48 NRs reported in the human genome (Robinson-Rechavi, Garcia and Laudet, 2003). Of the 40 NRs identified in B. glabrata, 15 were human NR homologues (For example DAX1, THR, TR etc.) and 19 were homologues of fly NRs. In Table 4.0.5, a cross-species comparison of nuclear receptors is shown with NRs from all the selected five species (humans, fly, worm, B. glabrata and L. gigantea). Table 4.0.4 show the tissue distribution of the NR genes in the RNA-Seq 12 tissue data and the pink shading in the table indicated the presence of a particular NR (identified by contig number) in the corresponding tissue. As evident from Table 4.0.4, B. glabrata NRs are most prominently expressed within the central nervous system (CNS) (34/40), Head foot region (Foot) (32/40) and amoebocyte producing organ (APO) (32/40). Twelve NRs show expression in all the 12 tissues. Most prominent of these are BgER, BgERR, BgFTZ-F1, Bg2DBDNRs, BgRAR, BgNR1D members and the unique NR (BgNRU4). An exhaustive search with mammalian and Drosophila NRs found no further sequence homologs. It should be noted that representative genes for each of the major nuclear receptor groups have been detected, suggesting the assembled genome provides a good first approximation of total gene content. Investigation of the genes in the final assembly with physical coverage of the genome being approximately 27.5X (GenBank: APKA00000000.1) identified
thyroid hormone receptor (BgTHR), which had not been picked up in the previous version (ver 4.01). Exhaustive search through the RNA-Seq transcript data and Vectorbase did not identify any of the subfamily 3C NRs thus supporting the previous search results.

Table 4.0.4: Distribution of putative neuropeptide precursors in the 12 tissues of adult BB02 snail. The 12 tissues are AG - albumen gland; BUC - buccal mass; CNS - central nervous system; DG_HP - digestive gland; Foot - headfoot; HAPO_APO – heart/amebocyte-producing organ; KID - kidney; MAN - mantle edge; OVO - ovotestis; SAL - salivary glands; STO - stomach and TRG – terminal genitalia.

<table>
<thead>
<tr>
<th>Contig</th>
<th>AG</th>
<th>BUC</th>
<th>CNS</th>
<th>DG-HP</th>
<th>APO</th>
<th>FOOT</th>
<th>OVO</th>
<th>SAL</th>
<th>STO</th>
<th>TRG</th>
<th>KID</th>
<th>MAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1481</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3667</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>751</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>469</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2702</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5703</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1318</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1866</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2718</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3695</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22311</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3827</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1296</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.151</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4.0.5: Sequence identification of all NRs for species used in this study.

<table>
<thead>
<tr>
<th>Group</th>
<th>C. elegans</th>
<th>Accession Number</th>
<th>D. melanogaster</th>
<th>Accession Number</th>
<th>B. glabrata</th>
<th>Contig Number</th>
<th>L. gigantea</th>
<th>Protein Id</th>
<th>H. sapiens</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0A</td>
<td>KNI</td>
<td></td>
<td>KNRL</td>
<td></td>
<td>EGON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0B</td>
<td>DAX1</td>
<td>182†</td>
<td>DAX1</td>
<td></td>
<td></td>
<td>153776</td>
<td></td>
<td>DAX1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A</td>
<td>THR**</td>
<td>3.151</td>
<td>THR</td>
<td></td>
<td></td>
<td>207623</td>
<td></td>
<td>THRa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B</td>
<td>CeODR-7</td>
<td>AAC46497</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1C</td>
<td>PPAR1</td>
<td>2052</td>
<td>PPAR2</td>
<td></td>
<td></td>
<td>174409</td>
<td></td>
<td>PPARa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D</td>
<td>NHR-85</td>
<td>AAO39185</td>
<td>E75</td>
<td></td>
<td></td>
<td>AAN11687</td>
<td></td>
<td>E75</td>
<td>Rev-erb</td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>E78</td>
<td></td>
<td>AAF51692</td>
<td>E78C</td>
<td></td>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1F</td>
<td>NHR-23</td>
<td>P41828</td>
<td>DHR3</td>
<td>AAA28461</td>
<td></td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1G</td>
<td>CNR14</td>
<td>AAA96982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1H</td>
<td>EcR</td>
<td></td>
<td>NP724456</td>
<td>EcR</td>
<td></td>
<td>1481</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1I</td>
<td>DAF-12</td>
<td>AAD34462</td>
<td>AAD31437</td>
<td>CAD36502</td>
<td></td>
<td>48*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1J</td>
<td>NHR-49</td>
<td>CAD57702</td>
<td>HNF4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1K</td>
<td>NHR-1</td>
<td>AAAC48174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2A</td>
<td>NHR-49</td>
<td>CAP16277</td>
<td>SVP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2B</td>
<td>USP</td>
<td></td>
<td>AAF45707</td>
<td>RXR</td>
<td></td>
<td>3695</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2C/D</td>
<td>NHR-41</td>
<td>NP_500073</td>
<td>DHR78</td>
<td>Q24142</td>
<td>TR</td>
<td>2718</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2E</td>
<td>FAX-1</td>
<td>AAD55066</td>
<td>CA97428</td>
<td>TLL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F</td>
<td>UNC55</td>
<td>CAP16277</td>
<td>SVP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A</td>
<td>ER</td>
<td>469</td>
<td>ER</td>
<td></td>
<td></td>
<td>132166</td>
<td></td>
<td>ERa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>ERR</td>
<td></td>
<td>AAL37554</td>
<td>ERR</td>
<td></td>
<td>751</td>
<td></td>
<td>ERRb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3C</td>
<td>MR</td>
<td></td>
<td>MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4A</td>
<td>NHR-6</td>
<td>CAA85271</td>
<td>DHR38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5A</td>
<td>NHR25</td>
<td>CAA91028</td>
<td>FTZ-F1</td>
<td></td>
<td></td>
<td>1318</td>
<td></td>
<td>FTZ-F1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5B</td>
<td>DHR39</td>
<td></td>
<td>HR3</td>
<td></td>
<td></td>
<td>3667</td>
<td></td>
<td>HR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6A</td>
<td>NHR91</td>
<td>CAB60329</td>
<td>DHR4</td>
<td></td>
<td></td>
<td>1866</td>
<td></td>
<td>HR4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2DBD</td>
<td>NR</td>
<td></td>
<td>2DBDNR1</td>
<td>2DBDNR2</td>
<td></td>
<td>304*</td>
<td></td>
<td>2DBDNR1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: NR indicates nuclear receptor, ** indicates additional information not shown in the table. The table includes the accession numbers from different species, highlighting the diversity and specificity of the identified NRs for various species used in the study.
NRs from fly, human and worm (NCBI accession numbers), compared to identified L. gigantea NRs (protein identification numbers: JGI genome portal version 1.1) and B. glabrata NRs (Contig numbers: Preliminary Bg Genomic Data (version 4.01). Of 40 putative NRs predicted in B. glabrata, 1 was short fragment, which could not be classified and were therefore not included in the table. *NRs not used in phylogenetic tree; + NRs are short fragments that could be multiple partial genes; ‡ NRs not found expressed in B. glabrata

**NR found in the final genome assembly.

4.4.2 Expression of nuclear receptor mRNAs

For *B. glabrata*, the presence of nuclear receptors was evaluated using PCR with receptor-specific primers, to demonstrate that the predicted genes were expressed as transcripts and to confirm their sequence. For some of the genes in *B. glabrata*, the DBD and LDB regions were amplified and sequenced separately to obtain fragments suitably sized for direct sequencing. Fragments were amplified for 31 of the 40 putative receptors from cDNA derived from whole adult snail, while 3 further nuclear receptors, BgTLX (contig 3), BgDSF (contig 42) and BgFAX1 (contig 148), were not expressed in adult snails but were isolated from 96hrs/120hrs embryos. These three NRs were also not expressed in any of the RNA Seq 12 tissues further confirming the results. These receptors belong to the NR2E sub family and may be involved in embryonic development as reported in *D. pulex* (Thomson *et al.*, 2009). All amplicons corresponded to the predicted sequence length and sequencing further confirmed the identity of the gene products. Thus mRNA expression was confirmed for 34 out of 40 identified receptor genes [GenBank accession: JZ390894-JZ390939]. Due to time constraints no PCR was conducted for BgTHR, which was identified from the final assembly. For *L. gigantea*, the available expressed sequence tags (EST) were screened using BlastN and identified transcripts confirmed the expression of the predicted genes and mRNA expression was confirmed for 22 out of 33 identified genes, (see Table 4.0.6). E-value of zero indicates a 100% similarity between the ESTs in the hit accession column and the corresponding NR.

<table>
<thead>
<tr>
<th>NR Unclassified</th>
<th>NRU2</th>
<th>NRU3</th>
<th>NRU4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>221</td>
<td>293</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250*‡</td>
</tr>
</tbody>
</table>
Table 4.0.6: Summary of all the top blast expressed sequence tag (EST) hits for L. gigantea NRs, including the EST description, Genbank accession number and E-value.

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Sequence description</th>
<th>E-Value</th>
<th>Hit ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LgDAX1</td>
<td>caxx lottia gigantea from radula and heart lottia gigantea cdna clone caxx4873 5 mma</td>
<td>0</td>
<td>FC610206</td>
</tr>
<tr>
<td>LgHIR</td>
<td>caxu lottia gigantea from female gonad lottia gigantea cdna clone caxuh659 5 mma</td>
<td>0</td>
<td>FC646813</td>
</tr>
<tr>
<td>LgKAR</td>
<td>cbgc lottia gigantea 15h 18h embryos lottia gigantea cdna clone cbgc7034 5 mma</td>
<td>0</td>
<td>FC805285</td>
</tr>
<tr>
<td>LgPFAR1</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgPFAR2</td>
<td>cbbn lottia gigantea larva lottia gigantea cdna clone cbbn9244 5 mma</td>
<td>0</td>
<td>FC775242</td>
</tr>
<tr>
<td>LgE75</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgRev_erb</td>
<td>cbbn lottia gigantea larva lottia gigantea cdna clone cbbn8629 5 mma</td>
<td>0</td>
<td>FC774887</td>
</tr>
<tr>
<td>LgE78C</td>
<td>caxx lottia gigantea from female gonad lottia gigantea cdna clone caxw9550 5 mma</td>
<td>0</td>
<td>FC678555</td>
</tr>
<tr>
<td>LgNR3</td>
<td>cbbn lottia gigantea larva lottia gigantea cdna clone cbbn11308 5 mma</td>
<td>0</td>
<td>FC759257</td>
</tr>
<tr>
<td>LgKOR</td>
<td>caxx lottia gigantea from male gonad lottia gigantea cdna clone caxx3392 5 mma</td>
<td>0</td>
<td>FC692720</td>
</tr>
<tr>
<td>LgEcr</td>
<td>caxx lottia gigantea from male gonad lottia gigantea cdna clone caxx3932 5 mma</td>
<td>0</td>
<td>FC701413</td>
</tr>
<tr>
<td>LgNR1J1</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgNR1J2</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax4376 5 mma</td>
<td>0</td>
<td>FC587866</td>
</tr>
<tr>
<td>LgNR1J3</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgHN4</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax2428 5 mma</td>
<td>0</td>
<td>FC584194</td>
</tr>
<tr>
<td>LgKKX</td>
<td>cbbn lottia gigantea larva lottia gigantea cdna clone cbbn11953 5 mma</td>
<td>0</td>
<td>FC760240</td>
</tr>
<tr>
<td>Lg1R</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax1708 5 mma</td>
<td>0</td>
<td>FC582901</td>
</tr>
<tr>
<td>LgFLX</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgDF</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgFXI</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgNR2E</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax1752 5 mma</td>
<td>0</td>
<td>FC592666</td>
</tr>
<tr>
<td>LgPNK</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgCOUP-1Fa</td>
<td>cbbn lottia gigantea larva lottia gigantea cdna clone cbbn11647 5 mma</td>
<td>0</td>
<td>FC759752</td>
</tr>
<tr>
<td>LgER</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax6663 5 mma</td>
<td>0</td>
<td>FC591742</td>
</tr>
<tr>
<td>LgBRK</td>
<td>cbgc lottia gigantea 15h 18h embryos lottia gigantea cdna clone cbgc18627 5 mma</td>
<td>0</td>
<td>FC782626</td>
</tr>
<tr>
<td>LgHR38</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgNR4a</td>
<td>caxx lottia gigantea from male gonad lottia gigantea cdna clone caxx755 5 mma</td>
<td>0</td>
<td>FC699273</td>
</tr>
<tr>
<td>LgFZ-F1</td>
<td>eDNA clone CBBN9480 5mRNA sequence 3.4.5.6.5d Larvae (M) Lottia gigantea</td>
<td>0</td>
<td>FC773593</td>
</tr>
<tr>
<td>LgBR4a</td>
<td>cbax lottia gigantea from radula and heart lottia gigantea cdna clone cbax4376 5 mma</td>
<td>0</td>
<td>FC587866</td>
</tr>
<tr>
<td>LgHR4b</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>LgHR39</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>Lg2DBDNR1</td>
<td>caxx lottia gigantea from male gonad lottia gigantea cdna clone caxx20789 5 mma</td>
<td>0</td>
<td>FC689646</td>
</tr>
<tr>
<td>Lg2DBDNR2</td>
<td>NA</td>
<td>--</td>
<td>---</td>
</tr>
</tbody>
</table>

4.4.3 Functional Annotation

Blast2go, InterProScan and GO annotation confirmed the identities of all the NRs, predicting the presence of specific conserved features. Table 4.0.7 shows the top protein Blast hits, their corresponding weight in terms of E-value, different protein signatures identified by InterProScan and GO annotation for the NRs. Overall InterProScan identified protein signatures specific to NRs like DNA binding Zinc finger (IPR001628) and ligand binding (IPR000536) and steroid hormone receptor signature (PR00398) in all the NRs and GO annotation identified biological activities like steroid hormone receptor activity; steroid hormone mediated signalling pathway in all the NRs. The biological process identified for COUP TFII was radial pattern formation/skeletal muscle tissue
development/neuron migration/anterior-posterior pattern specification, which is characteristic of COUP TFII NR (http://www.uniprot.org/uniprot/O09018). The structural annotation defined the gene boundaries and coding regions of the corresponding transcripts in GFF3 files that was manually edited using Artemis/Apollo.

4.4.4 Phylogenetic analyses

Phylogenetic analyses were performed using the conserved DBD and/or LBD regions of expressed NR sequences from *B. glabrata* and predicted sequences from *L. gigantea* aligned to *D. melanogaster, C. elegans* and *H. sapiens* NR sequences. Phylogenetic trees were constructed using 4 different approaches: Bayesian Inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML). All the models agreed at family and subfamily level (Figure 4.0.7) (Appendix 4.2, Appendix 4.3 and Appendix 4.4) but BI showed greatest resolution at the base of the phylogram. (Figure 4.0.7) shows the BI tree with nodal support generated from all the tree construction methods. Phylogenetic analysis of the nuclear receptor genes from *B. glabrata* and *L. gigantea* compared with *D. melanogaster, C. elegans* and *H. sapiens* (Figure 4.0.7) reveals two major clades (NR1/0 and NR2/3/4/5/6) each further dividing and sub-dividing into several sub-clades. All the trees predicted the position of BgNRU1/2/3 close to NR0A subfamily within the NR1 group; however the NR0 subfamily was derived to accommodate nuclear receptors that lack either the DBD or the LBD both of which are present in BgNRU1/2/3. These are different genes with overlapping regions but the exact relationship of each of these receptors to each other is also not resolved. Unidentified NRs have been previously reported in two cnidarian species (*Nematostella vectensis* and *Hydra magnipapillata*) (Reitzel et al., 2011; Reitzel and Tarrant, 2009) but our phylogenetic analysis using these NRs (assigned names same as in (Reitzel and Tarrant, 2009); NVNR1, NVNR2, NVNR3, (Data not shown) revealed that these do not cluster with BgNRU1/2/3 leaving open the question as to whether these NRs were a result of species-specific divergence. The NR1D1/2/3 receptors are partial fragments and, due to the fragmentary nature of the preliminary genome used for this study, we cannot currently determine if they are multiple partial genes or parts of the same transcript. Phylogenetic analyses (Bayesian Inference) of BgTHR were done using NR1 subfamily genes. The BgTHR clustered with human and Lottia THR and the branch is well supported with BMCMC posterior probability of 0.98 (Figure 4.0.8).

In subfamily 2, the position of members of Groups NR2C/D is also not resolved (Figure 4.0.7). Both mollusc genomes each contain one member of this clade that we have
designated LgTR/BgTR based on their sequence similarity with HsTRs, but they cluster with Ce_HR41 and Dm_HR78. The Bayesian tree predicts LgTR to be more closely related to Ce_HR41 with DM_HR78 as the most distal member but the ML tree shows both the mollusc sequences to be closely related. However the ML tree placed Ce_HR41 in NR2D, although this NR is usually classified in NR2C. Clearly, the NR2C/D groups seem to be related and not well defined.

Within the NR3 subfamily, only the position of the GR receptor showed non-concordance between the BI tree and the trees from the alternative methods. All of the trees were in agreement for the NR4/NR5/NR6 subfamilies, with one difference in the position of BgNR4a/LgNR4a and NR5B in the MP tree. The identified genes have been named in agreement with the unified nomenclature system for NRs based on their sequence homology and phylogenetic position.

Overall, the identified mollusc NRs encompass almost all the known NR subfamilies, with representatives divided between 21 groups (Table 4.0.5). Similarities and differences in the NR complement between the species used in the study were also seen. Eight of these sub-family groups 1D, 1F, 2A, 2E, 2F, 4A, 5A and 6A were represented in both invertebrates and vertebrates. The presence of these groups among protostomes and deuterostomes supports the suggestion that these receptors originate in a common ancestor of the bilateria before the divergence occurred (Bridgham et al., 2010; Bertrand et al., 2004). The mollusc genomes also contain NR homologues in 1E, 1J and 5B that are present in insects and nematodes but not in vertebrates. Representation of these groups in L. gigantea and B. glabrata shows that these nuclear receptors pre-date the ecdysozoa/lophotrochozoan split. Four groups (0B, 1A, 1C and 3A) are present in both molluscs and H. sapiens but not in D. melanogaster and C. elegans. The presence of these receptors in molluscs indicates the possibility of commonality in shared signalling pathways in molluscs and vertebrates, elucidating the evolutionary development of the endocrine system. The presence of additional receptors is most likely the result of gene or lineage-specific duplication events.
Figure 4.0.7: Phylogenetic relationships of NRs in molluscs, humans, fly and worm.

NRs from five species were subjected to phylogenetic comparisons using Bayesian Inference, Maximum Parsimony and Maximum Likelihood methods. The Bayesian tree is shown with posterior probability values from Bayesian tree and bootstrap values from Maximum Parsimony and Maximum Likelihood trees. The value of 1 on each node represents 100% posterior probability/bootstrap support; an X indicates an area of disagreement from the Bayesian tree (Appendix 4.2: expandable pdf format of this tree with all node support values). Notations Bg, Lg, Hs, Dm and Ce in association with receptor names denote sequences from B. glabrata (highlighted in red), L. gigantean (highlighted in blue), H. sapiens, D. melanogaster and C. elegans, respectively. Scale bar, 0.1 expected changes/site.
Figure 4.0.8: Phylogenetic relationships of THR\(s\) in molluscs, humans, fly and worm using Bayesian inference showing posterior probability values.

The value of 1 on each node represents 100% posterior probability. Notations Bg, Lg, Hs, Dm and Ce in association with receptor names denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster and C. elegans, respectively. Scale bar, 0.1 expected changes/site.
4.5 Discussion

4.5.1 Steroid receptors (Group 3)
Steroid receptors were originally thought to be a vertebrate-specific gene family but the identification of the genes with a clear homology to human ERR in Trichoplax (Baker, 2008), human ER in molluscs and SR in amphioxus (Baker et al., 2003) suggest that these receptors arose early in metazoan evolution and subsequently proliferated in vertebrates through a series of gene duplication events. We have identified vertebrate-like ERs in B. glabrata and L. gigantea, which have been previously reported in other molluscan species, including bivalves, Crassostrea gigas (Matsumoto et al., 2007), Mytilus edulis (Puinean et al., 2006) and Chlamys farreri (Zhang, Pan and Zhang, 2012), and gastropods, Aplysia californica (Thornton, Need and Crews, 2003), Lymnaea ollula (Kumkate et al., 2009), Marisa cornuarietis (Bannister et al., 2007), Thais clavigera (Kajiwara et al., 2006), Nucella lapillus (Castro et al., 2007), and cephalopod, Octopus vulgaris (Keay, Bridgham and Thornton, 2006). The phylogenetic position of the mollusc ER with the vertebrate ER is well supported with a BMCMC posterior probability of 99% and a bootstrap values of 0.95 (Figure 4.0.7), supporting the suggestion that vertebrate and invertebrate ER diverged from a common ancestor, before the evolution of the deuterostomes (Eick and Thornton, 2011; Baker, 2008; Thornton, Need and Crews, 2003). Structural similarities between ER DBDs of all molluscs, annelids, cephalochordates and vertebrates have been documented showing them to bind to and regulate transcription through estrogen response elements (Katsu et al., 2010; Keay and Thornton, 2009; Bridgham et al., 2008; Keay, Bridgham and Thornton, 2006). The ER LBD in non-vertebrates has unique functionalities, with the molluscs activating transcription constitutively in the absence of ligand (Bannister et al., 2013; Eick and Thornton, 2011; Matsumoto et al., 2007; Castro et al., 2007; Kajiwara et al., 2006; Keay, Bridgham and Thornton, 2006). In annelids the ER has been shown to activate transcription in presence of estrogens (Keay and Thornton, 2009) and in cephalochordates ER LBD transcriptional activity in response to estrogens is lost (Bridgham et al., 2008). No ERs have been identified in ecdysozoa (Maglich et al., 2001), echinoderms (Howard-Ashby et al., 2006) and urochordates (Dehal et al., 2002). One possible explanation could be that independent gain/loss of ligand binding ability has occurred during the evolution of these taxa; alternatively there could be a different origin of estrogen signalling mechanisms in molluscs and vertebrates (Eick and Thornton, 2011; Baker et al., 2003; Escriva, Delaunay and Laudet, 2000). We also identified a single estrogen-related receptor (ERR) in both B. glabrata and L. gigantea, which clusters with
Drosophila and human ERRs. In molluscs, ERRs (group 3B) have been reported previously in M. cornuarietis (Bannister et al., 2007), C. gigas [GenBank accession: EKC20050] and M. edulis (Puinean et al., 2006). ERRs have no known endogenous ligand, although they are thought to bind to estrogen response elements and may play a role in estrogen signalling and energy metabolism (Bardet, Laudet and Vanacker, 2006). The presence of another mammalian steroid hormone receptor, the androgen receptor (AR), in molluscs has been hotly contested, its existence having been inferred from the effects of androgens in several species of mollusc (Sternberg, Hotchkiss and LeBlanc, 2008; Koehler et al., 2007), rather than by identification of a homologous sequence. We identified no convincing orthologues in either of the two molluscs for the AR or for other members of group 3C (Figure 4.0.7), which also contains the glucocorticoid receptor (GR), mineralocorticoid receptor (MR) and progesterone receptor (PR). The absence of an AR sequence homologue, both in our systematic genomic searches, as well as specific lab-based searches in other molluscs (Santos et al., 2005) (an approach which has successfully identified other receptors) strongly suggests no vertebrate AR homologue exists in these two molluscs.

4.5.2 Group 1 and 2 nuclear receptors

The absence of a molluscan AR suggests that alternative pathways must exist for spermatogenesis/oogenesis in molluscs. I identified homologues of Group 2 C/D vertebrate testicular receptor 4 (TR4) in B. glabrata and L. gigantea designated as BgTR and LgTR, that in mice have been shown to control spermatogenesis (Mu et al., 2004) and folliculogenesis (Chen et al., 2008). TR4 can cross talk with other NRs like peroxisome proliferator-activated receptor (PPAR), retinoic acid receptor (RAR), thyroid hormone receptor (THR) and estrogen receptor (ER) to control its target genes (Collins et al., 2011). This suggests that there is some degeneracy in the alternative pathways for spermatogenesis and other processes, which exist in vertebrates and which may be important in invertebrates (Greenspan, 2005; Greenspan, 2001).

The identification of PPAR and retinoid X receptor (RXR) has particular significance in molluscs as a recent study of N. lapillus suggests PPAR pathways may induce imposex (Pascoal et al., 2012) in response to tributyltin (TBT), which binds to the RXR (le Maire et al., 2009) in this species. I identified two PPAR homologues in both L. gigantea and B. glabrata, both containing identical sequences to the P-box of human PPAR (CEGCKGFFRRTI). Convergence of 9-cis retinoic acid, a natural ligand for RXR, and PPAR signalling pathways through PPAR-RXRα heterodimerisation has been
demonstrated (Kliewer et al., 1992). *B. glabrata* RXR has been previously shown to strongly heterodimerize with vertebrate PPARα (Bouton et al., 2005), which might suggest the possibility of conservation of several important signalling pathways in molluscs.

A THR homologue has been identified in *B. glabrata* (Figure 4.0.8) that has been previously reported in *L. gigantea* and other invertebrates (Wu, Niles and LoVerde, 2007). This hormone receptor has been reported to be involved in the control of metamorphosis in vertebrates and in prenatal development of central nervous system (Paris et al., 2010; Bernal, 2007a). Besides THR, both the molluscs seem to have ecdysone receptor (EcR), considered play an important role in in *D. melanogaster* metamorphosis/molting (Li and Bender, 2000). Metamorphosis in most of the animals occurs in post-embryonic developmental stages and has been well studied in insects like drosophila for which ecdysteroids and juvenile hormones are the main regulators and in vertebrates where THR seems to be actively involved. On an evolutionary scale molluscs lie somewhere between the insects and vertebrates, so it would be quite interesting to explore the possible mechanisms for these hormone receptors homologues in them. Ecdysone signalling in insects is transduced by EcR and ultraspiracle (USP), which is an insect orthologue of the vertebrate RXR (King-Jones and Thummel, 2005), also present in molluscs. The presence of ecdysone has been reported in pulmonate snails, in *B. glabrata* and *Lymnaea stagnalis* (Whitehead and Sellheyer, 1982) but its origin (endogenous or exogenous) is still contested (Garcia, Griffond and Lafont, 1995). The interaction between the ligand binding pocket of *L. gigantea* EcR homolog and an ecdysone-related steroid and its strong expression in the testis (Laguerre and Veenstra, 2010), may indicate an important role in molluscan developmental processes, potentially by heterodimerizing with mollusc RXR.

Comparison of the NR complements in human, fly, worm and snails show that *L. gigantea* has an additional gene in the NR2E group. In vertebrates, the two members of this NR group, tailless (TLX) and the photoreceptor-specific nuclear receptor (PNR) have important roles in brain and retina development (Benoit et al., 2006). In *L. gigantea*, in addition to PNR, dissatisfaction (DSF), TLX and FAX1 (also present in *B. glabrata*), there is an additional NR2E gene. Since these genes function during development of the visual system in both *Drosophila* and humans (Miyawaki et al., 2004; Yu et al., 1994), it is tempting to speculate that an additional NR2E receptor in *L. gigantea* might assist in its relatively complex visual system as an adaption to the intertidal habitat. Functional analyses of the *L. gigantea* NR2E receptors will provide very interesting insights into the evolution of novel and shared roles for these genes.
The absence of NR1I homologues, which include vitamin D receptor (VDR) and pregnane X receptors (PXR), in *B. glabrata* and *L. gigantea*, is significant as these are considered as natural sensors and are involved in vertebrate response to xenobiotic stress (Benoit et al., 2006). NR1J group in protostomes shares similarity with vertebrate NR1I group (Snow and Larsen, 2000) and there is evidence that both NR1I/NR1J group share a common ancestor (Bertrand et al., 2004). Moreover, recent studies have indicated that NR1J members might regulate xenobiotic responses in *D. melanogaster* and *C. elegans* (King-Jones et al., 2006; Lindblom and Dodd, 2006). DHR96 in *D. melanogaster* plays a role in xenobiotic stress response (King-Jones et al., 2006) and also binds to cholesterol, regulating cholesterol metabolism (Horner et al., 2009). In *C. elegans*, NR1J receptor like NHR-8 plays a role in xenobiotic defence (Lindblom and Dodd, 2006). DAF-12 is involved in dauer formation, in which larval development is diverted as a response to adverse environmental conditions (Antebi et al., 2000). Our results show that *L. gigantea* has three representatives and *B. glabrata* three full-length sequences and one fragment (not shown in phylogenetic tree) (Table 4.0.5) representing NR1J, which segregated from the other invertebrate representatives to form a separate clade (Figure 4.0.7). The mollusc NR1J group members possess well-conserved base contact residues (ESCKAFFR) within the DBD, characteristic of the NR1J sub family (Thomson et al., 2009). In molluscs the NR1J group receptors may be able to perform similar functions to the closely related NR1I of vertebrates.

### 4.5.3 Group 5 and 6 nuclear receptors

Germ cell nuclear factor (GCNF) homologues (NR6) have previously been reported in bilateria (Enmark and Gustafsson, 2001) and we have identified them in both molluscs. BgHR4/LgHR4a/b cluster with the fly hormone receptor-like (Dm_HR4). During embryonic stages in vertebrates, GCNF can interfere with retinoic acid signalling affecting the expression of *cyp26A1*, which is essential for normal hindbrain patterning and early developmental stages (Chung and Cooney, 2003). In Group 5 I identified BgFTZ-F1/LgFTZ-F1 and BgHR39/LgHR39, homologues of *Drosophila* NR5 subfamily members, with a conserved stretch of 23 amino acids adjoining the C-terminal end of the zinc finger motif (AVRSDRMRGGRNKFGPMYKRDRA). This sequence is placed immediately after the DBD and plays an important role in high affinity interactions of the receptor with DNA (Ueda et al., 1992). All the receptors in NR5 and NR6 have an evolutionary origin preceding that of vertebrate receptors and are functionally active during embryonic development suggesting that developmental NRs are conserved in
bilateria (King-Jones and Thummel, 2005).

4.5.4 Group 0 receptors
The *L. gigantea* and *B. glabrata* genomes each contain one nuclear receptor homologue to the Human DAX1 receptor. Both mollusc receptors lack DBDs characteristic of the NR0B group. The predicted amino acid sequence is part of the conserved domain, LBD and shares 24 and 30% homology to human DAX1 and 64 and 68% similarity to the bivalve, *C. gigas* (EKC26839). DAX1 is a determinant for early mammalian embryonic development and development of embryonic germ cell lineages (Wang *et al.*, 2006) and regulates key steps for androgen production by repressing the transcriptional activation of SF-1 on the promoters of steroidogenic acute regulatory protein (*StAR*) (Lalli *et al.*, 1998). I was unable to generate an amplicon for *B. glabrata* DAX1 in either adult or embryonic stages (96hrs/120hrs). I cannot exclude the possibility that this gene is expressed during specific stages of development or under specific circumstances, not examined in this study. No NR0A members were identified in either of the mollusc genomes based upon the similarity to the invertebrate nuclear receptors of this group.

4.5.5 DBD nuclear receptors
I identified one 2DBD-NR in *B. glabrata* compared to two 2DBD-NRs reported from *L. gigantea*, three from *S. mansoni* [GenBank: XP_002582074] and one from *D. pulex* [JGI Genome Portal Protein Id: 442743]. These NRs have two tandem DNA binding domains, a well-conserved P-box sequence (CEACKK) in the first DBD and a P-box (CEGCK) in the second DBD. The P-box of the second DBD is similar to DBD of NR1 family but the P-box of the first DBD is unique to this NR and may determine novel targets specific to DNA (Wu *et al.*, 2007). The discovery of homologous receptors in both host (*B. glabrata*) and parasite (*S. mansoni*) may open up avenues to further characterize host-parasite interaction and for possible treatments to revert the parasite adaptation capabilities (Maule and Marks, 2006).

4.5.6 Novel Nuclear receptors
Differences in the NRs of *B. glabrata* and *L. gigantea*, indicate interspecies variation within the molluscs, which, given their evolutionary divergence is perhaps unsurprising. A unique subset of four NRs has been identified in *B. glabrata*, which did not associate with any specific NR group. These have been designated as BgNRU1, BgNRU2, BgNRU3, and BgNRU4 respectively and placed in a separate group (Table 4.0.5). These receptors may have originated from a specific duplication event in a *Biomphalaria* precursor, or
alternatively, they could be remnants of ancient NR subfamilies, whose representatives have been secondarily lost in the other represented species.

4.5.7 Molluscs as model organisms

Molluscan endocrinology is poorly understood, so further study to determine the functionality of these NRs promises new insights, especially concerning the many unanswered questions regarding the effects of steroids and other EDCs on molluscs. The identification of the NR genes in molluscs is a first step to understanding the evolutionary development of these important transcription factors. Our results suggest that the vertebrate-like NRs make molluscs a suitable model candidate to investigate functional relationships of individual receptors with their mammalian counterparts. The ultimate choice of model organisms lies not only with the biology of the organism, but also on its tractability for study and manipulation. Aside from the genome, the full range of genomic tools available for B. glabrata, such as RNAi (Jiang, Loker and Zhang, 2006), BAC library (Adema et al., 2006b), microarrays (Bayne, 2009b), as well as traditional approaches of enhancer and suppressor genetics and transgenics to explore regulatory networks, make B. glabrata a possible candidate. However, the advent of next generation sequencing makes transcriptome study from non-model organisms a real possibility and therefore mollusc species may ultimately make better model systems. In particular, this approach could be used to identify inputs from signal transduction pathways, potential hormone metabolism genes, co-activators, corepressors and other unknown factors that may impinge upon receptor activity. We have shown that there are NR groups with a single molluscan representative for which more than one homologue NRs may be present in humans. For these NR groups, B. glabrata could provide a simple model system, not only to study the development and diversification of endocrine system evolution, but also to study potential therapeutic connections. In the NR1 subfamily, both molluscs have single homologues of several significant NRs, such as the retinoic acid receptor (RAR), PPAR, and ROR. RARs have also been identified in the bivalve, C. gigas [GenBank: EKC30866] and the echinoderm, S. purpuratus [GenBank: XP_779976]. In vertebrates RAR binds retinoic acid, the biologically active form of vitamin A, which mediates cellular signalling in embryogenic anteroposterior patterning of the central nervous system (Chung and Cooney, 2003). Similar functional hypotheses have been made for the RAR of L. gigantea, and the annelid, Capitella teleta [P.id: 168520] (Albalat, 2009). The presence of THR in molluscs indicates that further investigation of molluscs as models of thyroid disease/dysfunction is warranted. Based on the absence of Group3C NRs, it can be
concluded that these two molluscs are inappropriate models for mammalian steroid hormone function mediated via NRs (with the possible exception of ERs), as the genes for several major steroid hormone receptors are not present in either species. However since steroid hormones have also been shown to act via non-genomic mechanisms in vertebrates, using membrane bound receptors (mPR and GPR30) (Levin, 2008) there is still potential to examine alternative pathways in these organisms.

4.6 Conclusions

We identified 40 nuclear receptor genes in the genome of *B. glabrata* and 33 in the *L. gigantea* genome, representing all seven principal vertebrate nuclear receptor groups. An interesting feature emerges when we examine the list of nuclear receptors that are unique to vertebrates. The absence of NRs such as VDR, CAR and PXR, as well as steroid hormone NRs, AR/PR/GR/MR for which we found no orthologues in the two mollusc species examined in this study, indicates that several significant signalling pathways are absent in molluscs. Nevertheless, we have identified an array of NRs common to both vertebrates/molluscs and molluscs/flies. These results could add weight to Thummel’s speculation about the convergent regulation of NRs in vertebrates and invertebrates (Thummel, 1996). Elucidation of NR targets in molluscs may unlock their potential as new model organisms allowing the discovery of new pathways leading to similar phenotypes as in vertebrates or, indeed, similar pathways that produce a different phenotype. The range of phenotypes, and their underlying genetic mechanisms, available for study in different species may enable the identification of alternative pathways mediated by NRs that might also be exploited (Greenspan, 2005). The potential of many invertebrate species, for endocrine study, is yet to be explored, but the underlying fundamental similarities and differences between molluscs and man may be the solution to determining not only the evolution of endocrine systems but also the full intricacies of our own.
Table 4.0.7: Results from Blast2go, InterProScan and GO annotations showing the best hits, their corresponding E-value and protein signatures identified using different InterProScan database.

<table>
<thead>
<tr>
<th>Contig number</th>
<th>Transcript (Vectorbase)</th>
<th>Sequence desc.</th>
<th>Hit description</th>
<th>E-Value</th>
<th>InterProScan</th>
<th>Go annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>534</td>
<td>BGLTMP002 793+ BGLTMP010 904</td>
<td>ec dysone-induced protein 78c-like</td>
<td>gi</td>
<td>52487384.0</td>
<td>ref</td>
<td>XP_005093662.1</td>
</tr>
<tr>
<td>2052</td>
<td>BGLTMP003 233</td>
<td>peroxisome proliferator-activated receptor gamma</td>
<td>gi</td>
<td>52491575.2</td>
<td>ref</td>
<td>XP_005112667.1</td>
</tr>
<tr>
<td>304</td>
<td>BGLTMP000 531</td>
<td>nuclear hormone receptor</td>
<td>gi</td>
<td>405971703.0</td>
<td>gb</td>
<td>EKC36526.1</td>
</tr>
<tr>
<td>1275</td>
<td>BGLTMP004 271</td>
<td>ec dysone in ducible protein 75</td>
<td>gi</td>
<td>52489744.3</td>
<td>ref</td>
<td>XP_005105182.1</td>
</tr>
<tr>
<td>1481</td>
<td>BGLTMP013 350</td>
<td>ec dysone receptor</td>
<td>gi</td>
<td>405951903.0</td>
<td>gb</td>
<td>EKC19773.1</td>
</tr>
<tr>
<td>1939</td>
<td>BGLTMP007 582</td>
<td>nuclear receptor subfamily 1 group d member 2</td>
<td>gi</td>
<td>52489023.1</td>
<td>ref</td>
<td>XP_005101663.1</td>
</tr>
<tr>
<td>1958</td>
<td>BGLTMP006 728</td>
<td>nuclear receptor subfamily 1 group d member 2</td>
<td>gi</td>
<td>443721770.0</td>
<td>gb</td>
<td>ELU10950.1</td>
</tr>
<tr>
<td>201</td>
<td>BGLTMP001 061</td>
<td>nuclear receptor ror-alpha</td>
<td>gi</td>
<td>52489832.8</td>
<td>ref</td>
<td>XP_005106164.1</td>
</tr>
<tr>
<td>2148</td>
<td>BGLTMP004 695</td>
<td>nuclear hormone receptor hr96</td>
<td>gi</td>
<td>52491666.4</td>
<td>ref</td>
<td>XP_005113105.1</td>
</tr>
<tr>
<td>221</td>
<td>BGLTMP012 725</td>
<td>nuclear hormone receptor e75</td>
<td>gi</td>
<td>52491559.7</td>
<td>ref</td>
<td>XP_005112592.1</td>
</tr>
<tr>
<td>2702</td>
<td>BGLTMP013 098</td>
<td>hepatocyte nuclear factor 4-beta-like</td>
<td>gi</td>
<td>405972081.0</td>
<td>gb</td>
<td>EKC36686.1</td>
</tr>
<tr>
<td>2842</td>
<td>BGLTMP005 503</td>
<td>retinoic acid receptor rxa-alpha</td>
<td>gi</td>
<td>52488728.2</td>
<td>ref</td>
<td>XP_005100202.1</td>
</tr>
<tr>
<td>293</td>
<td>BGLTMP013 426</td>
<td>nuclear hormone receptor e75</td>
<td>gi</td>
<td>524900029</td>
<td>ref</td>
<td>XP_005106570.1</td>
</tr>
<tr>
<td>3233</td>
<td>BGLTMP004 658+ BGLTMP004 659</td>
<td>vitamin d3 receptor b</td>
<td>gi</td>
<td>524864409</td>
<td>ref</td>
<td>XP_005089041.1</td>
</tr>
<tr>
<td>3667</td>
<td>BGLTMP013 844</td>
<td>nuclear hormone receptor ftz-f1 beta</td>
<td>gi</td>
<td>524911108</td>
<td>ref</td>
<td>XP_005110423.1</td>
</tr>
<tr>
<td>48</td>
<td>BGLTMP010 830</td>
<td>protein daf- isoform b</td>
<td>gi</td>
<td>524864713</td>
<td>ref</td>
<td>XP_005089191.1</td>
</tr>
<tr>
<td>5703</td>
<td>BGLTMP006 781+ BGLTMP006 780</td>
<td>coup transcription factor 2</td>
<td>gi</td>
<td>405977938</td>
<td>gb</td>
<td>EKC42362.1</td>
</tr>
<tr>
<td>954</td>
<td>BGLTMP003 373+ BGLTMP003 372</td>
<td>vitamin d3 receptor</td>
<td>gi</td>
<td>524900058</td>
<td>ref</td>
<td>XP_005106454.1</td>
</tr>
<tr>
<td>469</td>
<td>BGLTMP011 477</td>
<td>estrogen receptor</td>
<td>gi</td>
<td>525969631</td>
<td>ref</td>
<td>NP_001191648.1</td>
</tr>
<tr>
<td>751</td>
<td>BGLTMP006 863+ BGLTMP006 865</td>
<td>estrogen-related receptor gamma-like</td>
<td>gi</td>
<td>115394376</td>
<td>gb</td>
<td>AB197120.1</td>
</tr>
<tr>
<td>1318</td>
<td>BGLTMP002 005</td>
<td>nuclear receptor subfamily 5 group a member 2 isoform x3</td>
<td>gi</td>
<td>524871094</td>
<td>ref</td>
<td>XP_005092323.1</td>
</tr>
<tr>
<td>1866</td>
<td>BGLTMP004 549</td>
<td>isoform h</td>
<td>gi</td>
<td>212456767</td>
<td>gb</td>
<td>EFX67867.1</td>
</tr>
<tr>
<td>180</td>
<td>XXXXXXXX X</td>
<td>flocculation protein flol1-like</td>
<td>gi</td>
<td>524872016</td>
<td>ref</td>
<td>XP_005092775.1</td>
</tr>
<tr>
<td>2718</td>
<td>BGLTMP006 933+ BGLTMP006 934</td>
<td>nuclear receptor subfamily 2 group c member 1</td>
<td>gi</td>
<td>524864962</td>
<td>ref</td>
<td>XP_005089314.1</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Description</td>
<td>GO Terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BGLTMP009</td>
<td>retinoid x receptor</td>
<td>g</td>
<td>75017953</td>
<td>sp</td>
<td>Q8TS5C6.1</td>
<td>XR_R_BIOGLRe cName: Full=Retinoic acid receptor RXR; AltName: Full=Retinoid X receptor; Short=BgXRX &gt; g</td>
</tr>
<tr>
<td>BGLTMP006</td>
<td>nuclear receptor subfamily 0 group b member 1-like</td>
<td>g</td>
<td>524911207</td>
<td>ref</td>
<td>XP_005110471.1</td>
<td>PREDICTED: uncharacterised protein DDB_G0271670-like isoform X1 [Aplysia californica]</td>
</tr>
<tr>
<td>BGLTMP008</td>
<td>retinoid x</td>
<td>g</td>
<td>j</td>
<td>443700853</td>
<td>gb</td>
<td>ELT99619.1</td>
</tr>
<tr>
<td>BGLTMP006</td>
<td>nuclear receptor subfamily 1 group d member 2</td>
<td>g</td>
<td>524980544</td>
<td>ref</td>
<td>XP_005104404.1</td>
<td>PREDICTED: flocculation protein FLO11-like [Aplysia californica]</td>
</tr>
<tr>
<td>BGLTMP008</td>
<td>retinoic acid receptor beta isoform 1</td>
<td>g</td>
<td>j</td>
<td>294854043</td>
<td>gb</td>
<td>ADF43963.1</td>
</tr>
<tr>
<td>BGLTMP009</td>
<td>nuclear receptor subfamily 2 group f member 6-likes</td>
<td>g</td>
<td>j</td>
<td>443688970</td>
<td>gb</td>
<td>ELT919492.1</td>
</tr>
<tr>
<td>BGLTMP011</td>
<td>nuclear receptor ro-alpha</td>
<td>g</td>
<td>524900302</td>
<td>ref</td>
<td>XP_005106572.1</td>
<td>PREDICTED: putative GPI-anchored protein PB15E9.01c-like [Aplysia californica]</td>
</tr>
<tr>
<td>BGLTMP002</td>
<td>nuclear hormone receptor</td>
<td>g</td>
<td>j</td>
<td>405967531</td>
<td>gb</td>
<td>EKC32680.1</td>
</tr>
<tr>
<td>BGLTMP000</td>
<td>thyroid hormone receptor alpha isoform 1</td>
<td>g</td>
<td>b</td>
<td>6094470</td>
<td>sp</td>
<td>Q91241.1</td>
</tr>
<tr>
<td>BGLTMP009</td>
<td>e75 nuclear receptor</td>
<td>g</td>
<td>524980925</td>
<td>ref</td>
<td>XP_005102813.1</td>
<td>PREDICTED: mucin-17-like isoform X3 [Aplysia californica]</td>
</tr>
<tr>
<td>BGLTMP006</td>
<td>ecdysone-induced protein 78c</td>
<td>g</td>
<td>j</td>
<td>405954484</td>
<td>gb</td>
<td>EKC21910.1</td>
</tr>
<tr>
<td>BGLTMP011</td>
<td>nuclear receptor ro-beta</td>
<td>g</td>
<td>j</td>
<td>405950647</td>
<td>gb</td>
<td>EKC18621.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>receptor/Thyroid hormone receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----------------------------------</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BGLTMP002 448</td>
<td>nuclear receptor subfamily 2 group e member 1</td>
<td>gi</td>
<td>443724903</td>
<td>gb</td>
<td>ELU12704.1</td>
</tr>
<tr>
<td>22141</td>
<td>BGLTMP008 289</td>
<td>nuclear receptor</td>
<td>gi</td>
<td>524872628</td>
<td>ref</td>
<td>XP_005093075.1</td>
</tr>
<tr>
<td>42</td>
<td>BGLTMP000 234</td>
<td>nuclear receptor</td>
<td>gi</td>
<td>52466526</td>
<td>ref</td>
<td>XP_005090084.1</td>
</tr>
</tbody>
</table>
5 Chapter 5: Reproductive type neuropeptides in Biomphalaria glabrata - identification from the B. glabrata genome
5.1 Background

Gastropod molluscs are used extensively as model species for neuro-endocrine processes that regulate reproductive behaviour (Benjamin and Kemenes, 2013; Koene, 2010). There is strong evidence of neuro-endocrine control of reproduction in molluscs (York et al., 2012; Koene, 2010). Investigations, primarily in egg-laying species suggest that regulation is achieved not by any single peptide, but rather by a complex combination of neurohormones (Henry, Zatylny and Boucaud-Camou, 1999; Nässel, 1996; Van Golen et al., 1996; Geraerts, 1992). Some of the genes involved in reproduction such as conopressin, egg laying hormone, Ala-Pro-Gly-Trp-amide-related peptide (APGWa) and myomodulins have been identified in several molluscs (Veenstra, 2010). APGWamide has been found to aid peristalsis in the vas deferens of the great pond snail Lymnaea stagnalis (Koene, 2010; Van Golen et al., 1996) and act as a spawning stimulant in the male tropical abalone Haliotis asinine (Chansela et al., 2008). It antagonizes the effects of conopressin that causes the vas deferens to contract and ejaculate semen (Koene, 2010). The albumen gland produces epidermal growth factor and trypsin inhibitor, both of which are involved in embryonic development (Nagle et al., 2001). In L. stagnalis, the caudo-dorsal cell hormone (CDCH) contributes to oocyte release, egg packaging and oviposition (Geraerts, 1992; Dogterom, Bohlken and Geraerts, 1983) and the dorsal body hormone (DBH) regulates vitellogenesis and female cell maturation (Geraerts, WPM and Joosse, J, 1975) while the schistosomin peptide hormone acts antagonistically, preventing reproduction by inhibiting CDCH and DBH (Geraerts, 1992). There is evidence that cerebral ganglia containing dorsal body cells and caudo-dorsal cells may be the site for synthesis of steroid hormones and are not of neuronal origin (Van Minnen, Smit and Joosse, 1989; Boer, Slot and Van Andel, 1968). Moreover, gonadotropin-releasing hormone (GnRH) has been found to be important in controlling reproduction in both vertebrates and invertebrates (Gorbman and Sower, 2003). An endogenous homologue of vasopressin, a mammalian neuropeptide, has also been reported in several mollusc species (A. californica; L. stagnalis; O. vulgaris; Conus straitus) (Gruber, 2014; Cruz et al., 1987).

5.2 An introduction to neuropeptides

Neuropeptides are small proteinaceous signalling molecules produced and released by neurons through a regulated secretory pathway. These can act as transmitters, modulators of the central and peripheral nervous system in both vertebrates and invertebrates (Strand et al., 1991). They are involved in most, if not all, physiological processes in metazoans
such as embryonic development (Ugriumov, 2009), organogenesis (Pickart et al., 2006), disease-host response (Boldajipour et al., 2008; Merritt and Sood, 2007) and intercellular signalling (Caneparo et al., 2007). There is evidence that many peptides can have dual role, as neurohormones and neurotransmitters within the central nervous system (CNS). For example, caudo-dorsal cell hormone (CDCH) acts as a hormone in stimulating ovulation in L. stagnalis and also acts like a transmitter, organizing movements underlying egg-laying behaviour. About 70 genes encoding classical neuropeptides have been recognized, that have been grouped into 18 families based on their structural organisation or function (Burbach, 2011). Molecular mapping data from L. stagnalis studies on the central ganglia of the CNS show that approximately 2500 neurons of the total 20,000 neurons are peptidergic (secrete peptide hormones) (Benjamin and Kemenes, 2013). A recent peptidomic analysis of neuropeptides involved in copulatory behaviour indicates that the penis complex contains 44 different peptides arising from 10 different genes (El Filali et al., 2006).

Despite a range of functional roles, all classical neuropeptides are targeted to and processed by the regulated secretory pathway through conserved protein motifs (Burbach, 2011). Classically, neuropeptides are short sequence amino acids that are part of larger, inactive molecule precursor proteins (Southey, Sweedler and Rodriguez-Zas, 2008) and contain a signal peptide at the N-terminal that guides them through the ribosome into the lumen of rough endoplasmic reticulum (ER). The N-terminal signal peptide is a hydrophobic sequence of 17-30 amino acids, which is cleaved off upon entry into the rough ER by the action of the signal endopeptidase (Figure 5.1). Only those proteins lacking a transmembrane domain will pass through the ER and eventually be secreted outside of the outer cell membrane. The protein molecule with signal peptide removed is a biologically inactive pro-hormone composed of one peptide or a chain of several peptide (identical or dissimilar) (Hökfelt et al., 2000) which must be proteolytically processed into smaller active peptides. The pro-hormone is transferred to the Golgi apparatus where posttranslational processing starts.

In the Golgi apparatus, packaging of the pro-hormones into large dense core vesicles with processing enzymes (also called convertases) takes place and they undergo extensive cleavage prior to producing the actual bioactive peptides. Proprotein convertase cleaves the pro-hormone at specific cleavage sites that could be flanked by monobasic, dibasic or tetrabasic amino acid combination of lysine-arginine (KR); arginine-arginine (RR) or lysine-lysine (KK) from the C-terminal region, although some separations may occur at a single basic amino acid (Veenstra, 2000). After cleavage from the pro-hormone, the
resulting intermediate peptide contains carboxyl terminal basic amino acids, which are removed by a specific carboxyl peptidase. The carboxyl terminal glycine residue, if present is transformed into an amide (Figure 5.1). This amidation protects peptides from degradative enzymes and is essential for peptide bioactivity (Spijker et al., 1999). Some pre-hormones produce structurally related peptides possessing comparable activities, whereas others have been shown to yield peptides with different potencies (occasionally opposite) (Clynen et al., 2006). For example ghrelin precursor molecule produced two peptides with antagonistic roles in weight regulation (Zhang et al., 2005). L. stagnalis, on the other hand offers an example of encoding of structurally related amino acid motifs on one gene, 13 copies of F(X)RI amide encoded on the same gene (Benjamin and Kemenes, 2013). Therefore it is important to identify all peptides originating from a single precursor since each of these peptides might have a different physiological relevance.

Based on the structural and functional characteristics discussed above, a biologically active neuropeptide as defined by Veenstra (2000) is a peptide which has structural similarity to a known neuropeptide, or a peptide being produced in multiple copies which are identical and surrounded by typical neuropeptide convertase sites within the precursor that contains a signal peptide at the N-terminal of the peptide. This definition does not
include spacer peptides, which have often a large number of acidic amino acid residues and in general do not have well conserved copies on the same precursor, or proteins, which have repeated sequences but are unlikely to be neuropeptide precursors (Veenstra, 2000).

Although there is not much information regarding the impact of environmental chemicals on the expression of reproductive type neuropeptides in snails, TBT, a widespread contaminant that has been linked to the induction of imposex in molluscs, was found in higher concentration in nerves and ganglia of N. lapillus (Bryan et al., 1993). This led to the hypothesis that accumulation of TBT in ganglia may disturb reproductive-hormonal regulators acting through neuropeptides in molluscs (Oetken et al., 2004). In support of this, when Ilyanassa obsoleta were injected with \(10^{-12} - 10^{-10}\) M APGWamide/animal/per day, it caused imposex after seven days of exposure (Oberdorster, Romano and McClellan-Green, 2005; Oberdörster and McClellan-Green, 2000). Conversely, another study with Bolinus brandaris (a neogastropod) reported no imposex when exposed to 100 fold higher molar concentration of APGWamide (Santos et al., 2006). Castro et al. (2007) also found no effect after injecting \(10^{-9}\) M/g body weight APGWamide in Nucella Lapillus (Castro et al., 2007). The contradiction in these results could be due to the differences in the reproductive behaviour of the snails as discussed earlier (Chapter 2). Other contaminants, which may be found in urban effluents, include neuro-active chemicals like antidepressants. Fluoxetine (a drug used to treat depression in humans and pet animals) induces spawning and parturition in estuarine and freshwater bivalves (Fong and Ford, 2014; Metcalfe et al., 2010) and studies have demonstrated the potential of serotonin uptake inhibitors to disrupt reproductive activity in Spisula subtruncata clam populations in the vicinity of sewage effluent outfalls (de la Ossa Carretero, JA et al., 2008).

My fascination for bioinformatics and the availability of complete B. glabrata genome was an opportunity for me to gather knowledge regarding the reproductive type neuropeptide precursors in B. glabrata that would help to increase our knowledge of the basic reproductive biology and the possible underlying neuro-endocrine processes that might be regulating the physiological responses (in terms of reproductive output) of the snail when exposed to chemicals. The aim of this work was to identify putative reproductive type neuropeptides in B. glabrata and assess their structural organisation with known neuropeptides from other species.

5.3 Annotation of reproductive type neuropeptides
Annotation is a process of identifying ‘regions of interest/genes’ and attempting to assign them a name and/or function. It is a powerful approach to unravel and gather valuable biological information of genes and to assert the accuracy and reliability of the identified genes. As more genomes are sequenced and more annotations become available the evidence based approach to annotate genes becomes more powerful as the chances of homology increases.

The complete process of annotation describes manually validated gene names, and gene names transferred from other annotated genomes due to strong protein sequence homology. It includes:

- **Structural annotation**: This gives information about the physical location of the gene in the genome (chromosome, scaffold and contig) and its structural organisation (Intron-exon structure);
- **Functional annotation**: This predicts the different protein domains, links it to three functional signatures characteristics of animal physiology- reproduction, development and metabolism and suggests a gene name based on the presence of homologous/similar functional domains.

The main aim of annotation is to identify genes of interest for the annotators and make a record of this for the community. The work with neuropeptides has been performed in collaboration with Dr Scott Cummins (University of the Sunshine Coast, Queensland, Australia) and Dr Joris Koene (VU University, Amsterdam).

### 5.4 Methods and materials

#### 5.4.1 Data provided by the consortium

The data used for *B. glabrata* gene annotation was

- 12 tissues transcript data files from adult *B. glabrata* (BB02) having transcripts longer than 500 base pairs (bps). The 12 tissues were AG - albumen gland; BUC – buccal mass; CNS - central nervous system; DG_HP - digestive gland; Foot – head foot; HAPO_APO – heart/amoeocyte-producing organ; KID - kidney; MAN - mantle edge; OVO - ovotestis; SAL - salivary glands; STO - stomach and TRG – terminal genitalia.
- Data used from Vectorbase (https://www.vectorbase.org)
- BB02 Contigs (BglabB1.fa)
- BB02 Scaffold (Bglab1.fa)
BB02 Transcripts (BglaB1.fa)
BB02 Base features (BglaB1.o.GFF3)
BB02 Contig2Scaffold (BglaB1.agp)
BB02 Repeats.lib

- Expressed sequence tags (ESTs) from SnailDB.org
  Clustered
  Unclustered

The following methodological steps were performed to identify genes in *B. glabrata*

1) 12 ‘local’ databases were made from the 12 tissue transcript data FASTA files using Seqtools (http://www.seqtools.dk/).

2) *B. glabrata* 12 tissue transcriptome database in Seqtools were screened (BlastN) using GenBank ESTs (E-value: 1e-98; Descriptions: 10; Alignments: 2) and the same was done against the genome (Vectorbase E-value: 1e-80).

3) Identified sequences from different tissues were aligned using Muscle (Seqtools) and Clustal omega (http://www.clustal.org/) to generate a consensus sequence.

4). Exon-intron boundaries, domains, signal peptides, peptides and repeats were identified using a variety of web-based software.

5). Cross-species phylogenetic analysis of the newly identified genes was performed using Maximum Likelihood trees using MEGA5 (http://www.megasoftware.net). MEGA was used for making Maximum Likelihood due to problems with the older version of PAUP and non-availability of the newer version.

6). Identified gene information was aligned with the genomic sequence using Artemis and the edited GFF3 files uploaded in the Community Annotation Portal (CAP) of Vectorbase.

### 5.4.2 Identification of putative neuropeptides from *B. glabrata* transcript data files

The neuropeptide precursors datasets from *Aplysia californica*, *Lymnaea stagnalis*, *Crassostrea gigas* (Feng et al., 2009; Fleury et al., 2009; Moroz et al., 2006), *Lottia gigantea* (Veenstra, 2010), *Helix aspersa* and *Theba pisana* (Scott et al., personal communication) were compared to three Biolephalaria datasets, the genome assembly at Vectorbase, RNA-Seq transcripts in the 12 local libraries (corresponding to 12 tissue transcript data that was uploaded in Seqtools(8.4ver), and a dataset of all available *B. glabrata* ESTs at that time (October 2013), which consisted of 15224 entries from GenBank and 54304 (www.ncbi.nlm.nih.gov) entries from ESTs. The aim was to identify the sequence homologs of mollusc neuropeptide precursors and any additional NRs from the *B. glabrata* datasets with homology based searches.
TBLastN (protein query against the 6-frame translation of a nucleotide database) was used to search with default values and BLOSUM62 as a substitution matrix as search parameters. A secondary search was performed with the identified sequences with default parameters to identify if there were isoforms present. The identified transcripts were aligned using MUSCLE (Edgar, 2004) in Seqtools and ClustalΩ, analysed for splice variants and a consensus sequence obtained.

5.4.3 Computational analyses

The identified nucleotide sequences were translated using proteomics tools at EXPASY website (http://www.expasy.org). The translation process normally generated six possible reading frames. The frame that produced the longest reading frame with no stop codons was selected. The peptide sequences were analysed using SMART (http://smart.embl-heidelberg.de) confirming their conserved domains, repeats, transmembrane regions etc.

5.4.4 Gene Feature finding for neuropeptides

5.4.4.1 Identification of signal peptides

The translated reproductive type neuropeptide data set was analysed with TargetP1.1 (http://www.cbs.dtu.dk/services/TargetP/). All the peptide sequences (fasta format) were fed to TargetP using default parameters. Sequences were considered to be targeted to the secretory pathway if they had SP (secretory pathway) score >0.8, indicating they had a signal peptide. In order to ascertain the presence of signal peptides, SignalP3.0 (http://www.cbs.dtu.dk/), which implements the algorithm described by Bendtsen et al. (2004) was used (Dyrløv Bendtsen et al., 2004). This method uses both Neural Network method and the Hidden Markov Model methods and the criteria for selection were based on scoring method for discriminating between secretory and non-secretory proteins, length and composition of the sequence and cleavage site prediction of the protein (Appendix 5.1). Both Neural Network method and the Hidden Markov Model methods and the ‘short’ and ‘no graphics’ options were selected. Proteins were considered to have a signal sequence if SignalP returned a “YES” for four out of five scores for the Neural Network method and the Hidden Markov Model method predicted that it contained a signal peptide. Transmembrane domains were then predicted using TMHMM 2.0c (http://www.cbs.dtu.dk) using the default settings. Subsequently all the neuropeptides were analysed with the web version of Phobius (www.phobius.sbc.su.se) using a ‘short’ option (Appendix 5.1) that predicts transmembrane topology and signal peptide position. Consistency in predictions from each method was assessed to determine if the predictions
were robust.

5.4.4.2 Identification of the cleavage sites

Once the pro-hormone sequence was determined, this was analysed for possible cleavage sites with Neuropred (http://neuroproteomics.scs.illinois.edu) that is based on logistic regression models trained on experimentally verified cleavage information (Southey et al., 2006). Advanced search was done with the peptide sequences in fasta file format using following parameters:

- Model selection: Known motif and Mollusc;
- Output Selection tasks: Only predict cleavage site, Obtain mass of predicted sites and Model accuracy statistics;
- Post Translational Modifications: N- and O-glycosylation, Amidation, Acylation, phosphorylation and sulphonation.

The pro-hormone is processed to yield peptides by cleavage at specific cleavage sites preceded by a basic amino acid residue at position -4, -6 or -8 (Veenstra, 2000). These positions refer to the number of amino acid before the sessile bond, e.g. in the Lysine (Lys)-Arginine (Arg) pair, the Arg occupies position -1, the Lys -2. Generally in vertebrates and insects cleavage after a single Lys residue is very rare (Cohen, Rholam and Boussetta, 1995; Devi, 1991) but this seems to be more common in molluscs (Veenstra, 2000). In Aplysia amide gene, convertase cleavage occurs systematically at a single Lys residue supported by the Arg in the -4 position (Schaefer et al., 1985). In pro-hormone the amino acid residue in position +1, i.e. immediately after the cleavage site has a significant influence on cleavage (Cohen, Rholam and Boussetta, 1995; Nakayama et al., 1992; Devi, 1991). Generally cleavage of neuropeptide precursors seems to occur at basic amino acid residues but from Aplysia neuropeptide precursors there is good evidence that processing can also occur at other sites (Hummon et al., 2003).

5.4.4.3 Intron-exon boundaries

BlastN was performed in Vectorbase using default parameters using the RNA-Seq transcript identified from 12 tissues to identify the scaffold/contig and physical location of exons corresponding to the gene sequence of interest, in the genome. Then a local alignment was performed between the RNA-Seq transcript identified from 12 tissues, GenBank EST (if available), the predicted coding sequence from the Vectorbase and its corresponding scaffold using MUSCLE in Seqtools using default parameters. The splice sites were worked out using the transcript data. As a general rule, most introns start from
the sequence G (U/T) and end with the sequence AG (in the 5’ to 3’ direction) and identified as 5’ and 3’ splice sites (also referred to as splice donor and splice acceptor site) (Padgett et al., 1986) (Figure 5.2). Studies have shown that apart from the donor 5’ and acceptor 3’ splice sites, introns contain a branch site towards the 3’ end of the intron that becomes covalently bonded to the 5’ end during the splicing process (Guthrie and Patterson, 1988) and do not always follow GT-branch-AG rule (Jackson, 1991; Kiss et al., 1989). To overcome this, extra evidence based on homology-based approach was used to confirm the position of splice sites. The homology-based approach can often predict intron location by identifying gaps in the sequence alignment when the candidate gene is aligned to their homologs.

**Figure 5.2: Figure showing Standard gene structure.**
Starting with untranslated region (UTR) at 5’ end, the gene has coding region that translates into a protein, starts with start codon ATG and stops at stop codon (TAA, TAG, TGA) followed by UTR at the 3’ end. GT and AG are the standard splice sites present at the start and end of the introns.

### 5.4.5 Functional annotation of identified reproductive type neuropeptide

The predicted function of the identified genes was confirmed using Blast2go to search public databases, for gene homology as well as additional functional annotation including gene ontology (GO) classification and Kyoto Encyclopaedia of Genes and Genomes object identifiers (KEGG object IDs). FASTA files containing protein sequences for either neuropeptides were uploaded separately in Blast2go (http://www.blast2go.com) (BlastP E-value: 1e-06, word size (default): 3), BlastP searches were performed to identify homologs from different species with e-value less than or equal to 1e-06 and the results were used to map GO terms based on a look-up table matching UNIPROT results to GO terms. BLAST2GO also has the additional benefit of being able to incorporate InterProScan data, and adds this information during the GO annotation step. The InterProScan tool detects protein predictive models or signatures using the integrative data stored in the InterPro database (Hunter et al., 2009) which aggregates diverse information from multiple databases, including Gene3D, PANTHER (Protein Analysis Through Evolutionary Relationships), Pfam (Protein family), PIR (Protein Information Resource), PRINTS, ProDom (Protein Domain), ProSITE, SMART (Simple Modular Architecture Research Tool), HAMAP (High-quality Automated and Manual Annotation of microbial
Proteomes), SUPERFAMILY and TIGRFAM. In the 2014 release, close to 133026 different InterPro signatures were present in the database, and together with the over 25586 UniProtKB entries makes it a valuable tool for protein functional annotation.

The final set of proteins from *B. glabrata* was aligned with known neuropeptides from different species using ClustalΩ to identify the degree of conservation of peptide sequences. Phylogenetic analysis was performed using the alignment with the Maximum likelihood (ML) method. ML was carried out using MEGA5 (Hall, 2005), under Jones-Taylor-Thorton (JTT) substitution model (Jones, Taylor and Thornton, 1992) with a gamma distribution of rates between sites (eight categories, parameter alpha, estimated by the program). Due to lack of time, MrBayes was not used as it is more computer intensive and PAUP version 4.0b10, could not be used as there was some technical problem with the software CD. So I selected MEGA5 for doing phylogenetic tree analyses as it quite powerful and flexible interface and performs very fast.

5.4.6 Structural Annotation and final Gene prediction

Most of the genes predicted from the genome had their associated annotation files in GFF3 (Generic feature format v3) file containing the predicted genomic features such as exon-intron boundaries. A GFF3 file is a standard text file with nine tab delimited columns (Table 5.1) and was developed as part of the “Sequence Ontology Project” (SO) (Eilbeck et al., 2005), replacing the GFF2 file format developed at the Sanger Institute (www.sanger.ac.uk). The main advantage of GFF3 over its predecessors is the mechanism for representing hierarchical information for the gene with a three-tier system (See top left window in the Artemis gene builder in (Figure 5.3). The gene attribute is the parent to mRNA, which is then parent to the CDS.
Table 5.1: Explanation of the Gff3 file format edited during annotation process.

<table>
<thead>
<tr>
<th>Column</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequence id</td>
<td>Scaffold or contig number</td>
</tr>
<tr>
<td>2</td>
<td>Source</td>
<td>Free text qualifier that generated this feature, Ensembl in this case</td>
</tr>
<tr>
<td>3</td>
<td>Type</td>
<td>Term used to define the sequence such as gene, exon, CDS, mRNA etc.</td>
</tr>
<tr>
<td>4</td>
<td>Start</td>
<td>Starting base of the feature</td>
</tr>
<tr>
<td>5</td>
<td>End</td>
<td>Ending base of the feature</td>
</tr>
<tr>
<td>6</td>
<td>Score</td>
<td>A floating point number representing E-value or P-value.</td>
</tr>
<tr>
<td>7</td>
<td>Strand</td>
<td>Strand of the feature with respect to the sequence object, – for minus strand and . If not stranded</td>
</tr>
<tr>
<td>8</td>
<td>Phase</td>
<td>Reference to reading frame. Can be 0, 1 or 2 depending on the number of bases that are removed to reach the first base of next codon and only applicable to CDS.</td>
</tr>
<tr>
<td>9</td>
<td>Attributes</td>
<td>Description of the feature such as name; Parent; target etc</td>
</tr>
</tbody>
</table>

The two genome editing programs considered suitable for this task were Artemis (Rutherford et al., 2000) and Apollo (http://apollo.berkeleybop.org), both of which have been used by many genome projects and can work directly on GFF3 files. Apollo is a user-friendly editor that allows the annotator to update gene models based on the lines of evidence. The sequence aligner lets the annotator detect pseudogenes and highlight sequence discrepancies between expressed and genome sequences (Madupu et al., 2010). Moreover, Apollo can link directly on a remote database and so annotation can theoretically be performed anywhere and fed directly back to the database, useful for groups working in collaboration on similar gene families. Artemis, on the other hand, is java based and displays annotation derived from EMBL feature keys like repeats, promoters, coding sequence etc. (Carver et al., 2008). The Artemis Comparison Tool (ACT) allows comparison between complete and draft genome sequences and their associated annotation making it useful for comparison between different species. Both of the annotation methods were equally valid for this work but due to some technical issue with the web interface of Apollo, Artemis was chosen for annotating genes.

Structural annotation of the identified genes using Artemis enables the visualization of the predicted genes in the updated genome assembly and, using the available evidence (Blast results, ESTs, RNA-Seq data), editing of the exon-intron boundaries and identification of the complete coding sequence (CDS) with start and stop codons and untranslated regions.
UTRs), both 5-prime and 3-prime of the gene. UTRs are difficult to predict as they show high degree of sequence variation and sequence content, but evidence may be present in transcript data. GFF3 files were available for predicted genes, which were used directly in Artemis, in association with the appropriate scaffold sequence. The files were visualized, edited and saved with comments referring to the changes made and evidence used. Each gene annotation on Artemis is composed of 4 main elements (gene, transcript, CDS and exon) (Figure 5.3) that can be visualized in the gene builder in hierarchical order. Modifications were made including intron or exon elimination, exon creation or intron creation as necessary by adding coordinates of the exons in the Gene builder location window. For genes that were not predicted by the automated annotation pipeline, a new annotation element was created by inputting all co-ordinates for exons and introns in relevant reading frame, merging the exons to form the CDS and saving with a new identifier (For example, NEWGENESK01). A community portal for the B. glabrata genome project has been constructed (https://www.vectorbase.org/). The site is controlled with a username/-password login and could be accessed to upload the genes. Some of the putative neuropeptides were split over multiple scaffolds, which could be due to fragmented nature of the genome, making the process of gene finding more complex. For these genes editing was done in their respective GFF3 files (if predicted) and rather than uploading these gene directly on the community portal, they were emailed to the Vectorbase team informing them the order of the exons and the scaffolds in which they are located. Three of the RNA-Seq derived transcripts of pro-hormone neuropeptides did not map to the predicted genes in Vectorbase. For example putative FFamide was not found in Vectorbase or the 12 tissue data, but an EST was found in the GenBank entries. For such genes new GFF3 files were created and given an identifier as stated above.
Figure 5.3: Screen shot of the Artemis Gene Builder for APGWamide showing the hierarchical order of gene, transcript, CDS and exon. The location window shows the start and end base pair co-ordinates of all exons joined together. The Gene map shows the complete gene model with exon and introns.

5.5 Results and Discussion

5.5.1 Reproductive type neuropeptide genes

Here I describe an evidence-based approach that allowed identification of both evolutionarily conserved neuronal genes - the putative pro-hormones and neuropeptides in the genome of *B. glabrata*. Using computational *in silico* data mining, 17 putative reproductive-type neuropeptide genes and 4 insulin-like neuropeptides were identified from the genome and transcriptome databases for *B. glabrata*, encoding precursors that are predicted to yield a repertoire of small neuropeptides with over 124 peptide sequences similar to known bioactive peptides. Table 5.2 shows the tissue distribution of the neuropeptide precursor genes in the RNA-Seq 12 tissue data. The shaded portions in the table indicate the presence of a particular neuropeptide in the corresponding tissue.
As evident from Table 5.2, reproductive type neuropeptides are most prominently expressed within the terminal genitalia (TRG) (13/21) and central nervous system (CNS) (12/21). ELH-1 and myomodulin-3 are exclusive to the TRG while GnRH and Insulin like peptide -2 and -3 are exclusive to the CNS. Albumen gland, ovotestis and salivary glands expressed fewer of the identified neuropeptides with albumen gland expressing only pedal peptide-2. As previously reported in other mollusc species, I found that the B. glabrata genome contains gene families encoding multiple genes such as; pedal peptides (pedal peptide 1, 2, 3); insulin-like peptides (insulin-like peptide 1, 2, 3, 4) and myomodulin.

### Table 5.2: Distribution of putative neuropeptide precursors in the 12 tissues of adult BB02 snail.

<table>
<thead>
<tr>
<th>Neuropeptide Precursor gene</th>
<th>AG</th>
<th>BUC</th>
<th>CNS</th>
<th>DG-HP</th>
<th>APO</th>
<th>FOOT</th>
<th>OVO</th>
<th>SAL</th>
<th>STO</th>
<th>TRG</th>
<th>KID</th>
<th>MAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>APGWamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schistosomin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conopressin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMRFamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNRH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whitnin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal Peptide 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal Peptide 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal Peptide 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFRRamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myomodulin 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myomodulin 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myomodulin 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin-like peptide 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin-like peptide 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin-like peptide 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin-like peptide 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>1</td>
<td>7</td>
<td>12</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>13</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

The 12 tissues are AG -albumen gland; BUC - buccal mass; CNS - central nervous system; DG_HP - digestive gland; Foot - headfoot; HAPO_APO – heart/amebocyte-producing organ; KID - kidney; MAN - mantle edge; OVO - ovotestis; SAL - salivary glands; STO - stomach and TRG – terminal genitalia.
(myomodulin 1, 2, 3). All these genes are related but quite divergent from each other due to differences in cleavage products. An exhaustive search with known reproductive type neuropeptide precursors and peptides found no further sequence homologs.

### 5.5.2 Gene features of putative neuropeptides

The translated peptide sequences were analysed for gene features such as signal peptide, transmembrane regions, internal repeats and other conserved domains. All the neuropeptides had secretory pathway score >0.80, indicating the presence of signal peptide (TargetP1.1) (Emanuelsson, von Heijne and Schneider, 2001) and the position of signal peptide was identical between the different software used (see Appendix 5.1). No transmembrane domain was detected in any of the identified neuropeptides except for 3 (conopressin, GnRH and insulin-like peptide4), each of which had one transmembrane domain in the first 60 amino acids. Conopressin shows transmembrane domain from 1-25aa and the signal peptide identified is between 1-23aa. Similarly GnRH has a transmembrane domain from 1-23aa, which overlaps with the signal peptide region, and the insulin-like peptide4 transmembrane domain is from 1-29aa while the signal peptide is 1-25aa. In all the above cases the transmembrane region overlaps with (partially or fully) signal peptide. This shows that a single transmembrane domain identified in these precursor molecules may be a signal sequence falsely identified as a transmembrane domain.

Neuropred predicted a large number of peptides in *B. glabrata* pro-hormones. Of the predicted cleavage products, those that undergo α-amidation at the C-terminal were assumed to be the potential bioactive cleavage products (Spijker *et al.*, 1999). Table 5.3 shows the predicted amides (C-terminal amides arise from post-translational conversion of glycine) and other cleavage products from the *B. glabrata* pro-hormones which were identified by Neuropred and confirmed based on homology searches in other species by multiple alignments to identify conserved cleavage sites. A notable feature observed was cleavage occurring at monobasic cleavage site (Lysine or arginine residue) and supported by either Arg- or Lys- at -4 positions in *B. glabrata* neuropeptide genes, which has been previously reported in *Aplysia* (Schaefer *et al.*, 1985) Monobasic cleavage sites were found in FMRFamide, conopressin, ELH1 and myomodulin1. In FMRFamide, convertase cleavage occurs at a single Lys residue supported by Arg in -4 positions but in conopressin cleavage occurs at single arginine residue supported by lysine at -4 positions. In ELH1, there is one monobasic cleavage site at lysine supported by arginine at -8 positions. In myomodulin1 monobasic cleavage sites are common with single arginine residue
supported by lysine at -4 positions or single Lys residue supported by the Arg in the -4 position.
Table 5.3: Identified reproductive type neuropeptides from B. glabrata showing putative bioactive cleavage products and the number of peptides encoded in the prohormone.

<table>
<thead>
<tr>
<th>Neuropeptide</th>
<th>No of peptides encoded</th>
<th>Putative Bioactive Cleavage Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>APGWamide (AKA: C-terminally located anterior lobe peptide (CALP))</td>
<td>8</td>
<td>APGW[Amide]</td>
</tr>
<tr>
<td>Conopressin</td>
<td>1</td>
<td>CFIRNC[Amide]</td>
</tr>
<tr>
<td>ELH 1 (AKA: Caudodorsal cell hormone 1)</td>
<td>1</td>
<td>YPIGNDLELLANMVGIKKQQRQEVMDL[Amide]</td>
</tr>
<tr>
<td>ELH 2 (AKA: Caudodorsal cell hormone 2)</td>
<td>1</td>
<td>QPYGD[Amide]</td>
</tr>
<tr>
<td>FFamide</td>
<td>1</td>
<td>RVDSKDMIQQPLL[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>GVNPNLSLFF[Amide]</td>
</tr>
<tr>
<td>FMRFamide</td>
<td>8</td>
<td>FMRFamide</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>(X)nFLRFamide</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>FIRFamide</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>FIRFamide</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>QFYRamide</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>SDKGFYMRFamide</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>EPSRBYR[Amide]</td>
</tr>
<tr>
<td>GnRH</td>
<td>1</td>
<td>HMSQAQNYHFSNG[Amide]</td>
</tr>
<tr>
<td>Insulin-like peptide 1</td>
<td>1</td>
<td>EYDHTCVL[S][RPHN]</td>
</tr>
<tr>
<td>Insulin-like peptide 2</td>
<td>1</td>
<td>AADQPLS[Amide]</td>
</tr>
<tr>
<td>Insulin-like peptide 3</td>
<td>1</td>
<td>QSSFTVICCECFNCTVESEMIDLYC[Amide]</td>
</tr>
<tr>
<td>Insulin-like peptide 4</td>
<td>1</td>
<td>QDISADSVKGT[Amide]</td>
</tr>
<tr>
<td>LFRFamide</td>
<td>1</td>
<td>NTLFRF[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>QGAWFRF[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>AGTLR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>GT[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>GSLLRF[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TLRF[Amide]</td>
</tr>
<tr>
<td>Myomodulin 1</td>
<td>7</td>
<td>ASSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>PSSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>ASNFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>F[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>KSSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>SPSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>APSSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>VPSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>VSSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>SASFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>IPSSFVR[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>EQQSFVR[Amide]</td>
</tr>
<tr>
<td>Myomodulin 2</td>
<td>1</td>
<td>GLAMLRL[Amide]</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>SMKMLRL[Amide]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>QLMMLRL[Amide]</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>PMNMLRL[Amide]</td>
</tr>
<tr>
<td>Myomodulin 3</td>
<td>1</td>
<td>SMEMLSL[Amide]</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>QLMKLK[Amide]</td>
</tr>
</tbody>
</table>
5.5.3 Intronic-exonic boundaries

Compared to *B. glabrata* NR genes, where most of the genes have seven to eight exons, the number of intron/exon boundaries found in selected neurosecretory genes is relatively small, since there are few having more than two exons. Most of the exon-intron boundaries follow the consensus GT-AG rule except one (Pedal peptide 1) that has AC-AG between its two exons. The exons of eight neuropeptides span over two scaffolds and schistosomin has three exons that span over three scaffolds. In most instances the two adjacent exons overlapped by more than a few base pairs, of which only one is likely to be the real exon. Even the best gene predictors and genome annotation pipelines rarely exceed accuracies of 80% at the exon level which means at least one mis-annotated exon per gene (Reese et al., 1997) Automated annotation programs frequently have difficulty in choosing the correct ORF depending on the training data set or the settings and are most
likely to pick the longest ORF and it is here that manual curation becomes important.

5.5.4 Confirmation of peptide identity and function

Blast2go and InterProScan (http://www.ebi.ac.uk) confirmed the identities of all the reproductive type neuropeptides, predicting the presence of peptide specific conserved features. Table 5.4 shows the protein top Blast hits, their corresponding weight in terms of E-value and different protein signatures identified by InterProScan in the neuropeptides. 5 distinct protein families were detected and the sequences were associated with 9 Protein Information Resource (PIR) domain identifiers. The presence of transmembrane regions in conopressin, insulin-like peptide 4 and GnRH was reconfirmed by TMHMM database, which predicted transmembrane helices in the proteins. The GO annotation predicted myomodulins, APGWamide, FMRFamide, pedal peptides and NPY as part of neuropeptide signalling pathway (GO:0007218). Conopressin has neurophyseal domain, which is the characteristic of the family of neuropeptides showing neurophyseal hormone activity. For example, vasopressin in humans. Hormone activity was indicated for most of the neuropeptides except GnRH, Whitnin and LFRFamide (see Table 5.4).

The structural annotation that defined the gene boundaries and coding regions of the corresponding transcripts in GFF3 files was manually edited using Artemis. Three of the identified neuropeptides (FFamide, Insulin-like peptide1 and Pedal peptide1) did not have a predicted transcript in Vectorbase so a new gene model was created. All the proteins had a complete coding sequence with start and stop codons. UTRs were difficult to predict due to variation in the sequence but based on the 12 tissue sequence data. Only 5 precursor molecules were assigned a 5’UTR region.

Here I describe the structural characteristics of the _B. glabrata_ reproductive-type neuropeptides. For some of these detailed cross-species comparison of the peptides using multiple sequence alignment and phylogenetic analyses was performed to elaborate on the global organisation of mature peptides.
Table 5.4: Results from Blast2go, InterProScan and GO annotations showing the best hits, their corresponding E-value and protein signatures identified using different InterProScan database.

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Sequence description</th>
<th>Hit description</th>
<th>E-Value</th>
<th>InterProScan</th>
<th>Go annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>APGWamide</td>
<td>Cerebral peptide 1 precursor</td>
<td>gi</td>
<td>228795</td>
<td>orf</td>
<td></td>
</tr>
<tr>
<td>Schistosomin</td>
<td>Schistosomin-like precursor</td>
<td>gi</td>
<td>59145016</td>
<td>gb</td>
<td>ABW90145.1</td>
</tr>
<tr>
<td>Conopressin</td>
<td>Lys-conopressin-prepro-hormone</td>
<td>gi</td>
<td>59454</td>
<td>gb</td>
<td>AAA29289.1</td>
</tr>
<tr>
<td>FFamide</td>
<td>Hypothetical protein CGL_10013719</td>
<td>gi</td>
<td>40595030</td>
<td>gb</td>
<td>EKC18304.1</td>
</tr>
<tr>
<td>ELH1</td>
<td>Hypothetical protein</td>
<td>gi</td>
<td>516516102</td>
<td>rf</td>
<td>WP017904540.1</td>
</tr>
<tr>
<td>ELH2</td>
<td>Egg-laying hormone</td>
<td>gi</td>
<td>129299</td>
<td>sq</td>
<td>P06308.2</td>
</tr>
<tr>
<td>FMRFamide</td>
<td>FMRFamide partial</td>
<td>gi</td>
<td>169711</td>
<td>sq</td>
<td>P19802.2</td>
</tr>
<tr>
<td>GnRH</td>
<td>Gonadotropin-releasing hormone-like protein</td>
<td>gi</td>
<td>325296899</td>
<td>pe</td>
<td>NP_001191482.1</td>
</tr>
<tr>
<td>Whitnin</td>
<td>Whitnin precursor</td>
<td>gi</td>
<td>7108581</td>
<td>gb</td>
<td>AAF36485.1</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Description</td>
<td>Accession</td>
<td>GO Ids</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Myomodulins1</td>
<td>Myomodulin neuropeptides 1-like</td>
<td>gi</td>
<td>52488575</td>
<td>ref</td>
<td>XP_005099491.1</td>
</tr>
<tr>
<td>Myomodulins2</td>
<td>Myomodulin neuropeptide precursor</td>
<td>gi</td>
<td>359919</td>
<td>emb</td>
<td>CAA65635.1</td>
</tr>
<tr>
<td>Myomodulins3</td>
<td>Myomodulin-like precursor</td>
<td>gi</td>
<td>56567667</td>
<td>gb</td>
<td>AHU82757.1</td>
</tr>
<tr>
<td>LFRFatide</td>
<td>LFRFatide precursor</td>
<td>gi</td>
<td>54873678</td>
<td>gb</td>
<td>AAV41057.1</td>
</tr>
<tr>
<td>NPY</td>
<td>Neuropeptide Y</td>
<td>gi</td>
<td>39287257</td>
<td>gb</td>
<td>AFM85446.1</td>
</tr>
<tr>
<td>Pedal_peptide1</td>
<td>Pedal peptide-1 precursor</td>
<td>gi</td>
<td>52488460</td>
<td>ref</td>
<td>XP_005098921.1</td>
</tr>
<tr>
<td>Pedal_peptide2</td>
<td>Pedal peptide-1 precursor</td>
<td>gi</td>
<td>31745702</td>
<td>gb</td>
<td>AAP57098.1</td>
</tr>
<tr>
<td>Pedal_peptide3</td>
<td>Pedal peptide-1 precursor</td>
<td>gi</td>
<td>32529671</td>
<td>ref</td>
<td>NP_001191623.1</td>
</tr>
<tr>
<td>Insulin-like peptide 1</td>
<td>Insulin precursor</td>
<td>gi</td>
<td>32529675</td>
<td>ref</td>
<td>NP_001191615.1</td>
</tr>
<tr>
<td>Insulin-like peptide 2</td>
<td>Insulin-like peptide 5 precursor</td>
<td>gi</td>
<td>52488640</td>
<td>ref</td>
<td>XP_005090025.1</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------</td>
<td>---------------------------------</td>
<td>------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>9.55296E-15</td>
<td>no TM (SIGNALP_EUK)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like peptide 4</td>
<td>Insulin-like peptide precursor</td>
<td>gi</td>
<td>1836031</td>
<td>gb</td>
<td>AAB46831.1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>GO:0005179; GO:0005576; GO:0005185: neurohypophyseal hormone activity; GO:0005488: binding.</td>
<td></td>
</tr>
</tbody>
</table>
5.5.4.1 APGWamide

Bg-APGWamide encodes 8 APGWamide, one copy of SVDDYCQTLDQIMDEYIHKALEV amide, 4 spacer peptides and one C-terminal peptide (Fig 5.4) identical to previously reported APGWamide in *L. stagnalis*, *A. californica*, *L. gigantea* and *H. asinine* (York et al., 2012; Veenstra, 2000; Smit et al., 1993; Painter et al., 1989). Two APGWamide ESTs were found in the GenBank (EV823649, EV817552) and the 12 tissue RNA-Seq data shows that the gene is expressed in CNS, TRG, APO and KID (Figure 5.4). Multiple sequence alignment shows that the cleavage sites in Bg-APGWamide are conserved with other molluscan homologues and the Bg-APGWamide is most similar to *L. stagnalis* APGWamide (data not shown). Immunohistochemistry of APGWamide demonstrated the bioactivity of this tetrapeptide (Oberdörster and McClellan-Green, 2000; De Lange and Van Minnen, 1998; Smit et al., 1993) and APGWamide seems to control part of the male mating behavior by causing penis eversion (Benjamin, 2008) and its presence in abdominal tissue of imposex female *I. obsoleta* indicates its role in sexual reproduction (Oberdörster and McClellan-Green, 2000). In bivalves such as *Sepia officinalis*, a dipeptidyl aminopeptidase located in the central nervous system has been reported to process APGWamide to dipeptide GWamide that inhibits the contraction of distal oviduct (Henry, Favrel and Boucaud-Camou, 1997). Several related GWamide dipeptides such as TPGWamide, KPGWamide and RPGWamide have been identified from *S. officinalis* and *M. edulis* (Henry, Zatylny and Favrel, 2000; Henry, Favrel and Boucaud-Camou, 1997). In addition to its wide distribution in molluscs, APGWamide shows some structural similarity to crustacean neuropeptide precursor RPCH. RPCH consists of pQLNFSPGWamide (Kornthong et al., 2013; Martínez-Pérez et al., 2002) and the last three amino acids at the C-terminal of the amide (PGW) are identical (Martínez-Pérez et al., 2002b; De Lange and Van Minnen, 1998; Croll and Van Minnen, 1992). Although APGWamide and RPCH share some structural similarities, their functional diversity explains the independent evolution of their receptors (Martínez-Pérez et al., 2002; Henry, Zatylny and Boucaud-Camou, 1999).
Figure 5.4: PreproAPGWamide in B. glabrata.
Figure showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Spacer peptides are indicated in grey; lysine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.

5.5.4.2 Myomodulins
Myomodulins comprise of a family of peptides that control neuromuscular signaling in the specific muscle of the buccal mass controlling feeding system in A. californica (Miller et al., 1993a). They also have well-documented role in modulating various reproductive processes (Greenberg et al., 1997; Van Golen et al., 1996). In L. stagnalis myomodulins were shown to modulate neuromusculature of penis (Kellett et al., 1996) but its exact role in reproduction in H. asinine is not known (Cummins et al., 2009).

Three structurally related myomodulin-like neuropeptides have been identified from B. glabrata genome similar to those previously found in L. stagnalis (Kellett et al., 1996), A. californica (Miller et al., 1993b; Miller et al., 1993a), H. asinine (Cummins et al., 2009), L. gigantea (ESO97371), C. gigas (EKC28655) and Tritonia diomedea (ABU82757).

These neuropeptide precursors are characterised by multiple distinct but structurally related peptides cleaved during post-translational processing (Miller et al., 1993b). All the three Bg-myomodulin genes are differently expressed in the 12 tissues RNA Seq data (Table 5.2). Both Bg-myomodulin-1 and -2 are expressed in 7 different tissues in the snail but Bg-myomodulin 3 is only expressed in the terminal genitalia. The Biomphalaria myomodulin-1 precursor molecule is very similar to the FVRIamide neuropeptide precursor in the annelid, Platynereis dumerilii (Conzelmann et al., 2011; Jékely et al., 2008), and F(X)RIamide in L. stagnalis (El Filali et al., 2006). Since FVRI does not appear to be the dominant naming convention used in the model gastropod L. stagnalis as well as in other molluscs so we (Scott et al, personal communication) have followed similar naming convention. Bg-myomodulin-1 is 412aa sequence encoding 21 copies of XₙF(X)RI (where n can be any number between 3 and 31), one copy of TPMEGAGSYEDDSSEEPIGDD and SGADEVEDE, along with 8 spacer peptides (Figure 5.5). All the peptides have glycine at the C-terminus suggesting that the peptides are amidated after cleavage from the propeptide. Each of the XₙF(X)RI is flanked by monobasic Arg/Lys at C-terminal with another Arg at -4 position similar to FMRFamide.
(Schaefer et al., 1985). Bg-myomodulin-2 is the classical myomodulin precursor as described in other mollusc species (Kellett et al., 1996; Miller et al., 1993b; Miller et al., 1993a). 8 Bg-myomodulin ESTs were found that are overlapping partial fragments (ES490100; EX000262; EV819989; EX003480; ES746942; ES751452; EV820260; EX004598) for Bg-myomodulin-2. Based on the 12 tissue transcript data Bg-myomodulin-2 is a 299aa coding sequence consisting of five different myomodulin-like peptides (GLAMLRL, SMKMLRL, QLKMLRL, SLKMLRL and PMNMLRL) and 5 spacer peptides. The Bg-myomodulin-2 gene encodes 4 copies of highly conserved PMNMLRL as compared to 7 copies is H. asinine (York et al., 2012) and 9 to 10 copies in other gastropods (Veenstra, 2010; Kellett et al., 1996; Miller et al., 1993b) (Figure 5.5). These are flanked at the N-terminus by Lys-Arg basic amino acid pairs and have glycine followed by pair of basic amino acids except one of the SMKMLRL that has monobasic Arg- after glycine at the C-terminus. Bg-myomodulin-3 is the shortest transcript of all the three Bg-myomodulins with 164aa and 8 copies of QLKMLKL, one copy of SMEMLSL and one spacer peptide. As in Bg-myomodulin-2, Bg-myomodulin-3 is flanked at the N-terminus by Lys-Arg basic amino acid pairs and has glycine followed by pair of basic amino acids at the C-terminal except for SMEMLSL that has Arg- after glycine at the C-terminus (Figure 5.5). Peptides of this family exhibit considerable homology in their carboxy terminal sequence (MLRL) and unique N-terminus extension (Kellett et al., 1996; Van Golen et al., 1996; Hirata et al., 1989). Bg-myomodulin-3 peptides exhibit an exception as these peptides share a common C-terminal sequence (MLKL), where Arg- is replaced by Lys-and one of the peptide, SMEMLSL, in which Lysine is substituted for Serine. A multiple sequence alignment (Figure 5.6) highlights the conserved peptides and cleavage sites in B. glabrata, A. californica, L. stagnalis, L. gigantea, H. asinina and C. gigas.
Figure 5.5: Prepromyomodulin in B. glabrata.
Figure showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Spacer peptide are indicated in grey; Glysine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.
Figure 5.6: Multiple sequence alignment showing cleavage sites in prepromyomodulin in B. glabrata.

The accession numbers used in the alignment were: A. californica_1: XP_005099491; A. californica_2: NP_001191658; A. californica_3: NP_001191423; L. stagnalis: CAAG65635; C. gigas: EK28655; L. gigantea I: XP_009051543; L. gigantea II: XP_009051970; H. asinine: AFN20272.

5.5.4.3 Pedal peptides

“Pedal” peptide is a neuropeptide identified from the pedal ganglia from *A. californica* (Willows, Pavlova and Phillips, 1997; Lloyd and Connolly, 1989), which has four different pedal peptide precursors (NP_001191623, NP_001191625, NP_001191585, NP_001191585). One of the physiological roles of pedal peptide in *A. californica* is to control muscular contraction of the foot and thus mediate locomotory behavior (Hall and...
In *Tritonia diomedea*, the pedal peptides mediate the activity of ciliated cells responsible for locomotion and maybe egg transport (Willows, Pavlova and Phillips, 1997). Their function in *L. stagnalis* is still not known (Koene, 2010).

Three similar genes were identified in *B. glabrata* and named pedal peptide 1, 2 and 3 based on their sequence similarity with *Aplysia* pedal peptides (Figure 5.7). Bg-pedal peptide-1 pro-hormone contains 6 copies of PFDSISGSHGSLGFA as in *H. lucorum* (AAB51694) and *L. stagnalis* (AAP57098) and 1 copy of SVGTEDEVLDL and SMDEESLYGAYKESDDDIY (Figure 5.8). Bg-pedal peptide-2 contains 12 different peptides of which 5 are present in multiple copies and the consensus sequence is SIGS(X)FI. Bg-pedal peptide-3 has 14 different peptides with glycine at their C-terminal followed by a pair of basic amino acids. Of the identified peptides, only two peptides are present in 3 copies (PFDRIGTSSFTSF and PFDRIGNSAFTTF). Overall, Bg-pedal peptide-1 seems to lack C-terminal amide which is similar to *Aplysia* and *Lottia* pedal peptides (Veenstra, 2010). Bg-pedal peptide-2 has one C-terminal amide (PDDDEVEEDFENGQLVR) and Bg-pedal peptide-3 has 19 copies of C-terminal amides. This high variability in the peptides makes it difficult to find their homologs but multiple sequence alignment shows that the endoproteolytic cleavage sites are conserved in this gene family in mollusc species, all having dibasic amino acids except for Bg-pedal peptide-3 that has 5 tribasic cleavage sites (Figure 5.7). The multiple sequence alignment of the Bg-pedal peptides with other molluscs reveals conserved cleavage sites (Figure 5.8).
Figure 5.7: Prepropedal peptides in B. glabrata.

Figure showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Sparse peptide are indicated in grey; Glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.
Figure 5.8: Multiple sequence alignment showing cleavage sites in prepropedal peptides in B. glabrata.
The accession numbers used in the alignment were: A. californica_1: NP_001191585; A. californica_2: NP_001191623; A. californica_3: NP_001191625; L. stagnalis: AAP57098; T. diomedea: ABU82763; L. gigantea: XP_009053476; Helix lucorum: AAB51694.

5.5.4.4 FMRFamide
FMRFamide is a member of a large family of structurally related neuropeptides that was first isolated in the clam Macrocallista nimbosa and now has been isolated from species spanning all phyla of the animal kingdom (reviewed by (Greenberg and Price, 1992). Previous studies have shown that these peptides are often highly repeated within a protein
precursor and are found with a number of novel peptides (A. californica (Schaefer et al., 1985); D. melanogaster (Nambu et al., 1988; Schneider and Taghert, 1988); L. stagnalis (Saunders et al., 1991); Cepaea nemoralis (Rösser et al., 2006); C. elegans (Rosoff et al., 1993). The Bg-FMRFamide (Figure 5.9) consists of one copy of FIRFamide, one copy of a pentapeptide QFYRI, 8 copies of FMRF and 15 copies of (Xₐ) FLRF where n could be any number between n=0 to n=70, two copies of novel decapptide (EPSRISYPRY and SDKGPVYMRF) and 18 spacer peptides. Each F(X)RP propeptide is flanked by lysine-arginine or by a single arginine at the amino terminus with arginine/lysine at the carboxy terminus. Each FaRP propeptide has a glycine residue at the carboxy terminus, which is converted to an amide (Loi and Tublitz, 1997; Eipper, Stoffers and Mains, 1992). The tetrapeptide identified in B. glabrata have been previously reported in bivalves, cephalopods and gastropods, pentapeptide, QFYRI and heptapeptide (GDPFLRF) are found only in gastropods. The first decapeptide related to FaRPs was reported in Mytilus edulis and later in O. vulgaris and L. gigantean (Zatylny-Gaudin and Favrel, 2014) but the decapeptide identified in B. glabrata is different from that previously reported in other mollusc classes (Lopez-Vera et al., 2008).

In molluscs and other invertebrates FaRPs regulate various physiological activities. Studies have reported that FMRFamide related peptides also act in coordination with other neuropeptides like APGWamide in regulating male copulation behavior (Zatylny-Gaudin and Favrel, 2014; van Golen et al., 1995).
5.5.4.5  **LFRFamide**  

The LFRFamide precursor has been previously identified in *A. californica* (Cropper et al., 1994), *L. stagnalis* (Hoek et al., 2005), *L. gigantea* (Veenstra, 2010), *Sepia officinalis* (Zatylny-Gaudin et al., 2010) and is also present in *B. glabrata* genome. This is 187aa (Figure 5.10) prepro-hormone encoding one copy each of NTLFRF, QGAWFRY, AGTLLRF, GTLLRF, GGSLLRF and TLFRF and four spacer peptides (Figure 5.10). Of these, NTLFRF, GGSLLRF and TLFRF have been found in *L. stagnalis*. GSLLRF is the conserved peptide reported in all molluscs species investigated so far and TLFRF in *L. stagnalis* and *B. glabrata* seem to be the N-terminus truncated form of *Aplysia* hexapeptide STLFRF (Zatylny-Gaudin et al., 2010). Cross-species comparison shows that the cleavage sites are conserved with all the peptides having identical amino acids (RF) at their C-terminus except QGAWFRY where Arg- is substituted with Tyrosine (data not shown). Similar to FaRPs, LFRFamide peptides also regulate main physiological activities (Zatylny-Gaudin and Favrel, 2014). Functional characterization of *C. gigas* NPF like receptor identified LFRFamide as a specific ligand and was reported to be involved in reproduction and feeding (Bigot et al., 2014). In *L. stagnalis*, schistosome parasites induce LFRF gene expression to suppress host metabolism and reproduction (Hoek et al., 2005).

**MDHSTLLALTFTTAIWCHVYSEEMNHVSAINTLEDHQETHPOQKRSPASSPALYSEDLEADS**

**GLEEPMDDMDKKNTLRFKKRGQAWFROYCKHAGTLLRFGKRGTLLRFGKRGGSLLRFC**

**RGGNSDADFSEDDKRRTLRFGRSDLEEVREDALAREELNSNSPWNEMENPVKRGVNGFHW**

**GQESEN**

*Figure 5.10: PreproLFRFamide in B. glabrata.*  
*Figure showing signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Sparse peptides are indicated in grey; glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.*

5.5.4.6  **Insulin-like peptide**  

The *Biomphalaria* genome contains four insulin-like peptides (Figure 5.11) that are homologs of *Lymnaea* molluscan insulin-related peptides (MIP) (Smit et al., 1993; Li et al., 1992). The lymnaea MIPs encode a preprohormone with similar organization of preproinsulin. A typical insulin-related peptide is the member of insulin superfamily consisting of a signal peptide, a B chain, a C chain and an A chain (Smit et al., 1991) (Figure 5.11) and plays a vital role in regulating metabolism, growth, reproduction and
ageing processes (Claeys et al., 2002; Pertseva and Shpakov, 2002). Multiple sequence alignment with insulin-related peptides from mollusc species (L. stagnalis, A. californica, Tritonia diomedea, Haliotis corrugate), annelid (Platynereis dumerilii) and fly (D. melanogaster) exhibited an overall low sequence similarity with insulin of mammals. However, these peptides are principally similar by their structural organization with well-conserved residues essential for the basic insulin core structure (Figure 5.12). The spacing between the cysteine residues was conserved but the cysteine organisation in molluscs is unusual in the sense that they have four intramolecular disulphide bridges rather than the three typically found in the insulin family (Hamano et al., 2005). In B. glabrata three cysteine groups were present in the B chain and five cysteine groups were present in the A chain of the insulin-like peptides. Similar cysteine organisation has been reported in other mollusc species (Smit et al., 1998). This is unlike a typical vertebrate insulin family where two cysteine groups are present in the B chain and four cysteine groups are present in the A chain. Interestingly, there was a difference in the position of these extra cysteines in the Bg-insulin-like peptide-2 and the Lottia insulin-like peptide-4, which are closest to Drosophila insulin-like peptide-7 (Veenstra, 2010). Like the insect peptide (Li et al., 2008), both these genes (Bg-insulin-like peptide-2 and Lottia insulin-like peptide-4) have characteristic four amino acid residues between the second and the third Cys-residue of the A chain while classical insulin-like peptides typically have three amino acid residues at the similar position.

Another significant aspect of insulin-related peptides in invertebrates is the polymorphism of the genes encoding multiple isoforms, which is seldom seen in insulin of mammals (Rusakov et al., 2003). Polymorphism of insulin-related peptide in C. gigas was associated with the overall growth of the snail (Cong, Li and Kong, 2013).
INSULIN-LIKE PEPTIDE 1

**Signal peptide**

<table>
<thead>
<tr>
<th>B Chain</th>
<th>C Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAYYSRL LLLTLT AHTS VTVA EYDHI TNL VLSRPHRN RCRG GLR VDVTNL LNSVSGLLVRD</td>
<td>TTERVNE NKLKHIN KE AALSYTLK RERTGISV LE EYHT ISELKLY SYDYNKSHA AAY</td>
</tr>
<tr>
<td>TTTERNENKLKHIN KE AALSYTLK RERTGISV LE EYHT ISELKLY SYDYNKSHA AAY</td>
<td></td>
</tr>
</tbody>
</table>

**A Chain**

**PreproInsulin-like peptides in B. glabrata.**

Figure 5.11: Showing signal peptide highlighted in yellow; the different regions of the gene indicate as B Chain, C Chain and A Chain; likely endoproteolytic cleavage sites are indicated in red; Sparse peptide are indicated in grey; cysteine residues are made dark green; glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.

**Figure 5.12:** Multiple sequence alignment showing the conserved cysteine organisation in different...
The accession numbers used were: A. californica: NP_001191621; L. stagnalis I: P07223; L. stagnalis II: P25289; L. stagnalis III: P80090; L. stagnalis V: P31241; L. stagnalis VII: P91797; T. diomedeai I: ABU82759; T. diomedeai II: ABU82760; C. gigas: EKC18432; L. giganteai I: XP_009056008; L. giganteai II: XP_009063889; H. sapiens: AAA52543; D. melanogaster: NP_570070; H. asinine: AEW67132; H. corrugate: ACQ91106

5.5.4.7 Schistosomin

Schistosomin has been reported in four gastropods (A. californica, L. stagnalis, P. acuta and B. glabrata). Previous studies show that schistosomin in L. stagnalis inhibits reproduction and had a role in parasitic castration in the snail (Hordijk et al., 1992). But the B. glabrata schistosomin differed functionally from other molluscs as it was expressed in all life stages from embryonic development to sexually mature snails and seemed to help in development (Zhang et al., 2009).

The previously identified schistosomin B. glabrata (Zhang et al., 2009), is a 97 AA mature peptide with 18aa signal peptide, 8 cysteine residues that form 4 disulphide bonds (De Jong-Brink et al., 1997; Hordijk et al., 1991) and has a peptide (DNYRCPNGDAFECFESDATARFCVS) with glycine at its C-terminal followed by Lys-Arg pair forming cleavage sites. 4 ESTs for schistosomin were found in GenBank (ES488206; ES485584; DW474143; EU126802) and two amino acid sequences that differ by only two amino acids, one in the signal peptide region and one in the mature peptide (Figure 5.13). Extensive BLAST searches did not reveal significant identity of molluscan schistosomin to other known protein domains, suggesting that schistosomin are unique to the gastropods (Zhang et al., 2009) but its absence in L. gigantea narrows this neuropeptide to Euthyneura (Veenstra, 2010). One notable feature is that the tribasic cleavage site found in the schistosomin of B. glabrata is RRK, whereas in all other molluscan schistosomin it is RRR.

Preproschistosomin 1

MKTVFILALIVCAYVA^DNYRCPNGDAFECFESDATARFCVS^KAGVLGSKRKRKYEF
NGAKVSKKPEVE^REDWASTE^TSDNSDVPSVM*

Preproschistosomin 2

MKTVFILALIVCAYVA^DNYRCPNGDAFECFESDATARFCVS^KAGVLGSKRKRKYEF
NGAKVSKKPEVE^REDWASTE^TSDNSDVPSVM*

Figure 5.13: Preproschistosomin in B. glabrata.

Figure showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Sparse peptide are indicated in grey; cysteine residues are made dark green; glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue. The two amino acid sequences just differ by two amino acids highlighted in pink.
5.5.4.8 Gonadotropin-releasing hormone

Gonadotropin-releasing hormone (GnRH) is present throughout the bilateral animals regulating the production of follicle stimulating hormone and luteinizing hormone thereby regulating gametogenesis and steroidogenesis in vertebrates (Sower et al., 2012; Tsai, 2006). In gastropods, the first suggestion of a molluscan GnRH-like peptide was made in *Hellisoma trivolvis* using heterologous GnRH antibodies (Goldberg et al., 1993) and then in *L. stagnalis* (Young, Chang and Goldberg, 1999), *A. californica* (Tsai, Maldonado and Lunden, 2003) and *L. gigantea* (Veenstra, 2011). A comparative cross species analysis of the identified *B. glabrata* GnRH with *A. californica, L. stagnalis, C. gigas, L. gigantea, Octopus vulgaris, Homo sapiens, Uroteuthis edulis, Mizuhopecten yessoensis, Oncorhynchus masou and Schmidtea mediterranea*, showed that GnRH in all these species have a decapeptide with glycine at its C-terminus and followed by tribasic cleavage site. Overall, the GnRH identified in *B. glabrata* seems to exhibit similar organisation as in other molluscs with the cleavage sites and cysteine organisation being conserved (Figure 5.14). Mollusc GnRH differed from vertebrate GnRH in cysteine organisation with human GnRH_2 and *O. masou* GnRH having none and human GnRH_1 having one cysteine compared to two cysteines in other species in the study. Phylogenetic analysis shows that the vertebrate GnRH clustered separately from the invertebrate GnRH. *B. glabrata* GnRH is most closely related to *A. californica* with good branch support and is ancestral to other mollusc species.

\[
\text{MTNSWLSVLYVGLALLHMSQAKNYHESNGWYAGKR}GKR\text{SAPTYPSQSPIFDSETGSSSIDASCN}
\text{NISPESIFLINKLQEEVARILVVTSGPSSGLRALLESTDNS*}
\]
5.5.4.9 Egg-laying hormone

Egg laying hormones have been previously reported in *Aplysia* (McAllister *et al.*, 1983; Chiu *et al.*, 1979), *L. stagnalis* (e.g. Vreugdenhil *et al.*, 1988; Ebberink *et al.*, 1985) and *L. gigantea* (Veenstra, 2011) and a related hormone has also been identified in the leech *Theromyzon tessulatum* (Salzet *et al.*, 1997). This peptide induces ovulation, egg-mass formation and controls egg laying behavior (Li *et al.*, 1994; Nagle *et al.*, 1988; Vreugdenhil *et al.*, 1988). Two genes (Figure 5.15) coding ELH homologues were identified in *B. glabrata*. Both Bg-ELH precursors consist of a well-conserved region of 33 amino acids, near the carboxy terminus of the pro-hormone followed by glycine at the C-terminal of ELH causing amidation of ELH. This region is flanked on both sides by a
dibasic amino acid motif: Lys-Arg in Bg-ELH-2 but Bg-ELH-1 is flanked by monobasic lysine near the carboxy terminal (Figure 5.15). Additionally, the Bg-ELH-2 precursor consists of one more C-terminal peptide and has multiple spacer peptides flanked by mono-, di- and tribasic cleavage sites (Figure 5.15).

The number of cleavage sites (8) is conserved between Bg-ELH-1 and BgELH-2 found in A. californica and L. stagnalis, but there was a difference in the number of basic amino acids in endoproteolytic cleavage sites. The ELH domain (Pfam02323) identified in BgELH-2 was also found in Aplysia and Lymnaea but not found in Bg-ELH-1. No homologs for Bg-ELH-1 were found but the core basic organisation of the prepro-hormone is similar to other identified neuropeptides (Figure 5.15).

Egg-laying preprohormone 1
MLTTESRYKAFKFEKLFTGYTLGLYTSEQSVY(ESEFRARTRDLTDAVETPDEKLELTKQED
DTPGSAAPPLQISYVMPIGNDDLELLANIMGIKQOONRQEVMDMLLEEC

Egg-laying preprohormone 2
MLTYGGAFCPLSSONQFVGAMSVAASGDHVTVNADVKFEMNKIMASGGGEELMEIPQSP
EDETVAPNSVESELDDEFVDVESVESYVADKSRVQNNPKLRFKLVEYRLTSALRE
QANPPQYDDSEYQPEDSGAYERDLCPFPLFRHADYQASAYWGDAAADLDREDQFR
QISSSFARLYKVPIGNDELLALADLFLIERQKQAYSALKAMMDEAGCE

Figure 5.15: Characterization of B. glabrata ELH.
(A) Two PreproELH in B. glabrata showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Sparse peptide are indicated in grey; cysteine residues are made dark green; glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue. (B) Multiple sequence alignment of B. glabrata neuropeptide with different species where
alignments are shaded according to ClustalX colouring scheme, as detailed in Clamp et al. (2004). (C) Comparative cross species phylogenetic analyses using Maximum likelihood method with MEGAS5, under Jones-Taylor-Thorton (JTT) substitution model (Jones et al., 1992) with a gamma distribution of rates between sites (eight categories, parameter alpha, estimated by the program). (D) Schematics representation of neuropeptides precursors in different species showing signal peptide, cleavage sites, cysteine organisation, amidation sites and peptides bearing sequence identity to known bioactive peptides (conserved peptides). A. californica: AAA62380; A. californica_2: AAA27748; L. stagnalis: P06308; A. parvula: P17686; A. parvula_2: P17685; A. dactylomela: AC095731

5.5.4.10 Conopressin

Conopressin, a member of the vasopressin/oxytocin superfamily (also referred as neurohypophyseal hormones) was first identified in the venom of the fish hunting marine snail (Cruz et al., 1987). These have been previously identified in L. stagnalis (Van Kesteren et al., 1995b), Octopus vulgaris (Kanda et al., 2003) as well as Aplysia kurodai (BAB40371), A. californica (e.g. GD239415), C. gigas (AM854257 and AM853403), E. scolopes (DW266122), L. gigantea (Veenstra, 2010) and the leech Erpobdella octoculata (Salzet et al., 1993). Not only molluscs but also peptides of this family have been reported in annelids, insects and nematodes suggesting an ancient lineage of this family (Dutertre et al., 2008). In B. glabrata, one gene encoded a prepro-hormone of 157aa (Figure 5.16) and a propeptide consisting of a nine amino acid peptide (CFIRNCPKG), a neurophysin domain (Pfam00184) and a C-terminal copeptin-homologous extension of the neurophysin domain similar to that previously reported in mollusc conopressin and pro-hormones of the vertebrate vasopressin/oxytocin superfamily (Veenstra, 2010; Salzet et al., 1993). Comparative schematics show a well conserved basic structure and cysteine organisation between different mollusc species. All the identified species had 14 cysteines with similar location in the neuropeptide precursor. The cross-species phylogenetic analyses show that the Biomphalaria conopressin is closest to Lymnaea in terms of sequence similarity (Figure 5.16).

This is one of the well-characterized nonapeptide that seems to be involved in osmoregulation, neurotransmission and metabolism in molluscs and annelids (Gruber, 2014). In C. elegans this neuropeptide plays a role in reproduction. Similar function has also been reported in leeches molluscs and earthworms (Levoye et al., 2005). In L. stagnalis conopressin was isolated from penis complex and vas deferens suggesting a possible role of conopressin in controlling male copulatory behavior (Van Kesteren et al., 1995b).
Figure 5.16: Characterization of Bg-Conopressin.

(A) Bg-Conopressin showing the signal peptide highlighted in yellow; likely endoproteolytic cleavage sites are indicated in red; Sparse peptide are indicated in grey; cysteine residues are made dark green; glycine residues predicted to be converted to amide in green and likely bioactive biological peptides in light blue.

(B) Multiple sequence alignment of B. glabrata neuropeptide with different species where alignments are shaded according to ClustalX colouring scheme, as detailed in Clamp et al. (2004).

(C) Comparative cross species phylogenetic analyses using Maximum likelihood method with MEGA5, under Jones-Taylor-Thorton(JTT) substitution with a gamma distribution of rates between sites (eight categories, parameter alpha, estimated by the program).

(D) Schematics representation of neuropeptides precursors in different species showing signal peptide, cleavage sites, cysteine organisation, amidation sites and peptides bearing sequence identity to known bioactive peptides (conserved peptides). L. stagnalis: Q00945; A. kurodai: BAB40371; O. vulgaris: BAC82436; C. gigas: EKC41686; A. californica:NP_001191416; L. gigantea:ESO90380.

Overall, the structural organisation of B. glabrata reproductive type neuropeptides follows similar organisation to that found in other species. While homology searches provided a wealth of information when predicting neuropeptides known in other species, bioinformatics approaches and computational resources have proved most useful for identifying novel transcripts based on the structural organisation of these genes (presence of signal peptide; cysteine organisation; cleavage sites; post-translational modifications). Cross-species screening through multiple sequence alignment and phylogenetic tree construction revealed a low prepro-hormone sequence similarity but a high degree of conservation in the basic core structure of the neuropeptides in different species and this may represent evolutionary homology in these conserved peptides. Intron-exon boundaries were also analysed to determine if the conservation/variation was due to genomic structure but preliminary analyses did not reveal any
significant differences between the genomic organisation of each type of neuropeptide. One of the limitations of this whole process was that some of the neuropeptide precursors were split over multiple scaffolds and that could lead to error due to overlapping splice sites. A more complete genome would vastly improve the quality of the gene predictions, as identifying genes would be less error prone if they were lying on single scaffolds. Secondly, putative neuropeptides were predicted using *in silico* methods and then confirmed based on homology of the predicted peptides with the previously reported ones but it should be borne in mind that the proposed processing of the predicted neuropeptide are subject to the limitations of the splice-site predicting software.

### 5.6 Conclusion

Neuropeptides are a diverse class of chemical messengers instrumental in orchestrating complex physiological events which control the basic life processes of metabolism to growth to reproduction and immunity. This initiative has enhanced our knowledge of the neuropeptides present in the snail *B. glabrata* and the possible underlying neuro-endocrine processes that might be regulating reproductive activity. The overall analysis demonstrates conserved molluscan neuropeptide domains and precursor organisation as well as elucidating many previously unrecognized gastropod idiosyncrasies. This data of identified and conserved neuropeptides makes a significant contribution to understanding the evolution of hermaphroditic mode of reproduction in molluscs. This work is a useful addition to previous research as it has the potential to greatly aid the study of molluscan basic reproductive biology. The availability of the complete *B. glabrata* genome opens avenues for further exploring the biochemical pathways through which these neuropeptides exert their function. In this way *B. glabrata* could provide a simple model system to initially investigate the biodynamic interplay between neuropeptides and the nuclear receptors, leading eventually to identification of the mechanisms that might be involved in their interaction with the environment.
6 Chapter 6: General Discussion
6.1 Overview of the thesis

This thesis investigates the similarities between the molluscs and the mammals in their physiological responses to androgenic chemicals and examines if common genes and/or proteins in both molluscs and humans underlie similar mechanisms for male reproductive development. To answer these scientific questions an integrated strategy was developed involving two approaches. 1. *In vivo* exposure of the snails to vertebrate androgens. 2. *In silico* investigation in the snail genome for the presence of an androgen receptor as well as other genes present in the vertebrate androgen signalling pathway. The degree of conservation in snail responses and signalling molecules between snails and humans was expected to give insight into the conservation of functionality in the experimental model chosen.

The overall results showed that the biochemical pathways controlling androgen-dependent reproduction in this mollusc species are not similar to vertebrate androgen signalling pathways. A developmental exposure study using two androgenic chemicals did not reveal any comprehensive pattern of response in either snail growth or reproductive organ development. No significant changes in the weight of gonads (ovotestis) or sex accessory glands (albumen gland and glandular complex) as compared to controls/solvent controls were found in either of the chemical treatments. Therefore, these results were not similar to Hershberger’s prepubertal rodent assay where the accessory sex organs weight increase and the weight of testes decreases at the termination of the prepubertal exposure of rodents to androgenic chemicals (Moon *et al.*, 2009) (Chapter 3). *In silico* investigation into two snail genomes (*B. glabrata* and *L. gigantea*) for the presence/absence of androgen receptor identified a repertoire of 40 nuclear receptors in *B. glabrata* and 33 in *L. gigantea*, representing all major subfamilies of the nuclear receptor superfamily also found in vertebrates (Chapter 4). Of the mammalian sex steroid receptor homologs, the whole clade of sex steroid receptor subfamily (NR3C) was missing, although estrogen receptor-like and estrogen related receptor-like homologs were identified. Cross-species phylogenetic analyses with model organisms (fly, worm and human) showed that the test species shared NR homologs with both invertebrates and vertebrates. On an evolutionary scale, molluscs lie somewhere between the insects and vertebrates, exemplified by the presence of both a vertebrate-like THR and an insect-like EcR in molluscs. THR is reported to control metamorphosis and prenatal development of central nervous system in vertebrates (Paris *et al.*, 2010; Bernal, 2007b; Brown, 2005) and EcR plays an important role in *D. melanogaster* metamorphosis/moultning (Li and Bender, 2000). Both these receptors are
functionally similar (analogous) and are present in molluscs. It would be interesting to look whether either of these molecules control development in snails.

Neuropeptide genes controlling reproduction were also identified in our test species and exhaustive “data mining” came up with 17 reproductive type neuropeptide precursors and 4 insulin-like peptides that had 124 putative bioactive peptides in *B. glabrata*. The snail appears to have the molluscan homologues of vertebrate reproductive type neuropeptides (Chapter 5) revealing a similar structural organisation to that found in other species. Moreover, some of the reproductive type neuropeptides like GnRH, Conopressin, NPY and Insulin-like peptides show sequence similarities to their vertebrate counterparts. These findings set forth an avenue of research for the potential use of mollusc in understanding the neuro-endocrine evolution of vertebrate neuropeptides.

Taken together, this knowledge is significant given the intricacies of molluscan endocrine systems due to complexity of reproductive systems and life histories as it gives an insight into the evolution of endocrine control systems and their similarities/differences with the vertebrate endocrine mechanisms. Looking back at the available literature there are still many unanswered questions as previous studies have shown effects of synthetic androgens on molluscs and indicated probable or possible conservation in androgen signalling mechanisms. The following questions are foremost:

- Is the lack of chemical effect seen in *B. glabrata* my studies a species-specific response (pulmonates/prosobranch snails)?
- Is it to do with the dose of exposure? Was the dose used lower than that required to effect a response?
- What is controlling spermatogenesis and reproductive behaviour in *B. glabrata*?

Taking the first of these questions, differences in snail responses to chemicals were reported in the past. Reproductive output was inhibited in *P. antipodarum* at 160ng/l of MT (Duft *et al.*, 2007) whereas *L. stagnalis* when exposed to the same chemical at 100ng/l for 8 weeks showed weak aberration in the albumen gland and no dose dependent effect was seen on the ovotestis (Czech, Weber and Dietrich, 2001). In *M. Cornuarietis*, MT induced imposex and displayed inhibition of spermatogenesis in males at 100ng/l (Albanis *et al.*, 2006; Schulte-Oehlmann *et al.*, 2004). Interestingly it appears that 3 gastropod snails with different reproductive strategies (parthenogenic, hermaphrodite, gonochoric) exhibited different sensitivities towards chemicals (in this example androgen). Similarly, prosobranchs like *N. lapillus* and *Nassarius obsoletus* were affected at very low concentrations of TBT (3 ng/l) (Peña *et al.*, 1988) compared to the pulmonate *L.*
which was only affected at relatively high (230-3200 ng/L) concentrations, causing vacuolization and enlargement of the prostate gland in juvenile snails (Segner et al., 2003). In another study a LOEC of 1 μg/L for embryonic development in L. stagnalis was found during TBT exposure (Leung et al., 2007), whereas adult P. antipodarum was shown to have a LOEC of 50ng/L for 8-week exposure to TBT (Duft et al., 2007). These results suggest that prosobranchs might be more sensitive to androgenic chemicals than pulmonates.

Not only androgens but also estrogenic chemicals, such as ethinylestradiol (EE2) have been shown to have a greater effect at lower concentrations in prosobranchs (P. antipodarum) than in pulmonates. Embryo production was enhanced at 1ng/l in P. antipodarum (Jobling et al., 2003) while similar effects were seen in L. stagnalis at 500 ng/L (Segner et al., 2003; Anonymous). A one-day exposure of adult P. antipodarum to 100ng/L EE2 caused a significant increase in ER-mRNA expression (Stange et al., 2012) as compared to controls while no significant change was reported in adult Radix balthica exposed to similar concentration (Börjesson, 2012). Two species R. balthica and Bithynia tentaculata when exposed to identical concentrations of EE2 responded differently with B. tentaculata (a prosobranch) showing a significant inhibition of somatic growth and R. balthica (a pulmonate) showing enhanced somatic growth (Hallgren et al., 2012). It seems likely that prosobranchs are more sensitive to EDCs as compared to pulmonates.

Differences in the results of chemical exposures have been also reported between studies of the same species of snail. In contrast to my study, B. glabrata snails have previously been reported to be affected by semi-static MT exposure (concentration ranging between 0.01mg/l to 1.0mg/l) in a 4week exposure (Rivero-Wendt et al., 2013). However in this study the statistical comparisons were made between MT dosed and water control snails, rather than between MT dosed and solvent control (0.001% ethanol) exposed snails. The solvent control snails showed that largest average area occupied by sperm. Indeed the authors themselves suggested that the stimulatory effects on sperm production were due to the solvent (ethanol) used, which suggests that MT might not be the predominant factor (Rivero-Wendt et al., 2013). Moreover, the effective exposure concentration reported was much higher than what the snails in our laboratory were exposed to. It is possible that the effects seen in pulmonates at higher reported concentrations than in my experiments were due to an inflammatory effect (Czech, Weber and Dietrich, 2001) rather than due to endocrine disruption; endocrine disruption is generally considered as one of the modes of toxicological action that has potential to impact organisms at very low concentrations in the range of ng/l or lower μg/l concentrations (Oehlmann et al., 2007) depending on their
sensitivities and presence/absence of endogenous hormones. In another study using *L. stagnalis*, Segner et al. (2003) showed that adult survival was significantly reduced by TBT exposure, with an LC$_{50}$ at 21 days of 290 ng Sn/L (nominal concentration), which is in contrast to a 21-d exposure of juvenile *L. stagnalis* to TBT (in the range of 19ng Sn/L to 473ng/L), which did not result in significant mortality (Giusti *et al.*, 2013; Segner *et al.*, 2003). No significant mortality was observed after 49 days and 84 days of exposure of adult *L. stagnalis* to 100 ng Sn/L (Czech, Weber and Dietrich, 2001). Female *N. lapillus* showed a significant effect on penis growth in response to testosterone (Spooner *et al.*, 1991). Conversely, a more recent study found that testosterone had no effect at all on penis growth of *N. lapillus* (Castro *et al.*, 2007). One of the possibilities for the contradictory results obtained in the above studies devoted to the same species and the same compound could be that only a single dose (1µg/g body weight) was tested in the latter study even though it was within the range of doses used in the previous study (0.01-0.1,10µg/g body weight). Single dose effects can be interpreted in multiple ways, as it is difficult to explain random variation from ‘real’ biological effects. Moreover, statistical data analysis is more prone to type I errors or ‘false discovery’ (Fitts, 2011) as all animals in the same tank are not replicates to each other; infection or disease in one animal could skew the results and the absence of any supporting evidence to characterize these uncertainties in the data affects the reliability of the scientific interpretation. A review comparing sensitivity of three prosobranch snails with different habitats (fresh-, brackish- and marine waters) and different reproductive strategies (parthenogenesis, hermaphrodism and gonochorism) to TBT, reported their sensitivities to be species and sex-dependent (Ketata *et al.*, 2008). Freshwater gastropods (and freshwater species in general) are less sensitive than their marine relatives due to differences in chemical kinetics of TBT in freshwater as compared to marine water (Leung *et al.*, 2007; Leung *et al.*, 2004).

Given their evolutionary divergence in gastropods, interspecies variation within the phylum mollusca is perhaps unsurprising. Morphologically the phylum includes species as different as meiofaunal “worms”, now classed as molluscs, and the giant squid (Kocot, 2013). The extreme disparity in morphology among the major lineages within molluscs has led to several conflicting phylogenetic hypotheses (Haszprunar and Wanninger, 2012; Ponder and Lindberg, 2008). Even within gastropods, according to Smith et al. (2011), the different phylogenetic analyses studies completed to date have not been able to cover the diversity of 100,000 living species of this group (Smith *et al.*, 2011).

There are some species of molluscs in which chemicals seem to be having an effect not
only at the individual level, for example, TBT exposure causing imposex affecting molluscs not only at the individual level but also at the population level (Janer et al., 2006a; Oehlmann and Schulte-Oehlmann, 2003; Matthiessen and Gibbs, 1998) and so one might consider all the possible mechanisms/endpoints in which endocrine disrupting chemicals can affect the body’s signalling system (Matthiessen, 2008) in these classes of molluscs. Several hypotheses have been tested to investigate the mechanism through which TBT might affect the reproductive physiology in molluscs either through androgen receptor-mediated pathways (Oehlmann et al., 2007; Matthiessen and Gibbs, 1998a) or through retinoids (RXR) (Ciocan et al., 2012; Horiguchi et al., 2007; Nishikawa et al., 2004). Recently Pascoal et al. (2012) suggested the involvement of both retinoids and androgen signalling in inducing imposex in *N. lapillus* (Pascoal et al., 2012). There have been several studies trying to differentiate the mode of action of sex steroids in molluscs (Bannister et al., 2013; Stange et al., 2012; Benstead et al., 2011b; Bannister et al., 2007) but the absence of the whole clade of sex steroid receptors (androgen receptor, progesterone receptor, mineralocorticoids and glucocorticoids) identified in my studies in *B. glabrata* and *L. gigantea* reduces the possibility of finding an androgen receptor in gastropods and possibly in this phylum. If the AR is not involved in androgen signaling in molluscs, the alternative hypothesis, that retinoids are involved, may be valid. Retinoids have been suggested to be involved in inducing imposex in molluscs due to the ability of TBT to form dual ligand heterodimers RXRα:PPARγ (le Maire et al., 2009). The presence of these nuclear receptors in molluscs increases the chances of the retinoid signalling pathway being an active mechanism controlling reproductive development. It has also been shown that nanomolar concentrations of TBT can activate the RXRα:PPARγ heterodimer promoting adipocyte differentiation in amphibian, mouse and humans (le Maire et al., 2009; Iguchi et al., 2007; Grun and Blumberg, 2006; Kanayama et al., 2005) and cause deregulation of the aromatase gene expression (Kotake, 2012; Nakanishi and Nishikawa, 2009; Nakanishi, Nishikawa and Tanaka, 2006). TBT is one of the examples of endocrine disrupting chemicals causing adverse effects in both gastropods and mammals (Iguchi et al., 2008; Iguchi et al., 2007).

Alternatively pathways using membrane bound receptors could be an active mechanism in molluscs as these hormones/synthetic hormones can also trigger biological effects by activating and coordinating a number of transcription factors like cyclic AMP (cAMP) responsive element binding protein (CREB) via MAPK (Mitogen Activated protein Kinase) signaling pathway (Walker, 2010b; Porte et al., 2006; Canesi et al., 2004). Biological and genetic evidence points out that cAMP-dependent pathway is an essential
factor controlling myriad of biological systems like gonadal functions (Khotimchenko YS, 1989), development and homeostasis (Scheide and Dietz, 1986) and metabolism (Fabbri and Capuzzo, 2010). Scarce information on cAMP-dependent signaling was available for invertebrates when it was first reported in *M. edulis* (Stefano *et al.*, 1989) and later in freshwater snail *L. stagnalis* and the marine mollusc *A. californica* (Bailey CH, 2008). Various synthetic estrogenic chemicals (17β-estradiol, DES) were reported to affect haemocyte lysosomal function in *Mytilus* spp by rapid activation of p38 MAPK and STATs (Canesi *et al.*, 2006).

Understanding these pathways and identifying their components genes in molluscs through computational bioinformatics might bring forth interesting evolutionary aspects, guiding further investigation.

6.2 Agreement and disagreement with the previous mollusc exposure studies

The first report of the existence of vertebrate-like steroids in molluscs appeared in the 1950s (Rohlack, 1959; Hagerman, Wellington and Villee, 1957). Since then there have been many accounts in different species and different classes of molluscs showing the presence of these hormones in pg/g wet body weight, tissue distribution, and variation by sex and season. This is a short discussion of scientific studies conducted on different species of snails for endocrine disruptors looking at the experimental details and the interpretation of results. There is no disagreement with the fact that chemicals do have effects on some snail species (reviewed in (Pascoal *et al.*, 2012; Matthiessen, 2008). A closer examination of the experimental details, however, reveals that a few of these studies, looking into the effect of vertebrate-like steroids on growth, fecundity and development of secondary sexual characteristics in molluscs, have not performed any statistical analyses, eg (Sakr, Osman and Abo-Shafey, 1992) (*Theba pisana*); (Bettin, Oehlmann and Stroben, 1996) (*N. lapillus*); (Mori, Muramatsu and Nakamura, 1969) (*C. gigas*); (Tosti *et al.*, 2001) (*O. vulgaris*); (Takeda, 1979) (*Deroceras reticulatum*); (Moss, 1989) (*Mulinia lateralis*)) and so these cannot be taken into account when evaluating the weight of evidence. In some molluscan exposure studies, only one replicate has been used (Oehlmann *et al.*, 2006; Gagne and Blaise, 2005; Janer *et al.*, 2004; Osada *et al.*, 2003; Gagné *et al.*, 2002; Schulte-Oehlmann *et al.*, 2001; Czech, Weber and Dietrich, 2001; Gagné *et al.*, 2001). Such results are open to more than one interpretation because of the absence of any variability arising from examining more than one replicate. Moreover, it is
impossible to determine whether any difference between control and treatment represents a true effect of the treatment or of uncontrolled variables (Scott, 2012b). Any results observed could be due to the treatment, or conversely, could only be specific to the tank that holds the animals. If all animals in any one treatment are held together in the same container, it is impossible to determine intra- and inter-snail differences that might be contributing towards test endpoint variability (Forbes et al., 2007b; Forbes et al., 2007a). Tank effect could also be due to the position of the tank in the exposure room or could be due to an individual in the tank carrying some sort of infection. Replication of in vivo animal studies is a vital part of experimental design and should be done keeping in mind endpoint variability (Forbes et al., 2007b; Forbes et al., 2007a).

Molluscs are one of the most specious phyla and different scientific groups have been involved with different species. There are only a handful of papers that are based on studies carried out on any one species, and so there is no standardized protocol for mollusc chemical testing. Experiments are generally designed keeping in mind the life cycle, habitat and behaviour of the test species, even though efforts are being made to develop a standard protocol for mollusc life-cycle tests and partial life cycle tests (Matthiessen, 2008). Another major source of variability for all measured endpoints in in vivo mollusc studies is therefore due to variability within and among the test species in their response to the chemical. This complicates interpretation of the impacts of chemicals on molluscs as defined by the current literature. One of the many reasons for differences in the response seen could be difference in competition among animals in a group for food and space between different studies (Forbes et al., 2007b; Forbes et al., 2007a). For example, for TBT, there have been several similar studies (Stange et al., 2012; Pascoal et al., 2012; Lima et al., 2011; Castro et al., 2007; Santos et al., 2006; Bettin, Oehlmann and Stroben, 1996) carried out on one species (N. lapillus). There was no replication of the TBT treatment in many of these published experiments with this species (Stange et al., 2012; Castro et al., 2007; Bettin, Oehlmann and Stroben, 1996) and the snails were maintained at different densities in different experiments. Studies have shown that density affects growth and development of snails (Yoshida et al., 2013; Zachar and Neiman, 2013; Mangal TD, Paterson S, Fenton A, 2010). Other species of snails (Ocenebra erinacea, Ostrea edulis, Urosalpinx cinerea and Ocinebrina aciculate) have been tested with TBT. Female masculinization varied according to species and is a graded mechanism depended on the dose of exposure (Matthiessen, 2008).
6.3 Significance of the results

The results presented in this thesis have been obtained with careful scientific rigour, in order to overcome some of the problems associated with those studies mentioned above. This study showed clearly that reproductive physiology of *B. glabrata* is unaffected by androgenic chemicals at low doses. Mechanistically, these chemicals cannot act through androgen signalling pathways as claimed by many scientific studies previously reported, even though in different species (Pascoal *et al.*, 2012). No androgen receptor has been found in the fully sequenced genome of *B. glabrata, L. gigantea* and *C. gigas* (Vogeler *et al.*, 2014b). Taken together, these studies suggest that snails could be inappropriate models for mammalian steroid hormone function mediated via nuclear receptors. The whole clade of the NR3C subfamily is missing from both gastropod species investigated here and had also been shown to be absent in bivalve *C. gigas* (Vogeler *et al.*, 2014b) and it seems likely that this clade is vertebrate specific. This investigation has, however, raised the possibility of molecular conservation of some of the other metabolic and developmental pathways between snails and humans. Of the NRS identified in *B. glabrata*, 15

![Figure 6.1: Venn diagram showing the nuclear receptor homologues in B. glabrata, D. melanogaster and H. sapiens.](image)

*B. glabrata* has 40NRs of which 15 were homologues to human NRs and 19 were homologues to *Drosophila* NRs and 5 NRs were identified in all the three species.

NRs are found in both *B. glabrata* and humans while 5 NRs are found in all the three, *Drosophila, B. glabrata* and humans Figure 6.1. The mammalian homologs include retinoid receptors (RAR, ROR, RXR and PPAR); steroid hormone receptors like the
estrogen receptor and the estrogen related receptor; the thyroid hormone receptor; the testicular receptor (TR4) and DAX1.

Traditionally, only a few invertebrate animal models, primarily worms and flies, have been relied upon for investigation of many NR-controlled functions but this study reveals the snail’s potential as possibly a new model organism for research and treatment of some human diseases and disorders. The identification of a valid invertebrate model might, at least partially, replace rodent use reducing the use of animals in scientific research and chemical testing.

The absence of a molluscan AR and the constitutive expression of the ER suggest that alternative pathways must exist for spermatogenesis/oogenesis in molluscs and other nuclear receptors have been proposed as initiating these pathways for reproductive processes, some of which exist in vertebrates and which may be particularly important in invertebrates. Testicular receptors identified in both the mollusc species have been shown to control spermatogenesis in male mice (Mu et al., 2004) and folliculogenesis (Chen et al., 2008) in female mice. TR knockout mice show degenerated primary spermatocytes and some necrotic tubules (Liu, 2010). Moreover, the significance of RXR signalling in conjunction with PPAR signalling on reproductive physiology in molluscs has also been demonstrated. Although vertebrates may have subsequently evolved more complex processes involving steroid hormones, the presence of orthologous NRs involved in other pathways in invertebrates offers the opportunity to study these underlying/fundamental conserved networks in a simplified system. Another notable finding was the thyroid hormone receptor that plays key roles in growth development and metabolism in vertebrates and makes B. glabrata a potential model for thyroid hormone processes. These findings are significant for addressing fundamental biological questions as to the evolution and origin of nuclear receptors as these are the critical components of the hormonal system.

Not only nuclear receptors but also neuropeptides are the main players participating in the snail response to some chemicals. Recently a study reported proteomic analysis of the reproductive organs of L. stagnalis exposed to 19.6μg/l of chlordecone (weak binder to the estrogen receptor in vertebrates) causing an increased expression of ovopostatin (inhibits oviposition in L. stagnalis) and PIWI gene in reproductive organs (Giusti et al., 2014). PIWI is a key protein in germline stem cell differentiation (Cox, Chao and Lin, 2000) and has been associated with spermatogenesis in vertebrates as well as in invertebrates like hermaphrodites C. elegans (Juliano, Wang and Lin, 2011). This work has been culminated in collaboration with Dr Scott Cummins, classifying the secretory peptides employed by
B. glabrata for higher-order neural control of its functioning. This would be an additional knowledge to what is known about neuropeptides in molluscs facilitating the study of invertebrate neuropeptides with the hope of stimulating advances in model organism and parasite-host research.

6.4 Implications for unanswered questions in mollusc endocrinology

This thesis highlights the fact that molluscan endocrinology is not well understood. Even though we have moved a step forward in identifying the components of the neuro-endocrine system in gastropods there are still unanswered questions that need attention. Additional research will be necessary to understand the fundamental mechanisms and pathways that determine key developmental and physiological processes. In turn, this would help us to understand the morphological endpoints and improve tractability of the results in these studies. The fact that molluscs are the second most specious phyla inhabiting different habitats and have undergone different morphological adaptations is the most challenging aspect of research in this field. The evidence that endocrine disrupting chemicals affect molluscs is not very well understood but the phenotypic effects of some of these chemicals is still significant. The “take home message” is that we need more rigorous studies and there is a need for a widely agreed testing protocol for evaluating the strength of evidence associating exposure to chemicals and adverse effects seen in molluscs.

At the time when I started my PhD, only one mollusc genome was available (L. gigantea). Since then, however, there has been considerable progress in next generation sequencing, making sequencing of genomes cheaper and quicker than before. Concomitantly, there has been an increase in the number of snails with a fully sequenced genome and a comparative genomic analysis between snails with differing genetic diversities and this would deepen our knowledge related to their fundamental biological differences and similarities. This would facilitate identification of model systems in invertebrates, both to reveal the mechanistic causes of reproductive diseases and to develop test systems. The elucidation of the B. glabrata genome creates an invaluable resource to study the functional role of specific genes in these organisms. The conserved nature of the nuclear receptors across the metazoans helps to identify them, but the results should be treated with caution as molecular studies indicate that species-specific differences should be expected in ligand-receptor interaction (Daston et al., 2003). Not only receptor-mediated mechanisms but other mechanisms such as hormone synthesis, transport and metabolism, activation of nuclear receptors and gene methylation are
also equally important in endocrine disruption (Tabb and Blumberg, 2006).

6.5 Further Work

There exist a multitude of potential future avenues of investigation based on the work that I have described here. Nuclear receptors integrate between the central and peripheral nervous systems and communicate not only through established endocrine-hormone signalling but also various other physiological regulatory networks. This comprehensive analysis of the NR gene complement in this non-model species demonstrates the huge diversity of nuclear receptors found in these Lophotrochozoans, and indicates the scope of possible NR-mediated negative impacts of chemicals. Most importantly, the retinoid receptors, PPAR, ER, ERR, Thyroid receptor, COUP-TFII and Testicular orphan receptors identified in the snails represent the so called “ring of physiology” described in the mouse on the basis of gene expression of diverse anatomical systems (http://www.nursa.org/files/NURSA_Phase%20I_II_Final_Report.pdf). Studying the expression profile of NRs in snail tissues at different life stages would offer a simple but powerful way to obtain highly relational information about their ability to coordinate the transcriptional programs and their link with physiological pathways like reproduction, development and metabolism as an individual NR or as a superfamily as done in some recent rodent studies (Xie et al., 2009; Bookout et al., 2006; Yang et al., 2006). Statistical analyses of the hierarchical clustering of NR expression profiling correlating receptors based on their tissue distribution and biological functions, will bring together NR pairs with highest correlation. In particular, this approach could be used to identify hormone metabolic genes, co-activators, corepressors and other unknown factors affecting hormone activity from the signal transduction pathways. Furthermore, this would elucidate common sets of transcription regulators that govern the expression of specific NRs and the physiological processes directed by those networks. There have been some studies using this approach in different vertebrate species (Martinović-Weigelt et al., 2011; Holbeck et al., 2010; Szakács et al., 2006; Fu et al., 2005). This is a large-scale hypothesis free approach so the knowledge of the exact mechanism is not essential. The incorporation of a further expression profiling approach to study chemical-induced changes in these molecules would lead to characterization of basic biological networks that collectively define an organ’s response in an organism. Depending on the degree of conservation of the molecules in the network and on the conservation of their pathways, in snails and humans this could lead to the development of new model organisms for research into both host-parasite interaction and cross-species
toxicogenomics studies to understand off-target and mechanism based toxicities for chemical/drug testing purposes (Ulrich, 2003).

Given *B. glabrata*’s significance as an intermediate host to *S. mansoni*, and the discovery of homologous receptors in host (*B. glabrata*) and parasite (*S. mansoni*), this thesis adds more avenues to further characterize host-parasite interaction and for possible treatments to break the parasite life-cycle. Recent experimental evidence suggests that parasites can not only evade immune responses actively, but also exploit the hormonal microenvironment within the host to favour their growth, development and reproduction (Escobedo et al., 2005). Extensive analysis of the *S. mansoni* genome for the identification of potential novel drug targets emerged with a family of 21 nuclear receptors that included THR, RXR, COUP TFII and FTZ-F1, that could bind to the host hormones to direct downstream transcription and aid diverse developmental processes (Wu et al., 2007; de Mendonça et al., 2000). Immunization of BALB/c mice (albino, laboratory bred strain of house mice) with recombinant-THR-LBD extract from *S. japonicum* induced partial protection against schistosoma infection (27.52% of worm reduction and 29.5% of reduction in eggs produced in the liver by the parasite) and induced a higher level of T helper cell response in the mice (Qiu et al., 2012). Studies have shown that Thyroid hormone, Insulin and Interlukin 7, all contribute to a host environment favourable to schistosome development in the mouse (de Mendonça et al., 2000). One recent study in the fox tapeworm, *Echinococcus multilocularis* identified 17 NRs which broadly overlapped with the set of NRs expressed in *S. mansoni*. One of the unique receptors identified in the cestode belonged to NR1J subfamily of NRs that demonstrated a role in hormonal host-parasite cross-communication between NR signalling and TGF-β signalling during infection in the intermediate host ( Förster et al., 2011). DAF-12 is an evolutionarily conserved nuclear hormone receptor in several parasitic nematodes including *Strongyloides stercoralis*, *Ancylostoma spp.* and *Necator americanus*, involved in signaling pathway which governs development of the stage 3 infective larvae (iL3) (Wang et al., 2009). Administration of dafachronic acid (a steroid like molecule and a ligand to DAF-12) has been shown to markedly reduce the pathogenic iL3 population in *S. stercoralis*, indicating the potential use of DAF-12 ligands to treat disseminated strongyloidiasis (Rana and Misra-Bhattacharya, 2013; Wang et al., 2009). The *B. glabrata* genome data as well as the parasite data can be further exploited to identify differences between host-parasite metabolic pathways (believed to be regulated by some vertebrate orphan receptors) that make them dependent on the host. Reaction points in those metabolic pathways that produce or consume metabolites (chokepoints) could be
potential drug targets (Webster et al., 2010).

Moreover, aside from the genome, the different genomic tools available for *B. glabrata*, such as RNAi (Jiang, Loker and Zhang, 2006) and the traditional approaches of enhancer and suppressor genetics and transgenics make *B. glabrata* a candidate to demonstrate gene functions that are not based only on homology.

### 6.6 Conclusion

This thesis has shown that androgen-signalling pathways in the pulmonate gastropod molluscs, *B. glabrata*, and *L. gigantea* are not similar to vertebrates. *In vivo* exposures to androgens showed no effect on the snail reproductive physiology and the absence of an androgen receptor confirms the non-existence of the androgen signalling pathway in *L. gigantea* and *B. glabrata*. As a result, there is limited commonality with vertebrates with regards to the impacts of androgenic chemicals causing reproductive disorders in vertebrates, although there remains the possibility that other mollusc species might contain an AR considering the number of species this phylum holds. Wildlife is considered sentinel for human health, but care should be taken when extrapolating the impacts of chemicals between different species of organisms. There are a lot of variables that should be considered such as chemical kinetic, chemical dynamics, ADME (adsorption, distribution, metabolism and excretion) of the chemical in different species and genetic machinery to name a few. The elucidation of the *B. glabrata* genome and knowledge of nuclear receptors creates an invaluable resource to study the functional role of these transcription factors stimulating advances in cause-effect ‘interactome’ research.
Appendix 3.1

Recipe for minimal medium and other medium components for yeast androgenic screen obtained from Brunel University Laboratory.

Ingredients required for yeast screen and the method of their preparation:

Minimal Medium (pH 7.1)
Add 13.61 g KH$_2$PO$_4$, 1.98 g (NH$_4$)SO$_4$, 4.2 g KOH pellets, 0.2 g MgSO$_4$, 1 ml Fe$_2$(SO$_4$)$_3$ solution, 50 mg L-leucine, 50 mg L-histidine, 50 mg Adenine, 20 mg L-arginine-HCl, 20 mg L-methionine, 30 mg L-tyrosine, 30 mg L-isoleucine, 30 mg L-lysine-HCl, 25 mg L-phenylalanine, 100 mg L-glutamic acid, 150 mg L-valine, and 375 mg L-serine to 1 L double-distilled water. All these ingredients are placed on heated stirrer to dissolve and then it is dispensed in 45 ml aliquots into glass bottles. The glass bottles with the contents are sterilized at 121°C for 10 min, and store at room temperature.

D-(-)-Glucose
Prepare a 20% w/v solution (20 parts by weight of powder to 1000 parts by volume of liquid). Then sterilize in 20 ml aliquots at 121°C for 10 min and stored at room temperature.

L-Aspartic Acid
Make a stock solution of 4 mg/ml; sterilize in 20 ml aliquots at 121°C for 10 min and store at room temperature.

Vitamin Solution
Add 8 mg thiamine, 8 mg pyridoxine, 8 mg pantothenic acid, 40 mg inositol, and 20 ml biotin solution (2 mg/100 ml H$_2$O) to 180 ml double-distilled water. This is then filtered through a 0.2-μm pore size disposable filter, in a laminar air flow cabinet into sterile glass bottles in 10 ml aliquots and stored at 4°C.

L-Threonine
Prepare a solution of 24 mg/ml; sterilize in 10 ml aliquots at 121°C for 10 min and stored at 4°C.

Copper (II) Sulfate
Prepare a 20mM solution. This is then filtered through a 0.2-μm-pore size filter, in a laminar flow cabinet into sterile glass bottles in 5 ml aliquots and stored at room temperature.
**Chlorophenol red-β-D-galactopyranoside (CPRG)**

Make a 10mg/ml stock solution filtered through a 0.2-μm-pore size filter into sterile glass bottles, in a laminar flow cabinet and stored at room temperature.

**Method of preparation:**

We start by preparing the growth medium by adding 5ml glucose solution, 1.25ml L-aspartic acid solution, 0.5ml vitamin solution, 0.4ml L-threonine solution, and 125μl copper (II) sulfate solution to 45ml minimal medium. The mixture is then transferred to a sterile conical flask bringing the final volume to 50ml. To this mixture 125μl of 10X concentrated yeast stock from cryogenic vial stored at -20°C is added. The mixture is incubated for approximately 24hours at 28°C on an orbital shaker until it turns turbid.
Appendix 4.1
Alignment of all the sequences from human, fly, worm and owl limpet, used in the
investigation of nuclear receptors in B. glabrata.
#NEXUS
[Written Thu Feb 14 10:07:51 GMT 2013 by Mesquite
(build 566) at X-UXIFE125/134.83.70.53]

version 2.75

BEGIN TAXA;
TITLE Taxa;
DIMENSIONS NTAX=146;
TAXLABELS
HsAR
HsCOUP-TFa
HsCOUP-TFb
HsEAR2
CeDAF_12
Dm_E78c
Dm_E75B Dm_EGON HsERa HsERb HsERRa HsERRb HsERRg HsGR HsMR HsPR
HsGCNF HsHNF4 HsHNF4g DmFTZ_F1 HsLXRa HsLXRb HsFXR HSNGFIB HsNURR1
HsNOR1 HsPNR HsDAX1 HsSHP HsPPARa HsPPARb HsPPARg HSPXR HsCAR HSRARa
HsRARb HsRARg HsREV_ERBa HsRev_ERBb HsRORa HsRORg HsRORb HsRXRa
HsRXRb HsRXRg HsSF1 HsLRH1 HsTHRa HsTHRb HsTLX HsTR2 HsTR4 HsVDR
LgHNF4 LgNR1J1 Dm_SVP LgERR LgFTZ_f1 LgHR3 LgE78c LgDAX1 LgRev_erb
LgHR4a LgHR4b LgTR LgEcR LgFAX1a LgHR38 LgPPAR2 LgRXR LgNR2E1 LgTHR
LgNR1J3 LgCOUP_TFA LgHR39 LgRAR LgROR LgER LgTLX LgFAX1b LgPNR LgDSF
LgNR4a LgE75 LgPPAR1 LgNR2E2 LgNR1J2 BgCOUP_TFa BgNR4A BgERR BgHR39
BgDSF BgFAX1 BgDAX1 BgROR BgHR3 BgER BgRAR BgHNF4 BgNRU1 BgE75
BgRev_erb BgNR1J2 BgNR1J4 BgEcR BgHR4 BgFTZ_F1 BgPPAR2 BgPPAR1 BgNRU2
BgE78c BgTR BgNR1J3 BgRXR BgNRU3 BgHR38 Dmknirps Dmknirps_li Dm_ERR
Dm_HR3 Dm_USP Dm_HR78 Dm_TLL Dm_PNR Dm_HR38 Dm_HR39 Dm_HR4 Ce_ODR
Ce_HR23 Ce_NR14 Ce_HR8 Ce_HR1 Ce_HR49 Ce_UNC55 Ce_HR25 Ce_HR91
Dm_HNF4 Dm_HR96 Dm_EcR Dm_DSF Dm_FAX1 Ce_HR48 Ce_HR41 Ce_FAX1 Ce_HR6
Ce_NHR85
;
END;
BEGIN CHARACTERS;
TITLE Character Matrix;
DIMENSIONS NCHAR=308;
FORMAT DATATYPE = Protein GAP = - MISSING = ?;
MATRIX
HsAR
QKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQ--KYL-CAS-RNDCTIDKFRRKNCPSCRLRKCYEAGMTLGELGERQLVHVVKWAKALPGFRNLHVDDQMAVIQYSWMGL
MVF-AMGWRSFTNVN-----------------------SRMLYFAPDLV-----------FNEY-R---MHK-----SRM-YSQCVRMRHLSQ-EFG-WLQITPQEFLCMKALLLFSI-----IPV----------DGLKNQKFFDELRMNYIKELDRIIACKRK-NPTSCS-RRFYQLTKLLDSVQPIA
HsCOUP_TFa
HIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNL--TYT-CRA-NRNCPIDQHHRNQCQYCRLKKCLKVGMRREELAARLLFSAVEWARNIPFFPDLQITDQVSLLRLTWSEL
FVL-NAAQCSMPLHVA-PLLAAAG---------------------LH-A---------SPMSAD-----RVV-------AFMDHIRIFQEQVE-KLK-ALHVDSAEYSCLKAIVLFTS----------------DACGLSDAAHIESLQEKSQCALEEYVRSQYP----NQP-SRFGKLLLRLPSLRTVS
HsCOUP_TFb
HIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNL--SYT-CRA-NRNCPIDQHHRNQCQYCRLKKCLKVGMRREELAARMLFSAVEWARNIPFFPDLQITDQVALLRLTWSEL
FVL-NAAQCSMPLHVA-PLLAAAG---------------------LH-A---------SPMSAD-----RVV-------AFMDHIRIFQEQVE-KLK-ALHVDSAEYSCLKAIVLFTS----------------DACGLSDVAHVESLQEKSQCALEEYVRSQYP----NQP-TRFGKLLLRLPSLRTVS

200


FQVLPLDQLQLVELRNCWASLLML-ELAQDRLQFET-VEVSEPS-MLQKILTTR---
RRETGGNEP-------------------PVSASQVQAICKFLS-KCW-
SLNISTKEYA YLKTGVFLNP----------------DPVGLQ----------------CVKYGQLGWGTQIQILSEHRTM---
---HQ-GPHDFRIELENSTLFLR

HsSSHP

---------------------------------------------REALDVLAKTVAFRLNLS-
FWQLPQDQQRLLQCQGWPFLL-GLAQDAVTFE--AEAPVPS-ILKILLEE---
PSSSGSSQ-------------------LPD----------------RPQ----------------PSLAAVWLCQCLE-SEFW-
SLELSPKEYA CGLTFLNP----------------DQVGLQ----------------AAISHGLQEAHWVLCEVLEFWC-
P-AAAAA---GLRTVLLTASTLKLIP

HsSSHPa

NIEIRCGDKAGHYGVHACEGCKGFFRTIRLKL--VYDKCIRD----
SCKIQKNRKNCQYCRFHKCLSVGMSHHNTSCVETVTELFEAKAI PFGANLINDQVTLLKYGVEAIF
AMLASSVWKDGML--------------------------------VAYNGNI-----TREF---
LK-------SLRPK-FCDIEMPKFDFAKM-ADLEDDSDSLFVAACICG-DRPGPLL----
---N VGHIEKMQEGIVHVLRLHLSQNNH-P---D---FLFPKLLQKMADRLQVL

HsSSHPb

MMECRVCQDKAGHYGVHACEGCKGFFRIRKL-ELYEKER-----
SCKIQKNRKNCQYCRFQKCLALGMSHNTCTTETVTELFEAKISPSSLFLINDQVTLKLKYGVEAIF
AMLASIVKDGML--------------------------------VANGSGFV---TREF---
LK-------SLRPF-SCDIFMKEDFY-AK-ADLEDDSDSFLFAVILS---DRPGLL----
---N VNPVEIAEQDILRAFEHLQANHP-P---AQ-QLYFPKLLQKMADRLQVL

HsSSHPg

AIECRVCQDKAGHYGVHACEGCKGFFRIRKL--IYDRCDL---
NCRIHKSRRNKNCQYCRFQKCLAVGMSHHFRSVEAVQETEYASKIPGFVNLINDQVTLLKYGVEIY
TMLASMKDGVML-----------------------------------------------------------------------------
G----G----G-FOQ LLLMLKHY-MK-KLQHEEEYVLMQALISLPF---DRPGLV---
QHRVQDLQEQFAITLKSYIECRNRP-QP-AH-HLFLMKAMITELRIS

HsCAR

LRNPCQCVGDAQTGYHNALTCECGKGFRRTVKSGI--G-PTCFF---
AGCSVEKRSQKRHCAPRLQKCLDSGKKEDEMDTMYFEGÍSSFAKVISYRDLIEDQISLLKGAÁEL
CHIVLNTTFLCTQTN-------------------FLCGLP----------------RYTIEDGAV-
SPTV---G-FQVFELLELLHFG-TLR-KLQLEPEYVLLAAMALSFAPYLTDRTGV---
---QRBIDEQLQEQEALMTLQSYIKQOQR-RFR-DR-FLYAKLGLLAELRIS

HsSRa

KRFCVFQCDK SGSSHYGVSAACEGCKGFFRFSIQKGM--VYT-CHR-
DKNN InsetsRNRCYCRFQKCFEVGMSKTEELTKICIKITEFRAKLRPGFTTLIADQITLLKACLDI
LIRLCRTFYPEDQT----------------MTSFDGLT----------------LNRTQ---
MHN---AG-FGPLTDLVFAPAN-QLL-PELMDEAETGLLISAICLGC---DRQDLE---
---QPDRVMDLQ EPLEALKYYVRKR P-SC-HP-MFPPKLMKIIQDLRIS

HsSRb

KRFCVFQCDKSSHYGVSAACEGCKGFFRFSIQKGM--VYT-CHR-
DKNVKNTVRNRCYCRFQKCFEVGMSKTEELTKICIKITEFRAKLRPGFTTLIADQITLLKACLDI
LIRLCRTFYPEDQT----------------MTSFDGLT----------------LNRTQ---
MHN---AG-FGPLTDLVFAPAN-QLL-PELMDEAETGLLISAICLGC---DRQDLE---
---EPTKVDKLQ EPLEALKYYVRKR R-SC-HP-MFPPKLMKIIQDLRIS

HsSRag

KRFCVFQCDKSSHYGVSAACEGCKGFFRFSIQKGM--VYT-CHR-
DKNN InsetsRNRCYCRFQKCFEVGMSKTEELTKICIKITEFRAKLRPGFTTLIADQITLLKACLDI
LIRLCRTFYPEDQT----------------MTSFDGLT----------------LNRTQ---
MHN---AG-FGPLTDLVFAPAN-QLL-PELMDEAETGLLISAICLGC---DRMDE---
---EPEKVDKLQ EPLEALR LYYARR R-SC-HP-MFPPKLMKIIQDLRIS

HsReEv_Era

VLIVCVCGDASVQHYGVHACEGCKGFFRFSIQQNN---QYKDRKL--
NENCSIYRNIRNRCYCRFQKCLSVGSMDSTFPAVENKEVFAKHIPIFGDRLSQHQDVTLKAGTFE
LMVRFSALSFVNVKDT--------------------------------YLSQ---
LG-------AMG- MGDLSAMDFSE-KLN-SLALTEELGLTAVLVSA--DSSGME---
---NSASV EQLQTLRLRALK VKNRP-L-ET- SRTFKLKLKDPLRTLN

HsReEv_Erb

VLIVCVCGDASVQHYGVHACEGCKGFFRFSIQQNN---QYKDRKL--
NENCSIYRNIRNRCYCRFQKCLSVGSMDSTFPAVENKEVFAKHIPIFGDRLSQHQDVTLKAGTFE
LMVRFSALSFVNVKDT--------------------------------YLSQ---
LG-------AMG- MGDLSAMDFSE-KLN-SLALTEELGLTAVLVSA--DSSGME---
---NSASV EQLQTLRLRALK VKNRP-L-ET- SRTFKLKLKDPLRTLN

HsRORa

IIIPCKICGDSSSIHYGVITCECGKGFRRSQQ---SNATYS-CFR---
QKNCLIDRTSRNRCQHCRLQKCLAVGMSRDKitEAIQYVEFARIDGFMELOCQNDQIVLLKAGSLEV
VFIRMCRAFDSQNTT-------------------------VVFUNGK------------ASD-V-----
FKS---GCE-FFISVFFEFK-GL-SMCNLTEIFALFSQFVLMASA-DRSQWQP-----EKVRIKQ\KLQALQHQVKKNH---E-GLTNRKCIKSTICALC
hsRORg VIFPCICGDKSIGHGYVITCEGCKGFFRRSQR--CNAAYS-CTR--
QNCNIP1DSRTSRNRCQHCRLQKCLALMGRSRDHHTLEATIQYVEFARIDGFMELOCQNDQIVLLKAGAMVEV
VLMRMCAYNADNR------VVFEGK----------------GGM-E-I------
FRAK-----GCSE-LISSIFDFS-SL-AHIKSEDEILATLYALVINA------HRPGLQ---
-----EKRKEVQLYNLEAFHHHLCUTHR-Q-----S-I-LAKLP-PKGLRLSLC
hsRORb VIFPCICGDKSIGHGYVITCEGCKGFFRRSQR--NANASYS-CPR--
QRCLIDRTSRNRCQHCRLQKCLALMRGSRDIQTTHAIQYVEFARIDGFMELOCQNDQKLLKSGCLEV
VLMRMCAYNFLNNT-------------------VVFEGK----------------GGM-Q-M------
FKAL-----GDSD--LVNEAFDFPAK-NL-SDLQTEELFAIFASAVLISP-----DRAWLI--
-----EPRKVQKLQKEFIALQHVVQKKNL-D-----ETLAKLAIKTITAVC
hsRxRa KHIACICDRGGSCKHYSGCSCGKFFKRTIKRDL--TYT-CSR--
NKCLIDRKQRNRCQYCRYQCLAMGMKREQAADKQFLTVEWAKRIHFESSLPLDDQVILLRAWGNE
LIA-SFSLSRSIAK--DGILATG-LL-VHNS-------AH-----
-SAG--------VGAIFDRVLTLES-KMR-DMQMDKTELGCLRAIVLNFNP
DASKLSNPAEVEALREKVYASLEYCYKHK---EQP-GRAFKLLRLRPLRSLG
hsRxb KRLCAICGDRGSSCKHYSGCSCGKFFKRTIKRDL--TYS-CSR--
NKCDTVDKQRNRCQYCRYQCLAMGMKREAADKQFLTVEWAKRIHFESSLPLDDQVILLRAWGNE
LIA-SFSLSRSIDV--DGILATG-LL-VHNS-------AH-----
-SAG--------VGAIFDRVLTLES-KMR-DMQMDKTELGCLRAIVLNFNP
DASKLSNPEVEVLREKVYASLEYCYKHK---EQP-GRAFKLLRLRPLRSLG
hsRxc KHIACICDRGGSCKHYSGCSCGKFFKRTIKRDL--IYT-CSR--
NKCLIDRKQRNRCQYCRYQCLAMGMKREAADKQFLTVEWAKRIHFESSLPLDDQVILLRAWGNE
LIA-SFSLSRSVSO--DGILATG-LL-VHRS-------AH-----
-SAG--------VGAIFDRVLTLES-KMR-DMQMDKTELGCLRAIVLNFNP
DASKLSNPEVEVLREKVYASLEYCYKHK---EQP-GRAFKLLRLRPLRSLG
hsSf1 DELCVPCGKDSGSSCGHYGHCSCGKFFKRTQVNNK--HYT-CTE--
SQQSK1KOTKQRCFPCRQCLVTGRMLERADQTISFIDVWRARCVMFKEVANQIMLQHCWSEL
LVF-DHIYQVQHGK--EGSILVTQGEVE---------LTTV-A-------
-TAQ--------SLL-HSLVLRQLELV-QLL-ALQLDQREVFCLKFIIFLSEL-DLKFLN---
--HI----LVKDJAEKANAALLDLTYTLCY---P-HCG-DKFOQLLLCLVEVARS
hsLRH1 EELCVPCGVKDSGSSCGHYGHCSCGKFFKRTQVNNK--RYT-CIE--
NQNCQ1KOTKQRCFPCRQCLVTGRMLERADQTISFIDVWRARCVMFKEVANQIMLQHCWSEL
LIL-DHIYQVQHGK--EGSILVTQGEVE---------LYS-I---A-------
-SAQ--------ATL-NLNLASHELAL-QLR-QLDFQREVFCLKFLVIFLSEL-----DVKNLEN---
--FQ----LVEVGEQVNAALALLDLTYTMCNY---P-QOT-EKFQQQLLLREIRAS
hsTrRa DEQCVCGDKATGKHYHRYCITESCCEGKFFRRTIQQKNLHTYSTS-CKY--
DSCCVK1KTRTNQCLQCLFRCCKCIAVGMANKMTTPAIRTRVDFPAKLMPUBLICPEIQIIILKCGCMI
MSLRAAVRDPSNT------L twink---VKRE-Q-------
LKN------GG-LGVWSDAIFELGK-SLS-ANFLDLLTEVALLQAVLLMLT---DRESL----
----CVDKIEQKSEQAYLLTAFEHYNHVRK---N-IP-HFMPKLMLTVLDRMDMIG
hsTrRb DEQCVCPCGKATGKHYHRYCITESCCEGKFFRRTIQQKNLHTYSTS-CKY--
EGCVK1KTVRTRQCECFCRCKCIAVMANTTPAIRTRVDFPAKLMPUBLICPEIQIIILKCGCMI
MSLRAAVRDPESET------L圊NOMACA------TVRG-Q-------
LKN------GG-LGVWSDAIFELGK-SLS-ANFLDLLTEVALLQAVLLMLT---DRPGLLA----
----CVERIEYQDSSFLAHEHYINYRKH---VT-HFMPKLMLTVLDRMDMIG
hsTLX DIPCVPCGDSGSSCGHYGHCSCGKFFKRTQVNNK---TVY-CKN--
QQGCCPVCNHTRARMCLCKCIVMNMKESARALLMPSWAKSVPASTLSSLQDQMMLLEDANR
FVL-GIAQWAIPVAN-TNNLAVS----------------G-MNGDNDB---S------
-KLN------KISEIQALQVILLA-FRR-QRLLDATFACLKCIVTFKAVLTHS--G---
GSELRSFNAIAAIQAEDAILLTLNSYHTYR-----TQP-CRFGKLLRLRPLRSLG
hsTr2 FDLCVPCGDSGSSCGHYGHCSCGKFFKRTQVNNK---VYS-CSR--
SKDCIINKHRNRCQYCRYQCLVTGRMLERADQTISFIDVWRARCVMFKEVANQIMLQHCWSEL
FTL-GIAQWQMVNA-TLATLFOV------------------NCLHLNLDQDD---KM-STE-----
-RRK------LMLHFLIFLFECM-SMV-KLICIDEYALIKAIYFLFSP-------DDT-YLRSLRLLRLRPLRSLM
hsTr4 VEYCVCGDKASGRHYGAVCSCGKFFKRTQVNNK---TYS-CSR--
NQDCIINKHRNRCQYCRYQCLVTGRMLERADQTISFIDVWRARCVMFKEVANQIMLQHCWSEL
FTL-GIAQWQMVNA-TLATLFOV------------------NCLHLNLDQDD---KM-STE-----
-RRK------LMLHFLIFLFECM-SMV-KLICIDEYALIKAIYFLFSP-------DDT-YLRSLRLLRLRPLRSLM

204
FRI-GLAGAQVMSLS-TILAAIV------------------------NHLQNSIQEDE---KL-SGDE----
-RKK------QVMHEWKLQIEFCN-SMA-KLIDGKEYAYLKAIVLFSP----
DHPGLTSTQIEKFGEQAEQMELQDYVQKTYS----ETD-YLRALIVRLPARKVMS
Hs1VR PRICGVCGRATGHFNAMTCGCKGFFRSKMKRA---L-FTCPF--
NGDRCRTKNRRHQCACRLKRCVDMKMKEDLIVSYI3QKIGFAKMIPEFRLTDSETQ1IVLSSAIÆE
IMLRSNESFTMDMS------------------------WTGNDY--------KYRVSVD--
-TKA-----G-HSLELEPILFQYV-GLK-KINLHEEEHVLMAICIVSP-----DRPGVQ------
-----------DAALIAEIAIQDRLSNTLQTYRCHR---FP-SH-LLAKMIQKLADLRSIN
LgHNF4 SQYCAICGDKHGYAGASSCDCGCKGFFRSVKRS---VYT-CRF--
NRNCIVDKDRNCQHCYRCLKFCFAGMMKKEEMQQQLLIIIVKEIVPCFCELIDLDDQVALLRAHEG
LVL-GVAKRSLLEVND------------------------------------A-VIPRTAS-----
-------------GIIAIGVRLDELVQ-PLR-EIQIDTDEFAVICAIYFDP---
YARGNLNDQRIKMFYQVHVNLNEDYNQRO---DRG-FRGEILLPLQASIA
LgN1R1J ILKGCVCNDKALYNDFAIDCECKAFTFRNASKKN--G-YNCFS--
EGNCXKLDPTKHCSCGGLKKCFSMGMKKECNDTAFLRILIKGKFLPEIKLESQD1HILKGGIMEI
FILRSMFSDKNEK-----------------------------WYKSDTE----GKKQMDPR-V--
IATA-----LSPDVMSSHYFIVTSLLE-LIKSDRVYLLLIFVIELFSP-----DRQLVK------
-----------NKESSVTRSEQKSYNSLYKTVLESYIPE--VRK-AQ-TLYPKLMKLYEYRLSLG
Dm_SVP NQYCGVCDSKQGKHYQGTPCCECGSKFSLRVRNL---TYS-CRG--
SRNCPIQGHRRNCQCYCQRCLKCLKMQREELAALRLFSAPEWAKNPIFPEPELVQTDQVALLRLVNSL
FVL-NASQCMPLHVA-PLLAAG---------------------LH-A----------SPMAAD--
-RVV-AFMHDIDEQYVE-LLA-ALHVDSEAIESCLAIYLFITT--DACGLSDVTHIESLOQKSCALEYCRQYF------NPQ-TRFQGKLLRLPRRTVS
LgERR KRVLCVCNGASQHHGYVSGCSCRAAFFKRRTQIGNI---EYS-CPA--
SGECETIKRRRKAQACRFQKCLVGMREDLABRELITISWAKQGFGTSLNSQDMNLHSLSWVE
LCL-NLVRSCPYS--------------------------LY-CFAEDLKL----------VGSJ-L---
-VKTY-----NVP-PEFNLTRKLCK-KPT-LGLVSKEEYVLLACLITALCN------DVVA-E-----
-----------VTDVTRCLQDLQDSLMEITRQRYHG------GDM-KRQLQFLLLLIPITHK
LgFTZ_f1 EELCPVCQDGKVSYYALTCSCKGKFPRKTVQNNK---VYS-CVD--
NRSCH1DIQSKQRKCPYCFQKCLVGMRELADRELITISWAKQGFGTSLNSQDMNLHSLSWVE
LFL-DFVYQVRHDCA-----------------------------GSEIVLQNGKIN------------DL-DL--
-GBK------L-GDLKRTLLELIK-KTK-EVNLIDLEYACLPFIILLNP-----DVSQLEN------
-----------YVEQESRQVNASLLECLSFLY--P-DVK-DKFQGLLRLPEIRLS
LgHr3 VIPCKVCGRASQHHGYVTSCECGSKFSLRVRNL---GPVNYQ-CPR--
NKAC1DVRNRNCQCYCQRCLKCLMALGSRDDKTVLQVPPFISWAHIFFKMDQVDMQKLLNSQWEI
ALLQOCRAYVASNS------VFIGNSNS-------IPLA-D---
FHAL-----DAEEKMVMEKIEFQVF-TQV-IFKLEIEMALFSAILISS-----DRPGGLK------
-----------D7GDIQKHLKEIAKQLKEVGNHNS---DS--ETLAKLQVIPVRLNS
LgE78c VIPCKVCGRASQHHGYVTSCECGSKFSLRVRNL---GPVNYQ-CPR--
DKC7VTRLNRNCQCYCQRCLKCLAMSRDDKTVLQVPPFISWAHIFFKMDQVDMQKLLNSQWEI
WLTSMRSMFQRNNL-----------------------------VTFQEDGSM----------IHKE-D--
LVS-----VF-SPAVSSLFDVST-SPN-HLNNLDTEDVGLFTGIVATA------DRPGGLS------
-----------DSKGVEIQDQLKIEALKLQSVRNN---T---EE-NLFAISJIMKPFLRTLGL
LgDax1 ------------------------RVSDDLWVKVIFQAKSVE--
FCNLSSHNKTVLILLNSWSKILLL-FMAPPNFOQAV-----TLPSSSP--
VPT------EPV-A-SPDE------FSMKSVESIPTFR-KCQ--
HNLNIDSEYSPFRLYLLVLFNS-----GYIGL-----RADIDQDLNSTVQQLLEHVRATR--
-P-NDV-MHYSSLCLPCLTLYGTM
LgRev_erb SPHCKVCNCGGDSQGHYVTSCECGSKFSLRVRNL---THK-CAN--
EEKCE1TFPRNSQCFCRKLKCFVQGSMRESAMAPIEINVKFCFKIGFKEQDIQ1IKLKKGSSFV
LLARFTLLNVPQTEB--MMMMTPCTSL-----------RCPRH-V---
VEN------MP-MGMIKEFYGISE-VLPLDTRGELAIFSAILLISP-----DRPGK------
-----------NRMVTICKVGLHHMFYELLKTNPH-D-PD-NLFLLETSTIPQFQIN
LgHr4a MMICAIQDRCATLGHYGIITCCEGCCKGFRRKTVQNRK---IYS-CVA--
SGDCEINRAQNRNCQCFCRKKLCKMGVARIIDDDIVKLVHWIRNLFSEIPLKLQSMQMLTTKWHL
LLL-IMTAFSACP---LFTPLGQLNSKQCPYCYNLFTGNSLS----------TNQL-N--
-KEI-------GVPIKEVTHML-KFL-SLOLSQFFCVQLVILMSQ-----ND-NIN-N--
-----------EMSQSIIINIQERYMKVLQLYTQTFQ-C---S-SQ-S-YPFGEALLLLIPQIKDAS
LgHr4b DKGXCGCGRALGYNFDIAECSCKAFLFRRNAFLK---YFK-CFYP--
EEKCM6VSNRRFIFCRKLKCFEIGMRKIERGGDIVKLVHWIRNLFSEIPLKLQSMQMLTTKWHL
MILRLSFYRDPDLS--------------------------MVFSNKVC---INRE-Q---
LEE------GG-FGVLAAATINFAA-SLKL-SMSTDETEFAVLSALVCVSG---DRSGKL---
---ESVKRIEQLQPLFELAALKHYVRSSRE-E---QP-HVFAMLMRLTLDRIS
LgOR
VLKCLVCSKDSGGYHVGLACEGCKGFPFARALQNGPARK-CFY---
NKNCEINLQTRNRCQYRCALKCLALGMSRLNMRHHTQDVTIPFKPGPTGQLDQQDQISSLHKGCFFEV
ACCVCPVFDNSNT---IYLLNGT---LITRE-E---
MKI-----FFLGEHFVELNFLNC-RFN-APHLIDSEKALFSALVMIF------DRPGKL---
---NRDKVCKLQEMILQALQPEISDCHP-D---EV-GLFPRGLMISSSLRLG
LgER
SKVQCVNDSNASGHYGVWSECGKACKFPPRSIQGFV---DVY-CPA---
TNNTCTIKHRKRSCQACRLKCYEMKNKGLADRELVLINWAKHVFYDLSLSDQVHLIECCWME
LALL-NCAFGRSMDFH---GKRLVFPD---LDRIQ-L---
W-NL-----TGMDTEILEVAAASE-LMV-QYKISEEELLLQATVLCN---EVERRLA---
---SFSRIQDMVRVIDALVE-GRP-RA-DNS-RHVPSILLLTHIIQAG
LgTLX
DIPCKVCQDHSGGYHYACGCGAFFKRSIRK---HYI-CKG---
QGTCPDVKTHRNQCRLLKKEAEAMKDEAARAALLMLMSVKWTVPALSLFRPDQVILVEEAWREL
FVL-GAAQFPQIEAG-PLTLSA---------G-ISTEQSP---P---E---
-KV---SIMTEIMALQEVIT-KKF-AIQVDPTEYACLGIVLKFVTFFFFS---SQQ---FRFGLLLLPLSRLSG
LgFAX1b
PVSECVSCGDSYKGYHVFCDCGCGCFPKRSIRKNI---NYT-
CIEIGKGCCTCIIKARNWCPYCRLLKCLSLMNRRNCQYLIILMSLRLHYRSFLRSLYPADQMLVQN
TWSELFLL-SAAYW-PVDI--------QCIHHMA-KS---
TNDQ-5---C-K----------IKKILA-GQ-SLKVQMLFPELTAVLFVRL---GK--
---NNSSXKLSVNTLQDQALALAQYESSL-------DPDF-DSFRGKMILTPDLKLT
LgPNR
GLVCLVCGDTSKGGHYGIALCNCGSCGFRRSRRKL---IYR-CQA--
GTGMCNVDKAKHRNCQACRLKCIITGMNKDECSARLLFAMAQKSNLSPLNFLRDQVILVEEAWSE
LFLLL-CATQWMSMPEA-CPLFLFD---------HGSHTQ---
---GKV--------SYSPDMLVQETVT-RFK-AVCDPANSACLSKALRZKFP---
ETRGLKPSQVENLQDQAPQMGLQHIRAQHP---------TRFRFRGLLTLFAPRYIP
LgDSF
DIPCVKCDGGSSGYHYGICGSGCGCFPKRSIRKNI---AYT-CKA--
MKGCMPDKTTHRQSRCLRCLKCDADMKNEVAMLRTTIGWQKIPAFMLTEADQVIT
LTTSWDCLFLL-GIQVQWGPLF---NGHRSI-NF---
LGL-----------------NQGEMP-TEEQLRECTIA-RLR-EMALDPTETLCLKAILCFKS---DG---
---TNLVDKTVENQLEAHMLTEHILQ---YPRQ-RMVRGFRMLMLVISQKLT
LgNR4a
GDKLCLVCEDRASGHHYGVQMCLGCSGCGCFPKRSIRKNI---
FGQVCVFTRNLSCQHCRYKQVAGMKEYQYLELKFIEAWKVGPFQEMSMDDQMLCLLASFIMEL
NVRLLAYSMLNDRC---------IKFAEGV--------IPTV----AOG------MGWGRELVNATIEPTS-RLR-ELKIDRTFANLALVTFP---DAIQV---
---DKLQVLQGLTQKIDALLRTRYSNRP---T---E-RYRGKILLRPALRTP-
LgE75
TVCLRVCWGKDSASHGHYHVACEGCGCFPKRFRQ---SYRQPKL---
NQQCNMRVNNRCQRCYRLKCICVSMGRSDSTFPAISKVSVDAGGPIGUNTQDQVTLKAGKFE
LLVRLSCLPDFTDNT-------LMPTGRLR---------FRPQ---
SNNTN-NISC-GAPFLDLEMDPFAE-RFN-KLNLDEEALSFASIVLLSP---DRPLGR---
---NLEQEVIKQNALTESLHNININYK-E----DN-TRFAMKLMTDLRTL
LgFPAR1
MKGCMLMRISRNRCQOFLRNCSVGMVSHFLPCICVICTTTFAKEIQPFLDLPSDDQRLLWVGGCCLL
AAIADHSTHILEEK-WE-DEDK---FRD----
TKH-W-NEYL-GLEFVHPMYYDVG-KLR-KYELTVLSLCCMVFCEP-------DREGKL---
---TVKMLENLDMLSMACQLLNHNG-N---GS-MLFVRCIETVVALTIRIT
LgNR2E2
DALVCQVCDGSSGYHVACEGCGCFPKRSIRKNI---DYV-CKE---
NGNTTVDVARNQQCACFRKCLDKMNKETAARLLFMSKVRANISPIQLFPFDQAILLEESWEL
FIL-SASQSWLPAAD-ALLAMSS---------G-VTSENSTA---T---E---
-KSV-----YISSQIRSPQIIN-RFN-SLVRDTEYACLKLALFLFVSIFSK---I---
FSERTGRDLPSQQTDLQDQAPQLMMHEIIQSHHP---GNK-ARFGLLLLPLSRLAS
LgNR1J2
KYCQVCVCDGALYNFASCEKCAFRRNAKPG---DYFCFPY---
EKKCDMVSNRFPRCRKRLSFEIQMNKIAEISVRVFIDSSMKKSKFALSQTDQISLLKGSIE
LIIRSVFIDREKQH---FLPDYN---GKYAMTAEQ---
LKM------EGSLFEDHMKVFSLAM-DLKADETVLLLVISSFSP---DRPSIS---
---DKTYIAKEQDKYALHHLYESKYP-VSVR-TYIPKLMKLTIDIRLN
BgCOP_tFa
QIEVCVCDGKSSGKHYGQFTCEGCKFPPFRALQNGPARK---TET-CRG---
SRNCPEQDHRHRNCQRCYRLLKCLSKSMRREELALRRFLSSAVEWANIPFPDPDQVTQYDALLRGLWEL
FVL-NAAQCSMPLHVA-PLLAAAG---LH-A----------SPMAEA
VLL-SAGYDTE---------ISFTPQVNLRLQNRLSLVGLKEIP---------LEHV-A---
-KEA---------GPLVTRFFLLH-SFS-NKLKSFAYAVCIAKTLHLL----SADSTLD---
--RSIIKDVKLTQDFVKLQIQLHQLQPG--EDAQITSLAQILDWLPDLRNAS
Dm_HNF4
PTVACICGDRATKHYGASSCDGCDFRFRS5VKNH--QYT-CRF--
ARNCVVDKDRNRNCYCRCLRFCKFRAGKMKKESSMQQQLTVWAKQPAPNLQELDQLDVALLRAHAGEH
LL-GLSRSRPH-LKD-VLiLSNN-------C-VIRHC---DPLVP----
-NLD--------ISRIGAIIDEVT-VMT-DVGIDTFACIKALVFDFD------------
NAKGNLEPFRKLSRLHQLNNLEDYISDQY-----Esr-GRFGEILLILPVQLQST---

Dm_Hr96
PKNCACGCDQLGYNFNAVCECSKAFFRRNALAK--QFT-CFF--
NQNCIDTVTRRFQCKCRKLRCMKMSENLVARAIKRLIKM4KITAFFMCQEDQVALLKGGGTEM
MIMRSVMIYDRAA--------WKVPHTKENVM---GNIR-TDMLKFAEG---------I---YEEHQMFF-TFDEKWRMDENIIIMCAIVLFTS---ARSRV
IHKD--------VIRLEQNSYYLRRLESVYS-GC-EAR-NAFKIKIQIKISVERL
Dm_EcR
EELCLVCGRASGHYNALTCEGCKGFFRSTKVSA---VYC-CKF--
GRACEMMNYMRKQECRLKCLAVGMRPEEITILTVQLIVEFAKLPAFTKIPQEDQITLLKACSEV
MMLMRARHYDSDS--------IFFANNRS-----YTRD-S--
YKMD----AGMA-DNIEDDLLHFCR-QMF-SMVKDNVEYALLTAIVFS-----DRPQLE------
------------------KAQLVLEEAIQSYYIDTLRY1LNLHC--GDMSML-VFYAKLLSILTELTG

Ce_Hr48
NKVCRCGDKAFSYNFINVITESCSKAFFRRANKEK---EIR-CFF--
NEQCEINMVSRFCQRCLTKCFAVGKKE------------

Ce_Hr41
GELCVGCDKASGRHYGAVSCEGCKGFFKRSIKRI---GYV-CRS--
QKDCPVTKFKHRNRQCYCRKLKCSMGMRSE--

Ce_Fax1
TLCCAVCGBVSSGKYGLACNGCSCGGFFKRSVRR---IYR-CQA-
GTVNCVVDKAKHNRQOCARCRLKKCNKMGND-

Ce_Hr6
DVDCRVCEDHSSGKHYISIFSCDCGAGFFKRSIRRHR---QYT-
CKNKSEGQCKVTHRNRQCACRLKCLEGMNKD----------

Ce_Nhr85
-----------------------------------------------
NFIVPLKAAIDFANS1PAFLSITQRTQVRHLLLQNSVFDVMLL-ASASAASQHFPFG--
 Lanka------------LT--------YDQS-S-AN-------PIIPQAIQSISA-RIR-
 QL----PFPQITLAIVACQQA----DLPESQ-----------QPMLAERLWVLGKLG-GIQSAL-
 ----TAP----------SLLADVTLR

END;
Appendix 4.2: Maximum Parsimony tree showing nuclear receptors in molluscs, human, fly and worm.
The nuclear receptors from five different species were subjected to phylogenetic comparison using Maximum Parsimony method with PAUP version 4.0b10. Notations Bg, Lg, Hs, Dm and Ce in association with the receptor name denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster and C. elegans, respectively.
Appendix 4.3

Appendix 4.3: Maximum Likelihood tree showing nuclear receptors in molluscs, human, fly and worm.

The nuclear receptors from five different species were subjected to phylogenetic comparison using Maximum Likelihood method using Jones-Taylor-Thorton (JTT) substitution model with PHYML v2.4.4. Notations Bg, Lg, Hs, Dm and Ce in association with the receptor name denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster and C. elegans, respectively.
Appendix 4.4

Appendix 4.4: Bayesian Inference tree showing nuclear receptors in molluscs, human, fly and worm.
This nuclear receptor phylogenetic tree is the expanded for of Figure 4.0.7. Bg, Lg, Hs, Dm and Ce in association with the receptor name denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster and C. elegans, respectively.
## Appendix 5.1

### TargetP v1.1 prediction results

Number of query sequences: 21
Cleavage site predictions included.
Using NON-PLANT networks.

<table>
<thead>
<tr>
<th>Name</th>
<th>Len</th>
<th>mTP</th>
<th>SP</th>
<th>other</th>
<th>Loc</th>
<th>RC</th>
<th>TPlen</th>
</tr>
</thead>
<tbody>
<tr>
<td>APGWamide</td>
<td>213</td>
<td>0.029</td>
<td>0.961</td>
<td>0.029</td>
<td>S</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Schistosomin</td>
<td>96</td>
<td>0.022</td>
<td>0.964</td>
<td>0.093</td>
<td>S</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Conopressin</td>
<td>157</td>
<td>0.030</td>
<td>0.946</td>
<td>0.056</td>
<td>S</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>ELH1</td>
<td>117</td>
<td>0.192</td>
<td>0.528</td>
<td>0.069</td>
<td>S</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>ELH2</td>
<td>230</td>
<td>0.338</td>
<td>0.622</td>
<td>0.085</td>
<td>S</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>FFamide</td>
<td>91</td>
<td>0.012</td>
<td>0.978</td>
<td>0.075</td>
<td>S</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>FMRFamide</td>
<td>292</td>
<td>0.028</td>
<td>0.963</td>
<td>0.035</td>
<td>S</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>GnRH</td>
<td>106</td>
<td>0.068</td>
<td>0.916</td>
<td>0.032</td>
<td>S</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>Whitnin</td>
<td>114</td>
<td>0.020</td>
<td>0.961</td>
<td>0.054</td>
<td>S</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Myomodulin1</td>
<td>412</td>
<td>0.024</td>
<td>0.928</td>
<td>0.140</td>
<td>S</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Myomodulin2</td>
<td>291</td>
<td>0.026</td>
<td>0.946</td>
<td>0.075</td>
<td>S</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Myomodulin3</td>
<td>164</td>
<td>0.032</td>
<td>0.919</td>
<td>0.055</td>
<td>S</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>LFRFamide</td>
<td>187</td>
<td>0.021</td>
<td>0.879</td>
<td>0.139</td>
<td>S</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>NPY</td>
<td>90</td>
<td>0.014</td>
<td>0.975</td>
<td>0.057</td>
<td>S</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Pedal_peptide1</td>
<td>610</td>
<td>0.035</td>
<td>0.918</td>
<td>0.154</td>
<td>S</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>Pedal_peptide2</td>
<td>200</td>
<td>0.024</td>
<td>0.920</td>
<td>0.072</td>
<td>S</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Pedal_peptide3</td>
<td>508</td>
<td>0.026</td>
<td>0.918</td>
<td>0.144</td>
<td>S</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>133</td>
<td>0.077</td>
<td>0.947</td>
<td>0.022</td>
<td>S</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>183</td>
<td>0.035</td>
<td>0.926</td>
<td>0.046</td>
<td>S</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>118</td>
<td>0.065</td>
<td>0.916</td>
<td>0.029</td>
<td>S</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>168</td>
<td>0.008</td>
<td>0.978</td>
<td>0.095</td>
<td>S</td>
<td>1</td>
<td>25</td>
</tr>
</tbody>
</table>
Cut-off 0.780 0.000 0.730

# SignalP-NN euk predictions

<table>
<thead>
<tr>
<th># name</th>
<th>Cmax</th>
<th>pos ?</th>
<th>Ymax</th>
<th>pos ?</th>
<th>Smax</th>
<th>pos ?</th>
<th>Smean</th>
<th>D</th>
<th># name</th>
<th>Cmax</th>
<th>pos ?</th>
<th>Sprob ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>APGWamide</td>
<td>0.695</td>
<td>31 Y</td>
<td>0.766</td>
<td>31 Y</td>
<td>0.989</td>
<td>13 Y</td>
<td>0.872</td>
<td>Y</td>
<td>0.819 Y</td>
<td>APGWamide</td>
<td>0.744</td>
<td>31 Y</td>
</tr>
<tr>
<td>Schistosomin</td>
<td>0.846</td>
<td>18 Y</td>
<td>0.853</td>
<td>18 Y</td>
<td>0.993</td>
<td>2 Y</td>
<td>0.953</td>
<td>Y</td>
<td>0.903 Y</td>
<td>Schistosomin</td>
<td>0.970</td>
<td>18 Y</td>
</tr>
<tr>
<td>Conopressin</td>
<td>0.670</td>
<td>26 Y</td>
<td>0.695</td>
<td>26 Y</td>
<td>0.995</td>
<td>13 Y</td>
<td>0.872</td>
<td>Y</td>
<td>0.783 Y</td>
<td>Conopressin</td>
<td>0.487</td>
<td>26 N</td>
</tr>
<tr>
<td>ELH1</td>
<td>0.825</td>
<td>36 Y</td>
<td>0.747</td>
<td>36 Y</td>
<td>0.963</td>
<td>23 Y</td>
<td>0.622</td>
<td>Y</td>
<td>0.684 Y</td>
<td>ELH1</td>
<td>0.960</td>
<td>36 Y</td>
</tr>
<tr>
<td>ELH2</td>
<td>0.522</td>
<td>23 Y</td>
<td>0.457</td>
<td>23 Y</td>
<td>0.935</td>
<td>1 Y</td>
<td>0.575</td>
<td>Y</td>
<td>0.516 Y</td>
<td>ELH2</td>
<td>0.523</td>
<td>23 Y</td>
</tr>
<tr>
<td>FFamide</td>
<td>0.493</td>
<td>23 Y</td>
<td>0.623</td>
<td>23 Y</td>
<td>0.985</td>
<td>13 Y</td>
<td>0.895</td>
<td>Y</td>
<td>0.819 Y</td>
<td>FFamide</td>
<td>0.769</td>
<td>23 Y</td>
</tr>
<tr>
<td>FMRFamide</td>
<td>0.713</td>
<td>24 Y</td>
<td>0.742</td>
<td>24 Y</td>
<td>0.987</td>
<td>17 Y</td>
<td>0.940</td>
<td>Y</td>
<td>0.841 Y</td>
<td>FMRFamide</td>
<td>0.743</td>
<td>24 Y</td>
</tr>
<tr>
<td>GnRH</td>
<td>0.975</td>
<td>24 Y</td>
<td>0.925</td>
<td>24 Y</td>
<td>0.996</td>
<td>14 Y</td>
<td>0.940</td>
<td>Y</td>
<td>0.933 Y</td>
<td>GnRH</td>
<td>0.997</td>
<td>24 Y</td>
</tr>
<tr>
<td>Whitnin</td>
<td>0.690</td>
<td>22 Y</td>
<td>0.735</td>
<td>22 Y</td>
<td>0.959</td>
<td>11 Y</td>
<td>0.896</td>
<td>Y</td>
<td>0.815 Y</td>
<td>Whitnin</td>
<td>0.557</td>
<td>24 Y</td>
</tr>
<tr>
<td>Myomodulin1</td>
<td>0.805</td>
<td>20 Y</td>
<td>0.749</td>
<td>20 Y</td>
<td>0.960</td>
<td>1 Y</td>
<td>0.781</td>
<td>Y</td>
<td>0.765 Y</td>
<td>Myomodulin1</td>
<td>0.983</td>
<td>20 Y</td>
</tr>
<tr>
<td>Myomodulin2</td>
<td>0.577</td>
<td>23 Y</td>
<td>0.622</td>
<td>23 Y</td>
<td>0.954</td>
<td>6 Y</td>
<td>0.780</td>
<td>Y</td>
<td>0.701 Y</td>
<td>Myomodulin2</td>
<td>0.609</td>
<td>23 Y</td>
</tr>
<tr>
<td>Myomodulin3</td>
<td>0.813</td>
<td>25 Y</td>
<td>0.795</td>
<td>25 Y</td>
<td>0.986</td>
<td>13 Y</td>
<td>0.877</td>
<td>Y</td>
<td>0.836 Y</td>
<td>Myomodulin3</td>
<td>0.639</td>
<td>25 Y</td>
</tr>
<tr>
<td>LFRFamide</td>
<td>0.833</td>
<td>23 Y</td>
<td>0.744</td>
<td>23 Y</td>
<td>0.891</td>
<td>13 Y</td>
<td>0.704</td>
<td>Y</td>
<td>0.724 Y</td>
<td>LFRFamide</td>
<td>0.824</td>
<td>23 Y</td>
</tr>
<tr>
<td>NPY</td>
<td>0.612</td>
<td>22 Y</td>
<td>0.683</td>
<td>22 Y</td>
<td>0.997</td>
<td>13 Y</td>
<td>0.924</td>
<td>Y</td>
<td>0.804 Y</td>
<td>NPY</td>
<td>0.690</td>
<td>22 Y</td>
</tr>
<tr>
<td>Pedal_peptide1</td>
<td>0.218</td>
<td>24 N</td>
<td>0.335</td>
<td>24 Y</td>
<td>0.950</td>
<td>9 Y</td>
<td>0.669</td>
<td>Y</td>
<td>0.502 Y</td>
<td>Pedal_peptide1</td>
<td>0.258</td>
<td>24 N</td>
</tr>
<tr>
<td>Pedal_peptide2</td>
<td>0.451</td>
<td>25 Y</td>
<td>0.594</td>
<td>25 Y</td>
<td>0.977</td>
<td>16 Y</td>
<td>0.841</td>
<td>Y</td>
<td>0.717 Y</td>
<td>Pedal_peptide2</td>
<td>0.737</td>
<td>25 Y</td>
</tr>
<tr>
<td>Pedal_peptide3</td>
<td>0.637</td>
<td>19 Y</td>
<td>0.710</td>
<td>19 Y</td>
<td>0.934</td>
<td>9 Y</td>
<td>0.814</td>
<td>Y</td>
<td>0.762 Y</td>
<td>Pedal_peptide3</td>
<td>0.716</td>
<td>19 Y</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>0.914</td>
<td>23 Y</td>
<td>0.865</td>
<td>23 Y</td>
<td>0.992</td>
<td>1 Y</td>
<td>0.898</td>
<td>Y</td>
<td>0.882 Y</td>
<td>Insulin_like_peptide1</td>
<td>0.919</td>
<td>23 Y</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>0.908</td>
<td>22 Y</td>
<td>0.855</td>
<td>22 Y</td>
<td>0.982</td>
<td>13 Y</td>
<td>0.892</td>
<td>Y</td>
<td>0.874 Y</td>
<td>Insulin_like_peptide2</td>
<td>0.898</td>
<td>22 Y</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>0.898</td>
<td>20 Y</td>
<td>0.855</td>
<td>20 Y</td>
<td>0.995</td>
<td>13 Y</td>
<td>0.940</td>
<td>Y</td>
<td>0.897 Y</td>
<td>Insulin_like_peptide3</td>
<td>0.634</td>
<td>20 Y</td>
</tr>
<tr>
<td>Insulin_like_peptide</td>
<td>0.746</td>
<td>26 Y</td>
<td>0.739</td>
<td>26 Y</td>
<td>0.973</td>
<td>12 Y</td>
<td>0.805</td>
<td>Y</td>
<td>0.772 Y</td>
<td>Insulin_like_peptide4</td>
<td>0.893</td>
<td>26 Y</td>
</tr>
</tbody>
</table>

# TMHMM 2.0c predictions

<table>
<thead>
<tr>
<th># APGWamide Length: 213</th>
</tr>
</thead>
<tbody>
<tr>
<td># APGWamide Number of predicted TMHs: 0</td>
</tr>
<tr>
<td># APGWamide Exp number of AAs in TMHs: 12.75856</td>
</tr>
<tr>
<td># APGWamide Exp number, first 60 AAs: 12.75856</td>
</tr>
<tr>
<td># APGWamide Total prob of N-in: 0.51992</td>
</tr>
<tr>
<td># APGWamide POSSIBLE N-term signal sequence</td>
</tr>
<tr>
<td>APGWamide TMHMM2.0 outside 1 213</td>
</tr>
</tbody>
</table>

# Schistosomin Length: 96

<table>
<thead>
<tr>
<th># Schistosomin Number of predicted TMHs: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td># Schistosomin Exp number of AAs in TMHs: 0.64892</td>
</tr>
<tr>
<td># Schistosomin Exp number, first 60 AAs: 0.64892</td>
</tr>
</tbody>
</table>
# Schistosomin Total prob of N-in: 0.08835
Schistosomin TMHMM2.0 outside 1 96

# Conopressin Length: 157
# Conopressin Number of predicted TMHs: 1
# Conopressin Exp number of AAs in TMHs: 16.65291
# Conopressin Exp number, first 60 AAs: 16.65269
# Conopressin Total prob of N-in: 0.08581
# Conopressin POSSIBLE N-term signal sequence
Conopressin TMHMM2.0 outside 1 4
Conopressin TMHMM2.0 TM helix 5 27
Conopressin TMHMM2.0 inside 28 157

# ELH1 Length: 117
# ELH1 Number of predicted TMHs: 0
# ELH1 Exp number of AAs in TMHs: 8.89296
# ELH1 Exp number, first 60 AAs: 8.89113
# ELH1 Total prob of N-in: 0.37967
ELH1 TMHMM2.0 outside 1 117

# ELH2 Length: 230
# ELH2 Number of predicted TMHs: 0
# ELH2 Exp number of AAs in TMHs: 0.11823
# ELH2 Exp number, first 60 AAs: 0.11085
# ELH2 Total prob of N-in: 0.02623
ELH2 TMHMM2.0 outside 1 230

# FFamide Length: 91
# FFamide Number of predicted TMHs: 0
# FFamide Exp number of AAs in TMHs: 5.81686
# FFamide Exp number, first 60 AAs: 5.81669
# FFamide Total prob of N-in: 0.15353
FFamide TMHMM2.0 outside 1 91

# FMRFamide Length: 292
# FMRFamide Number of predicted TMHs: 0
# FMRFamide Exp number of AAs in TMHs: 3.21426
# FMRFamide Exp number, first 60 AAs: 3.21426
# FMRFamide Total prob of N-in: 0.18827
FMRFamide TMHMM2.0 outside 1 292

# GnRH Length: 106
# GnRH Number of predicted TMHs: 1
# GnRH Exp number of AAs in TMHs: 16.72643
# GnRH Exp number, first 60 AAs: 16.72384
# GnRH Total prob of N-in: 0.68137
# GnRH POSSIBLE N-term signal sequence
GnRH TMHMM2.0 inside
GnRH TMHMM2.0 TMhelix
GnRH TMHMM2.0 outside

# Whitnin Length: 114
# Whitnin Number of predicted TMHs: 0
# Whitnin Exp number of AAs in TMHs: 6.74482
# Whitnin Exp number, first 60 AAs: 6.74482
# Whitnin Total prob of N-in: 0.02809
Whitnin TMHMM2.0 outside

# Myomodulin1 Length: 412
# Myomodulin1 Number of predicted TMHs: 0
# Myomodulin1 Exp number of AAs in TMHs: 0.23007
# Myomodulin1 Exp number, first 60 AAs: 0.22997
# Myomodulin1 Total prob of N-in: 0.01416
Myomodulin1 TMHMM2.0 outside

# Myomodulin2 Length: 291
# Myomodulin2 Number of predicted TMHs: 0
# Myomodulin2 Exp number of AAs in TMHs: 0.13538
# Myomodulin2 Exp number, first 60 AAs: 0.13538
# Myomodulin2 Total prob of N-in: 0.01472
Myomodulin2 TMHMM2.0 outside

# Myomodulin3 Length: 164
# Myomodulin3 Number of predicted TMHs: 0
# Myomodulin3 Exp number of AAs in TMHs: 3.00215
# Myomodulin3 Exp number, first 60 AAs: 3.00215
# Myomodulin3 Total prob of N-in: 0.16742
Myomodulin3 TMHMM2.0 outside

# LFRFamide Length: 187
# LFRFamide Number of predicted TMHs: 0
# LFRFamide Exp number of AAs in TMHs: 4.54339
# LFRFamide Exp number, first 60 AAs: 4.54301
# LFRFamide Total prob of N-in: 0.05176
LFRFamide   TMHMM2.0   outside   1   187
# NPY Length: 90
# NPY Number of predicted TMHs: 0
# NPY Exp number of AAs in TMHs: 14.12354
# NPY Exp number, first 60 AAs: 14.12351
# NPY Total prob of N-in: 0.53621
# NPY POSSIBLE N-term signal sequence
NPY   TMHMM2.0   outside   1   90

# Pedal_peptide1 Length: 610
# Pedal_peptide1 Number of predicted TMHs: 0
# Pedal_peptide1 Exp number of AAs in TMHs: 0.08516
# Pedal_peptide1 Exp number, first 60 AAs: 0.08516
# Pedal_peptide1 Total prob of N-in: 0.00445
Pedal_peptide1   TMHMM2.0   outside   1   610

# Pedal_peptide2 Length: 200
# Pedal_peptide2 Number of predicted TMHs: 0
# Pedal_peptide2 Exp number of AAs in TMHs: 9.26723
# Pedal_peptide2 Exp number, first 60 AAs: 9.26723
# Pedal_peptide2 Total prob of N-in: 0.50095
Pedal_peptide2   TMHMM2.0   outside   1   200

# Pedal_peptide3 Length: 508
# Pedal_peptide3 Number of predicted TMHs: 0
# Pedal_peptide3 Exp number of AAs in TMHs: 0.03298
# Pedal_peptide3 Exp number, first 60 AAs: 0.03298
# Pedal_peptide3 Total prob of N-in: 0.00244
Pedal_peptide3   TMHMM2.0   outside   1   508

# Insulin-like_peptide1 Length: 133
# Insulin-like_peptide1 Number of predicted TMHs: 0
# Insulin-like_peptide1 Exp number of AAs in TMHs: 0.18019
# Insulin-like_peptide1 Exp number, first 60 AAs: 0.18019
# Insulin-like_peptide1 Total prob of N-in: 0.05416
Insulin-like_peptide1   TMHMM2.0   outside   1   133

# Insulin-like_peptide2 Length: 183
# Insulin-like_peptide2 Number of predicted TMHs: 0
# Insulin-like_peptide2 Exp number of AAs in TMHs: 1.0999
# Insulin-like_peptide2 Exp number, first 60 AAs: 1.0999
# Insulin-like_peptide2 Total prob of N-in: 0.08493
Insulin-like_peptide2 TMHMM2.0 outside 1 183

# Insulin-like_peptide3 Length: 118
# Insulin-like_peptide3 Number of predicted TMHs: 0
# Insulin-like_peptide3 Exp number of AAs in TMHs: 0.40773
# Insulin-like_peptide3 Exp number, first 60 AAs: 0.4066
# Insulin-like_peptide3 Total prob of N-in: 0.06364
Insulin-like_peptide3 TMHMM2.0 outside 1 118

# Insulin-like_peptide4 Length: 168
# Insulin-like_peptide4 Number of predicted TMHs: 1
# Insulin-like_peptide4 Exp number of AAs in TMHs: 17.61363
# Insulin-like_peptide4 Exp number, first 60 AAs: 17.60696
# Insulin-like_peptide4 Total prob of N-in: 0.81821
Insulin-like_peptide4 POSSIBLE N-term signal sequence
Insulin-like_peptide4 TMHMM2.0 inside 1 6
Insulin-like_peptide4 TMHMM2.0 TM helix 7 29
Insulin-like_peptide4 TMHMM2.0 outside 30 168

# Phobius predictions

ID   APGWamide
FT   SIGNAL   1   30
FT   REGION   1   9   N-REGION.
FT   REGION   10  25   H-REGION.
FT   REGION   26  30   C-REGION.
FT   TOPO_DOM 31  213   NON CYTOPLASMIC.
//
ID   Schistosomin
FT   SIGNAL   1   17
FT   REGION   1   2   N-REGION.
FT   REGION   3   13   H-REGION.
FT   REGION   14  17   C-REGION.
FT   TOPO_DOM 18  96   NON CYTOPLASMIC.
//
ID   Conopressin
FT   SIGNAL   1   23
FT   REGION   1   3   N-REGION.
FT  REGION   4    18       H-REGION.
FT  REGION   19   23       C-REGION.
FT  TOPO_DOM  24   157     NON CYTOPLASMIC.

//
ID   ELH1
FT  SIGNAL   1    35
FT  REGION   1    14       N-REGION.
FT  REGION   15   26       H-REGION.
FT  REGION   27   35       C-REGION.
FT  TOPO_DOM  36   117     NON CYTOPLASMIC.

//
ID   ELH2
FT  SIGNAL   1    27
FT  REGION   1    3        N-REGION.
FT  REGION   4    15       H-REGION.
FT  REGION   16   27       C-REGION.
FT  TOPO_DOM  28   230     NON CYTOPLASMIC.

//
ID   FFamide
FT  SIGNAL   1    22
FT  REGION   1    4        N-REGION.
FT  REGION   5    17       H-REGION.
FT  REGION   18   22       C-REGION.
FT  TOPO_DOM  23   91      NON CYTOPLASMIC.

//
ID   FMRFamide
FT  SIGNAL   1    23
FT  REGION   1    6        N-REGION.
FT  REGION   7    18       H-REGION.
FT  REGION   19   23       C-REGION.
FT  TOPO_DOM  24   292     NON CYTOPLASMIC.

//
ID   GnRH
FT  SIGNAL   1    23
FT  REGION   1    5        N-REGION.
FT  REGION   6    18       H-REGION.
FT  REGION   19   23       C-REGION.
FT  TOPO_DOM  24   106     NON CYTOPLASMIC.

//
ID   Whitnin
FT  SIGNAL   1    21
FT REGION  1  4    N-REGION.
FT REGION  5  16   H-REGION.
FT REGION 17  21   C-REGION.
FT TOPO_DOM 22 114 NON CYTOPLASMIC.
//
ID  Myomodulin1
FT SIGNAL   1  19
FT REGION  1  4    N-REGION.
FT REGION  5  14   H-REGION.
FT REGION 15 19    C-REGION.
FT TOPO_DOM 20 412 NON CYTOPLASMIC.
//
ID  Myomodulin2
FT SIGNAL   1  22
FT REGION  1  5    N-REGION.
FT REGION  6  17   H-REGION.
FT REGION 18 22    C-REGION.
FT TOPO_DOM 23 291 NON CYTOPLASMIC.
//
ID  Myomodulin3
FT SIGNAL   1  24
FT REGION  1  5    N-REGION.
FT REGION  6  17   H-REGION.
FT REGION 18 24    C-REGION.
FT TOPO_DOM 25 164 NON CYTOPLASMIC.
//
ID  LFRFamide
FT SIGNAL   1  22
FT REGION  1  5    N-REGION.
FT REGION  6  17   H-REGION.
FT REGION 18 22    C-REGION.
FT TOPO_DOM 23 187 NON CYTOPLASMIC.
//
ID  NPY
FT SIGNAL   1  17
FT REGION  1  3    N-REGION.
FT REGION  4  12   H-REGION.
FT REGION 13 17    C-REGION.
FT TOPO_DOM 18 90 NON CYTOPLASMIC.
//
ID  Pedal_peptide1
FT SIGNAL   1  16
FT REGION 1 2 N-REGION.
FT REGION 3 11 H-REGION.
FT REGION 12 16 C-REGION.
FT TOPO_DOM 17 610 NON CYTOPLASMIC.

//
ID Pedal_peptide2
FT SIGNAL 1 24
FT REGION 1 3 N-REGION.
FT REGION 4 16 H-REGION.
FT REGION 17 24 C-REGION.
FT TOPO_DOM 25 200 NON CYTOPLASMIC.

//
ID Pedal_peptide3
FT SIGNAL 1 17
FT REGION 1 4 N-REGION.
FT REGION 5 12 H-REGION.
FT REGION 13 17 C-REGION.
FT TOPO_DOM 18 508 NON CYTOPLASMIC.

//
ID Insulin-like_peptide1
FT SIGNAL 1 22
FT REGION 1 6 N-REGION.
FT REGION 7 14 H-REGION.
FT REGION 15 22 C-REGION.
FT TOPO_DOM 23 133 NON CYTOPLASMIC.

//
ID Insulin-like_peptide2
FT SIGNAL 1 21
FT REGION 1 3 N-REGION.
FT REGION 4 14 H-REGION.
FT REGION 15 21 C-REGION.
FT TOPO_DOM 22 183 NON CYTOPLASMIC.

//
ID Insulin-like_peptide3
FT SIGNAL 1 19
FT REGION 1 1 N-REGION.
FT REGION 2 12 H-REGION.
FT REGION 13 19 C-REGION.
// Insulin-like_peptide4

FT  TOPO_DOM  20  118  NON CYTOPLASMIC.

FT  SIGNAL  1  25
FT  REGION  1  6  N-REGION.
FT  REGION  7  20  H-REGION.
FT  REGION  21  25  C-REGION.

FT  TOPO_DOM  26  168  NON CYTOPLASMIC
REFERENCES


disrupters-Phylogenetic approach and common principles focussing on androgenic/antiandrogenic compounds-COMPRENDO: Final published report, .


receptor (RXR) from the mollusk *Biomphalaria glabrata* transactivates transcription in the presence of retinoids, *Journal of Molecular Endocrinology*, 34(2).


De Jong-Brink, M., Hoek, R.M., Lageweg, W. and Smit, A.B. (1997) 'Schistosome Parasites Induce Physiological Changes in their Snail Hosts by Interfering with Two Regulatory Systems, the Internal Defense System and the Neuroendocrine System', in *Parasites and*
Pathogens. Springer, pp. 57-75.


Felsenstein, J. (1978) 'Cases in which parsimony or compatibility methods will be positively misleading', Systematic Biology, 27(4), pp. 401-410.


Forman, B.M., Casanova, J., Raaka, B.M., Ghysdael, J. and Samuels, H.H. (1992) 'Half-site spacing and orientation determines whether thyroid hormone and retinoic acid receptors and related factors bind to DNA response elements as monomers, homodimers, or heterodimers', *Molecular endocrinology (Baltimore, Md.)*, 6(3), pp. 429-442.


Lehrbuch der Speziellen Zoologie.


Harry, H. (1965) 'Evidence of a gonadal hormone controlling the development of the accessory reproductive organs in Taphius glabratrus (Say)(Gastropoda, Basommatophora)', *Transactions*.
of the American Microscopical Society, 84, pp. 157.


Heindel, J.J. (2006) 'Role of exposure to environmental chemicals in the developmental basis of reproductive disease and dysfunction', Seminars in reproductive medicine. Published in 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA., 168-177.


strains of *Biomphalaria glabrata* following exposure to the digenetic trematodes *Schistosoma mansoni* and *Echinostoma paraensei*, *Dev. Comp. Immunol.*, 29(4), pp. 295-303.


Jobling, S., Casey, D., Rodgers-Gray, T., Oehlmann, J., Schulte-Oehlmann, U., Pawlowski, S.,


of methyltestosterone on the sexual development and reproduction of adult medaka (Oryzias latipes), Aquatic toxicology, 87(1), pp. 37-46.


Kim, H.S., Shin, J.H., Moon, H.J., Kim, T.S., Kang, I.H., Seok, J.H., Kim, I.Y., Park, K.L. and


invertebrates', Ecotoxicology, 16(1), pp. 131-143.


red flour beetle Tribolium castaneum', *Genome research*, 18(1), pp. 113-122.


Nässel, D.R. (1996) 'Peptidergic neurohormonal control systems in invertebrates', *Current opinion*


Newton WL. 'The establishment of a strain of Australorbis glabratus which combines albinism and high susceptibility to infection with Schistosoma mansoni.', J Parasitol, 41, pp. 526-8.


b)iology, 92(4), pp. 307-315.


Schulte-Oehlmann, U., Oetken, M., Bachmann, J. and Oehlmann, J. (2004) 'Effects of ethinyloestradiol and methyltestosterone in prosobranch snails', in *Pharmaceuticals in the*


Scott, A.P. (2012b) 'Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical review of the evidence that steroids have biological effects', *Steroids*, .


Van Kesteren, R.E., Smit, A.B., De Lange, R.P., Kits, K.S., Van Golen, F.A., Van Der Schors,


Vianey-Liaud, M., Nassi, H., Lancaster, F. and Dupouy, J. (1989) 'Duration of pairing and use of


Identification of the nuclear receptor DAF-12 as a therapeutic target in parasitic nematodes', *Proceedings of the National Academy of Sciences of the United States of America*, 106(23), pp. 9138-9143.


Wu, W., Niles, E.G. and LoVerde, P.T. (2007) 'Thyroid hormone receptor orthologues from invertebrate species with emphasis on *Schistosoma mansoni*', *BMC Evol Biol*, 7(1), pp. 150.


Websites


Artemis. www.sanger.ac.uk.


Blast2GO. http://www.blast2go.com/b2ghome.

BUAV. http://www.buav.org/international.


The COMPRENDO Project report.


InterProScan5. http://www.ebi.ac.uk/Tools/pfa/iprscan5/.


MEGA. http://www.megasoftware.net.


NC3Rs. http://www.nc3rs.org.uk.


NURSA report (2012).


OECD (2009).
www.oecd.org/document/62/0,2340,en_2649_34377_2348606_1_1_1_1,00.html.


Primer3. http://bioinfo.ut.ee/primer3-0.4.0/.


SMART. http://smart.embl-heidelberg.de.


UK Home Office report on animal testing (2011).